Supplemental Material

TP53-inducible putative long non-coding RNAs encode functional polypeptides that suppress cell proliferation

Wenli Xu[1, 2,†], Chang Liu[1,†], Bing Deng[1,†], Penghui Lin[1,†], Zhenghua Sun[3], Anrui Liu[1], Jiajia Xuan[1], Yuying Li[3], Keren Zhou[1], Xiaoqin Zhang[4], Qiaojuan Huang[1], Hui Zhou[1], Qingyu He[3,*], Bin Li[1,*], Lianghu Qu[1,*], and Jianhua Yang[1, 5,*]

[1]MOE Key Laboratory of Gene Function and Regulation, State Key Laboratory for Biocontrol, School of Life Sciences, Sun Yat-sen University, Guangzhou, 510275, China; [2]Department of Infectious Diseases, the Third Affiliated Hospital of Sun Yat-sen University, Guangzhou, 510630, China; [3]Key Laboratory of Functional Protein Research of Guangdong Higher Education Institutes, Institute of Life and Health Engineering, College of Life Science and Technology, Jinan University, Guangzhou 510632, China; [4]School of Medicine, South China University of Technology, Guangzhou, 510006, China; [5]The Fifth Affiliated Hospital of Sun Yat-sen University, Zhuhai, 519000, China.

***Corresponding author. Email:**
yangjh7@mail.sysu.edu.cn
lssqlh@mail.sysu.edu.cn
libin73@mail.sysu.edu.cn
tqyhe@email.jnu.edu.cn

Table of Contents

Supplemental data:

Supplemental Fig S5. Identification of peptides encoded by TP53-regulated lncRNAs and their subcellular localization.

Supplemental Fig S6. Other unique peptides of TP53LC02 (A), TP53LC03 (B) and TP53LC04 (C) were identified using multiple reaction monitoring-mass spectrometry (MRM-MS) in HepG2-ADR^{+} cells.

Supplemental Fig S7. The genomic view of *TP53LC04* locus.

Supplemental Fig S8. Characteristics of the expression and location of TP53LC04.

Supplemental Fig S9. Supplementary evidence showing the selected lncRNA candidates and TP53LC04 are TP53-inducible.

Supplemental Fig S10. Alternative frameshift mutations showing that the TP53LC04 peptide, not its original lncRNA, functions in regulating cell proliferation.

Supplemental Fig S11. The TP53-inducible TP53LC02 peptide also participates in controlling cell proliferation.

Supplemental Fig S12. The RNA-seq analysis of stable TP53LC04-silenced HepG2 cells upon DNA damage shows the involvement of proliferation pathway.

Supplemental Tables:

Supplemental Table S1. Transcripts differentially expressed upon DNA damage treatment with ADR identified by the RNA-seq analysis.

Supplemental Table S2. TP53-regulated genes in response to DNA damage identified by the RNA-seq integrative analysis.

Supplemental Table S3. LncRNAs differentially expressed upon DNA damage treatment with ADR identified by the RNA-seq analysis.

Supplemental Table S4. All lncRNAs containing the translated ORFs upon DNA damage treatment with ADR identified by the Ribo-seq analysis.

Supplemental Table S5. 249 novel lncRNA-encoded ORFs upon DNA damage treatment with ADR identified by the Ribo-seq analysis.

Supplemental Table S6. List of 59 lncRNA candidates used in this study.

Supplemental Table S7. The results of the shotgun and PRM MS analysis on unique peptides of the new microproteins.

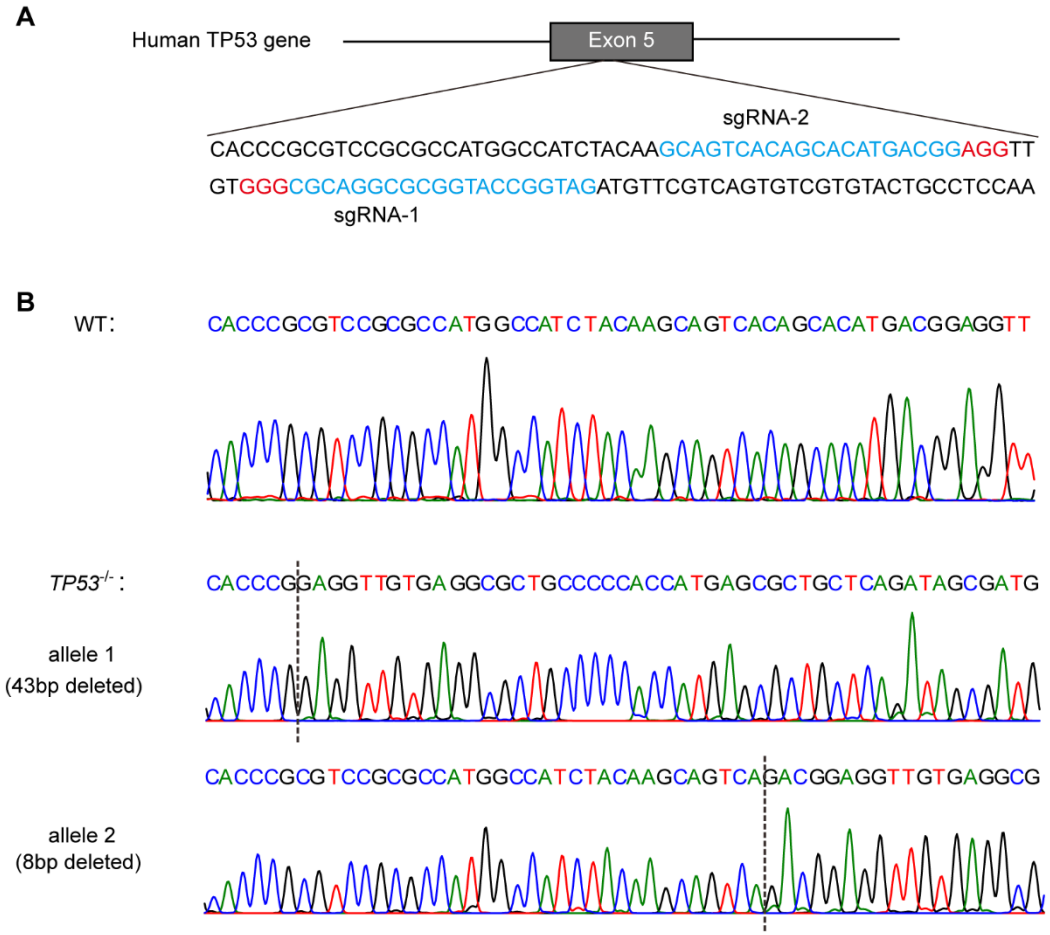
Supplemental Table S8. Overview of all MS-detected new microproteins.

Supplemental Table S9. Transcripts differentially expressed in stable TP53LC04-silenced HepG2 cells upon DNA damage treatment with ADR identified by the RNA-seq analysis.

