Identify non-mutational p53 loss of function in human cancers

Qianpeng Li[1,2,3,*], Yang Zhang[1,2,3,*], Zhang Zhang[1,2,3], Ann L. Oberg[4], David E. Kozono[5], Hua Lu[6], Jann N. Sarkaria[7], Lina Ma[1,2,3,#], Liguo Wang[4,8,9,#]

[1] China National Center for Bioinformation, Beijing 100101, China
[2] National Genomics Data Center & CAS Key Laboratory of Genome Sciences and Information, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing 100101, China
[3] University of Chinese Academy of Sciences, Beijing 100101, China
[4] Division of Computational Biology, Mayo Clinic College of Medicine and Science, Rochester, MN 55905, USA
[5] Department of Radiation Oncology, Dana-Farber Cancer Institute and Brigham and Women's Hospital, Boston, MA 02215, USA
[6] Department of Biochemistry & Molecular Biology and Tulane Cancer Center, Tulane University School of Medicine, New Orleans, LA 70112, USA
[7] Department of Radiation Oncology, Mayo Clinic College of Medicine and Science, Rochester, MN, 55905, USA
[8] Department of Biochemistry and Molecular Biology, Mayo Clinic College of Medicine and Science, Rochester, MN, 55905 USA
[9] Bioinformatics and Computational Biology Graduate Program, University of Minnesota Rochester, Rochester, MN 55904, USA

*Equal contributions

Author email:
Qianpeng Li: liqianpeng2019d@big.ac.cn
Yang Zhang: zhangyang@big.ac.cn
Zhang Zhang: zhangzhang@big.ac.cn
Ann L. Oberg: Oberg.Ann@mayo.edu
David E. Kozono: dkozono@bwh.harvard.edu
Hua Lu: hlu2@tulane.edu
Jann N. Sarkaria: Sarkaria.Jann@mayo.edu

[#]Corresponding Authors:

Lina Ma, Ph.D.
Associate Professor
National Genomics Data Center
China National Center for Bioinformation / Beijing Institute of Genomics, Chinese Academy of Sciences
NO.1 Beichen West Road, Chaoyang District, Beijing 100101, China
Phone: 86-10-84097845
E-mail: malina@big.ac.cn

Liguo Wang, Ph.D.
Associate Professor
Division of Computational Biology
Mayo Clinic College of Medicine and Science
200 First Street SW Rochester, MN 55905, United States of America
Phone: 1-507-284-8728
Fax: 1-507-284-0745
E-mail: Wang.Liguo@mayo.edu

Abstract

An accurate assessment of *TP53* status is critical for cancer genomic medicine. Tumors with non-mutational p53 inactivations that failed to identify from DNA sequencing are largely misclassified as p53-normal, which leads to inaccurate prognosis and downstream association analyses. Here we built the support vector machine (SVM) models to systematically evaluate p53's functional status in multiple TCGA cancer cohorts. Cross-validation using independent samples demonstrated the excellent performance of the SVM models with mean AUC = 0.9845, precision = 0.9844 and recall = 0.9825. Our model revealed that most (87–100%) wild-type *TP53* (*TP53*^{WT}) tumors are loss-of-function. Further analyses revealed that these genetically normal but functionally impaired tumors (*TP53*^{WT}-pLoF) exhibit similar genomic characteristics as those p53 mutants with significantly increased tumor mutation burden, copy number variation burden, aneuploidy score, and hypoxia score. Clinically, compared to p53-normal tumors, patients with *TP53*^{WT}-pLoF tumors have significantly shortened overall survival and disease-free survival and exhibited increased sensitivity to platinum-based chemotherapy and radiation therapy. The increased sensitivity to radiation therapy is further validated in glioblastoma patient-derived xenografts models.

Introduction

Genomic variants identified from tumor DNA sequencing not only improved our understanding of cancer genetics (Alexandrov et al., 2013; Stephens et al., 2009; Vogelstein et al., 2013), but also revolutionized health care by providing diagnoses, informing prognoses, and guiding precision medicine. For example, breast cancers with loss-of-function mutations in *BRCA1*/*BRCA2* are sensitive to PARP inhibitors (Bryant et al., 2005; Farmer et al., 2005). Gain or loss of gene function due to genetic alterations is one of the main drivers of cancer development, and the identification of causal and targetable genetic aberrations remains a high priority in cancer genomic studies. However, genetic alteration is not the only mechanism to activate or disrupt gene function; alternative mechanisms besides direct genetic alterations include DNA methylation and RNA methylation (Agirre et al., 2003; Glodzik et al., 2020; Tian, Lai, Yu, Li, & Chen, 2020), histone methylation and acetylation (Zhang & Reinberg, 2001), chromatin topology and nucleosome positioning (Gorkin, Leung, & Ren, 2014; Jiang & Pugh, 2009), RNA editing (Ruan et al., 2022), RNA splicing (Surget, Khoury, & Bourdon, 2013), activation/inactivation of upstream regulators (Schoenfelder & Fraser, 2019), protein-protein interaction (Momand, Zambetti, Olson, George, & Levine, 1992; Teufel, Bycroft, & Fersht, 2009), and post-translational modifications such as phosphorylation, acetylation, ubiquitination, and SUMOylation (Ashcroft, Kubbutat, & Vousden, 1999; Igelmann, Neubauer, & Ferbeyre, 2019; Karve & Cheema, 2011). These molecular mechanisms are prevalent but remain a formidable challenge to characterize due to heterogeneous causes and a significant impediment for molecular diagnostics, clinical management, and therapeutic selection for cancer patients.

Tumor suppressor p53 (encoded by gene *TP53*) is a transcription factor that binds to its consensus DNA element and regulates downstream transcriptional programs, and thereby plays critical roles in preventing tumorigenesis and tumor progression by inducing cell cycle arrest, cell death (apoptosis, ferroptosis, necrosis, and autophagy), senescence, and DNA damage repair (Liu & Gu, 2022;

Mantovani, Collavin, & Del Sal, 2019; Vousden & Lane, 2007). *TP53* is one of the most extensively studied genes, and its transcriptional targets have been well characterized, including many that have been experimentally validated (Fischer, 2017). Meanwhile, it stands out as the most commonly mutated tumor suppressor gene that has been genetically inactivated in 20 cancer types (Jia & Zhao, 2019), and the somatic mutation frequencies are greater than 50% in ovarian, esophageal, pancreatic, lung, colorectal, uterine, head and neck, oral (gingivobuccal), soft tissue (leiomyosarcoma), gastric and biliary tract cancers (ICGC: https://dcc.icgc.org/genes/ENSG00000141510/mutations). *TP53* has become a promising therapeutic target for human cancers. It is found that the restoration of p53 function in established tumors causes tumor regression *in vivo*, which represents a new and promising therapeutic approach (Martins, Brown-Swigart, & Evan, 2006; Ventura et al., 2007). p53 activities have been frequently implicated in chemo- and radiation therapy (Bertheau et al., 2002; Bertheau et al., 2007; Concin et al., 2000; Lowe et al., 1994; Skinner et al., 2012; Tchelebi, Ashamalla, & Graves, 2014), and p53 dysfunction is also associated with immunosuppression and immune evasion (Cortez et al., 2016; Guo, Yu, Xiao, Celis, & Cui, 2017; Maddalena et al., 2021). Therefore, an accurate assessment of p53 functional status is critical for prognosis and personalized medicine.

The p53 protein is a potent roadblock to tumor development, therefore, it is generally considered that p53 function is inactivated even in genetically *TP53* wild-type tumors. Such concept is supported by sporadic studies (Bosari et al., 1995; Isaacs, Hardman, Carman, Barrett, & Weissman, 1998; Moll, LaQuaglia, Benard, & Riou, 1995; Moll, Riou, & Levine, 1992; Tominaga et al., 1993), but has not been evaluated systematically in large cohorts of clinical samples. In this study, we aim to predict p53 functional status from the expression profile of its transcriptional targets, particularly for those tumors with wild-type *TP53*. To this end, we analyzed the whole-exome sequencing (WES) and RNA-seq data generated from TCGA breast cancer (BRCA), lung cancer (LUNG), esophagus cancer (ESCA), and colon cancer (COAD) cohorts. We chose these four cancer types because of abundant cases of

TP53 truncating mutations that can be considered as the ground truth of loss of function (LoF) and used to train the SVM models.

We calculated the composite expression of curated p53 targets as a surrogate measure of p53 activity, and trained SVM models to predict p53's status. Specifically, we first cataloged cancer type-specific *TP53* targets through a systematic literature review and performed meta-analyses of p53 ChIP-seq and RNA-seq data. Then, we calculated the composite expression score (CES) from p53 activated and repressed genes using algorithms including GSVA (gene set variation analysis)(Hanzelmann, Castelo, & Guinney, 2013), ssGSEA (single sample gene set enrichment analysis) (Barbie et al., 2009), combined Z-score (E. Lee, Chuang, Kim, Ideker, & Lee, 2008; Tomfohr, Lu, & Kepler, 2005), and the first principal component of PCA (principal component analysis). Finally, we trained SVM models using CES scores from non-cancerous normal tissues (designated as the "NT" group in which p53 function is assumed to be normal) as well as tumor samples harboring *TP53* truncating mutations (designated as the "TM" group in which p53's tumor suppressor function is assumed to be lost). Ten-fold cross-validation demonstrated the excellent performance of our SVM models in discriminating NT and TM samples.

We re-evaluated p53 function status of TCGA tumors with the SVM models and compared genome and chromosome instability measurements to examine the prediction results. We found that almost all missense mutation samples (termed the "MM" group hereafter) were predicted to be LoF as expected. Most *TP53* wild-type samples (termed the "WT" group hereafter) were also predicted to be LoF (termed $TP53^{WT}$-pLoF), suggesting the functional p53 deficiency despite genetic intactness. Further analyses suggested these $TP53^{WT}$-pLoF tumors manifest distinctive genomic characteristics compared to other $TP53^{WT}$ samples that are predicted to be normal (termed $TP53^{WT}$-pN). $TP53^{WT}$-pLoF tumor genomes have dramatically increased tumor mutation burden (TMB), copy number variation burden,

aneuploidy score, and hypoxia score, consistent with p53's function as a DNA damage repair regulator and the "guardian of the genome".

Clinically, *TP53*^{WT}-pLoF patients have significantly shortened overall or disease-free survival compared to *TP53*^{WT}-pN patients. Also, our data demonstrate that *TP53*^{WT}-pLoF tumors exhibited significantly increased sensitivities to radio-and chemotherapy, providing useful information for personalized medicine. The increased sensitivity of *TP53*^{WT}-pLoF tumors to radiation therapy is further confirmed by the glioblastoma patient derived xenografts models (PDX).

Results

Identification of p53 target genes

Thousands of transcriptional targets of p53 have been reported (Fischer, 2017). Through systematic literature review, meta-analyses of the 32 p53 ChIP-seq datasets curated by the ReMap 2020 database (Cheneby et al., 2020), and the RNA-seq data of 54 tissues generated from the GTEx project, we identified 147 high-confidence transcriptional targets of p53 (**Figure 1 a**). These 147 genes were expressed across different tissues and showed p53 binding to their promoters (**Supplementary Table 1**). To further confirm these p53 target genes, we reanalyzed RNA-seq and ChIP-seq data prior to and after p53 activation in MCF-7 breast cancer cells harboring the wild-type *TP53*. Expectedly, there was no or very weak p53 binding at the promoters of p53 target genes (such as *ATF3*, *BTG2*) before p53 was activated by gamma irradiation, while prominent p53 bindings were observed, and the expression of these targets was consistently increased after p53 activation (Hafner et al., 2017) (**Supplementary Figure 1**).

p53 regulated transcriptional programs vary by cell type (Bouvard et al., 2000; Burns, Bernhard, & El-

Deiry, 2001; Karsli Uzunbas, Ahmed, & Sammons, 2019). To define cancer-specific p53 targets, we performed gene expression analyses between NT and TM groups (samples assigned to the testing set were excluded) using TCGA RNA-seq data and overlapped the identified differentially expressed genes (DEGs) with the previously selected 147 genes (**Figure 1 a**). We selected 47, 55, 36, and 50 p53 targets for BRCA, LUNG, ESCA, and COAD, respectively. 16 genes (*ABCB1*, *ANLN*, *BIRC5*, *CCNB1*, *CCNB2*, *CDC20*, *CDC25C*, *CDK1*, *CKS2*, *ECT2*, *FAM13C*, *NEK2*, *PCNA*, *PLK1*, *PMAIP1*, *PRC1*) were shared by all four cancer types. As p53 core targets, these 16 genes are significantly enriched in cell cycle control and DNA damage response pathways (**Supplementary Figure 2**).

