

# Identification and analysis of large intergenic non-coding RNAs regulated by p53 family members through a genome-wide analysis of p53-binding sites

Masashi Idogawa<sup>1</sup>, Tomoko Ohashi<sup>1</sup>, Yasushi Sasaki<sup>1</sup>, Reo Maruyama<sup>2</sup>, Lisa Kashima<sup>1</sup>, Hiromu Suzuki<sup>2</sup> and Takashi Tokino<sup>1,\*</sup>

<sup>1</sup>Department of Medical Genome Sciences, Research Institute for Frontier Medicine and <sup>2</sup>Department of Molecular Biology, Sapporo Medical University School of Medicine, Sapporo 060-8556, Japan

Received December 3, 2013; Revised December 24, 2013; Accepted December 30, 2013

**p53 is one of the most important known tumor suppressor genes, and it is inactivated in approximately half of human cancers. p53 family members execute various functions, such as apoptosis induction and cell cycle arrest, by modulating transcriptional regulation. Therefore, the direct transcriptional targets of the p53 family must be explored to elucidate the functional mechanisms of family members. To identify the direct transcriptional targets of p53 family members, we performed chromatin immunoprecipitation together with next-generation sequencing (ChIP-seq) and searched for p53-binding motifs across the entire human genome. Among the identified ChIP-seq peaks, approximately half were located in an intergenic region. Therefore, we assumed large intergenic non-coding RNAs (lincRNAs) to be major targets of the p53 family. Recent reports have revealed that lincRNAs play an important role in various biological and pathological processes, such as development, differentiation, stemness and carcinogenesis. Through a combination of ChIP-seq and *in silico* analyses, we found 23 lincRNAs that are upregulated by the p53 family. Additionally, knockdown of specific lincRNAs modulated p53-induced apoptosis and promoted the transcription of a gene cluster. Our results suggest that p53 family members, and lincRNAs constitute a complex transcriptional network involved in various biological functions and tumor suppression.**

## INTRODUCTION

p53 is one of the most important known tumor suppressor genes. In approximately half of all human cancers, p53 is inactivated as a direct result of mutations in the p53 gene (1). Furthermore, mutation or deletion of p53 is related to a poor prognosis and resistance to chemotherapy and radiation (2). The p53 protein is activated by a variety of cell stresses, such as DNA damage, oncogene activation, spindle damage and hypoxia. Activated p53 transactivates a number of target genes, many of which are involved in DNA repair, cell cycle arrest and apoptosis (3). Two other p53 family members, p63 and p73, also induce cell cycle arrest and apoptosis and play an important role in development and differentiation (4). Dominant negative forms of p63 and p73 are overexpressed in some types of cancers (5,6). At least 30% of head and neck squamous cell carcinomas harbor mutations in genes that regulate squamous differentiation,

including p63 (7). Furthermore, p53 transactivates not only coding genes but also non-coding RNA genes, including miRNAs, which are major mediators of gene suppression induced by p53 (8). A recent study revealed that large intergenic non-coding RNAs (lincRNAs) play important roles in various biological and pathological processes such as development, differentiation, stemness and carcinogenesis (9–11). Several lincRNAs are transactivated by p53 and mediate gene suppression (12).

Many researchers are working to identify the target genes of p53 and its family members because these proteins execute various important functions, mainly through transcriptional regulation. One useful common approach to achieve this goal is to compare mRNA expression in the presence and absence of p53 activation using a cDNA microarray. However, not all genes whose expression is altered by p53 activation are controlled by p53 directly; their expression may also be affected

\*To whom correspondence should be addressed at: Department of Medical Genome Sciences, Research Institute for Frontier Medicine, Sapporo Medical University School of Medicine, S1W17, Chuo-Ku, Sapporo 060-8556, Japan. Tel: +81 116112111; Fax: +81 116183313; Email: tokino@sapmed.ac.jp

by a secondary mediator. To identify the direct targets of p53 family members, the direct binding of p53 to regulatory regions near or inside candidate target genes should be confirmed via chromatin immunoprecipitation (ChIP). ChIP-cloning and ChIP-chip methods have been developed for the comprehensive detection of p53-binding regions (13,14). The recent development of next-generation sequencing has enabled the direct sequencing of ChIP-DNA (ChIP-seq). Several algorithms for the computational prediction of p53-binding sites (p53BSs) have been reported concurrently (14–16).

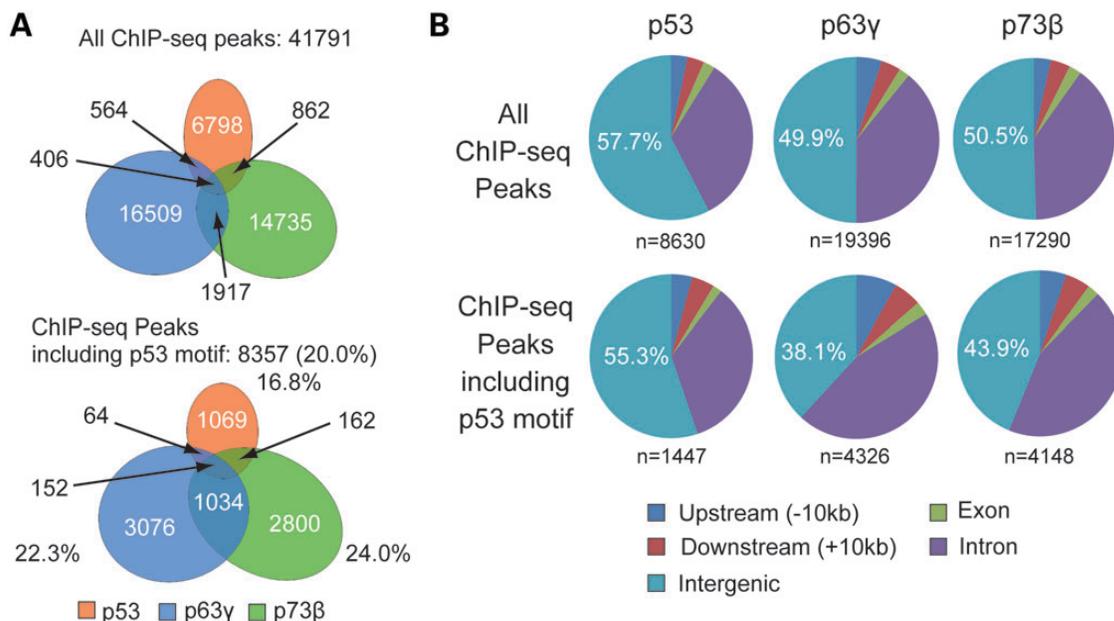
In this study, we performed a comprehensive characterization of p53BS across the entire human genome through a combination of ChIP-seq analysis and an *in silico* p53-binding motif (p53 motif) search. We showed that several lincRNAs are transactivated by p53 family members and modulate p53-induced apoptosis.

## RESULTS

### Correlation of ChIP-seq peaks and *in silico* p53 motifs in the human genome

To identify potential binding sites of p53 family members as comprehensively as possible, we infected H1299 lung cancer cells (p53 null) with adenoviruses expressing FLAG-tagged p53, p63 $\gamma$  or p73 $\beta$ . Among the isoforms of these proteins, we selected TAp63 $\gamma$  and TAp73 $\beta$  to maximize the coverage of binding sites because these proteins exhibit higher transcriptional activity than the other isoforms (17). The cells were then subjected to ChIP, and the genomic DNA bound to each FLAG-tagged protein was sequenced with a next-generation sequencer

(ChIP-seq). We identified a total of 41 791 p53 family ChIP-seq peaks (Fig. 1A, top and Supplementary Material, Table S1). These peaks indicate the physiological binding potential of p53 family members and include some false-positive noise due to non-specific binding caused by the excess of proteins present and the experimental process of immunoprecipitation. If these ChIP-seq peaks represent the direct binding of p53 family proteins, the peak sequences should include the canonical p53 motif, RRRCWWGYYY + spacer + RRRCWWGYYY because p63 and p73 can also potentially bind to the consensus p53 motif. Therefore, we focused on the ChIP-seq peaks that included canonical p53 motifs to exclude false-positive noise. We found a total of 636 233 p53 motifs in the human genome using the *in silico* system. Among the ChIP-seq peaks identified for all three p53 family members, 20.0% (8357/41 791) included these p53 motifs (Fig. 1A, bottom; Supplementary Material, Table S2). This result indicates that p53 motifs were enriched by approximately 50-fold in the ChIP-seq peaks compared with the whole human genome. These peaks included not only known p53 target genes, such as p21 (CDKN1A) and MDM2, but also AKR1B10 (18) and CLCA2 (19), both of which were recently identified as novel p53 targets (Supplementary Material, Table S3). Approximately half of the ChIP-seq peaks for the three p53 members were found within 10 kb of coding genes. The remaining ChIP-seq peaks were located in intergenic regions (57.7, 49.9 and 50.5% for p53, p63 $\gamma$  and p73 $\beta$ , respectively) (Fig. 1B, top). However, when only the ChIP-seq peaks that included p53 motifs were considered, the percentage of intergenic regions decreased to 38.1 and 43.9% for p63 $\gamma$  and p73 $\beta$ , while the percentage of p53 ChIP-seq peaks including p53 motifs remained at 55.3% (Fig. 1B, bottom). These results suggest that the p63 and