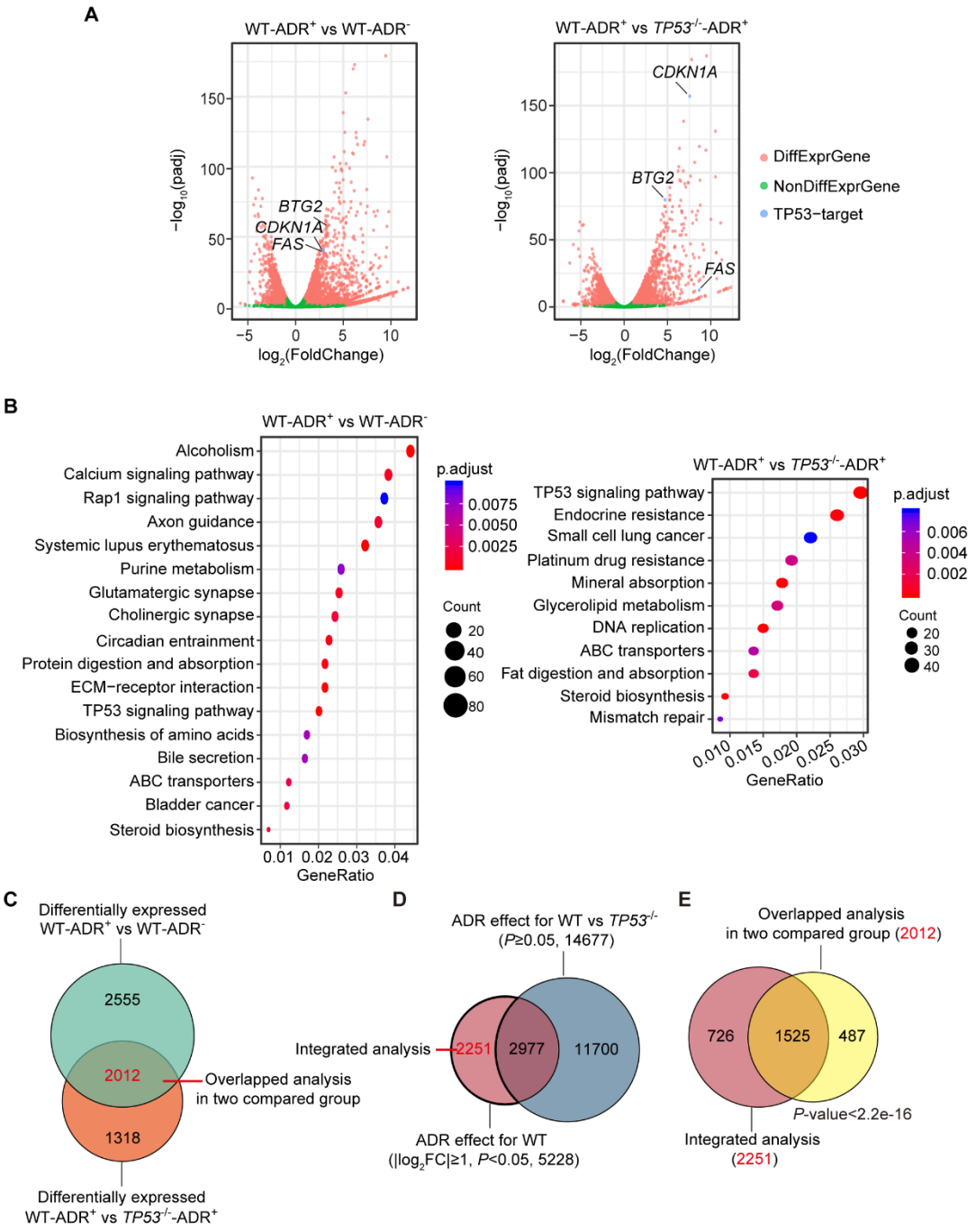
Supplemental Table S10. Sequences of primers and oligos used in this study.

Supplemental Table S11. List of the corresponding TP53LC gene ORFs and frameshift mutation sequences used in this study.

Supplemental Figures

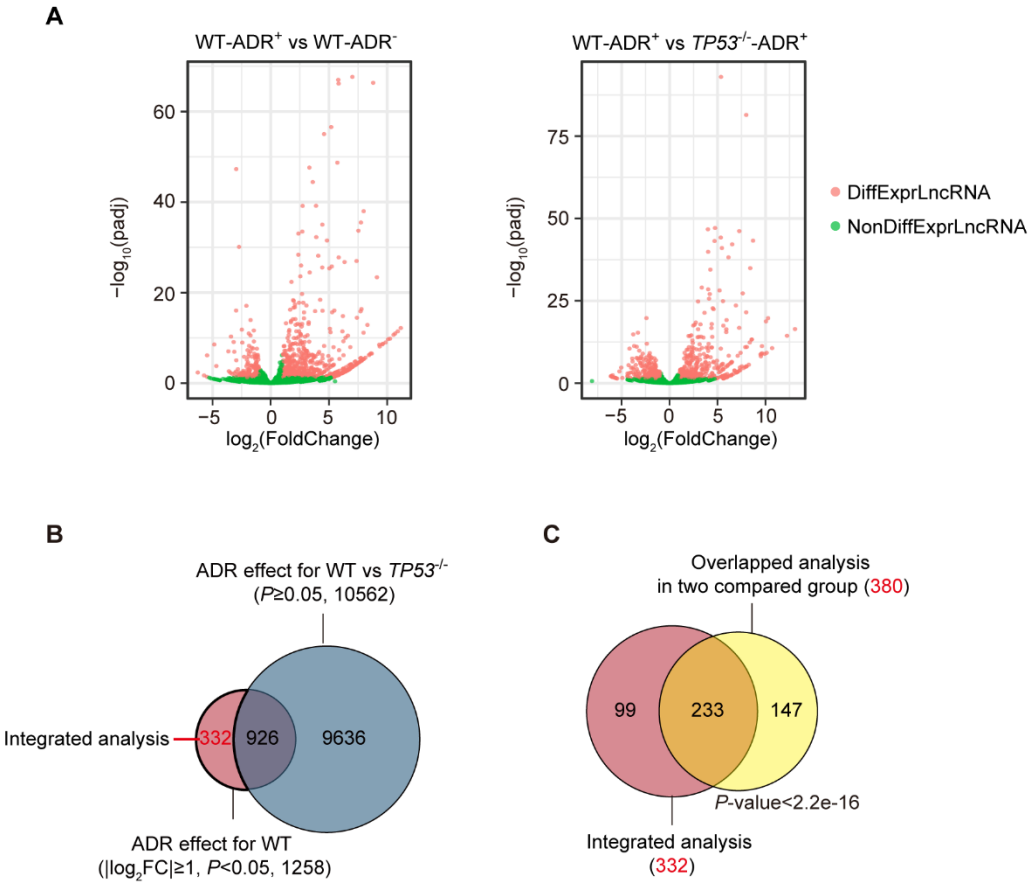


Supplemental Fig S1. Deletion of human TP53 gene in HepG2 cells using the CRISPR-Cas9 system. (**A**) The locus and sequences of two specific sgRNAs targeting the exon 5 of the human TP53 gene. Gray box, the exon 5 of the human TP53 gene; black line, intron; blue font, single guide RNA (sgRNA); red font, the PAM motif. (**B**) Sanger sequencing showing the deletion mutation of two TP53 alleles in HepG2$^{TP53-/-}$ cells with the corresponding sequences of the wild-type TP53 gene in HepG2 cells for comparison. The dashed line indicated the locus of the deleted sequences in two TP53 alleles of HepG2$^{TP53-/-}$ cells.