Composite expression of p53 target genes

We interrogated the expression profiles of p53 target genes in four groups, including NT, WT, MM and TM. As expected, p53-activated and p53-repressed genes exhibited the opposite trends; p53-activated genes were significantly decreased in the TM group, whereas p53-repressed genes were significantly increased in the TM group, consistent with the p53 LoF status in this group (**Figure 1 b-d**). p53 target gene expression patterns were similar between the MM and TM groups (**Figure 1 b-d**), suggesting impaired p53 function in the MM samples, and that missense/in-frame mutations and truncating mutations seem to have similar impact on p53's cellular activity. This is consistent with the fact that most missense mutations were located within the DNA binding domain, disrupting the ability of p53 to bind to DNA and transactivate its downstream targets (Kato et al., 2003).

In the LUNG and BRCA cohorts, the WT group exhibited an intermediate state between the NT and TM groups, which suggests that p53 function might be compromised in majority of these samples, despite their intact DNA sequences (**Figure 1 b-c**). In the COAD cohort, the expression profile of the WT group was virtually indistinguishable from those of the mutant groups (MM + TM) (**Figure 1 d**). Collectively, these data indicated p53 lost its tumor suppressor function in $TP53^{WT}$ samples.

Training and evaluating SVM model based on the expression of p53 targets

We hypothesized that both mutational and non-mutational p53 inactivation could be predicted from the altered expression of its transcriptional targets. Toward this end, we calculated the composite expression score (CES) of p53 target genes using four algorithms, including GSVA (Hanzelmann et al., 2013), ssGSEA (Barbie et al., 2009), Z-score (E. Lee et al., 2008; Tomfohr et al., 2005), and PCA (**Supplementary Table 2**). Compared to the expression of individual genes, CES not only provides a combined and stable measure of p53 activity but also reduces the dimensionality of the expression data and mitigates potential overfitting of the SVM model.

Using LUNG as an example, we observed significant inverse correlations between CES scores calculated from p53-activated and p53-repressed genes; the Pearson's correlation coefficients were 0.996, -0.998, and -0.749 for GSVA, ssGSEA, and combined Z-score, respectively (**Supplementary Figure 3 a-c**). In addition, there were significant positive correlations (Pearson's correlation coefficients ranged from 0.874 to 0.996) among CES scores calculated by different algorithms, indicating high concordance of these algorithms (**Supplementary Figure 3 d-i**). When comparing the CES scores across the NT, WT, MM, and TM groups, we found CES scores of the NT samples were closely congregated in a narrow range. In contrast, CES scores calculated from the tumor samples (including WT, MM, and TM groups) showed a higher degree of dispersion, suggesting the increased heterogeneity of p53 activity in tumor samples (**Supplementary Figure 4**). Consistent with the gene-level expression data (**Figure 1 b**), CES scores of WT samples were intermediate between those of NT and TM, and the CES scores of MM samples resembled those of TM samples (**Figure 2 a-j**). These results again suggest that p53 function is compromised in most WT samples.

Next, we trained SVM models using CES scores from the NT group (assuming p53 is functionally normal) and TM group (assuming p53 is LoF) separately for the four TCGA cancer types, including

LUNG, BRCA, COAD, and ESCA. Ten-fold cross-validation demonstrated the excellent performance of these SVM models. As exemplified by LUNG cohort, the mean precision, recall, F1-score, and AUROC are 0.988 ± 0.013, 0.986 ± 0.02, 0.987 ± 0.016, and 0.986 ± 0.02, respectively (**Figure 2 k**). Our SVM model achieved a similar performance for the BRCA cohort (**Supplementary Table 3**). The smaller cohort sizes and particularly the limited number of truncating mutation cases in other TCGA cancer types discouraged us from training cancer-specific SVM models. Instead, we compiled a pan-cancer cohort (n = 5160) consisting of 9 cancer types with high *TP53* mutation rates (**Supplementary Figure 5**), including BLCA, BRCA, COAD, ESCA, HNSC, LUNG, LIHC, STAD, and UCES. Using the same procedures, we selected 27 shared p53 repressed genes and 32 shared p53 activated genes for this pan-cancer cohort (**Supplementary Figure 5 a and b**), calculated CES scores, and trained an SVM classifier using the pooled NT (n = 453) and TM (n = 882) samples (**Supplementary Figure 5 c and d**). Despite the intrinsic heterogeneity of different cancer types, our SVM model demonstrated outstanding performance with high precision (0.957 ± 0.093), recall (0.971 ± 0.018), macro-averaged F1-score (0.942 ± 0.080), and AUROC (0.988 ± 0.006) (**Supplementary Figure 5 e, Supplementary Table 3**).

Predicting p53 status using SVM models

We applied the trained SVM models to predict the p53 status of $TP53^{WT}$ and $TP53^{MM}$ tumors that had been excluded intentionally from the training and testing data. The results showed that most $TP53^{WT}$ samples (87%-100%) and almost all the $TP53^{MM}$ samples (98%-100%) were predicted to be LoF for all five cohorts, including LUNG, BRCA, COAD, ESCA, and pan-cancer (**Supplementary Figure 6**). For example, in the LUNG cohort, 94 % of the $TP53^{WT}$ samples and 100 % of $TP53^{MM}$ samples were predicted to be pLoF (**Figure 3 a)**. These results suggest the prevalence of p53 inactivation regardless the *TP53* genetic status, presumably by non-mutational mechanisms (Cross et al., 2011; Gogna, Madan, Kuppusamy, & Pati, 2012; Sasaki, Nie, & Maki, 2007) .

To demonstrate the prognostic value of the SVM prediction, we divided the *TP53*^{WT} tumors into two subgroups including *TP53*^{WT}-pLoF, which represents samples predicted to be LoF, and *TP53*^{WT}-pN, which represents samples predicted to be normal. We then compared the overall survival (OS) of patients from the LUNG cohort. The 10-year OS rates for patients with *TP53*^{WT}-pN (n = 20), *TP53*^{WT}-pLoF (n = 310), *TP53*^{MM} (n = 419), and *TP53*^{TM} (n = 254) lung cancers were 77.0%, 30.2%, 24.3%, 16.7%, respectively (**Figure 3 b**). Therefore, the survival rate of lung cancer patients with *TP53*^{WT}-pLoF tumors was dramatically decreased by 61% compared to patients with *TP53*^{WT}-pN tumors after adjusting for demographic variables including sex, age, and smoking status ($P = 0.016$, Cox regression). In contrast, the overall survival of patients with *TP53*^{WT}-pLoF tumors showed no difference from that of patients with *TP53*^{MM} or *TP53*^{TM} tumors. For the BRCA cohort, 84% (71 out of 85) of the *TP53*^{WT}-pN tumors are classified as Luminal-A subtype, but we did not observe a significant difference in progression-free survival (PFS) of patients with these tumors. The remaining 15% (13 out of 85) of *TP53*^{WT}-pN cases are normal-like breast tumors, among which the *TP53*^{WT}-pLoF patients also exhibited significantly reduced prognosis compared to the *TP53*^{WT}-pN group ($P = 0.017$, log rank test) (**Supplementary Figure 7 f**). Normal-like tumors were generally considered as “artifacts” of having high percentage of normal specimens or slow-growing basal-like tumors (Parker et al., 2009). Our data indicated the PFS is significantly reduced when p53 is inactivated, thus supporting the normal-like as a distinct tumor subtype rather than the normal tissue contamination. In summary, in both lung and breast (normal-like) cancers, *TP53*^{WT}-pLoF tumors exhibit a significantly worse prognosis compared to the *TP53*^{WT}-pN tumors.

Since p53 plays a prominent role in DNA damage repair (Lane, 1992; Levine, 1997), the p53-defective cells would accumulate more DNA damages than p53 wild-type cells. To test this hypothesis, we compared genome and chromosome instability measurements, including TMB, copy number variation burden, and aneuploidy score, between the *TP53*^{WT}-pLoF and *TP53*^{WT}-pN groups. We found the

genome instabilities were indeed significantly increased in the *TP53*^{WT}-pLoF group compared to the *TP53*^{WT}-pN group, suggesting deficient DNA damage repair in *TP53*^{WT}-pLoF tumors (**Figure 3 c-e**). Similar trends were observed in the other TCGA cohorts (**Supplementary Table 4**). Hypoxia is a physiological inducer of p53 (An et al., 1998), and p53 is frequently reported to decrease cell hypoxia by inhibiting HIF1A activity (Ravi et al., 2000; Sermeus & Michiels, 2011; Yamakuchi et al., 2010). Consistent with these findings, the Buffa hypoxia score (Buffa, Harris, West, & Miller, 2010) was significantly lower in the *TP53*^{WT}-pN tumors compared to the *TP53*^{WT}-pLoF, *TP53*^{MM} and *TP53*^{TM} tumors (**Figure 3 f**). Collectively, compared to *TP53*^{WT}-pN tumors, those *TP53*^{WT}-pLoF tumors resembled the p53 mutant tumors by manifesting significantly worse prognosis, increased genomic instability, and higher Buffa hypoxia score. These data not only functionally reaffirm our SVM predictions but also highlight the necessity of re-stratifying the *TP53* status determined from the DNA sequencing.

Increased sensitivity of *TP53*^{WT}-pLoF tumors to chemo-and radiation therapy

The role of mutant p53 in response to chemo-and radiation therapy remains controversial. Mutation in p53 has been linked to reduced sensitivity to chemo-and radiation therapy (Gurtner et al., 2020; Hutchinson, Mierzwa, & D'Silva, 2020; J. M. Lee & Bernstein, 1993; Lowe et al., 1994; Skinner et al., 2012), while other studies showed that p53 inactivation increases tumor's sensitivities to chemo- and radiation therapy (Bertheau et al., 2002; Bertheau et al., 2007; Concin et al., 2000). Direct comparisons of radiation or chemotherapy efficacy using TCGA samples are not possible due to the lack of clinical responsiveness data. Instead, we evaluated the chemo- and radiation sensitivity using previously published gene signatures (**Supplementary Table 5**) and studied the radiation therapeutic effects with preclinical animal models. Specifically, chemotherapy sensitivity of lung and breast cancer were measured by the RPS (recombination proficiency score). The RPS was positively correlated with the DNA recombination proficiency and was negatively correlated with sensitivity of platinum-based

chemotherapy, which was clinically validated in breast cancer and NSCLC patients (Pitroda et al., 2014). Tumor radiation sensitivity of breast cancers were measured by the RSS (radiation sensitivity signature) score. RSS score was calculated from the 51-gene signature reported by Speers *et al*. and also validated in patients (Speers et al., 2015) (see Methods). We found that one gene (*RAD51*) of the RPS signature and four genes of the RSS signature are overlapped with the p53 targets selected for LUNG and BRCA (**Supplementary Figure 8 a**). To explore the relationship between predicted p53 status and the radiation therapy (RT) response *in vivo*, we analyzed an independent dataset of 35 patient-derived xenografts (PDX) models that recapitulate many salient genetic and phenotypic features of glioblastoma (GBM) patients (Vaubel et al., 2020).

When comparing the RPS score between groups with different p53 statuses in the LUNG cohort, we found that tumors in the NT group display the highest RPS, while tumors in the $TP53^{MM}$ and $TP53^{TM}$ groups displayed the lowest RPS. Expectedly, the RPS scores of $TP53^{WT}$-pN tumors were close to those of the NT group and were significantly higher than the $TP53^{WT}$-pLoF tumors ($P = 2.8\times10^{-9}$, two-sided Wilcoxon test). Similar results were observed in the BRCA cohort. Importantly, four BRCA $TP53^{MM}$ tumors that were predicted to be p53 normal also had significantly higher RPS scores than the remaining $TP53^{MM}$ samples ($P = 0.029$, two-sided Wilcoxon test) (**Figure 4 a-b)**. The lower RPS score was linked to increased mutagenesis, adverse clinical features, and inferior patient survival rates, but such adverse prognosis could be counteracted by adjuvant platinum-based chemotherapy (Pitroda et al., 2014). Although mutations in *BRCA1*/*BRCA2* are generally considered to be principal drivers of homologous recombination deficiency (HRD), strong positive associations between *TP53* mutation ratio and HRD scores have been observed in TCGA pan-cancer analysis (Takamatsu et al., 2021), and convergent studies demonstrated the involvement of p53 in regulating homologous recombination (Gatz & Wiesmuller, 2006; Sengupta & Harris, 2005). These data suggested that platinum-based chemotherapy is more beneficial to $TP53^{WT}$-pLoF tumors than $TP53^{WT}$-pN tumors.