**Figure 1.** Correlation of ChIP-seq peaks and *in silico* p53 motifs in the human genome. (A) The overlap between all ChIP-seq peaks (top) and ChIP-seq peaks containing p53 motifs (bottom) immunoprecipitated with p53, p63 $\gamma$  and p73 $\beta$  proteins is displayed in Venn diagrams. The percentage of ChIP-seq peaks containing p53 motifs is also indicated. (B) The ChIP-seq peaks are categorized by their genomic position; exon, intron, 10 kb upstream from a TSS, 10 kb downstream from a transcription termination site (TTS) and intergenic regions. The proportions of each position in all ChIP-seq peaks (top) and ChIP-seq peaks including p53 motifs (bottom) are displayed as pie diagrams.

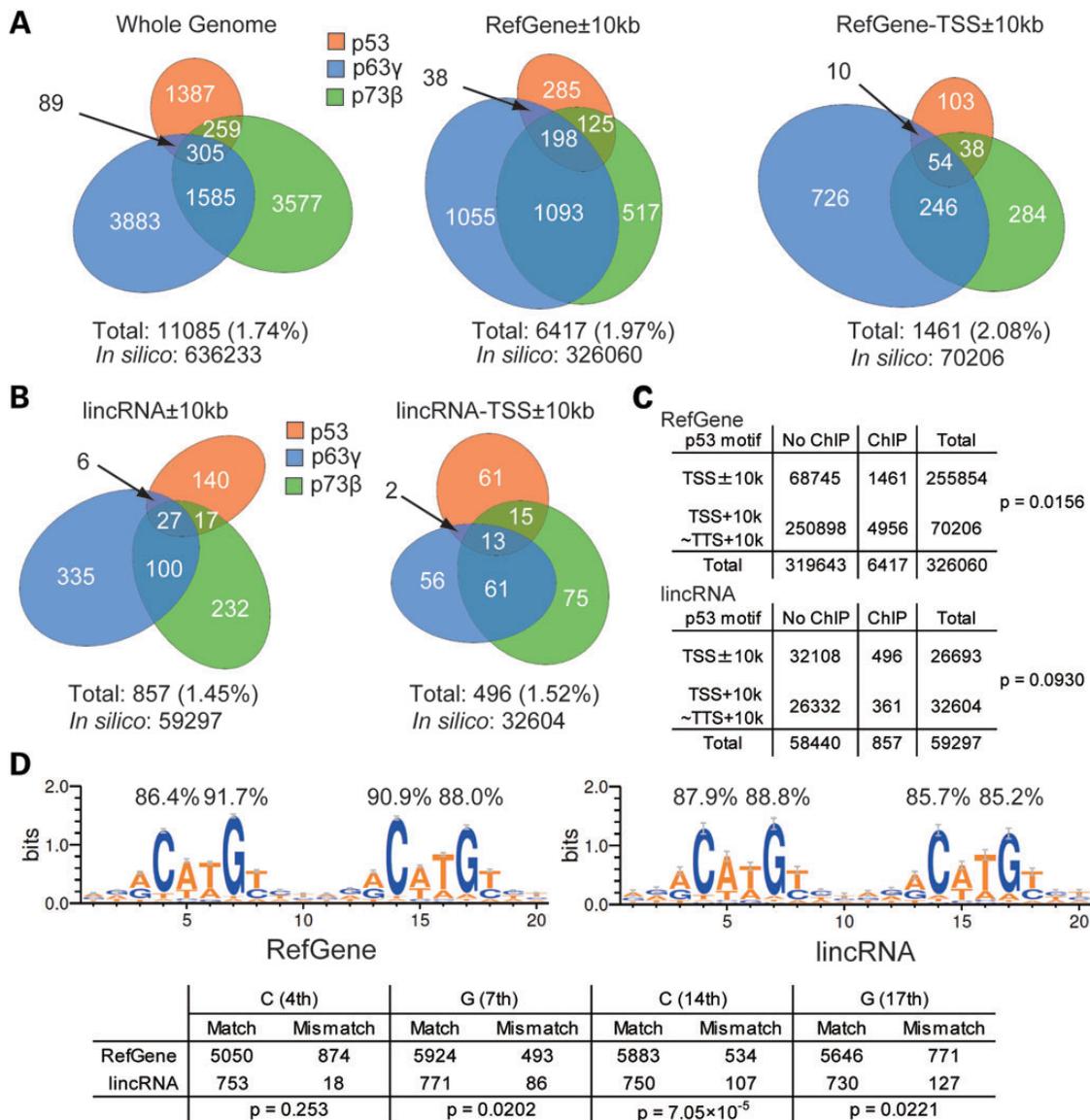
p73 proteins prefer to directly bind near coding genes, while half of the p53 protein binds in intergenic regions.

### Proportion of *in silico* p53 motifs included in ChIP-seq peaks among p53 family members in RefGenes and lincRNAs

Among all of the potential p53BSs in the human genome identified *in silico*, only 1.74% (11 085/636 233) were included in ChIP-seq peaks (Fig. 2A, left). The proportion of p53 motifs located in ChIP-seq peaks (ChIP-seq p53 motifs) increased to 1.97% (6417/326 060) when we focused solely on the p53 motifs found within  $\pm 10$  kb of NCBI RefSeq genes (RefGenes) (Fig. 2A, middle). Furthermore, when we investigated the p53

motifs that were identified *in silico* near a transcription start site (TSS)  $\pm 10$  kb, the proportion of ChIP-seq p53 motifs increased to 2.08% (1461/70 206) (Fig. 2A, right). This increase was statistically significant (Fig. 2C, top)

However, approximately half of all ChIP-seq peaks were located in intergenic regions (Fig. 1B). It was recently reported that many long non-coding RNAs are found in intergenic regions; these sequences are now annotated as lincRNAs (9). Therefore, we suspected that p53 family members might also transactivate these lincRNAs. We found 59 297 unique p53 motifs in lincRNA genes  $\pm 10$  kb through *in silico* analysis, 1.45% (857/59 297) of which were included in p53 family ChIP-seq peaks (Fig. 2B, left). As observed for RefGenes,



**Figure 2.** The number of ChIP-seq p53 motifs for different p53 family members and the motif character. (A) The numbers of p53 motifs included in ChIP-seq peaks for p53, p63 $\gamma$  and p73 $\beta$  in the whole human genome (left), RefSeq genes (RefGenes)  $\pm 10$  kb (middle) and within the TSS of RefGenes  $\pm 10$  kb (right) are displayed as Venn diagrams. (B) The numbers of ChIP-seq p53 motifs in lincRNAs  $\pm 10$  kb (left) and within the TSS of lincRNAs  $\pm 10$  kb (right) are also displayed as Venn diagrams. (C) The numbers of p53 motifs in TSS  $\pm 10$  kb and TSS + 10 kb  $\sim$  TTS + 10 kb. *P*-values were calculated using Pearson's  $\chi^2$ -test. (D) The relative frequencies of bases at each position of the p53BSs are reflected in the character height presented in this motif diagram, generated by WebLogo3 (<http://weblogo.threeplusone.com/>). The percentage indicates the frequency of matched bases at each position (top). *P*-values were calculated using Pearson's  $\chi^2$ -test (bottom).