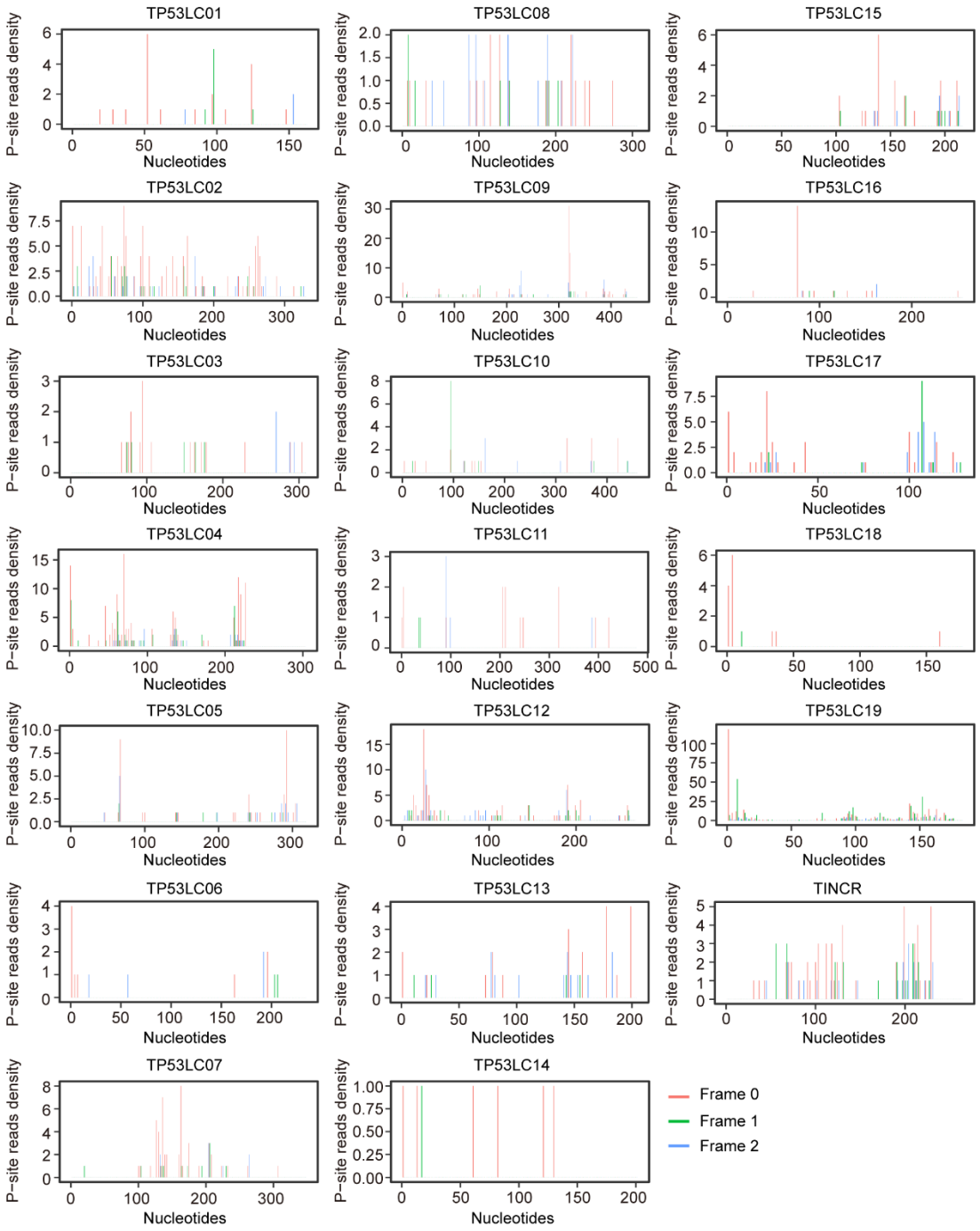


Supplemental Fig S2. The RNA-seq analysis of HepG2 cells and HepG2$^{TP53-/-}$ cells upon DNA damage shows the involvement of TP53 pathway. (**A**) The volcano plots showing the fold-change and significance of all the expressed protein-coding genes in the two comparison groups (left, WT-ADR$^+$ vs W-ADR$^-$; right, WT-ADR$^+$ vs *TP53*$^{-/-}$-ADR$^+$). The canonical TP53 targets (*CDKN1A*, *FAS* and *BTG2*) were pointed out with lines. The magenta spots and the green represent the differentially expressed protein genes and non-differentially expressed protein genes, respectively. The abscissa represents the difference between the two data sets following $\log_2$ conversion and the ordinate represents the -$\log_{10}$(*P*-value). (**B**) KEGG enrichment analysis for

differentially expressed protein-coding genes in the groups above. (**C**) Venn diagram showing the overlap of protein genes between the above two comparison groups. The overlap genes represented TP53-dependent protein genes in response to DNA damage. (**D**) Venn diagram showing the TP53-dependent protein genes in response to DNA damage by excluding ADR effect for WT vs *TP53*$^{-/-}$ set from ADR effect for WT set in the integrative analysis. (**E**) Comparison of the TP53-dependent protein genes in response to DNA damage identified by previous two comparison groups and integrative analysis. The *P*-value was calculated by Fisher's test.



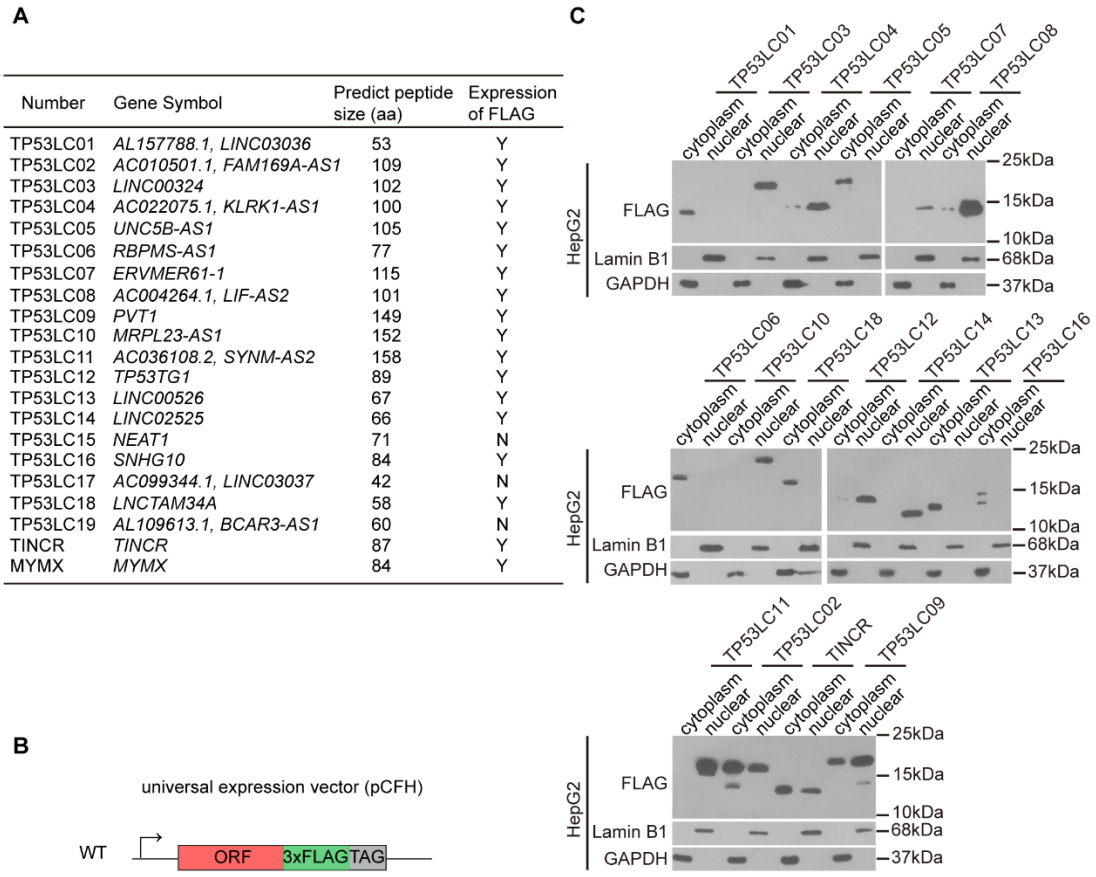
Supplemental Fig S3. The identification of TP53-dependent lncRNAs in response to DNA damage. (**A**) The volcano plots showing the fold-change and significance of all the expressed lncRNAs in the two comparison groups above. The magenta spots and the green represent the differentially expressed lncRNAs and nondifferentially expressed lncRNAs, respectively. (**B**) Venn diagram showing the TP53-dependent lncRNAs in response to DNA damage by excluding ADR effect for WT vs *TP53*$^{-/-}$ set from ADR effect for WT set in integrative analysis. (**C**) Comparison of the TP53-dependent lncRNAs in response to DNA damage identified by previous two comparison groups and integrative analysis. The *P*-value was calculated by Fisher's test.