We calculated radiosensitivity scores after dividing the RSS gene signature into positive (i.e., genes positively correlated with radiosensitivity) and negative (genes negatively correlated with radiosensitivity) subsets. We only analyzed the TCGA BRCA cohort since the RSS signature was derived from breast cancer. Compared to the *TP53*^{WT}-pN tumors, RSS scores of *TP53*^{WT}-pLoF tumors were significantly increased when measured by positive genes (**Figure 4 c**), but significantly decreased when measured by negative genes, suggesting the increased radiosensitivity of *TP53*^{WT}-pLoF tumors (**Figure 4 d**). In contrast, the radiosensitivity of *TP53*^{MM}-pN tumors was significantly decreased, as compared to tumors in the *TP53*^{MM}-pLoF tumors (**Figure 4 c-d**). Importantly, we observed similar results even after removing the genes that overlapped with p53 signature (**Supplementary Figure 8 b-e**). Collectively, these results suggest that tumors with the predicted p53 loss-of-function, in accordance with p53 mutant tumors, are associated with increased sensitivity to platinum-based chemotherapy and radiation therapy. This is likely because p53 inactivation compromises the DNA damage repair, and therefore sensitizes tumor cells to chemo and radiation therapies.

We further confirmed this result using 35 PDX models. We first defined 17 p53-activated and 19 p53-repressed genes from TCGA GBM cohort (**Supplementary Table 1**). We were unable to train an SVM model to predict the p53 status, due to the insufficient sample sizes of both normal and tumors harboring *TP53* truncating mutation. Alternatively, we employed the unsupervised clustering to predict the p53 function status using CES scores calculated from RNA-seq data of PDX mouse models (**Supplementary Table 6**). To determine the RT responsiveness, we calculated the ratio of median survival days between the RT group and the placebo group, and PDX models with ratios >= 1.52 were considered as RT-responsive (**Figure 4 e, Supplementary Table 6**). Out of 20 PDXs that were predicted to be LoF, 15 respond to RT. As a comparison, out of 15 PDXs that were predicted to be p53 normal, only 4 are responders. These results indicate RT sensitivity was significantly increased in pLoF group (Fisher's exact test, P = 0.0068) (**Figure 4 f and g**). Although *TP53* somatic mutations

(determined from WES) were enriched in the pLoF group (**Figure 4 f)**, the association between *TP53* genetic mutation and RT response was not significant (Fisher's exact test, $P = 0.13$). In summary, these data suggested that our *in silico* predictions successfully uncovered the significant association between p53 status and RT response, which would be otherwise missed by examining the *TP53* genetic status alone.

Dissecting predicted *TP53*^{WT}-pLoF

Next, we sought to explore the potential mechanisms and factors that could explain why most of the *TP53*^{WT} tumors were predicted to be LoF by the SVM model. We investigated the RNA and protein expression, re-evaluated all the *TP53* missense mutations (reported from WES) by RNA-seq, and examined the alteration status of p53 upstream regulator MDM2 and MDM4. We did not detect significant changes in protein abundance between *TP53*^{WT}-pN and *TP53*^{WT}-pLoF tumors, and *TP53* RNA expressions were even increased in the *TP53*^{WT}-pLoF tumors (**Supplementary Figure 9,** taking LUNG as an example). This is consistent with previous findings that mutant p53 is associated with increased *TP53* mRNA expression (Donehower et al., 2019). Also, this result suggests that the impaired p53 function in the *TP53*^{WT}-pLoF group is not likely caused by reduced *TP53* mRNA and protein abundances.

We then reassessed all the *TP53* missense mutations (such as R249 and R273) of *TP53*^{WT}-pLoF tumors using the RNA-seq sequence data of the LUNG and BRCA cohorts. Surprisingly, about 10.3% (32 out of 310) LUNG and 10.2% (57 out of 561) BRCA *TP53*^{WT}-pLoF tumors are genuine p53 mutants as demonstrated by the large number of RNA-seq reads carrying the mutant alleles (**Supplementary Table 7)**. For example, two lung adenocarcinoma samples (TCGA-55-6987 and TCGA-55-8621) were reported as *TP53*^{WT} by TCGA, but the mutant allele fractions (MAF) at 17: g.7577118C>A and 17: g.7577559G>A were 45% (44 out of 98) and 35% (35 out of 100), respectively. It is worth noting

that the mutant alleles of these two mutations were also detectable from the WES sequence data albeit with much fewer supporting reads (2 and 5 reads, respectively), which explains why they were omitted by the TCGA somatic variant caller (**Supplementary Figure 10**). The significantly increased MAFs in RNA-seq data is probably due to the preferential expression of the mutant alleles. Interestingly, compared to the amino acid location of missense mutations reported from WES data, those mutations identified using RNA-seq data were significantly enriched ($P = 1.05\times10^{-6}$, two-sided Fisher exact test) at the p53 R249 position in both lung and breast cancers (**Figure 5 a and b**). The missense mutation at R249 is generally considered as a structural mutation destabilizing p53 protein due to the altered 3D structure (Joerger & Fersht, 2016). Also, these mutant tumors rescued from the RNA-seq data exhibited similar genomic characteristics (tumor mutation burden, copy number variation burden, aneuploidy score, and Buffa hypoxia score) compared to the $TP53^{MM}$ and $TP53^{TM}$ groups (**Supplementary Figure 11**).

As p53 negative regulators, MDM2 and its homolog MDMX (encoded by the *MDM4* gene) inhibit p53 activity via different mechanisms, including transcriptional repression and proteasomal degradation (Haupt, Maya, Kazaz, & Oren, 1997; Katz et al., 2018; Wasylishen & Lozano, 2016). It was found that amplification of *MDM2* and/or *MDM4* is mutually exclusive with *TP53* mutations in lung cancer (odds ratio = 4.32, $P = 2.2\times10^{-16}$, χ^2 test) and breast cancer (odds ratio = 1.46, $P = 2.0\times10^{-16}$, χ^2 test) (**Figure 5 c**). As expected, all of the MDM2 and/or MDM4 amplification samples (LUNG, n = 38; BRCA, n = 84) were pLoF, and accounted for 12.3% and 15.0% of all $TP53^{WT}$-pLoF samples in the LUNG and BRCA cohorts, respectively (**Figure 5 d**).

Conclusion

In this study, we defined p53 transcriptional targets and calculated their CES as a surrogate measure of p53 activity, trained and validated SVM models using CES from NT and TM groups, which

represented p53-normal and p53-LoF, respectively. We demonstrated that p53 status could be accurately predicted using *in silico* approach, and showed that non-mutational p53 inactivation is pervasive in human malignancies. Our analyses further revealed that the predicted *TP53*^{WT}-pLoF tumors exhibited a similar level of genomic instability as those with genetic inactivation–including significantly increased mutations, copy number alterations, and aneuploidy. Similar to p53 mutants, those patients with *TP53*^{WT}-pLoF tumors had much worse overall survival than those with *TP53*^{WT}-pN tumors, highlighting the prognostic value of our prediction. When evaluated using clinically validated signatures, *TP53*^{WT}-pLoF tumors exhibited significantly increased sensitivity to platinum-based chemotherapy and radiation therapy. This observation was confirmed in our GBM preclinical animal models. Finally, *TP53*^{WT}-pLoF could be partially explained by false negative mutation calls, or *MDM2* and *MDM4* amplifications.

Discussion

Genomic instability, self-sufficiency in terms of growth signals, insensitivity to anti-growth signals, and evasion of programmed cell death are hallmarks of cancer (Hanahan & Weinberg, 2000). These biological processes are orchestrated by oncogenes (e.g., *EGFR*, *PIK3A*, and *KRAS*) and tumor suppressor genes (e.g., *TP53*, *RB1*, *PTEN*, and *CDKN2A*). During tumorigenesis, genomic amplification (or copy gain) and missense mutation are observed in oncogenes, while genomic deletion (or copy loss), non-sense, and frameshift mutations are associated with tumor suppressor genes (Soussi & Wiman, 2015). Indeed, *RB1*, *PTEN*, and *CDKN2A* are generally inactivated by homozygous deletions, which directly leads to little or no protein expression (**Supplementary Figure 12**). Unlike other tumor suppressors, the predominant genomic alteration for *TP53* is missense mutation, which leads to an altered or faulty protein (**Supplementary Figure 12**). In addition, multiple lines of evidence have characterized the oncogenic role of mutant p53 (Govindan & Weber, 2014), including the over-expression in p53 mutants, accumulation of mutant p53 protein in the nucleus and the cytoplasm, as

well as functional studies *in vivo* and *in vitro* systems (Bieging, Mello, & Attardi, 2014; Freed-Pastor & Prives, 2012; Hollstein, Sidransky, Vogelstein, & Harris, 1991; Oren & Rotter, 2010). The primary goal of this study is to investigate the loss of p53's tumor suppressor function in tumors that may be thought to be p53 functionally normal due to lack of *TP53* genetic alterations. First, our selected p53 target genes are highly enriched in cell division, cell cycle control, DNA integrity checkpoint, and DNA damage repair pathways (**Supplementary Figure 2 b**). Second, we trained SVM models using normal tissues and tumor samples with truncating *TP53* mutations that result in shortened protein. Therefore, those *TP53* missense mutations that were predicted to be LoF ($TP53^{MM}$-pLoF) should be interpreted as the "loss of p53's tumor suppressor function". We could not rule out the possibilities that the mutant p53 protein in the $TP53^{MM}$-pLoF sample have gained oncogenic functions.

According to our analyses, among $TP53^{WT}$ tumors that were predicted to be LoF ($TP53^{WT}$-pLoF), 22-25% could be explained by false-negatives (i.e., *TP53* mutations that failed to be detected from the WES assay) or *MDM2* and *MDM4* amplification. However, the underpinning mechanisms for the remaining 75-78% are unknown and warrant further investigation. We analyzed the DNA methylation data generated from the Infinium HumanMethylation450 BeadChip array, but did not detect significant differences between $TP53^{WT}$-pLoF and $TP53^{WT}$-pN tumors among all nine CpG sites that mapped to *TP53* (cg02087342, cg06365412, cg10792831, cg12041075, cg12041429, cg13468400, cg16397722, cg18198734, cg22949073). Presumably, p53 function is compromised by other transcriptional, post-transcriptional and post-translational mechanisms (Bode & Dong, 2004; Dai & Gu, 2010). For those $TP53^{WT}$ samples that were predicted to be p53 normal, we did not detect significant differences in tumor purities (measured by IHC) between $TP53^{WT}$-pN and other tumor samples in LUNG ($P = 0.460$, Wilcoxon rank sum test) and BRCA cohorts ($P = 0.066$, Wilcox rank sum test), suggesting that these $TP53^{WT}$-pN tumors are not attributed to lower tumor cellularity. We also observed four $TP53^{MM}$ samples that were predicted to be p53 normal from the BRCA cohort, including TCGA-E2-A1B1,

TCGA-LD-A74U, TCGA-BH-A1FE, and TCGA-EW-A1P1 (**Supplementary Table 8 c**). All reported mutations from these four samples could be confirmed from RNA-seq alignments with high mutant allele frequencies, which rules out the possibility of false positives or cross-sample contaminations. The potential reasons remain unclear, and the limited sample size precludes systematic evaluations.

Gene expression reflects the downstream effects of transcriptional regulation. This study demonstrated that composite expression calculated from a group of logically connected genes is a reliable measure of the activity of their common upstream regulator. A large number of truncating mutations in *TP53*, providing a reliable way to identify loss of function cases, made this analysis particularly approachable. When obtaining sufficient training data is not practical, unsupervised machine-learning approaches such as K-means clustering, semi-supervised learning, principal component analysis (PCA), and neural networks could be used as an alternative to the SVM model.