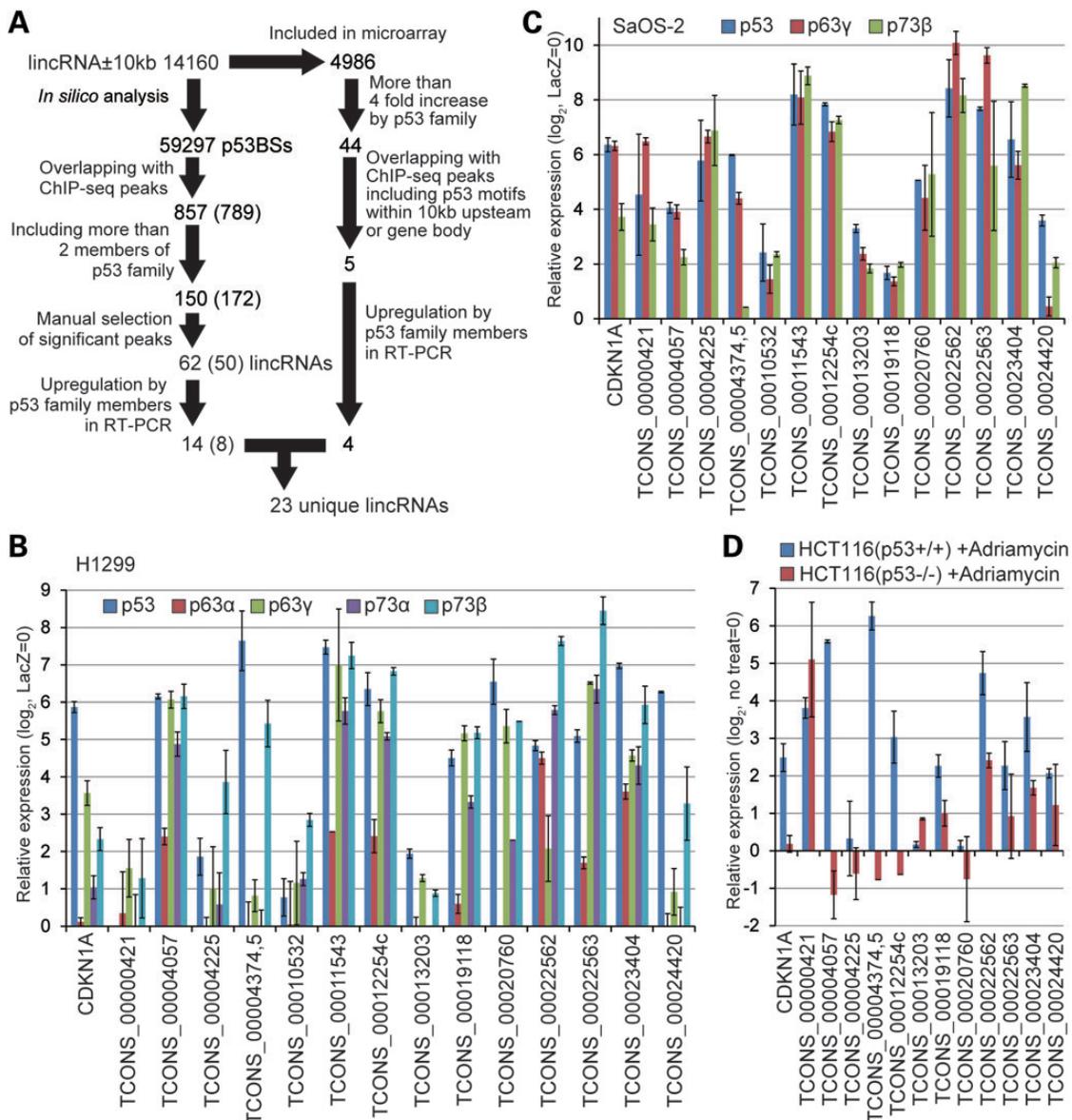
when we focused on the p53 motifs identified *in silico* within the TSS  $\pm$  10 kb of lincRNAs, we found that the percentage of p53 motifs included in the ChIP-seq peaks increased to 1.52% (496/32 604, Fig. 2B, right). These data indicate that p53 family members prefer to bind p53 motifs near TSSs in lincRNAs as well as RefGenes.

We further characterized the pattern of p53 motifs selected through ChIP-seq analysis in comparison with the consensus motif (RRRCWWGYYY RRRCWWGYYY). The consensus p53 motif includes two sets of core C and G sequences (i.e. C in the 4th and 14th positions, and G in the 7th and 17th positions), which are critical for the binding and transactivation of target genes. Therefore, mismatches of the core C and G sequences in the p53BSs of most target genes are rare. To clarify the

mismatches of core C and G sequences, we analyzed the incidence of bases at each position. We found that three core C and G sequences, the G at position 7, the C at position 14 and the G at position 17, were not as strictly conserved in the p53 motifs of lincRNAs as in the p53 motifs of RefGenes (Fig. 2D). This result indicates that the p53 transcriptional machinery and spectrum of target genes may differ between coding genes and lincRNAs.

### Transactivation of lincRNAs by p53 family members

Based on comprehensive analysis of p53BSs in the whole human genome, we sought to identify the novel lincRNAs targeted by p53 family members (Fig. 3A, left). *In silico*, we identified



**Figure 3.** Induction of lincRNAs by p53 family members. (A) Flowchart of lincRNA selection. At 24 h after infection with adenoviruses expressing LacZ (control), p53, p63 $\alpha$ , p63 $\gamma$ , p73 $\alpha$  and p73 $\beta$  at a multiplicity of infection (MOI) of 25, the expression of the indicated lincRNAs and the positive control CDKN1A were quantified via RT-qPCR in H1299 cells (B). The expression of the indicated lincRNAs was also quantified in SaOS-2 cells infected with LacZ (control), p53, p63 $\gamma$  and p73 $\beta$  at a multiplicity of infection (MOI) of 25 (C) and HCT116 (p53+/+) or (p53-/-) cells in the presence or absence of adriamycin (0.5  $\mu$ g/ml) for 24 h (D). The averages of three experiments are indicated in  $\log_2$ , with LacZ or no treatment = 0. Error bars indicate the SD. TCONS\_00012254c is a cluster of five lincRNAs.

59 297 unique p53 motifs among 14 160 lincRNAs  $\pm$  10 kb. Among these p53 motifs, 857 (1.45%) were captured by ChIP-seq for at least one p53 family member (Fig. 2B, left). We focused on 150 p53 motifs that were captured by ChIP-seq for more than two p53 family members. We infected the H1299 lung cancer and SaOS-2 osteosarcoma cell lines (both p53 null) with adenovirus vectors expressing p53 family members or LacZ as a control and compared the expression levels of 62 lincRNAs via conventional reverse transcription-polymerase chain reaction (RT-PCR) (Supplementary Material, Table S4). We found that the expression of 14 lincRNAs was significantly upregulated by p53, p63 $\gamma$  or p73 $\beta$  (Table 1 and Supplementary Material, Fig. S1). To salvage p53 family-targeted lincRNAs that escaped these screening conditions, we performed the same screen with high-sensitivity peak detection (Fig. 3A, left, in parentheses) and identified eight additional lincRNAs (Table 1; Supplementary Material, Fig. S1, Table S4).

Furthermore, we infected H1299 cells with adenoviral vectors expressing p53 family members, including p63 $\alpha$  and p73 $\alpha$ , or with LacZ as a control and compared lincRNA expression using a cDNA microarray including probes for 4986 lincRNAs (Fig. 3A, right). We identified 44 lincRNAs whose expression was increased by at least one p53 family member by >4-fold compared with LacZ (Supplementary Material, Table S5). Among these lincRNAs, five exhibited ChIP-seq peaks located either 10 kb upstream of a gene or within the gene body. However, three of the lincRNAs had been identified previously through p53-motif screening. As a result of the cDNA microarray screening, we identified one additional lincRNA as a