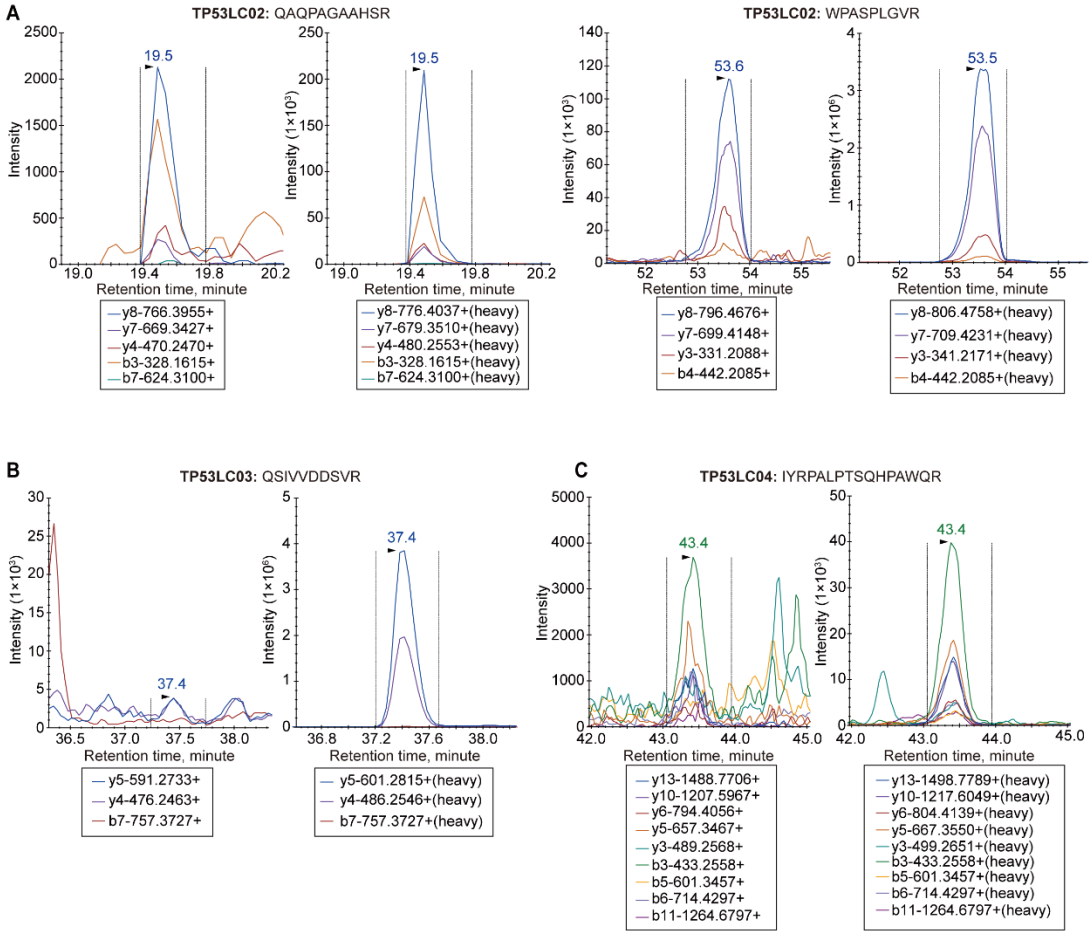
Supplemental Fig S4. The 3-nt periodicity of the translatable ORFs in lncRNAs. The distribution of 20 lncORFs for 3-nt periodicity.