TP53 is the most mutated gene in human cancers. Our analyses further revealed that most *TP53*^{WT} tumors are genetically normal but phenotypically deficient in p53 function. This finding demonstrates an inherent limitation in the stratification and segmentation of patients based on DNA markers alone and encourages the development of "ensemble" approaches based on genomic, epigenomic, transcriptomic and proteomic data.

Materials and Methods

Data collection

We downloaded somatic mutation data for *TP53* from the LUNG (LUSC + LUAD), BRCA, COAD, ESCA, BLCA, HNSC, LIHC, STAD, and UCES datasets on cBioPortal (https://www.cbioportal.org/)

along with pre-calculated meta scores including the fraction of genome altered (FGA), mutation count, aneuploidy score, and Buffa hypoxia score. TCGA WES and RNA-seq BAM files for BRCA and LUNG were downloaded from the GDC data portal (https://portal.gdc.cancer.gov/), which were used to re-evaluate p53 mutation status. TCGA level-3 RNA-seq expression data of selected cancers and the corresponding demographic and survival data were downloaded from the University of California Santa Cruz's Xena web server (https://xena.ucsc.edu/). Pre-calculated raw RNA-seq read counts were used to identify differentially expressed genes between the NT and TM groups, log2-transformed FPKM (i.e., Fragments Per Kilobase of exon per Million mapped fragments) was used to calculate CES scores, and TPM (i.e., Transcript Per Million) was used to calculate GSVA to evaluate chemo- and radiotherapy sensitivities. ChIP-seq data (p53 binding peaks) were downloaded from the ReMap 2020 database (Cheneby et al., 2020) (https://remap.univ-amu.fr/). Gene expression data (TPM) of normal tissues were downloaded from the GTEx (release V8) data portal (https://gtexportal.org/home/).

Identification of p53 target genes across different cancer types

To identify p53 targets for each cancer type, we employed a series of steps. First, we started with 147 genes that have been experimentally validated as p53 targets, including 116 genes activated by p53 and 31 genes repressed by p53 (Fischer, 2017). We further analyzed the publicly available ChIP-seq data from ReMap 2020 (Cheneby et al., 2020) and found that 99% of the p53 target genes from the Fischer *et al.* study have at least one p53 peak within the promotor region, which is defined as +/- 1 kb around the transcription start site (TSS). The basal level expression of p53 targets in normal tissues was evaluated using the GTEx (Consortium, 2015) RNA-seq data. Specifically, we first calculated the median TPM across all samples in each tissue type, and then calculated the mean of median TPMs across all tissue types. The result showed that the mean of median TPMs of each p53 target gene was > 1 TPM, indicating that these 147 p53 targets were reliably expressed across GTEx normal tissues and thus were included in our downstream analyses.

Since p53 might regulate a specific set of genes in different cancer types (Fischer, 2017), and to define cancer-specific p53 targets, we first performed differential expression analysis in each cancer by comparing the NT group (considered to be p53 functionally normal) to the TM group (considered to be p53 functionally impaired). Note, samples assigned to the testing set were not used for the differential expression analysis. Then, the differentially expressed genes (adjusted *p*-value < 0.05 and |fold change| > 2) were intersected with the 147 genes defined previously. We defined p53-activated and p53-repressed genes as those downregulated and upregulated in the TM group, respectively. The differential analyses were conducted using the R package DESeq2 on the raw RNA-seq reads counts (Love, Huber, & Anders, 2014).

Computation of composite expression scores

CES is a single enrichment score that provides intuitive and stable measurement of p53 activity in each cancer. We used four algorithms to calculate the CES including (1) Gene Set Variation Analysis (GSVA), a gene set enrichment method that estimates variation of pathway activity over a sample population in an unsupervised manner (Hanzelmann et al., 2013); (2) single-sample GSEA (ssGSEA), an extension of GSEA, which calculates separate enrichment scores for each pairing of a sample and gene set (Barbie et al., 2009); (3) The first Principal Component (PC1) score of the principal component analysis (PCA); and (4) combined Z-score, a normalized value based on the mean and standard deviation (E. Lee et al., 2008; Tomfohr et al., 2005). To calculate the combined Z-score, gene expression values (log2(FPKM)) of each sample were converted into Z-scores by $Z = (x - \mu)/\sigma$, where μ and σ is the average and standard deviation of log2(FPKM) across all samples of a gene. Given a gene set $\gamma = \{1,\ldots,k\}$ with standardized values $z_1,\ldots,z_k$ for each gene in a specific sample, the combined Z-score Z_γ for the gene set γ is defined as:

$$Z_\gamma = \frac{\sum_{i=1}^{k} Z_i}{\sqrt{k}}$$

GSVA, ssGSEA, and Z-score were calculated for p53 activated and repressed targets separately, while PC1 was calculated using all p53 targets, which resulted in a total of seven CES scores for each sample. The GSVA package (https://www.bioconductor.org/packages/release/bioc/html/GSVA.html) was used to compute GSVA, ssGSEA, and Z-score. The Python package sklearn (https://scikit-learn.org/stable/) was used to perform PCA analyses.

Training and evaluating the performance of SVM models

Since the prediction of p53 status is a binary classification, we chose to use the SVM model–a widely used, supervised machine-learning method. SVM models with linear kernel were trained using the CES scores of NT (normal tissue, coded as "0") and TM (truncating mutations, coded as "1") groups. GridSearch was used to pick the proper C and gamma parameters for the SVM model. We used TCGA data for both training and testing; no separate or external validation set was used. The number of samples available for training/testing depended on the cancer type. The number of training and testing samples (n) were 362, 246, 99, 64, and 1335 for the LUNG, BRCA, COAD, ESCA, and pan-cancer cohorts, respectively (**Supplementary Table 8 a**). A total of seven features ($GSVA_{activated}$, $GSVA_{repressed}$, $ssGSEA_{activated}$, $ssGSEA_{repressed}$, $Zscore_{activated}$, $Zscore_{repressed}$, and PC1) were used in the SVM models. Therefore, $p << n$ for all SVM models built in this study, where p represents the number of features (predictor variables) and n represents the number of samples.

We applied ten-fold cross-validation to evaluate the performance of the SVM model. Specifically, we used the "*train_test_split*" function from the "*sklearn.model_selection*" class to split the samples (i.e, NT and TM samples) into training set (75%) and testing set (25%). We performed the differential

expression for gene selection and trained the SVM models with the training set, and then used the testing set to evaluate the performance. This enforced that the training and test sets were independent. In each iteration, a confusion matrix was made, and performance measurements (i.e., sensitivity/recall, precision, accuracy) were calculated. Finally, we summarized the performance of models with the mean of the measurement scores and the ROC (receiver operating characteristic) curves. The performance measurements are defined as below:

$$\text{recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

$$\text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

$$\text{accuracy} = \frac{\text{TP} + \text{FN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

$$\text{F1} = \frac{2}{\text{recall}^{-1} + \text{precision}^{-1}}$$

Where FN = false negative; FP = false positive; TN = true negative; TP = true positive. We used the Scikit-learn (www.scikit-learn.org) Python package for SVM modeling and ten-fold cross-validation. We used pandas (https://pandas.pydata.org/), numpy (https://numpy.org/), and scipy (https://www.scipy.org/) for data processing and numerical analyses. Details of SVM models are available in **Supplementary Table 9**. All TCGA barcodes (i.e., sample identifiers) for training and testing used in this study are listed in **Supplementary Table 8 b-f.** Python source code is available from https://github.com/liguowang/epage.

Re-evaluation of *TP53* mutation status using RNA-seq

The genomic locations of *TP53* somatic mutations in LUNG and BRCA samples were downloaded from the cBioPortal (https://www.cbioportal.org/). The reference and mutant allele counts were then calculated from the RNA-seq BAM file using Samtools' pileup engine and an in-house Python script

(Li et al., 2009). The samples with fewer than 10 read counts at a given genome site were discarded (n=41), and a cut-off of minor allele frequency = 0.1 was employed to determine the genotype. A sample was determined as *TP53* mutated if there are at least five high-quality reads (Phred-scaled sequencing quality and mapping quality > 30) supporting the mutant allele. 32 and 57 $TP53^{WT}$ samples met this criterion in LUNG and BRCA samples, respectively. The Integrative Genomics Viewer (IGV) was used to manually inspect the variants and visualize the alignments.

Investigation of two independent treatment-related signatures

Tumor chemotherapy sensitivity (i.e., RPS score) was estimated from the 4 genes involved in DNA repair (*RIF1*, *PARI/F2R*, *RAD51*, and Ku80/*XRCC5*) reported by Pitroda *et al.* (Pitroda et al., 2014). Tumor radiation sensitivity signature (RSS) scores in breast cancer were calculated from the 51-gene signature reported by Speers *et al.* (Speers et al., 2015), and the genes were divided in to "positive" and "negative" groups according to the correlation between their expression level and radiation sensitivity. The R package GSVA (Hanzelmann et al., 2013) was used to calculate the RPS and RSS scores. Specifically, according to the study of Pitroda *et al.*, the original RPS was defined as the sum of the expression levels times -1 after log2 transformation and Robust Multi-array Average (RMA)-normalization. In this study, RPS was represented by -GSVA of TPM.

GBM derived xenografts

Glioblastoma patient-derived xenografts (PDX) were generated by the Mayo PDX National Resource (Vaubel et al., 2020). Mice with established orthotopic tumors were randomized in groups of 5–10 mice and treated with placebo and radiation therapy (RT). To determine the RT responsiveness of PDX, we first calculate the ratio of median survivals between the RT group and placebo group. A cutoff value (1.52) was determined as the changing point on the non-parametric kernel density curve (**Supplementary Table 6)**.

DNA (WES) and RNA (RNA-seq) sequencing were performed by Mayo Clinic medical genomic facility or Translational Genomics Institute (TGen) using the SureSelect (Agilent, Santa Clara, CA) or TGen Strexome V2 capture kits. Raw sequencing reads were aligned to both human (hg38) and mouse (mm10) reference genomes to remove any reads potentially originated from the mouse tissue. Then, filtered WES and RNA-seq data were analyzed using Mayo's MAP-RSeq and GenomeGPS workflow, respectively. The raw sequencing data were submitted to the NCBI Sequence Read Archive with accession # PRJNA543854 and PRJNA548556 for WES and RNA-seq, respectively. Annotated genomic and transcriptomic data are also publicly accessible through the cBioPortal (https://www.cbioportal.org/study/summary?id=gbm_mayo_pdx_sarkaria_2019).

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

We downloaded somatic mutation data for *TP53* from the LUNG (LUSC + LUAD), BRCA, COAD, ESCA, BLCA, HNSC, LIHC, STAD, and UCES datasets on cBioPortal (https://www.cbioportal.org/) along with pre-calculated meta scores including the fraction of genome altered (FGA), mutation count, aneuploidy score, and Buffa hypoxia score. TCGA WES and RNA-seq BAM files for BRCA and LUNG were downloaded from the GDC data portal (https://portal.gdc.cancer.gov/), which were used to re-evaluate p53 mutation status. TCGA level-3 RNA-seq expression data of selected cancers and

the corresponding demographic and survival data were downloaded from the University of California Santa Cruz's Xena web server (https://xena.ucsc.edu/). Pre-calculated raw RNA-seq read counts were used to identify differentially expressed genes between the NT and TM groups, log2-transformed FPKM (i.e., Fragments Per Kilobase of exon per Million mapped fragments) was used to calculate CES scores, and TPM (i.e., Transcript Per Million) was used to calculate GSVA to evaluate chemo- and radiotherapy sensitivities. ChIP-seq data (p53 binding peaks) were downloaded from the ReMap 2020 database (Cheneby et al., 2020) (https://remap.univ-amu.fr/). Gene expression data (TPM) of normal tissues were downloaded from the GTEx (release V8) data portal (https://gtexportal.org/home/). All of the results generated in this study are included as supplementary data sets (supplementary table 1-8).

Python source code of our p53 status prediction method is available from https://github.com/liguowang/epage.

Competing interests

The authors declare no competing interests.

Funding

This work is partially supported by the US National Institute of Health [U10-CA180882-07], the Center for Individualized Medicine of Mayo Clinic, the Strategic Priority Research Program of the Chinese Academy of Sciences [XDB38030400], and the Youth Innovation Promotion Association of Chinese Academy of Sciences [2019104].