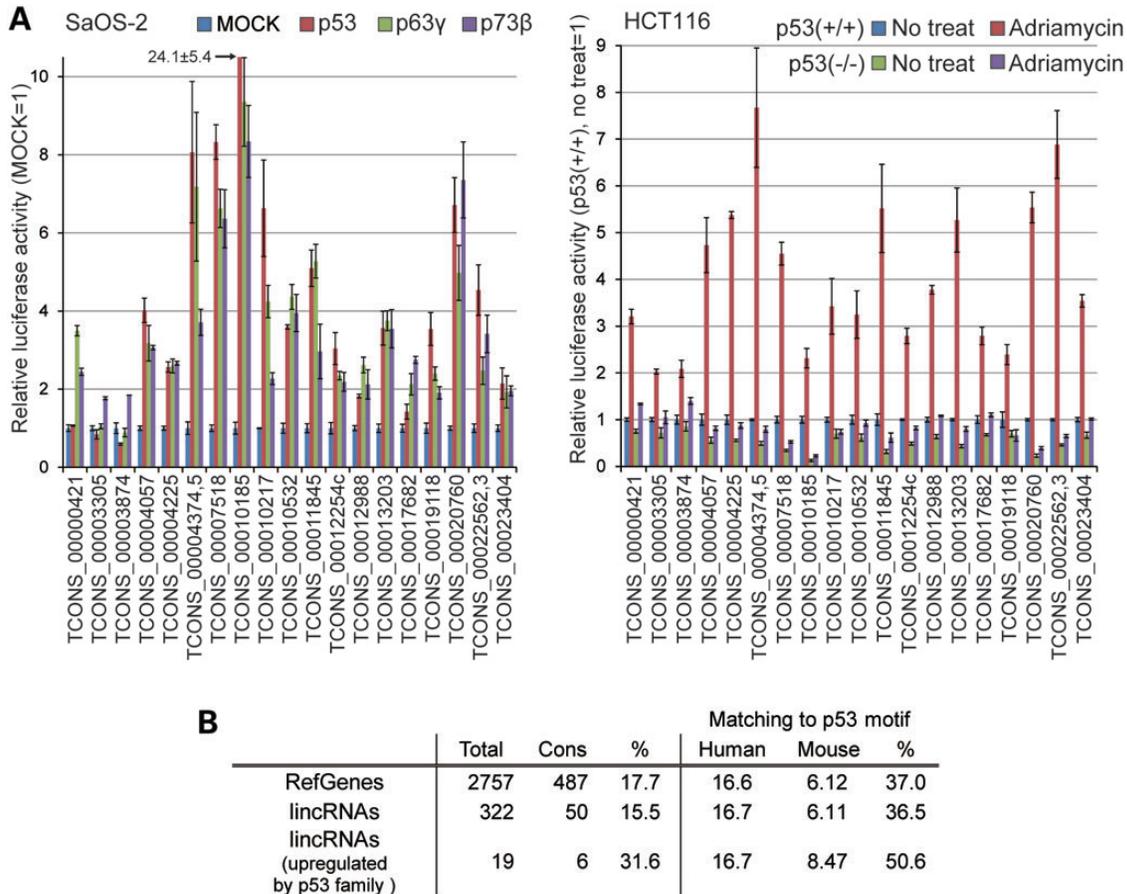
target of p53 family members (Table 1; Supplementary Material, Table S4). In these lincRNAs, the binding of p53 family members to the ChIP-seq p53 motifs was also confirmed via ChIP-PCR (Supplementary Material, Fig. S2, Table S6).

We then analyzed the expression of these lincRNAs through quantitative RT-PCR (RT-qPCR). We successfully validated the p53 family member-induced upregulation of 14 lincRNAs (Fig. 3B and C). Furthermore, we quantified the expression of these 14 lincRNAs in isogenic cancer cell lines (HCT116 (p53+/+) and HCT116 (p53-/-)) (Fig. 3D). Seven of the 14 lincRNAs were significantly upregulated following adriamycin treatment in HCT116 (p53+/+) compared with HCT116 (p53-/-) cells, demonstrating that this upregulation depends on p53. However, TCONS\_00000421 was upregulated to almost the same level in both the HCT116 (p53+/+) and (p53-/-) cell lines, indicating that its upregulation is likely dependent on p63 or p73, rather than p53, in these cells. Figure 3B also shows that the p53-induced transactivation of TCONS\_00000421 was weak in H1299 cells, which is consistent with our findings in isogenic HCT116 cells. The expression of other lincRNAs was not detected in HCT116 (p53+/+) or (p53-/-) cells. This lack of expression may be due to the specific tissue type used in the experiments, as some lincRNAs are expressed only in certain tissues (9). We also performed a reporter assay for 19 p53BSs identified among lincRNAs through ChIP-seq p53-motif screening in SaOS-2 and HCT116 cells (Fig. 4A). The transcriptional activity of most of the p53BSs examined in this reporter assay was significantly elevated by both the expression of p53 family members in SaOS-2 cells

**Table 1.** Characterization of p53BS sequences in lincRNAs transactivated by p53 family members

lincRNA	Chr	No. exon	Length of transcript	p53-motif sequences	Match	Core mismatch	Distance from TSS	Conserved	Screening	
TCONS_00000421	chr1	2	421	GGACAgGCTT	AAACAgGTTT	18	0	1464	–	S
TCONS_00003305	chr2	2	2928	GGGCTgGCCa	AGACcTGCCt	16	0	–426	–	H
TCONS_00003874	chr2	3	713	tAACTTGCCC	AAGCTTtCTC	18	1	–2798	–	S
TCONS_00004057	chr2	4	646	GGACgTGgTg	GtGCATGCCT	16	0	12666	–	S
TCONS_00004225	chr2	3	942	AAGCTTGTgC	AGACATGTaT	18	0	28892	16	S
TCONS_00004374	chr2	2	2671	tGcCTTGCaT	GGGCTTGTCT	17	0	–4686	–	S
TCONS_00004375	chr2	2	2347							S
TCONS_00007518	chr4	2	553	GGGCAAGTCg	GGACTTGagg	16	0	–8972	–	S
TCONS_00010185	chr5	2	377	GGACTgGTTC	AGGCATGCCa	18	0	13442	–	H
TCONS_00010217	chr5	2	508	tGACATGTAg	AGACcAGTaa	14	0	–9776	–	S
TCONS_00010532	chr5	2	277	GGACcTGCCC	AGACAgGCCg	17	0	–7460	18	H
TCONS_00011845	chr6	3	1209	GGACcTGCCt	AGACcTGCCC	18	0	788	–	H, M
TCONS_00011543	chr6	2	327					17438		S
TCONS_00012254	chr6	4	1859	GGACAAGCCC	AGGgATGgCa	17	1	16892	16	S, M
TCONS_00012988	chr7	3	230	GctCTTGCCC	ctACTTGTTc	16	0	16246	14	S
TCONS_00013203	chr7	3	510	tAACATGTCa	tGACAgGCCC	16	0	801	10	S
TCONS_00017682	chr10	4	1393	AAACTAGTgg	tGGCTTGCCC	17	0	5626	12	S
TCONS_00019118	chr11	2	1112	tAACTAGCag	AGACTTgATC	16	0	–1524	13	H
TCONS_00020760	chr12	5	2707	GAACATGCaC	cAGCTAGCCt	18	0	–528	15	S
TCONS_00022562	chr14	2	686	AGACATGTCC	AttcTGTCC	16	1	–222	16	H
TCONS_00022563	chr14	2	610					–5601		H, M
TCONS_00023404	chr15	2	535	AtACATGTTT	ctGCTTaTCC	16	1	–6259	15	H
TCONS_00024420	chr16	2	5163	AAACATGTCT	tAACATGTCT	19	0	–231	–	M

Chr, chromosome; no. exon, number of exons; p53-motif sequences, bases matching the consensus p53 motif (RRRCWWGYYY RRCWWGYYY) are displayed in upper case. Match: the number of nucleotides matching the consensus p53 motif. Core mismatch: the number of nucleotide mismatches in the four core sequences (C in the 4th and 14th positions and G in the 7th and 14th positions). Conserved: the number of nucleotides in the corresponding mouse genome regions matching the consensus p53 motif. Screening: the screening method used for identification. *In silico* p53 motif and ChIP-seq with stringent (S) or high-sensitivity (H) peak detection or cDNA microarray (M).



**Figure 4.** Transcriptional activity and sequence conservation of p53BSs in lincRNAs. (A) Relative luciferase activity of reporter vectors including a p53BS linked to the indicated lincRNA gene transfected into SaOS-2 cells (p53 null) with plasmids expressing MOCK, p53, p63γ and p73β (left). The relative luciferase activity of the same reporters is shown in HCT116 (p53+/+) or (p53-/-) cells with or without adriamycin treatment (0.5 μg/ml) (right). The averages of three experiments are presented; MOCK = 1. Error bars indicate SD. (B) The conservation of p53BSs between the human and mouse genomes is displayed. Total: the number of p53BSs that associated with more than two p53 family members in ChIP-seq. Cons: the number of p53BSs with a corresponding region of the mouse genome that matches >15 bases of the consensus p53 motif. Matching the p53 motif: the average number of matching nucleotides in human p53BSs (human) and the corresponding mouse genome region (mouse) of the consensus p53 motif.