A

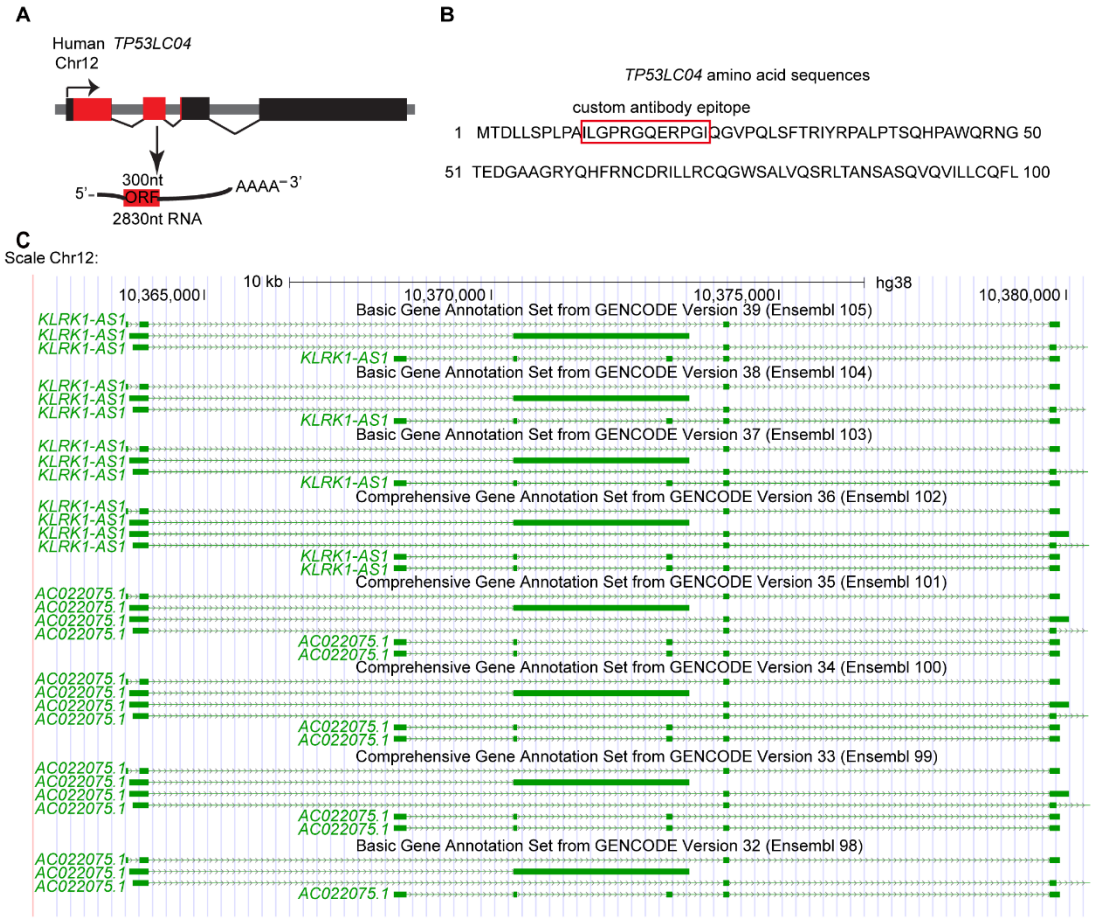
Number	Gene Symbol	Predict peptide size (aa)	Expression of FLAG
TP53LC01	*AL157788.1, LINC03036*	53	Y
TP53LC02	*AC010501.1, FAM169A-AS1*	109	Y
TP53LC03	*LINC00324*	102	Y
TP53LC04	*AC022075.1, KLRK1-AS1*	100	Y
TP53LC05	*UNC5B-AS1*	105	Y
TP53LC06	*RBPMS-AS1*	77	Y
TP53LC07	*ERVMER61-1*	115	Y
TP53LC08	*AC004264.1, LIF-AS2*	101	Y
TP53LC09	*PVT1*	149	Y
TP53LC10	*MRPL23-AS1*	152	Y
TP53LC11	*AC036108.2, SYNM-AS2*	158	Y
TP53LC12	*TP53TG1*	89	Y
TP53LC13	*LINC00526*	67	Y
TP53LC14	*LINC02525*	66	Y
TP53LC15	*NEAT1*	71	N
TP53LC16	*SNHG10*	84	Y
TP53LC17	*AC099344.1, LINC03037*	42	N
TP53LC18	*LNCTAM34A*	58	Y
TP53LC19	*AL109613.1, BCAR3-AS1*	60	N
TINCR	*TINCR*	87	Y
MYMX	*MYMX*	84	Y



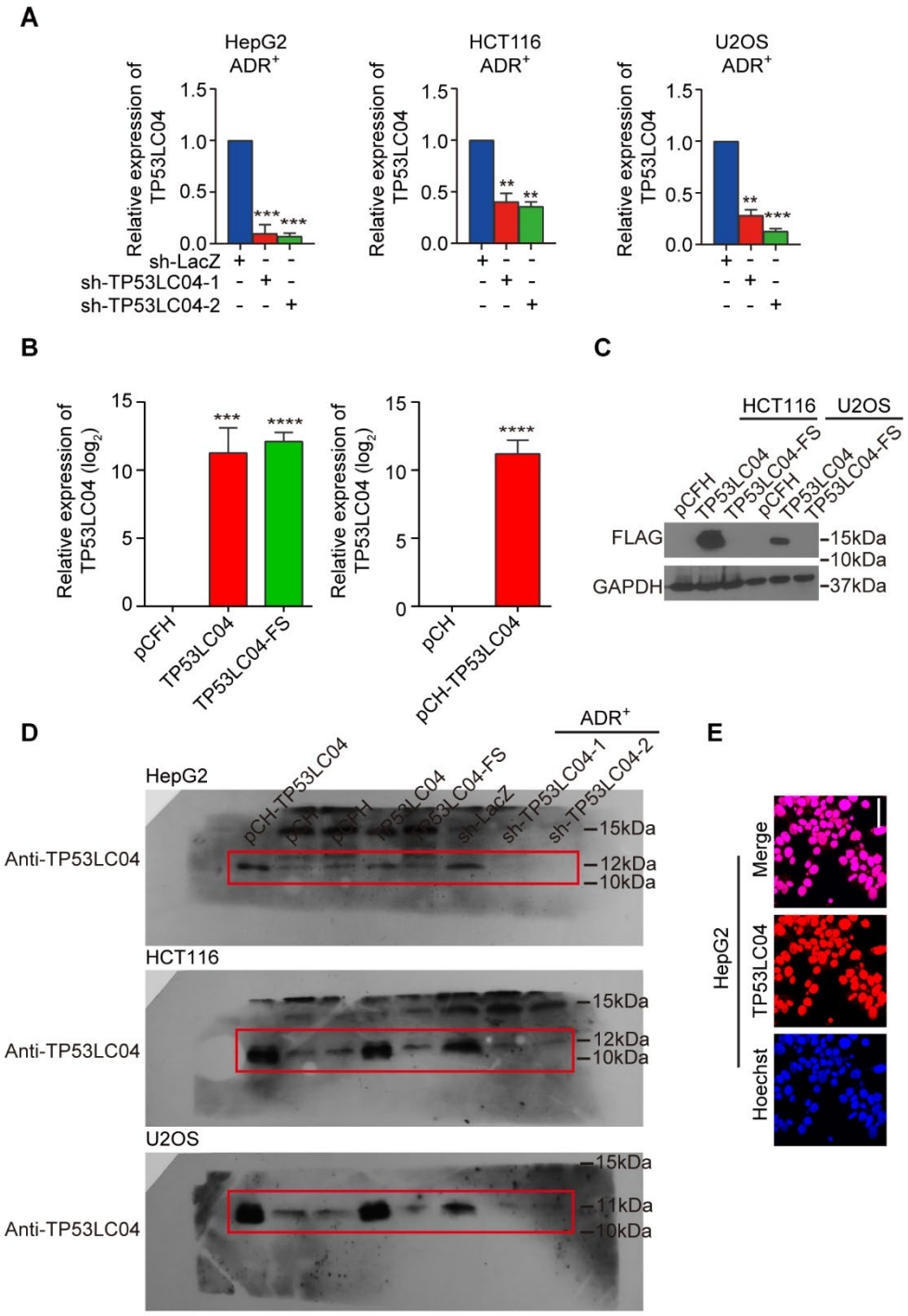
Supplemental Fig S5. Identification of peptides encoded by TP53-regulated lncRNAs and their subcellular localization. (**A**) The list of candidate TP53-regulated lncRNAs (TP53LCs). LncRNAs annotated by GENCODE v30 and HGNC were used in the list. The predicted peptide length and the coding ability of these annotated lncRNAs were also showed in the list. (**B**) Schematic diagram of the C-terminally 3 × FLAG-tagged vector (named as pCFH). (**C**) Subcellular localization of the verified peptides in HepG2 cells.



Supplemental Fig S6. Other unique peptides of TP53LC02 (**A**), TP53LC03 (**B**) and TP53LC04 (**C**) were identified using multiple reaction monitoring-mass spectrometry (MRM-MS) in HepG2-ADR^{+} cells. Mass chromatograms of the corresponding endogenous peptides are showed in left, and the corresponding heavy-peptides are indicated in right.