Author's Contributions

L.W., L.M., and H.L. conceived the study. Q.L. and Y.Z. implemented the SVM model and performed the bioinformatics analyses. Z.Z., A.O., J.N.K., and D.E.K. assisted and supervised model building

and analysis. L.W., L.M., Q.L., and Y.Z. took the lead in writing the manuscript. All authors provided critical feedback and helped shape the research, analyses, and manuscript.

Acknowledgments

The results shown in this study are in part based upon data generated by the TCGA Research Network (https://www.cancer.gov/tcga). We thank TCGA's specimen donors and research groups that make genomic data publicly available.

Abbreviations

AUROC, area under the receiver operating characteristic
BAM, binary version of SAM (Sequence Alignment/Map) file
BRCA, breast invasive carcinoma
CES, composite expression score
ChIP-seq, chromatin immunoprecipitation sequencing
COAD, colon adenocarcinoma
ESCA, esophageal carcinoma
FGA, fraction of genome altered
FN, false negative
FP, false positive
FPKM, Fragments Per Kilobase of exon per Million mapped fragments
GoF, gain of function
GSVA, gene set variation analysis
GTEx, genotype-tissue expression
HNSC, head and neck squamous cell carcinoma
HRD, homologous recombination deficiency
ICGC, International Cancer Genome Consortium
IHC, immuno histochemistry staining
LIHC, liver hepatocellular carcinoma
LoF, loss of function
MM, missense mutation
MAF, mutant allele fractions
NSCLS, non-small cell lung carcinomas
NT, normal tissue
OR, odds ratio
PCA, principal component analysis
PDX, patient derived xenografts
ROC, receiver operating characteristic
RPS, recombination proficiency score
RSS, radiation sensitivity signature
ssGSEA, single sample gene set enrichment analysis
STAD, stomach adenocarcinoma
SVM, support vector machine
TCGA, The Cancer Genome Atlas
TPM, transcript per million
TM, truncating mutation
TMB, tumor mutation burden
TN, true negative
TP, true positive
UCES, uterine corpus endometrial carcinoma
WES, whole exome sequencing
WT, wildtype

Reference

Agirre, X., Vizmanos, J. L., Calasanz, M. J., Garcia-Delgado, M., Larrayoz, M. J., & Novo, F. J. (2003). Methylation of CpG dinucleotides and/or CCWGG motifs at the promoter of TP53 correlates with decreased gene expression in a subset of acute lymphoblastic leukemia patients. *Oncogene, 22*(7), 1070-1072. doi:10.1038/sj.onc.1206236

Alexandrov, L. B., Nik-Zainal, S., Wedge, D. C., Aparicio, S. A., Behjati, S., Biankin, A. V., . . . Stratton, M. R. (2013). Signatures of mutational processes in human cancer. *Nature, 500*(7463), 415-421. doi:10.1038/nature12477

An, W. G., Kanekal, M., Simon, M. C., Maltepe, E., Blagosklonny, M. V., & Neckers, L. M. (1998). Stabilization of wild-type p53 by hypoxia-inducible factor 1alpha. *Nature, 392*(6674), 405-408. doi:10.1038/32925

Ashcroft, M., Kubbutat, M. H., & Vousden, K. H. (1999). Regulation of p53 function and stability by phosphorylation. *Mol Cell Biol, 19*(3), 1751-1758. doi:10.1128/MCB.19.3.1751

Barbie, D. A., Tamayo, P., Boehm, J. S., Kim, S. Y., Moody, S. E., Dunn, I. F., . . . Hahn, W. C. (2009). Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature, 462*(7269), 108-112. doi:10.1038/nature08460

Bertheau, P., Plassa, F., Espie, M., Turpin, E., de Roquancourt, A., Marty, M., . . . de The, H. (2002). Effect of mutated TP53 on response of advanced breast cancers to high-dose chemotherapy. *Lancet, 360*(9336), 852-854. doi:10.1016/S0140-6736(02)09969-5

Bertheau, P., Turpin, E., Rickman, D. S., Espie, M., de Reynies, A., Feugeas, J. P., . . . de The, H. (2007). Exquisite sensitivity of TP53 mutant and basal breast cancers to a dose-dense epirubicin-cyclophosphamide regimen. *PLoS Med, 4*(3), e90. doi:10.1371/journal.pmed.0040090

Bieging, K. T., Mello, S. S., & Attardi, L. D. (2014). Unravelling mechanisms of p53-mediated tumour suppression. *Nat Rev Cancer, 14*(5), 359-370. doi:10.1038/nrc3711

Bode, A. M., & Dong, Z. (2004). Post-translational modification of p53 in tumorigenesis. *Nat Rev Cancer, 4*(10), 793-805. doi:10.1038/nrc1455

Bosari, S., Viale, G., Roncalli, M., Graziani, D., Borsani, G., Lee, A. K., & Coggi, G. (1995). p53 gene mutations, p53 protein accumulation and compartmentalization in colorectal adenocarcinoma. *Am J Pathol, 147*(3), 790-798.

Bouvard, V., Zaitchouk, T., Vacher, M., Duthu, A., Canivet, M., Choisy-Rossi, C., . . . May, E. (2000). Tissue and cell-specific expression of the p53-target genes: bax, fas, mdm2 and waf1/p21, before and following ionising irradiation in mice. *Oncogene, 19*(5), 649-660. doi:10.1038/sj.onc.1203366

Bryant, H. E., Schultz, N., Thomas, H. D., Parker, K. M., Flower, D., Lopez, E., . . . Helleday, T. (2005). Specific killing of BRCA2-deficient tumours with inhibitors of poly(ADP-ribose) polymerase. *Nature, 434*(7035), 913-917. doi:10.1038/nature03443

Buffa, F. M., Harris, A. L., West, C. M., & Miller, C. J. (2010). Large meta-analysis of multiple cancers reveals a common, compact and highly prognostic hypoxia metagene. *Br J Cancer, 102*(2), 428-435. doi:10.1038/sj.bjc.6605450

Burns, T. F., Bernhard, E. J., & El-Deiry, W. S. (2001). Tissue specific expression of p53 target genes suggests a key role for KILLER/DR5 in p53-dependent apoptosis in vivo. *Oncogene, 20*(34), 4601-4612. doi:10.1038/sj.onc.1204484

Cheneby, J., Menetrier, Z., Mestdagh, M., Rosnet, T., Douida, A., Rhalloussi, W., . . . Ballester, B. (2020). ReMap 2020: a database of regulatory regions from an integrative analysis of Human and Arabidopsis DNA-binding sequencing experiments. *Nucleic Acids Res, 48*(D1), D180-D188. doi:10.1093/nar/gkz945

Concin, N., Zeillinger, C., Stimpfel, M., Schiebel, I., Tong, D., Wolff, U., . . . Zeillinger, R. (2000). p53-dependent radioresistance in ovarian carcinoma cell lines. *Cancer Lett, 150*(2), 191-199. doi:10.1016/s0304-3835(99)00393-6

Consortium, G. T. (2015). Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science, 348*(6235), 648-660. doi:10.1126/science.1262110

Cortez, M. A., Ivan, C., Valdecanas, D., Wang, X., Peltier, H. J., Ye, Y., . . . Welsh, J. W. (2016). PDL1 Regulation by p53 via miR-34. *J Natl Cancer Inst, 108*(1). doi:10.1093/jnci/djv303

Cross, B., Chen, L., Cheng, Q., Li, B., Yuan, Z. M., & Chen, J. (2011). Inhibition of p53 DNA binding function by the MDM2 protein acidic domain. *J Biol Chem, 286*(18), 16018-16029. doi:10.1074/jbc.M111.228981

Dai, C., & Gu, W. (2010). p53 post-translational modification: deregulated in tumorigenesis. *Trends Mol Med, 16*(11), 528-536. doi:10.1016/j.molmed.2010.09.002

Donehower, L. A., Soussi, T., Korkut, A., Liu, Y., Schultz, A., Cardenas, M., . . . Wheeler, D. A. (2019). Integrated Analysis of TP53 Gene and Pathway Alterations in The Cancer Genome Atlas. *Cell Rep, 28*(5), 1370-1384 e1375. doi:10.1016/j.celrep.2019.07.001

Farmer, H., McCabe, N., Lord, C. J., Tutt, A. N., Johnson, D. A., Richardson, T. B., . . . Ashworth, A. (2005). Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. *Nature, 434*(7035), 917-921. doi:10.1038/nature03445

Fischer, M. (2017). Census and evaluation of p53 target genes. *Oncogene, 36*(28), 3943-3956. doi:10.1038/onc.2016.502

Freed-Pastor, W. A., & Prives, C. (2012). Mutant p53: one name, many proteins. *Genes Dev, 26*(12), 1268-1286. doi:10.1101/gad.190678.112

Gatz, S. A., & Wiesmuller, L. (2006). p53 in recombination and repair. *Cell Death Differ, 13*(6), 1003-1016. doi:10.1038/sj.cdd.4401903

Glodzik, D., Bosch, A., Hartman, J., Aine, M., Vallon-Christersson, J., Reutersward, C., . . . Staaf, J. (2020). Comprehensive molecular comparison of BRCA1 hypermethylated and BRCA1 mutated triple negative breast cancers. *Nat Commun, 11*(1), 3747. doi:10.1038/s41467-020-17537-2

Gogna, R., Madan, E., Kuppusamy, P., & Pati, U. (2012). Re-oxygenation causes hypoxic tumor regression through restoration of p53 wild-type conformation and post-translational modifications. *Cell Death Dis, 3*, e286. doi:10.1038/cddis.2012.15

Gorkin, D. U., Leung, D., & Ren, B. (2014). The 3D genome in transcriptional regulation and pluripotency. *Cell Stem Cell, 14*(6), 762-775. doi:10.1016/j.stem.2014.05.017

Govindan, R., & Weber, J. (2014). TP53 mutations and lung cancer: not all mutations are created equal. *Clin Cancer Res, 20*(17), 4419-4421. doi:10.1158/1078-0432.CCR-14-0899

Guo, G., Yu, M., Xiao, W., Celis, E., & Cui, Y. (2017). Local Activation of p53 in the Tumor Microenvironment Overcomes Immune Suppression and Enhances Antitumor Immunity. *Cancer Res, 77*(9), 2292-2305. doi:10.1158/0008-5472.CAN-16-2832

Gurtner, K., Kryzmien, Z., Koi, L., Wang, M., Benes, C. H., Hering, S., . . . Krause, M. (2020). Radioresistance of KRAS/TP53-mutated lung cancer can be overcome by radiation dose escalation or EGFR tyrosine kinase inhibition in vivo. *Int J Cancer, 147*(2), 472-477. doi:10.1002/ijc.32598

Hafner, A., Stewart-Ornstein, J., Purvis, J. E., Forrester, W. C., Bulyk, M. L., & Lahav, G. (2017). p53 pulses lead to distinct patterns of gene expression albeit similar DNA-binding dynamics. *Nat Struct Mol Biol, 24*(10), 840-847. doi:10.1038/nsmb.3452

Hanahan, D., & Weinberg, R. A. (2000). The hallmarks of cancer. *Cell, 100*(1), 57-70. doi:10.1016/s0092-8674(00)81683-9

Hanzelmann, S., Castelo, R., & Guinney, J. (2013). GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics, 14*, 7. doi:10.1186/1471-2105-14-7

Haupt, Y., Maya, R., Kazaz, A., & Oren, M. (1997). Mdm2 promotes the rapid degradation of p53. *Nature, 387*(6630), 296-299. doi:10.1038/387296a0

Hollstein, M., Sidransky, D., Vogelstein, B., & Harris, C. C. (1991). p53 mutations in human cancers. *Science, 253*(5015), 49-53. doi:10.1126/science.1905840

Hutchinson, M. N. D., Mierzwa, M., & D'Silva, N. J. (2020). Radiation resistance in head and neck squamous cell carcinoma: dire need for an appropriate sensitizer. *Oncogene, 39*(18), 3638-3649. doi:10.1038/s41388-020-1250-3

Igelmann, S., Neubauer, H. A., & Ferbeyre, G. (2019). STAT3 and STAT5 Activation in Solid Cancers. *Cancers (Basel), 11*(10). doi:10.3390/cancers11101428

Isaacs, J. S., Hardman, R., Carman, T. A., Barrett, J. C., & Weissman, B. E. (1998). Differential subcellular p53 localization and function in N- and S-type neuroblastoma cell lines. *Cell Growth Differ, 9*(7), 545-555.