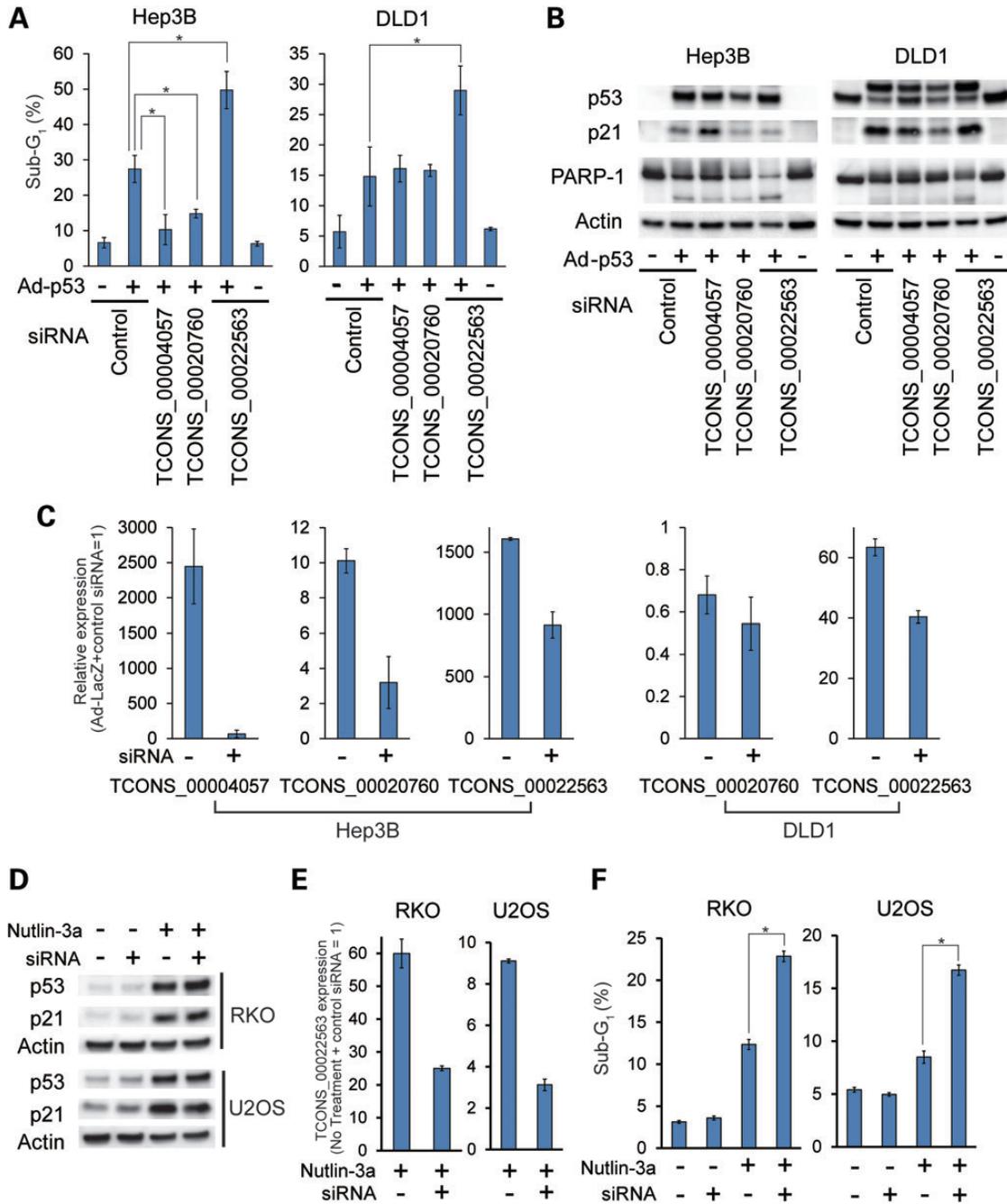
and adriamycin treatment in HCT116 (p53+/+) cells. The reporter activity of p53BSs showing one core mismatch (TCONS\_00003874, TCONS\_00012254c, TCONS\_00022562, TCONS\_00022563 and TCONS\_00023404) was relatively weak in SaOS-2 cells (Fig. 4A), while the induction of these lincRNAs was strong (Fig. 3B and C). This discrepancy suggests that this reporter assay does not always reflect the endogenous transcription of lincRNAs and that the transcriptional machinery used for lincRNAs may differ from that used for coding genes. This finding corroborates the observation presented in Figure 2D, i.e. that lincRNA p53 motifs exhibit more mismatches in core C and G sequences than do RefGenes.

Among the 19 p53BSs identified among lincRNAs through ChIP-seq p53-motif screening, 31.6% (6/19) were conserved between the human and mouse genomes. However, only 15.5% (50/322) of the p53BSs in lincRNAs captured by ChIP-seq for more than two p53 family members were conserved between the human and mouse genomes (Fig. 4B). The average number of consensus p53-motif nucleotide matches in the corresponding mouse genome was also higher in the p53BSs of lincRNAs upregulated by p53 family members than in other lincRNAs

(Fig. 4B). These results indicate that the p53BSs found among lincRNAs that are upregulated by p53 family members are highly conserved across species, supporting their biological importance.

### Modulation of apoptosis and gene expression by p53-induced lincRNAs

To investigate the biological function of p53-induced lincRNAs, we selected three lincRNAs that were transactivated by all p53 family members in both H1299 and SaOS-2 cells (Supplementary Material, Fig. S3) and performed siRNA knockdown of these lincRNAs. We used two cell lines, Hep3B (p53 null) and DLD1 (p53 mutant), that we had employed to evaluate p53-induced apoptosis in a previous study (20), and the level of p53-induced apoptosis was quantified by evaluating the sub-G<sub>1</sub> population in these cells (Fig. 5A). Interestingly, knockdown of TCONS\_00022563 significantly enhanced p53-induced apoptosis in both cell types. Two other indicators of apoptosis, PARP-1 cleavage and increased caspase-3 activity, were also observed in these cells (Fig. 5B and Supplementary Material,



**Figure 5.** Knockdown of lincRNAs modulates p53-induced apoptosis. (A) Hep3B and DLD1 cells were transfected with siRNAs targeting the indicated lincRNAs. At 4 h after transfection, these cells were infected with an adenovirus expressing p53 (Ad-p53: +) or LacZ (Ad-p53: -) as a control at an MOI of 100 (Hep3B) or 50 (DLD1). At 24 h after infection, the siRNAs were transfected into these cells again. The cells were analyzed via flow cytometry 24 h after the second transfection. The percentages of cells in sub-G<sub>1</sub> (the average of three independent experiments) are indicated. (B) Under the same conditions as in (A), the expression of the indicated protein was evaluated by western blotting. (C) Under the same conditions as in (A), lincRNA expression was quantified via RT-qPCR in Ad-p53-infected cells transfected with control siRNA (siRNA: -) or siRNAs targeting the indicated lincRNAs (siRNA: +). The averages of three experiments are indicated; Ad-LacZ + control siRNA = 1. Error bars indicate SD. RKO and U2OS cells were transfected with control siRNA (siRNA: -) or siRNAs targeting TCONS\_00022563 (siRNA: +). At 4 h after transfection, these cells were treated with 50 μM Nutlin-3a. At 24 h after treatment, the siRNAs were transfected into these cells again. At 24 h after the second transfection, the expression of the indicated protein was evaluated by western blotting (D), and the TCONS\_00022563 expression was quantified via RT-qPCR (E). Under the same conditions, the cells were analyzed via flow cytometry. The percentages of cells in sub-G<sub>1</sub> are indicated (F). In (A), (C), (E) and (F), the error bars indicate SD. Asterisk indicates *P*-values of <0.05 according to *t*-tests.