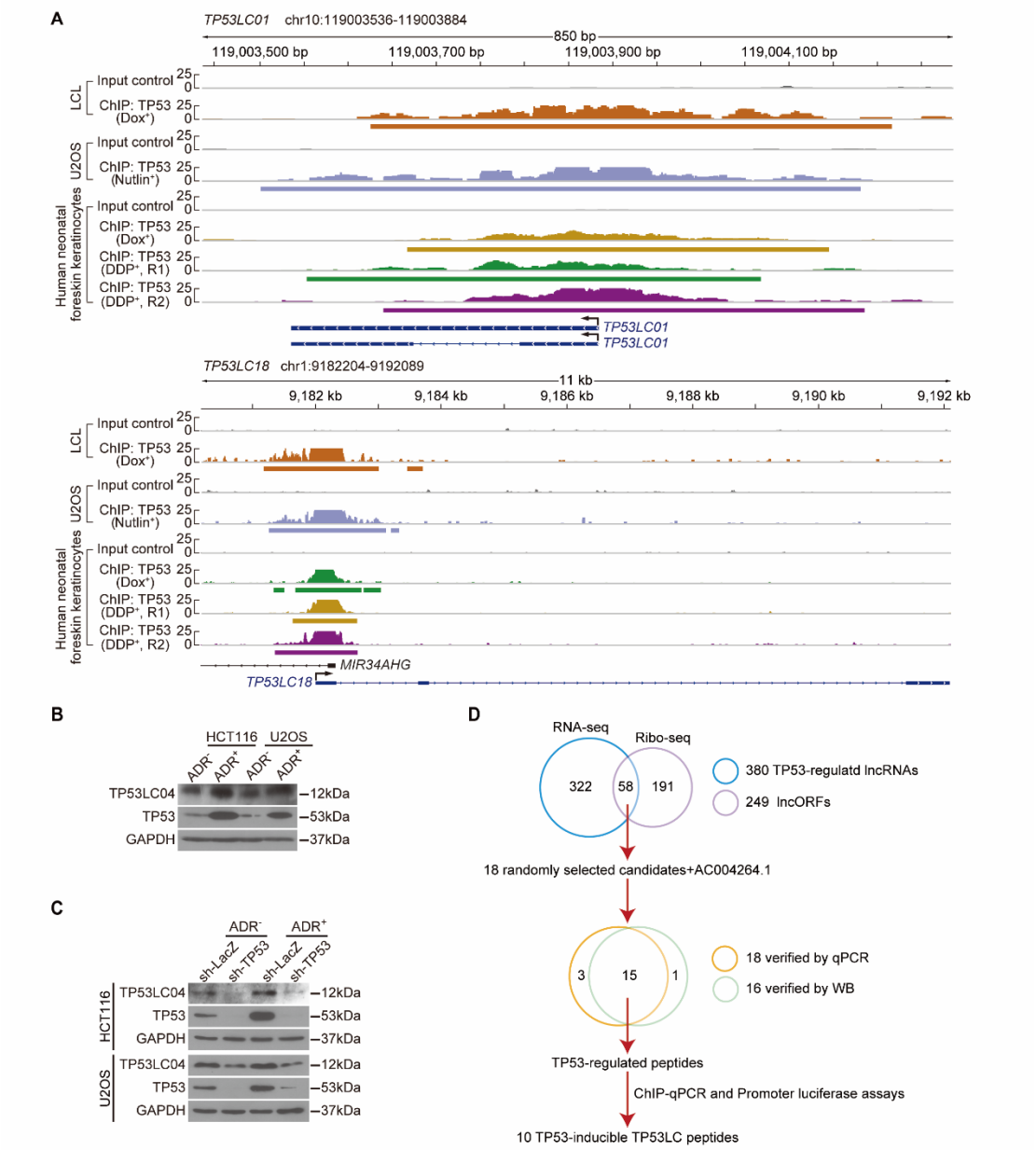


Supplemental Fig S7. The genomic view of *TP53LC04* locus. (**A**) The structure schematic of the transcript and ORF of the TP53LC04 peptide. The boxes represent exons and the lines represent introns or intergenic regions, the red box represents the ORF of the TP53LC04 peptide. And the length of the transcript is indicated below. (**B**) The amino acid sequences of the human TP53LC04 peptide and the customized antibody were generated against the sequences in the red box. (**C**) Genomic view of *TP53LC04* locus in different GENCODE versions.