Jia, P., & Zhao, Z. (2019). Characterization of Tumor-Suppressor Gene Inactivation Events in 33 Cancer Types. *Cell Rep, 26*(2), 496-506 e493. doi:10.1016/j.celrep.2018.12.066

Jiang, C., & Pugh, B. F. (2009). Nucleosome positioning and gene regulation: advances through genomics. *Nat Rev Genet, 10*(3), 161-172. doi:10.1038/nrg2522

Joerger, A. C., & Fersht, A. R. (2016). The p53 Pathway: Origins, Inactivation in Cancer, and Emerging Therapeutic Approaches. *Annu Rev Biochem, 85*, 375-404. doi:10.1146/annurev-biochem-060815-014710

Karsli Uzunbas, G., Ahmed, F., & Sammons, M. A. (2019). Control of p53-dependent transcription and enhancer activity by the p53 family member p63. *J Biol Chem, 294*(27), 10720-10736. doi:10.1074/jbc.RA119.007965

Karve, T. M., & Cheema, A. K. (2011). Small changes huge impact: the role of protein posttranslational modifications in cellular homeostasis and disease. *J Amino Acids, 2011*, 207691. doi:10.4061/2011/207691

Kato, S., Han, S. Y., Liu, W., Otsuka, K., Shibata, H., Kanamaru, R., & Ishioka, C. (2003). Understanding the function-structure and function-mutation relationships of p53 tumor suppressor protein by high-resolution missense mutation analysis. *Proc Natl Acad Sci U S A, 100*(14), 8424-8429. doi:10.1073/pnas.1431692100

Katz, C., Low-Calle, A. M., Choe, J. H., Laptenko, O., Tong, D., Joseph-Chowdhury, J. N., . . . Prives, C. (2018). Wild-type and cancer-related p53 proteins are preferentially degraded by MDM2 as dimers rather than tetramers. *Genes Dev, 32*(5-6), 430-447. doi:10.1101/gad.304071.117

Lane, D. P. (1992). Cancer. p53, guardian of the genome. *Nature, 358*(6381), 15-16. doi:10.1038/358015a0

Lee, E., Chuang, H. Y., Kim, J. W., Ideker, T., & Lee, D. (2008). Inferring pathway activity toward precise disease classification. *PLoS Comput Biol, 4*(11), e1000217. doi:10.1371/journal.pcbi.1000217

Lee, J. M., & Bernstein, A. (1993). p53 mutations increase resistance to ionizing radiation. *Proc Natl Acad Sci U S A, 90*(12), 5742-5746. doi:10.1073/pnas.90.12.5742

Levine, A. J. (1997). p53, the cellular gatekeeper for growth and division. *Cell, 88*(3), 323-331. doi:10.1016/s0092-8674(00)81871-1

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., . . . Genome Project Data Processing, S. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics, 25*(16), 2078-2079. doi:10.1093/bioinformatics/btp352

Liu, Y., & Gu, W. (2022). p53 in ferroptosis regulation: the new weapon for the old guardian. *Cell Death Differ*. doi:10.1038/s41418-022-00943-y

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol, 15*(12), 550. doi:10.1186/s13059-014-0550-8

Lowe, S. W., Bodis, S., McClatchey, A., Remington, L., Ruley, H. E., Fisher, D. E., . . . Jacks, T. (1994). p53 status and the efficacy of cancer therapy in vivo. *Science, 266*(5186), 807-810. doi:10.1126/science.7973635

Maddalena, M., Mallel, G., Nataraj, N. B., Shreberk-Shaked, M., Hassin, O., Mukherjee, S., . . . Oren, M. (2021). TP53 missense mutations in PDAC are associated with enhanced fibrosis and an immunosuppressive microenvironment. *Proc Natl Acad Sci U S A, 118*(23). doi:10.1073/pnas.2025631118

Mantovani, F., Collavin, L., & Del Sal, G. (2019). Mutant p53 as a guardian of the cancer cell. *Cell Death Differ, 26*(2), 199-212. doi:10.1038/s41418-018-0246-9

Martins, C. P., Brown-Swigart, L., & Evan, G. I. (2006). Modeling the therapeutic efficacy of p53 restoration in tumors. *Cell, 127*(7), 1323-1334. doi:10.1016/j.cell.2006.12.007

Moll, U. M., LaQuaglia, M., Benard, J., & Riou, G. (1995). Wild-type p53 protein undergoes cytoplasmic sequestration in undifferentiated neuroblastomas but not in differentiated tumors. *Proc Natl Acad Sci U S A, 92*(10), 4407-4411. doi:10.1073/pnas.92.10.4407

Moll, U. M., Riou, G., & Levine, A. J. (1992). Two distinct mechanisms alter p53 in breast cancer: mutation and nuclear exclusion. *Proc Natl Acad Sci U S A, 89*(15), 7262-7266. doi:10.1073/pnas.89.15.7262

Momand, J., Zambetti, G. P., Olson, D. C., George, D., & Levine, A. J. (1992). The mdm-2 oncogene product forms a complex with the p53 protein and inhibits p53-mediated transactivation. *Cell, 69*(7), 1237-1245. doi:10.1016/0092-8674(92)90644-r

Oren, M., & Rotter, V. (2010). Mutant p53 gain-of-function in cancer. *Cold Spring Harb Perspect Biol, 2*(2), a001107. doi:10.1101/cshperspect.a001107

Parker, J. S., Mullins, M., Cheang, M. C., Leung, S., Voduc, D., Vickery, T., . . . Bernard, P. S. (2009). Supervised risk predictor of breast cancer based on intrinsic subtypes. *J Clin Oncol, 27*(8), 1160-1167. doi:10.1200/JCO.2008.18.1370

Pitroda, S. P., Pashtan, I. M., Logan, H. L., Budke, B., Darga, T. E., Weichselbaum, R. R., & Connell, P. P. (2014). DNA repair pathway gene expression score correlates with repair proficiency and tumor sensitivity to chemotherapy. *Sci Transl Med, 6*(229), 229ra242. doi:10.1126/scitranslmed.3008291

Ravi, R., Mookerjee, B., Bhujwalla, Z. M., Sutter, C. H., Artemov, D., Zeng, Q., . . . Bedi, A. (2000). Regulation of tumor angiogenesis by p53-induced degradation of hypoxia-inducible factor 1alpha. *Genes Dev, 14*(1), 34-44.

Ruan, H., Li, Q., Liu, Y., Liu, Y., Lussier, C., Diao, L., & Han, L. (2022). GPEdit: the genetic and pharmacogenomic landscape of A-to-I RNA editing in cancers. *Nucleic Acids Res, 50*(D1), D1231-D1237. doi:10.1093/nar/gkab810

Sasaki, M., Nie, L., & Maki, C. G. (2007). MDM2 binding induces a conformational change in p53 that is opposed by heat-shock protein 90 and precedes p53 proteasomal degradation. *J Biol Chem, 282*(19), 14626-14634. doi:10.1074/jbc.M610514200

Schoenfelder, S., & Fraser, P. (2019). Long-range enhancer-promoter contacts in gene expression control. *Nat Rev Genet, 20*(8), 437-455. doi:10.1038/s41576-019-0128-0

Sengupta, S., & Harris, C. C. (2005). p53: traffic cop at the crossroads of DNA repair and recombination. *Nat Rev Mol Cell Biol, 6*(1), 44-55. doi:10.1038/nrm1546

Sermeus, A., & Michiels, C. (2011). Reciprocal influence of the p53 and the hypoxic pathways. *Cell Death Dis, 2*, e164. doi:10.1038/cddis.2011.48

Skinner, H. D., Sandulache, V. C., Ow, T. J., Meyn, R. E., Yordy, J. S., Beadle, B. M., . . . Myers, J. N. (2012). TP53 disruptive mutations lead to head and neck cancer treatment failure through inhibition of radiation-induced senescence. *Clin Cancer Res, 18*(1), 290-300. doi:10.1158/1078-0432.CCR-11-2260

Soussi, T., & Wiman, K. G. (2015). TP53: an oncogene in disguise. *Cell Death Differ, 22*(8), 1239-1249. doi:10.1038/cdd.2015.53

Speers, C., Zhao, S., Liu, M., Bartelink, H., Pierce, L. J., & Feng, F. Y. (2015). Development and Validation of a Novel Radiosensitivity Signature in Human Breast Cancer. *Clin Cancer Res, 21*(16), 3667-3677. doi:10.1158/1078-0432.CCR-14-2898

Stephens, P. J., McBride, D. J., Lin, M. L., Varela, I., Pleasance, E. D., Simpson, J. T., . . . Stratton, M. R. (2009). Complex landscapes of somatic rearrangement in human breast cancer genomes. *Nature, 462*(7276), 1005-1010. doi:10.1038/nature08645

Surget, S., Khoury, M. P., & Bourdon, J. C. (2013). Uncovering the role of p53 splice variants in human malignancy: a clinical perspective. *Onco Targets Ther, 7*, 57-68. doi:10.2147/OTT.S53876

Takamatsu, S., Brown, J. B., Yamaguchi, K., Hamanishi, J., Yamanoi, K., Takaya, H., . . . Matsumura, N. (2021). Utility of Homologous Recombination Deficiency Biomarkers Across Cancer Types. *JCO Precis Oncol, 5*. doi:10.1200/PO.21.00141

Tchelebi, L., Ashamalla, H., & Graves, P. R. (2014). Mutant p53 and the response to chemotherapy and radiation. *Subcell Biochem, 85*, 133-159. doi:10.1007/978-94-017-9211-0_8

Teufel, D. P., Bycroft, M., & Fersht, A. R. (2009). Regulation by phosphorylation of the relative affinities of the N-terminal transactivation domains of p53 for p300 domains and Mdm2. *Oncogene, 28*(20), 2112-2118. doi:10.1038/onc.2009.71

Tian, S., Lai, J., Yu, T., Li, Q., & Chen, Q. (2020). Regulation of Gene Expression Associated With the N6-Methyladenosine (m6A) Enzyme System and Its Significance in Cancer. *Front Oncol, 10*, 623634. doi:10.3389/fonc.2020.623634

Tomfohr, J., Lu, J., & Kepler, T. B. (2005). Pathway level analysis of gene expression using singular value decomposition. *BMC Bioinformatics, 6*, 225. doi:10.1186/1471-2105-6-225

Tominaga, O., Hamelin, R., Trouvat, V., Salmon, R. J., Lesec, G., Thomas, G., & Remvikos, Y. (1993). Frequently elevated content of immunochemically defined wild-type p53 protein in colorectal adenomas. *Oncogene, 8*(10), 2653-2658.