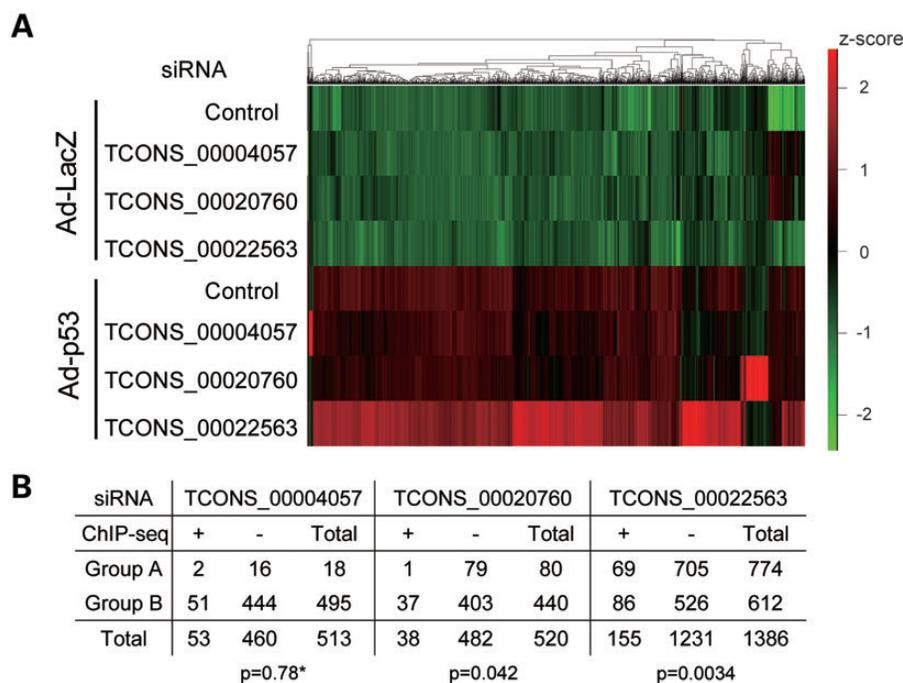
Fig. S4A). In contrast, knockdown of TCONS\_00004057 or TCONS\_00020760 significantly reduced the sub-G<sub>1</sub> population in Hep3B cells, but not in DLD1 cells. In each knockdown assay,

p53 was expressed at comparable levels (Fig. 5B). In Hep3B cells, all three lincRNAs were significantly upregulated by p53 expression and effectively knocked down by the corresponding

siRNAs (Fig. 5C, left). In DLD1 cells, TCONS\_00020760 was not upregulated by p53 expression; TCONS\_00004057 was undetectable, regardless of p53 expression, and TCONS\_00022563 was upregulated by p53 expression and knocked down by its siRNA (Fig. 5C, right). Thus, the observed changes in the expression of each lincRNA were consistent with the alteration of apoptosis levels shown in Figure 5A. Furthermore, we evaluated the level of apoptosis induction under conditions involving the activation of endogenous p53 by Nutlin-3a, which stabilizes the p53 protein through MDM2 inhibition in RKO colon cancer cells and U2OS osteosarcoma cells (wild-type p53 in both cell lines). The amount of p53 protein was increased by Nutlin-3a (Fig. 5D), and the inhibition of TCONS\_00022563 by siRNAs was confirmed via RT-qPCR (Fig. 5E). Knockdown of TCONS\_00022563 significantly enhanced p53-induced apoptosis in both cell types (Fig. 5F and Supplementary Material, Fig. S4B). These results strongly support a role for these p53-transactivated lincRNAs in modulating p53-induced apoptosis.

Although the diverse functions of lincRNAs are still being explored, several studies have revealed that lincRNAs regulate the transcriptional machinery (12,21). Therefore, we knocked down three lincRNAs and evaluated the subsequent changes in gene expression through microarray analysis. We selected 1551 genes whose expression was increased >4-fold in any of the p53-infected cells compared with the controls and performed

hierarchical clustering (Fig. 6A). Gene clusters that increased only following the knockdown of a given lincRNA when p53 was expressed (Ad-p53) were detected. Furthermore, knockdown of TCONS\_00022563 increased the total number of upregulated genes (from 750 in cells transfected with control siRNA to 1386). To confirm whether the upregulation of these genes was caused by direct p53 transactivation, the selected 1551 genes were divided into two groups: Group A showed a >2-fold increase in expression due to the presence of p53 (i.e. the difference between Ad-p53 and Ad-LacZ) when each lincRNA was knocked down, compared with the transfection of control siRNA; in Group B, the change was <2-fold. Additionally, we evaluated whether the genes in each group contained ChIP-seq p53 motifs within their gene body  $\pm 10$  kb (Fig. 6B; Supplementary Material, Table S7). Interestingly, the proportion of ChIP-seq p53 motifs (+), or probable direct p53 targets, was low in Group A (dependent on lincRNA knockdown) compared with Group B (independent of lincRNA knockdown); this difference was statistically significant for TCONS\_00020760 and TCONS\_00022563 knockdown (Fig. 6B). The representative p53 target genes p21 (CDKN1A) and MDM2 were classified into Group B (+) (independent of lincRNA knockdown and probable direct p53 targets) under the knockdown of all three lincRNAs (Supplementary Material, Table S7). To test the direct interaction between the p53 protein and these lincRNAs, we performed an RNA immunoprecipitation assay with a



**Figure 6.** Knockdown of lincRNAs modulates p53-induced gene expression. (A) Hep3B cells were transfected with siRNAs targeting the indicated lincRNAs. At 4 h after transfection, these cells were infected with adenovirus expressing p53 (Ad-p53) or LacZ (Ad-LacZ) as a control at an MOI of 100. At 24 h after infection, the cells were again transfected with the siRNAs. mRNA expression was analyzed using microarrays 24 h after the second transfection. Genes showing >4-fold increases in expression in p53-infected cells compared with controls were selected, and hierarchical clustering analysis was performed. The red and green colors in the heat map indicate positive and negative Z-scores, respectively. (B) Genes for which the expression was increased by >4-fold in p53-infected cells (Ad-p53) compared with corresponding control cells (Ad-LacZ) were evaluated in the context of lincRNA knockdown. The evaluated genes were divided into two groups: genes in Group A showed a >2-fold increase of expression due to p53 (i.e. the difference between Ad-p53 and Ad-LacZ) following the knockdown of each lincRNA compared with control siRNA, while the genes in Group B presented a change of <2-fold. ChIP-seq: the number of genes with (+) or without (-) ChIP-seq p53 motifs within their gene body  $\pm 10$  kb. *P*-values were calculated using Pearson's  $\chi^2$ -test with Yates' continuity correction. Asterisk indicates the number in Group A is too small to allow the calculation of an exact *P*-value using  $\chi^2$ -test.

FLAG antibody in FLAG-p53-overexpressing Hep3B cells, but we did not detect an interaction under standard experimental conditions (data not shown). These results suggest that knockdown of these lincRNAs induces the transactivation of a gene cluster (such as Group A (-) in Figure 6B) in a manner that is p53 dependent but does not rely on direct p53-induced transactivation. We speculate that these lincRNAs inhibit the expression of specific gene clusters that are potentially transactivated by p53 target genes or by other co-factors in coordination with p53 and that the knockdown of these lincRNAs reverses this inhibition of expression, resulting in the transactivation of the gene clusters and the modulation of apoptosis.

## DISCUSSION

In this study, we performed a ChIP-seq assay in H1299 cells overexpressing p53 family members to identify binding sites for p53 family members across the entire human genome. Although overexpression experiments may have some artificial effects compared with ChIP with endogenous proteins, we chose to use an overexpression system because broad coverage of binding sites was needed to allow the comprehensive detection of the genomic regions bound to p53 family members, and this system allowed us to cover physiological target genes whose binding affinity for p53 family members is too weak to be detected via ChIP using endogenous proteins. Furthermore, the exogenous expression of FLAG-tagged proteins and the use of the same anti-FLAG antibody enabled direct comparisons between p53 family members because of the comparable immunoprecipitation efficiency achieved.

However, although the overexpression system enables high-detection sensitivity, it also causes false-positive noise. Therefore, we classified the ChIP-seq peaks into groups with or without p53 motifs and focused on the peaks containing p53 motifs to exclude non-specific noise caused by the excess of proteins and the experimental process of immunoprecipitation. As a result, we identified more ChIP-seq peaks indicating the physiological-binding potential of p53 family members in this study compared with previous ChIP-seq studies using endogenous proteins (22,23). In addition, this classification enabled us to recharacterize the canonical motifs of p53BSs. Thus, we identified a difference in the motif patterns of the p53BSs of RefGenes and lincRNAs (Fig. 2D) that was not detectable through *de novo* motif analysis.

The conservation of each nucleotide in the p53 motifs was clearly different between the motifs found in RefGenes and lincRNAs. Interestingly, several lincRNAs were found to be strongly transactivated by p53 family members, despite their p53BSs showing a mismatch in the core C and G sequences (Fig. 2D). These results led to the speculation that the associated transcriptional machinery and chromatin modifications are different between RefGenes and lincRNAs. In fact, it has been reported that many non-coding RNAs are transcribed not only by RNA polymerase II, which transcribes most mRNAs, but also by RNA polymerase III (24).