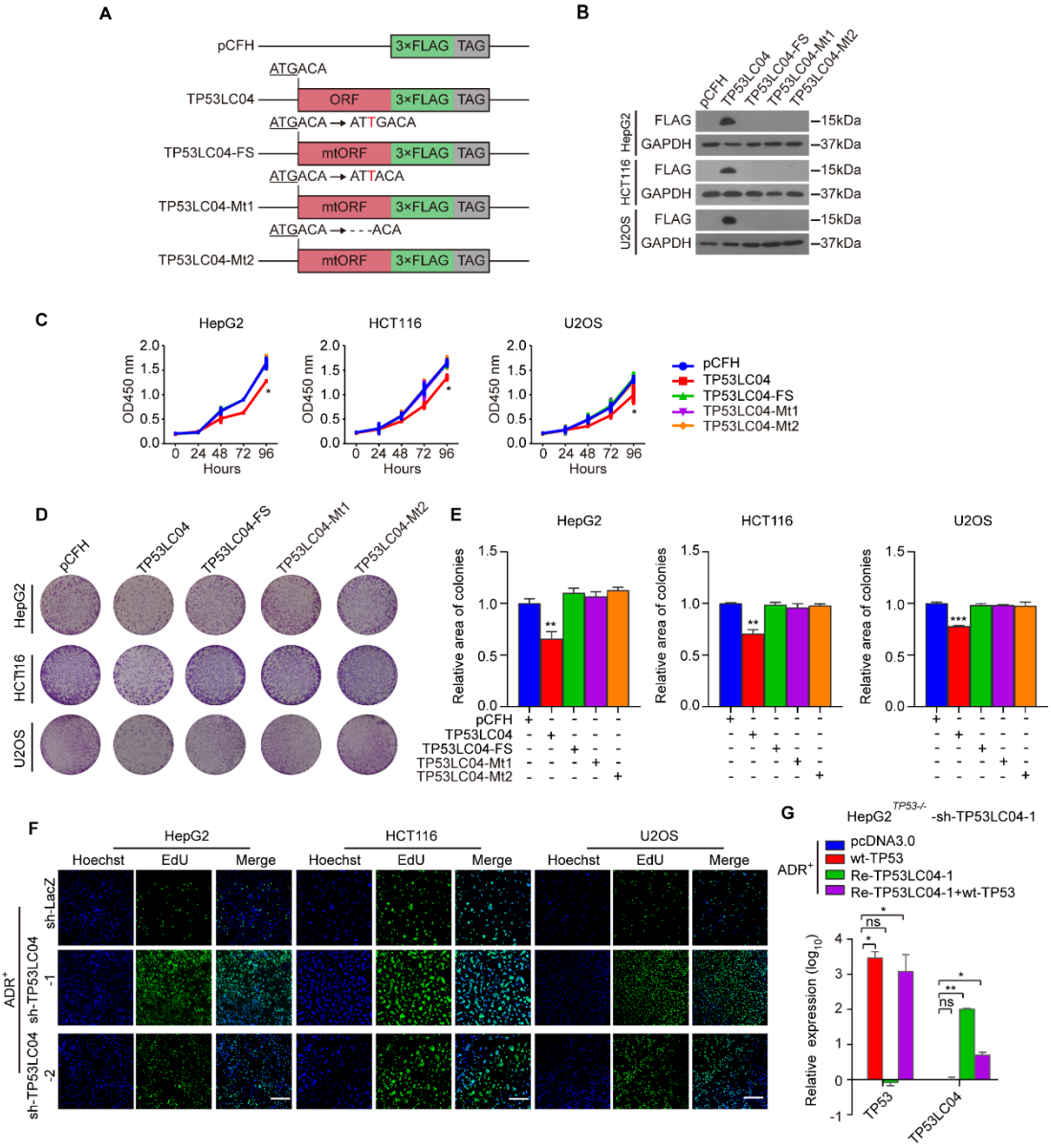


Supplemental Fig S8. Characteristics of the expression and location of TP53LC04. (**A**) The relative RNA expression of TP53LC04 detected by qRT-PCR in HepG2, HCT116 and U2OS cells that stably silenced TP53LC04 under the DNA-damage conditions. (**B**) The relative RNA expression of TP53LC04 detected by qRT-PCR in HepG2 cells overexpressing TP53LC04. (**C**) The protein expression of TP53LC04 detected by western blotting in HCT116 and U2OS overexpressing TP53LC04-3 × FLAG and its mutant vectors. (**D**) Full Western blot of the new antibody against TP53LC04. The red box represents the main band position of the TP53LC04 antibody. (**E**) The

immunofluorescence assays showing the subcellular location of the TP53LC04 peptide in HepG2 cells. HepG2 cells were immuno-stained using the anti-TP53LC04 antibody. Scale bar, 10 μm.

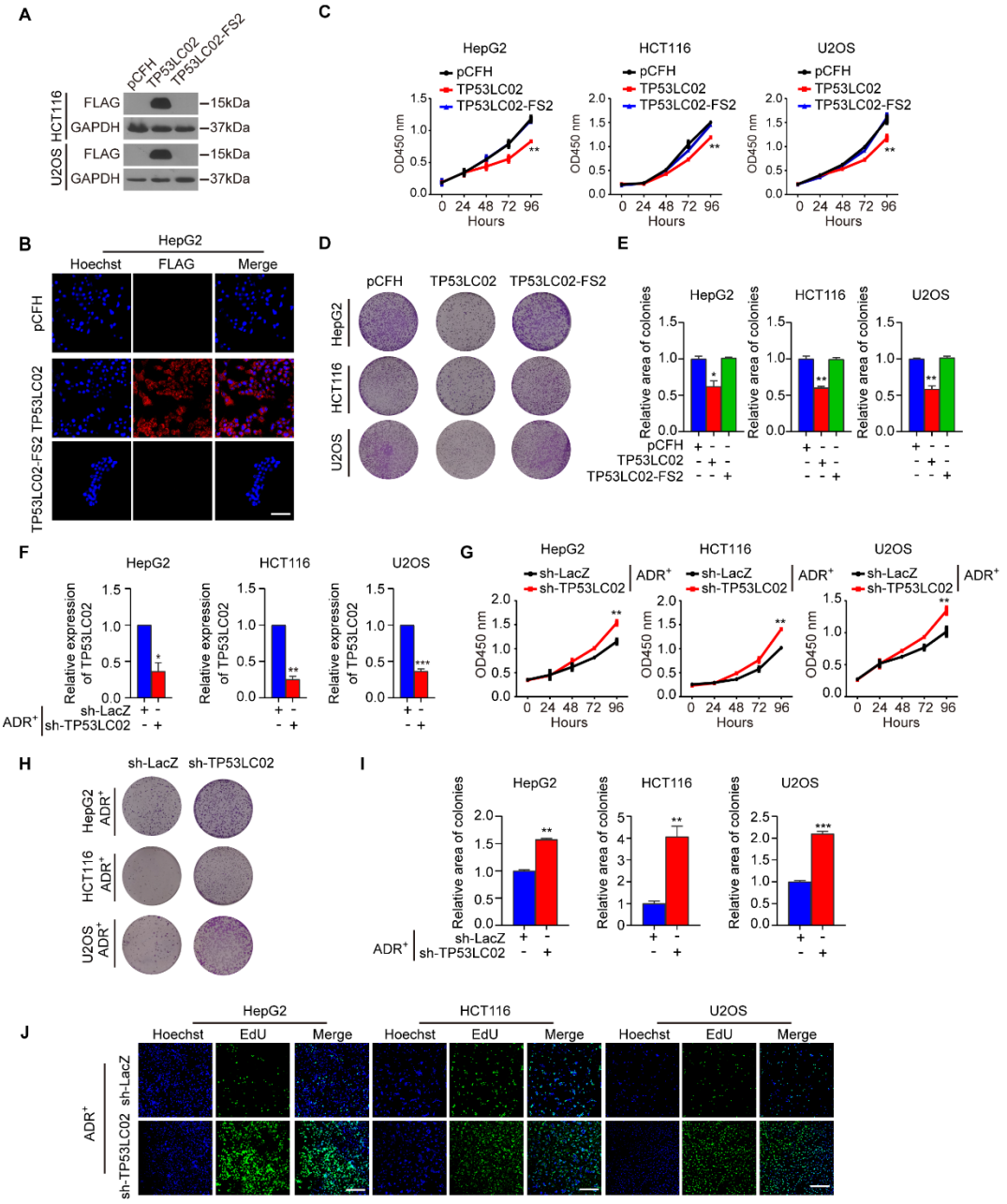


Supplemental Fig S9. The selected lncRNA candidates and TP53LC04 are TP53-inducible. (**A**) ChIP-seq tracks of normalized sequence tags show enrichment for TP53 along the *TP53LC01* (also known as *AL157788.1*) and *TP53LC18* (also known as *LNCTAM34A*) loci in relative treatment visualized on IGV software. The box under track represents the TP53 enrichment peak, and "Dox^{+}" represented the samples treated with doxorubicin (dox) also called adriamycin (ADR), and "DDP^{+}" represented the samples treated with cis-platinum. (**B**) The indicated protein levels in HCT116 and U2OS cells with or without ADR treatment. (**C**) The protein levels of TP53LC04 and TP53 in HCT116 and U2OS cells that stably knocked down TP53, with or without ADR treatment. (**D**) The diagram of the screening and validation process.