Vaubel, R. A., Tian, S., Remonde, D., Schroeder, M. A., Mladek, A. C., Kitange, G. J., . . . Sarkaria, J. N. (2020). Genomic and Phenotypic Characterization of a Broad Panel of Patient-Derived Xenografts Reflects the Diversity of Glioblastoma. *Clin Cancer Res, 26*(5), 1094-1104. doi:10.1158/1078-0432.CCR-19-0909

Ventura, A., Kirsch, D. G., McLaughlin, M. E., Tuveson, D. A., Grimm, J., Lintault, L., . . . Jacks, T. (2007). Restoration of p53 function leads to tumour regression in vivo. *Nature, 445*(7128), 661-665. doi:10.1038/nature05541

Vogelstein, B., Papadopoulos, N., Velculescu, V. E., Zhou, S., Diaz, L. A., Jr., & Kinzler, K. W. (2013). Cancer genome landscapes. *Science, 339*(6127), 1546-1558. doi:10.1126/science.1235122

Vousden, K. H., & Lane, D. P. (2007). p53 in health and disease. *Nat Rev Mol Cell Biol, 8*(4), 275-283. doi:10.1038/nrm2147

Wasylishen, A. R., & Lozano, G. (2016). Attenuating the p53 Pathway in Human Cancers: Many Means to the Same End. *Cold Spring Harb Perspect Med, 6*(8). doi:10.1101/cshperspect.a026211

Yamakuchi, M., Lotterman, C. D., Bao, C., Hruban, R. H., Karim, B., Mendell, J. T., . . . Lowenstein, C. J. (2010). P53-induced microRNA-107 inhibits HIF-1 and tumor angiogenesis. *Proc Natl Acad Sci U S A, 107*(14), 6334-6339. doi:10.1073/pnas.0911082107

Zhang, Y., & Reinberg, D. (2001). Transcription regulation by histone methylation: interplay between different covalent modifications of the core histone tails. *Genes Dev, 15*(18), 2343-2360. doi:10.1101/gad.927301

Figures

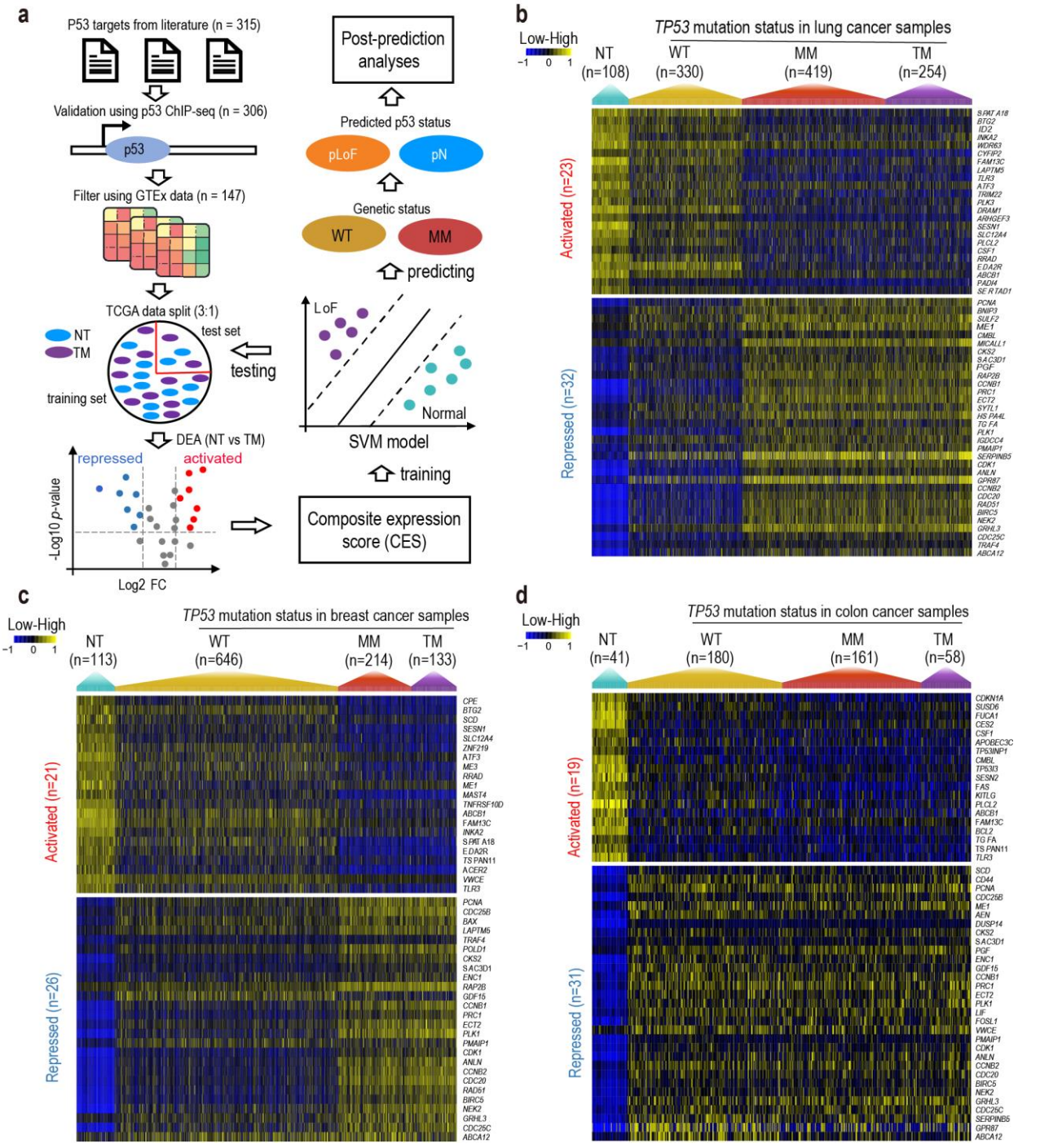


Figure 1. Identification of p53 target genes and their expression profiles in TCGA lung, breast, and colon cancers. (**a**) The analytic workflow to identify p53 target genes, train and validate SVM models. (**b-d**) Heatmaps show the expression profiles of p53-activated and repressed genes in TCGA lung, breast, and colon cancers. Samples in each cohort were divided into four groups according to TCGA designation, including NT (adjacent normal tissue), WT (tumor tissue with wild-type *TP53*), MM (tumor tissue with *TP53* missense mutation or in-frame mutation), and TM (tumor tissue with *TP53* truncating mutation). CES, composite expression score; DEA, differential expression analysis; FC, fold change; GTEx, Genotype-Tissue Expression; pLoF, predicted loss of function; pN, predicted normal; SVM, support vector machine.

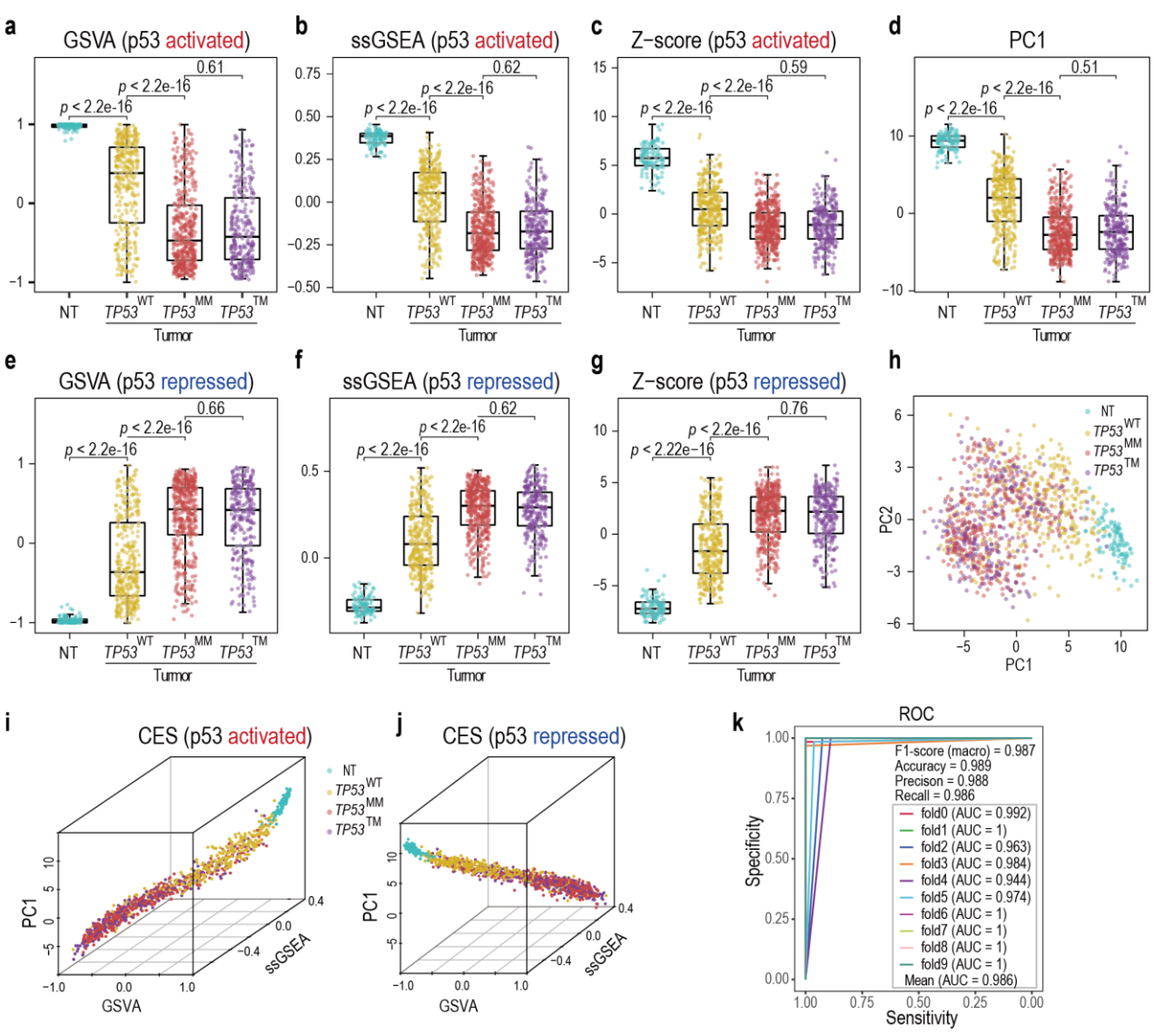


Figure 2. **Composite expression scores (CES) and the ROC curve of the SVM model built from the TCGA lung cancer cohort.** (**a-c**) CES scores of p53-activated genes calculated by GSVA, ssGSEA, and Z-score. In each panel, the blue, yellow, red, and purple box-and-whisker plots indicate the distribution of CES scores in NT, WT, MM, and TM groups, respectively. (**d**) CES score calculated by the first principal component scores (PC1). The p53-activated and p53-repressed genes were combined. (**e-g**) CES scores of p53-repressed genes calculated by GSVA, ssGSEA, and Z-score. (**h**) A two-dimensional PCA plot shows the clusters of NT, WT, MM, and TM groups. (**i-j**) Three-dimensional scatter plots show the combinatorial effects of CES score calculated from p53-activated genes and repressed genes, respectively. (**k**) The performance of the SVM model was evaluated using ten-fold cross-validation. GSVA, gene set variation analysis; ssGSEA, single-sample gene set enrichment analysis; PCA, principal component analysis; AUC, area under the curve; ROC, receiver operating characteristic; NT, normal tissue; *TP53*^{WT}, samples with wild-type *TP53*; *TP53*^{MM}, samples with *TP53* missense mutation or in-frame mutation; *TP53*^{TM}, samples with *TP53* truncating mutation.