We selected and analyzed lincRNAs that contained p53 motifs and were associated with more than two p53 family members via ChIP-seq. Interestingly, p53 transactivated several lincRNAs (Fig. 3) that presented p53BSs which were captured by

ChIP-seq using both p63 $\gamma$  and p73 $\beta$ , but not p53 (Table 1). We speculate that p53 binds to p53BSs less stably than p63 or p73 and tends to be lost in the process of ChIP-seq, resulting in false negatives. The fact that ChIP-seq p53 motifs captured using p63 $\gamma$  or p73 $\beta$  also frequently respond to p53 should be taken into consideration in the analysis of ChIP-seq data.

We showed that several lincRNAs were transactivated by p53 family members (Fig. 3) and that the knockdown of three lincRNAs induced the expression of a gene cluster (Fig. 6) resulting in the modulation of apoptosis (Fig. 5; Supplementary Material, Fig. S4). These results suggest that p53 transactivates several lincRNAs and that these lincRNAs then modulate the expression of their target genes, forming a complex transcriptional network. However, the precise mechanisms by which these lincRNAs modulate these functions remain unclear. One cancer-related lincRNA, HOX antisense intergenic RNA, mediates the silencing of the HOXD gene by recruiting the PRC2-LSD1 complex to the HOXD promoter (21). lincRNA-p21 is transactivated by p53 and suppresses several genes by binding to their promoter regions, along with the hnRNP-K protein (12). Therefore, the identification of proteins binding to the lincRNAs that we identified as targets of p53 family members in this study may provide important clues to their diverse functions.

We mainly analyzed lincRNAs that were transactivated by all three p53 family members (Figs 5 and 6). However, lincRNAs that were specifically transactivated by only one member of the p53 family remain to be analyzed (Fig. 2B). These lincRNAs may play unique roles corresponding to the specific functions of each p53 family member. Missense mutations in p53 not only result in loss of its normal functions in tumor suppression (loss of function) but also lead to the acquisition of new oncogenic functions (gain of function, GOF). The aberrant transactivation of genes that are not normally transactivated by wild-type p53 is one cause of GOF (2). We speculate that mutant p53 may also aberrantly transactivate several lincRNAs that are not targets of wild-type p53, resulting in gain of oncogenic functions.

## MATERIALS AND METHODS

### Cell culture

Human H1299 lung cancer, SaOS-2 osteosarcoma, DLD1 colorectal cancer and Hep3B hepatocellular cancer cells were purchased from the American Type Culture Collection and the Japan Collection of Research Bioresources. The HCT116 (p53 $+/+$ ) colon cancer cell line and its derivative HCT116 (p53 $-/-$ ) were kindly provided by Dr Bert Vogelstein (Howard Hughes Medical Institute, Johns Hopkins University, Baltimore, MD, USA). The construction, purification and infection of replication-deficient recombinant adenoviruses containing FLAG-tagged p53 (Ad-p53), TAp63 $\gamma$  (Ad-p63 $\gamma$ ) and TAp73 $\beta$  (Ad-p73 $\beta$ ) or the bacterial lacZ gene (Ad-LacZ) were described previously (25).

### *In silico* analysis of p53-responsive elements

We searched the whole human genome (hg18 assembly) for the canonical p53-binding sequence motif (p53 motif), RRR CWGYYY + spacer (0–13 nucleotides) + RRR CWGYYY. Motifs that met the criteria suggested by the p53scan algorithm

(14) were selected. NCBI Reference Sequences (RefSeq) and the Human lincRNA Catalog (BROAD Institute) (9) were used to annotate coding genes and lincRNAs, respectively.

### Chromatin immunoprecipitation and sequence analysis

H1299 cells were infected with Ad-p53, p63 $\gamma$  and p73 $\beta$  at an MOI of 25. At 24 h after infection, these cells were subjected to ChIP with a FLAG antibody, and fragment libraries were prepared from immunoprecipitated DNA using the SOLiD ChIP-Seq Kit (Life Technologies). These libraries were sequenced using the SOLiD4 system (Life Technologies), and the sequencing reads (50-bp) were obtained (number of reads: 36 089 861 for p53, 38 843 003 for p63 $\gamma$ , 43 406 760 for p73 $\beta$ , and 17 610 657 in the input sample). The sequencing reads were aligned to the human genome sequence (hg18 assembly) using the Bowtie program (26), with the option of no mismatches in the first 28 bases (alignment rates: 48.9% in p53, 45.8% in p63 $\gamma$ , 42.4% in p73 $\beta$  and 35.1% in the input sample). Peaks were detected using MACS (27) (Supplementary Material, Table S1). Under the high sensitivity of the applied conditions, peaks detected using USeq (28) and CCAT (29) were merged. The ChIP-seq data were deposited into the DDBJ sequence read archive (accession number: DRA000614).

### Antibodies and reagents

Adriamycin was purchased from Sigma. An anti-FLAG (M2) mouse antibody was obtained from Sigma. Anti-p53 (DO-1) and anti-p21 (F-5) mouse antibodies were procured from Santa Cruz Biotechnology. An anti-actin mouse antibody was purchased from Millipore, and an anti-PARP-1 mouse antibody was purchased from BD Pharmingen.

### Western-blot analysis

Total cell lysates were extracted at 4°C with RIPA buffer (150 mM NaCl, 1% NP40, 0.5% sodium deoxycholate, 0.1% SDS and 50 mM Tris-HCl, pH 8.0). The samples were fractionated via SDS-PAGE and transferred onto Immobilon-P membranes (Millipore). Immunoreactive proteins were detected using enhanced chemiluminescence (GE Healthcare).

### RT-PCR

Total RNA was prepared from cell lines using the RNeasy Mini Kit (Qiagen). For RT-PCR analysis, cDNA was synthesized from 5  $\mu$ g of total RNA with SuperScript III (Life Technologies). GoTaq (Promega) was used for PCR and quantitative PCR (qPCR). In the qPCR analyses, the mean value of three replicates was normalized using GAPDH. All primer sequences are described in Supplementary Material, Table S4.

### Luciferase reporter assay

The p53-motif  $\pm$  10 bp (Supplementary Material, Table S4) was subcloned into the pGL3-promoter plasmid (Promega). Cells were transiently transfected in triplicate with one of the luciferase reporters, phRG-TK (Promega) and the expression vector for FLAG-tagged p53, p63 $\gamma$ , p73 $\beta$  or MOCK inserted into

pCMV-Tag2-FLAG (Stratagene) using Lipofectamine 2000 (Life Technologies). Luciferase activity was measured with the Dual-Luciferase Reporter Assay System (Promega). Renilla luciferase activity was used as an internal control.

### siRNAs

Pairs of synthetic siRNAs targeting lincRNAs (TCONS\_00004057, TCONS\_00020760, TCONS\_00022563) were purchased from Sigma. Their sequences were as follows: TCONS\_00004057: GCGUCGUGUUCGUCUCUdTdT and AGACAGACGAACACGACGCdTdT, CAAUUUCUCUAUCAAUGUdTdT and ACAUUGAUAGGAGAAAUU GdTdT; TCONS\_00020760: GACAUUCAAGCAGCACUAU dTdT and AUAGUGCUGCUUGAAUGUCdTdT, GGUGUU AUGCUUAUUACAUDdTdT and AUGAAUAAGCAUAA-CACCDdTdT; TCONS\_00022563: CUAGGACCGUGGCAGG-CUUDdTdT and AAGCCUGCCACGGUCCUAGdTdT, CGGC CAAGACCAUAGACCUdTdT and AGGUCUAUGGUCUU GGCCGDdTdT. A universal-negative control siRNA was also purchased from Sigma.

### Flow cytometry

For flow cytometry,  $1 \times 10^6$  cells were plated into 60 mm plates. At 24 h after plating, the cells were transfected with siRNA using Lipofectamine RNAiMAX (Life Technologies). Four hours after transfection, the cells were infected with adenovirus; at 24 h after infection, the cells were again transfected with siRNA. The cells were analyzed in a FACSCalibur flow cytometer (BD Bioscience) 48 h after infection, as previously described (20). The experiments were repeated at least three times, and 50 000 events were examined for each sample. The resulting data were analyzed using FlowJo software (Tree Star).

### Microarray analysis

Total RNA was labeled with Cy3 and hybridized to a microarray (Agilent SurePrint G3 Human GE v2 for lincRNA screening and SurePrint G3 Human GE in Fig. 6) and then scanned with Agilent SureScan according to the manufacturer's protocols. The obtained data were normalized using Limma (R package). The expression levels were converted to Z-scores and subjected to hierarchical clustering based on the average Euclidean distance using gplots (R package). The microarray data were deposited in the NCBI Gene Expression Omnibus (accession number: GSE39773, GSE52739).

## SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* online.

## ACKNOWLEDGEMENTS

We thank Professor Satoru Miyano (Institute of Medical Science, University of Tokyo) for critical comments and constructive suggestions which have greatly improved this manuscript.

*Conflict of Interest statement.* None declared.

## FUNDING

This research was supported in part by Grants-in-Aid for Cancer Research from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

## REFERENCES

- Vogelstein, B., Lane, D. and Levine, A.J. (2000) Surfing the p53 network. *Nature*, **408**, 307–310.
- Brosh, R. and Rotter, V. (2009) When mutants gain new powers: news from the mutant p53 field. *Nat. Rev. Cancer*, **9**, 701–713.
- Riley, T., Sontag, E., Chen, P. and Levine, A. (2008) Transcriptional control of human p53-regulated genes. *Nat. Rev. Mol. Cell Biol.*, **9**, 402–412.
- Stiewe, T. (2007) The p53 family in differentiation and tumorigenesis. *Nat. Rev. Cancer*, **7**, 165–168.
- Cam, H., Griesmann, H., Beitzinger, M., Hofmann, L., Beinoraviciute-Kellner, R., Sauer, M., Huttinger-Kirchhof, N., Oswald, C., Friedl, P., Gattenlohner, S. *et al.* (2006) p53 family members in myogenic differentiation and rhabdomyosarcoma development. *Cancer Cell*, **10**, 281–293.
- Hibi, K., Trink, B., Patturajan, M., Westra, W.H., Caballero, O.L., Hill, D.E., Ratovitski, E.A., Jen, J. and Sidransky, D. (2000) AIS is an oncogene amplified in squamous cell carcinoma. *Proc. Natl. Acad. Sci. USA*, **97**, 5462–5467.
- Stransky, N., Egloff, A.M., Tward, A.D., Kostic, A.D., Cibulskis, K., Sivachenko, A., Kryukov, G.V., Lawrence, M.S., Sougnez, C., McKenna, A. *et al.* (2011) The mutational landscape of head and neck squamous cell carcinoma. *Science*, **333**, 1157–1160.
- He, L., He, X., Lim, L.P., de Stanchina, E., Xuan, Z., Liang, Y., Xue, W., Zender, L., Magnus, J., Ridzon, D. *et al.* (2007) A microRNA component of the p53 tumour suppressor network. *Nature*, **447**, 1130–1134.
- Cabili, M.N., Trapnell, C., Goff, L., Koziol, M., Tazon-Vega, B., Regev, A. and Rinn, J.L. (2011) Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev.*, **25**, 1915–1927.
- Guttman, M., Amit, I., Garber, M., French, C., Lin, M.F., Feldser, D., Huarte, M., Zuk, O., Carey, B.W., Cassady, J.P. *et al.* (2009) Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature*, **458**, 223–227.
- Gibb, E.A., Brown, C.J. and Lam, W.L. (2011) The functional role of long non-coding RNA in human carcinomas. *Mol. Cancer*, **10**, 38.
- Huarte, M., Guttman, M., Feldser, D., Garber, M., Koziol, M.J., Kenzelmann-Broz, D., Khalil, A.M., Zuk, O., Amit, I., Rabani, M. *et al.* (2010) A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. *Cell*, **142**, 409–419.
- Wei, C.L., Wu, Q., Vega, V.B., Chiu, K.P., Ng, P., Zhang, T., Shahab, A., Yong, H.C., Fu, Y., Weng, Z. *et al.* (2006) A global map of p53 transcription-factor binding sites in the human genome. *Cell*, **124**, 207–219.
- Smeenk, L., van Heeringen, S.J., Koepfel, M., van Driel, M.A., Bartels, S.J., Akkers, R.C., Denissov, S., Stunnenberg, H.G. and Lohrum, M. (2008) Characterization of genome-wide p53-binding sites upon stress response. *Nucleic Acids Res.*, **36**, 3639–3654.
- Hoh, J., Jin, S., Parrado, T., Edington, J., Levine, A.J. and Ott, J. (2002) The p53MH algorithm and its application in detecting p53-responsive genes. *Proc. Natl. Acad. Sci. USA*, **99**, 8467–8472.
- Maruyama, R., Aoki, F., Toyota, M., Sasaki, Y., Akashi, H., Mita, H., Suzuki, H., Akino, K., Ohe-Toyota, M., Maruyama, Y. *et al.* (2006) Comparative genome analysis identifies the vitamin D receptor gene as a direct target of p53-mediated transcriptional activation. *Cancer Res.*, **66**, 4574–4583.
- Sasaki, Y., Negishi, H., Idogawa, M., Yokota, I., Koyama, R., Kusano, M., Suzuki, H., Fujita, M., Maruyama, R., Toyota, M. *et al.* (2011) p53 negatively regulates the hepatoma growth factor HDGF. *Cancer Res.*, **71**, 7038–7047.
- Ohashi, T., Idogawa, M., Sasaki, Y., Suzuki, H. and Tokino, T. (2013) AKR1B10, a transcriptional target of p53, is downregulated in colorectal cancers associated with poor prognosis. *Mol. Cancer Res.*, **11**, 1554–1563.
- Sasaki, Y., Koyama, R., Maruyama, R., Hirano, T., Tamura, M., Sugisaka, J., Suzuki, H., Idogawa, M., Shinomura, Y. and Tokino, T. (2012) CLCA2, a target of the p53 family, negatively regulates cancer cell migration and invasion. *Cancer Biol. Ther.*, **13**, 1512–1521.
- Idogawa, M., Sasaki, Y., Suzuki, H., Mita, H., Imai, K., Shinomura, Y. and Tokino, T. (2009) A single recombinant adenovirus expressing p53 and p21-targeting artificial microRNAs efficiently induces apoptosis in human cancer cells. *Clin. Cancer Res.*, **15**, 3725–3732.
- Gupta, R.A., Shah, N., Wang, K.C., Kim, J., Horlings, H.M., Wong, D.J., Tsai, M.C., Hung, T., Argani, P., Rinn, J.L. *et al.* (2010) Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature*, **464**, 1071–1076.
- Smeenk, L., van Heeringen, S.J., Koepfel, M., Gilbert, B., Janssen-Megens, E., Stunnenberg, H.G. and Lohrum, M. (2011) Role of p53 serine 46 in p53 target gene regulation. *PLoS ONE*, **6**, e17574.
- Kouwenhoven, E.N., van Heeringen, S.J., Tena, J.J., Oti, M., Dutilh, B.E., Alonso, M.E., de la Calle-Mustienes, E., Smeenk, L., Rinne, T., Parsaulian, L. *et al.* (2010) Genome-wide profiling of p63 DNA-binding sites identifies an element that regulates gene expression during limb development in the 7q21 SHFM1 locus. *PLoS Genet.*, **6**, e1001065.
- Dieci, G., Fiorino, G., Castelnovo, M., Teichmann, M. and Pagano, A. (2007) The expanding RNA polymerase III transcriptome. *Trends Genet.*, **23**, 614–622.
- Oshima, Y., Sasaki, Y., Negishi, H., Idogawa, M., Toyota, M., Yamashita, T., Wada, T., Nagoya, S., Kawaguchi, S. and Tokino, T. (2007) Antitumor effect of adenovirus-mediated p53 family gene transfer on osteosarcoma cell lines. *Cancer Biol. Ther.*, **6**, 1058–1066.
- Langmead, B., Trapnell, C., Pop, M. and Salzberg, S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.*, **10**, R25.
- Zhang, Y., Liu, T., Meyer, C.A., Eeckhoutte, J., Johnson, D.S., Bernstein, B.E., Nussbaum, C., Myers, R.M., Brown, M., Li, W. *et al.* (2008) Model-based analysis of ChIP-Seq (MACS). *Genome Biol.*, **9**, R137.
- Nix, D.A., Courdy, S.J. and Boucher, K.M. (2008) Empirical methods for controlling false positives and estimating confidence in ChIP-Seq peaks. *BMC Bioinformatics*, **9**, 523.
- Xu, H., Handoko, L., Wei, X., Ye, C., Sheng, J., Wei, C.L., Lin, F. and Sung, W.K. (2010) A signal-noise model for significance analysis of ChIP-seq with negative control. *Bioinformatics*, **26**, 1199–1204.