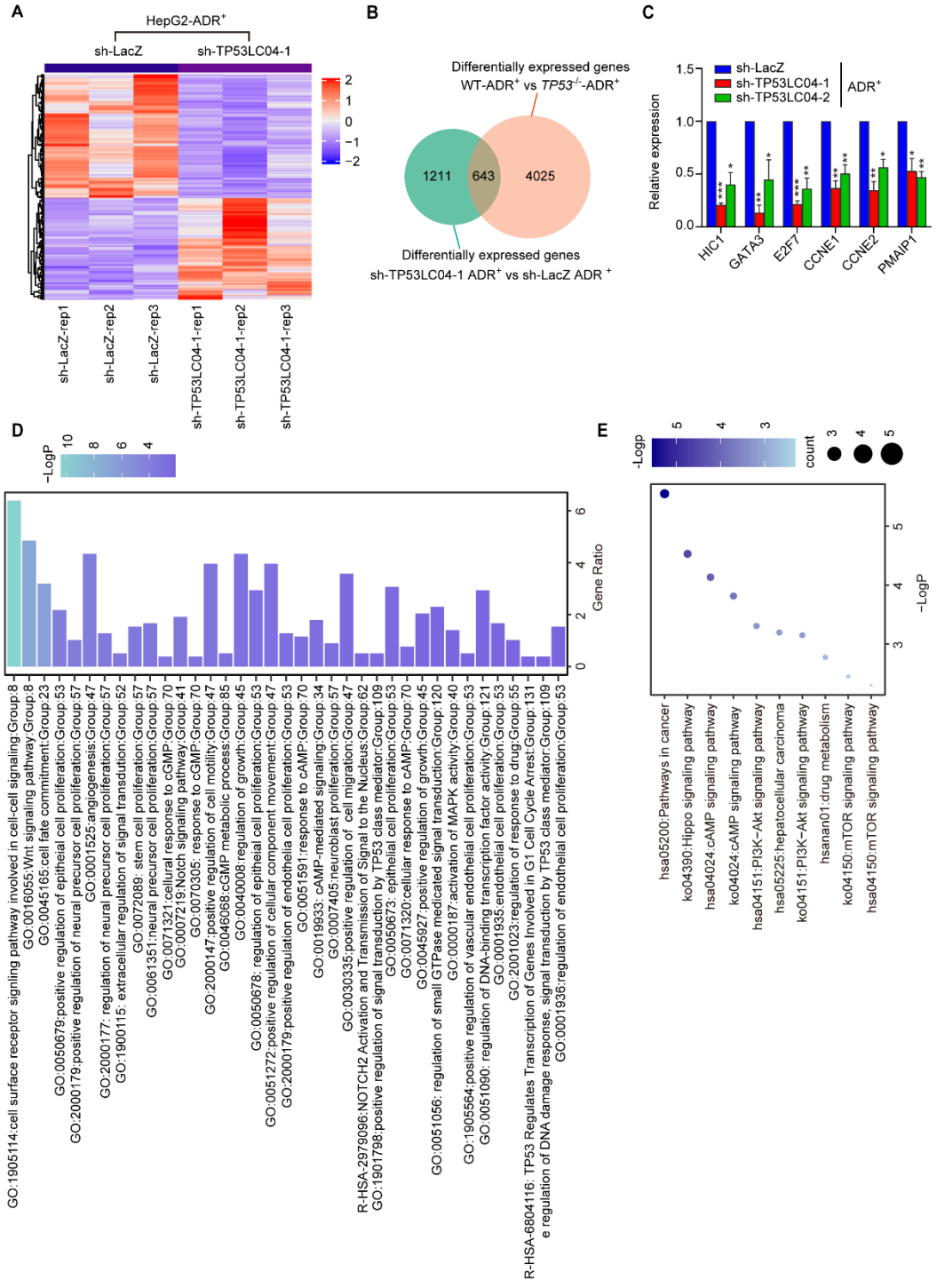
Supplemental Fig S10. Alternative frameshift mutations showing that the TP53LC04 peptide, not its original lncRNA, functions in regulating cell proliferation. (**A**) Schematic diagram of the 3 × FLAG-tagged vectors inserted with wild-type TP53LC04 ORF or three different frameshift mutants. (**B**) The protein levels of TP53LC04-3 × FLAG in various cell lines overexpressing wild-type and mutated vectors. (**C**) CCK-8 assays showing the cell viability in various cell lines transfected with the indicated vectors above. (**D-E**) Colony formation assays showing the abilities of colony formation in various cell lines transfected with the indicated vectors above. The representative pictures of colonies were showed in the (**D**) and the relative area of colonies was counted in the (**E**). This figure is an extended version of Figure 6, with two additional columns compared to the original figure. (**F**) EdU-based assay showing the cell proliferation of various cell lines stably knocked down TP53LC04, with ADR

treatment. 10 ×, Scale bar, 10 μm. The sh-LacZ control data presented here represent the shared control group from the TP53LC02 and TP53LC04 silencing experiments, both of which were conducted under identical conditions. The representative images shown were captured from three independent biological replicates. (**G**) The qPCR assays showing the expression of TP53 and TP53LC04 in sh-TP53LC04 HepG2$^{TP53-/-}$ restoring wild-type TP53 and sh-resistant TP53LC04, with ADR treatment. Data are represented as mean ± SEM. *$P < 0.05$, **$P < 0.01$ and ***$P < 0.001$.



Supplemental Fig S11. The TP53-inducible TP53LC02 peptide participates in controlling cell proliferation. (**A**) The protein levels of TP53LC02-3 × FLAG in HCT116 and U2OS cells overexpressing wild-type and mutated vectors. (**B**) The immunofluorescence assays showing the overexpression and subcellular location of the TP53LC02 peptide in HepG2 cells. HepG2 cells were transfected with the indicated vectors and then the TP53LC02 peptide was immuno-stained using the anti-FLAG antibody. Scale bar, 10 μm. (**C**) CCK-8 assays showing the cell viability in HepG2, HCT116 and U2OS cells transfected with the indicated TP53LC02 vectors. (**D-E**) Colony formation assays showing the abilities of colony formation in HepG2, HCT116 and U2OS cells transfected with the indicated TP53LC02 vectors. The representative pictures of colonies were showed in the (**D**) and the relative area of colonies was

counted in the (**E**). (**F**) The relative RNA expression of TP53LC02 detected by qRT-PCR in HepG2, HCT116 and U2OS cells that stably silenced TP53LC02 under the DNA-damage conditions. (**G**) CCK-8 assays showing the cell viability in HepG2, HCT116 and U2OS cells stably silenced TP53LC02 with the DNA-damage drug ADR treatment for 24 h. (**H-I**) Colony formation assays showing the abilities of colony formation in HepG2, HCT116 and U2OS cells stably silenced TP53LC02. The representative pictures of colonies were showed in the (**H**) and the relative area of colonies was counted in the (**I**). (**J**) EdU-based assay showing the cell proliferation of the above three ADR^+ cells that stably silenced TP53LC02. 10 ×, Scale bar, 10 μm. The sh-LacZ control data are shared with those shown in Figure S10F and were generated under identical experimental conditions.



Supplemental Fig S12. The RNA-seq analysis of stable TP53LC04-silenced HepG2 cells upon DNA damage shows the involvement of proliferation pathway. (**A**) The differentially expressed genes of the RNA-seq analysis in HepG2 cells that stable-silenced TP53LC04, compared with the control sample (sh-LacZ), both cells treated with 500 ng/mL ADR. (**B**) The Venn diagram showing the overlap of differentially expressed protein genes between two comparison groups (sh-TP53LC04-1-ADR^{+} vs sh-LacZ-ADR^{+}; WT-ADR^{+} vs *TP53*$^{-/-}$-ADR^{+}). (**C**) The selected differentially expressed

genes were determined by qPCR. (**D-E**) GO and KEGG pathway enrichment analysis for downregulated genes correspond to proliferation in the groups above.