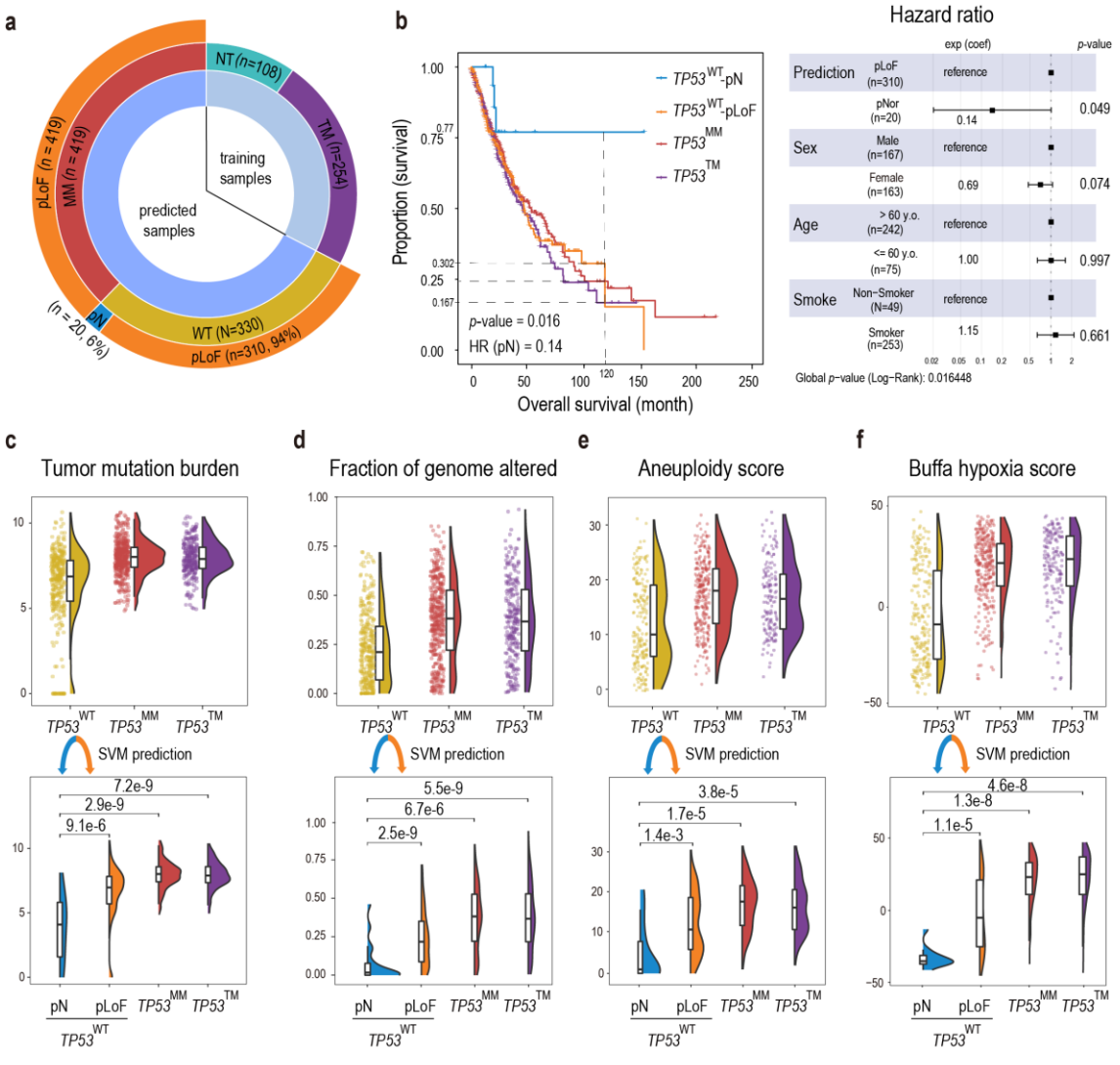


Figure 3. **Evaluation of the SVM predictions in the TCGA lung cancer cohort.** (**a**) A sunburst chart shows the breakdown of the lung cancer samples before (inner and middle layer circles) and after (outer layer circle) prediction. Samples used to train the SVM model include NT (adjacent normal tissue) and *TP53*^{TM} (tumor tissue with *TP53* truncating mutation) groups, representing p53-normal and p53-LoF status, respectively. Samples that were subjected to SVM prediction include *TP53*^{WT} (tumor tissue with wild-type *TP53*) and *TP53*^{MM} (tumor tissue with *TP53* missense mutation or in-frame mutation) groups. Samples in *TP53*^{WT} and *TP53*^{MM} groups were predicted to be normal (pN) or LoF (pLoF) by the SVM model. (**b**) Left: overall survival of patients with *TP53*^{MM}, *TP53*^{TM}, *TP53*^{WT}-pN (wild type TP53 which were predicted to be normal) and *TP53*^{WT}-pLoF (wild-type TP53 which were predicted to be loss of function) tumors. Right: forest plot of a Cox Proportional-Hazards model showing *p*-values, hazard ratios, and the 95% confidence intervals for covariates. (**c**) Upper panel: Comparison of tumor mutational burden (TMB) among *TP53*^{WT}, *TP53*^{MM}, and *TP53*^{TM} groups. Lower panel: the *TP53*^{WT} tumors were further divided into two subgroups (*TP53*^{WT}-pN and *TP53*^{WT}-pLoF) according to the SVM prediction. (**d**) Similar to (c), comparison of the copy number variation burden measured by "Fraction of genome altered". (**e**) Similar to (c), comparison of the aneuploidy score. (**f**) Similar to (c), comparison of hypoxia score.

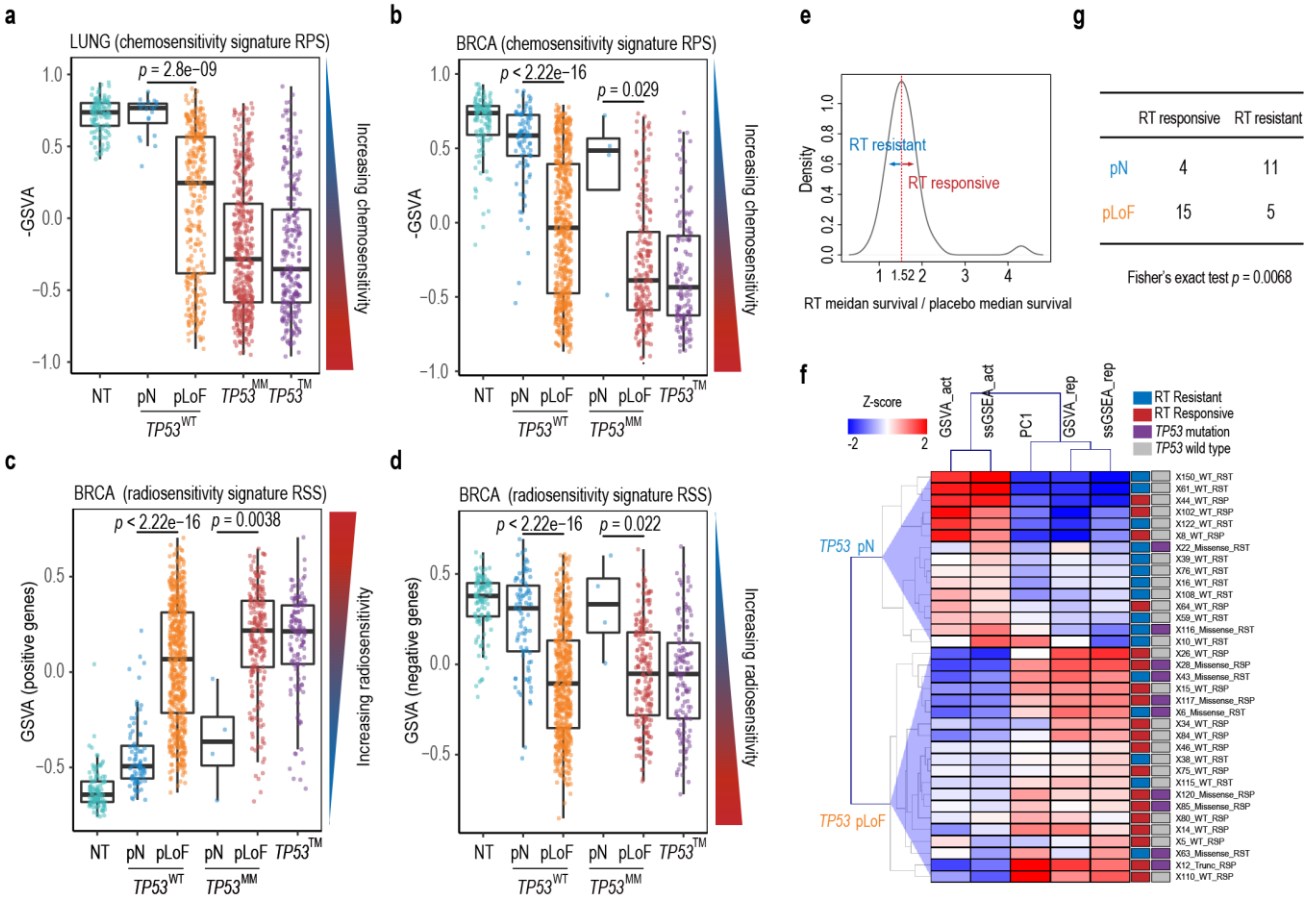


Figure 4. Relationships between p53 status and chemo-and radiation therapy sensitivities. (**a-b**) Comparison of the RPS scores (represented by -GSVA here, see Methods) amongst the NT, *TP53*^{WT}-pN, *TP53*^{WT}-pLoF, *TP53*^{MM}, and *TP53*^{TM} groups in the TCGA LUNG and BRCA cohorts, respectively. (**c-d**) Comparison of the RSS scores of positive and negative genes amongst NT, *TP53*^{WT}-pN, *TP53*^{WT}-pLoF, *TP53*^{MM}-pN, *TP53*^{MM}-pLoF, and *TP53*^{TM} groups in the BRCA cohort, respectively. RSS was calculated from genes positively associated with radiosensitivity using the GSVA algorithm. (**e**) Distribution of ratio of median survival days between the RT group and the placebo group. PDX models with ratios >= 1.52 were considered as RT-responsive. (**f**) A heatmap shows the p53 function status predicted by unsupervised clustering using composite expression (GSVA, ssGSEA, and PC1) of GBM specific p53 target genes in PDX data. RT response and *TP53* mutation status of each sample are color coded on the right side. (**g**) A contingency table shows that pLoF samples are associated with RT responsive.

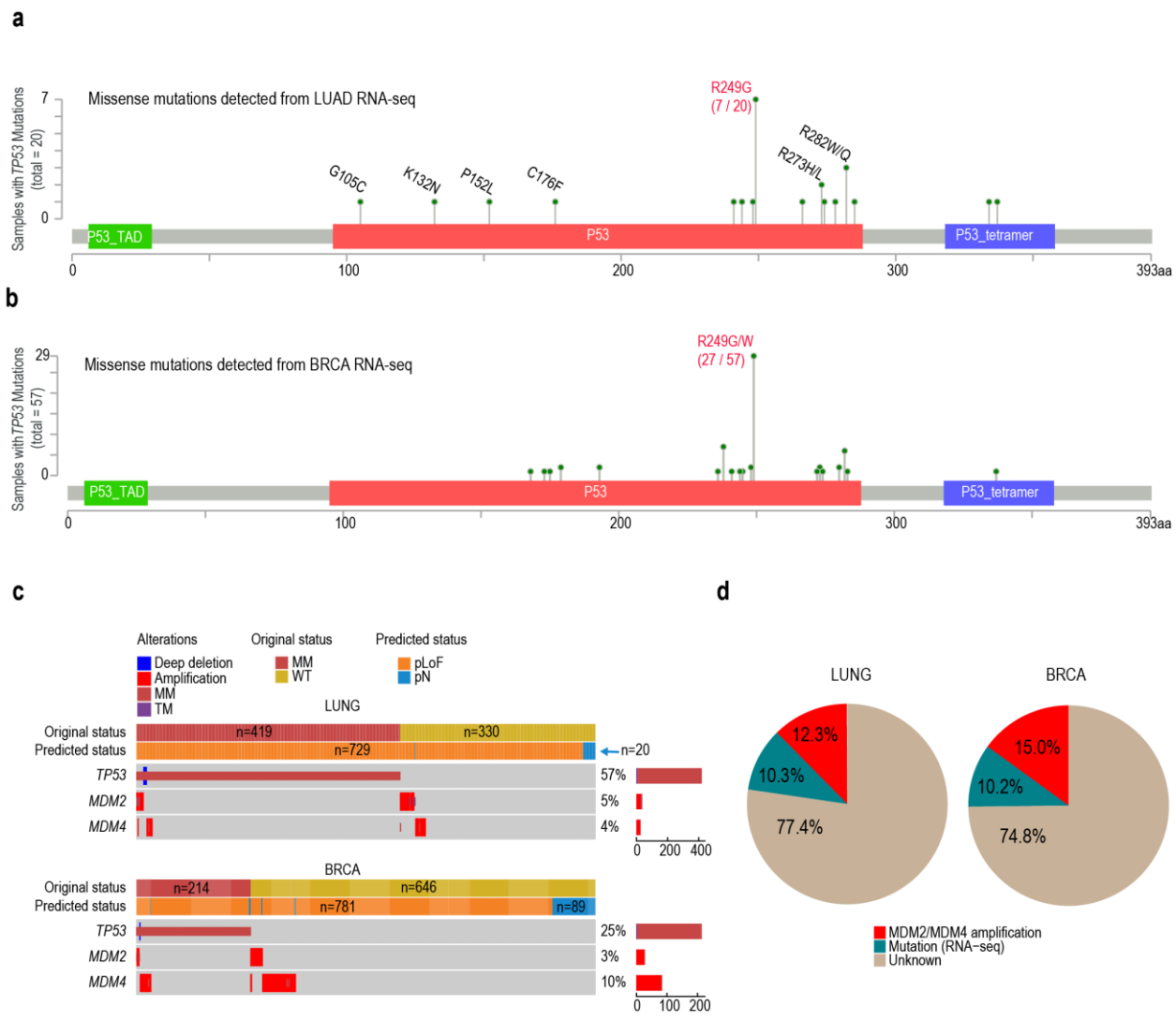
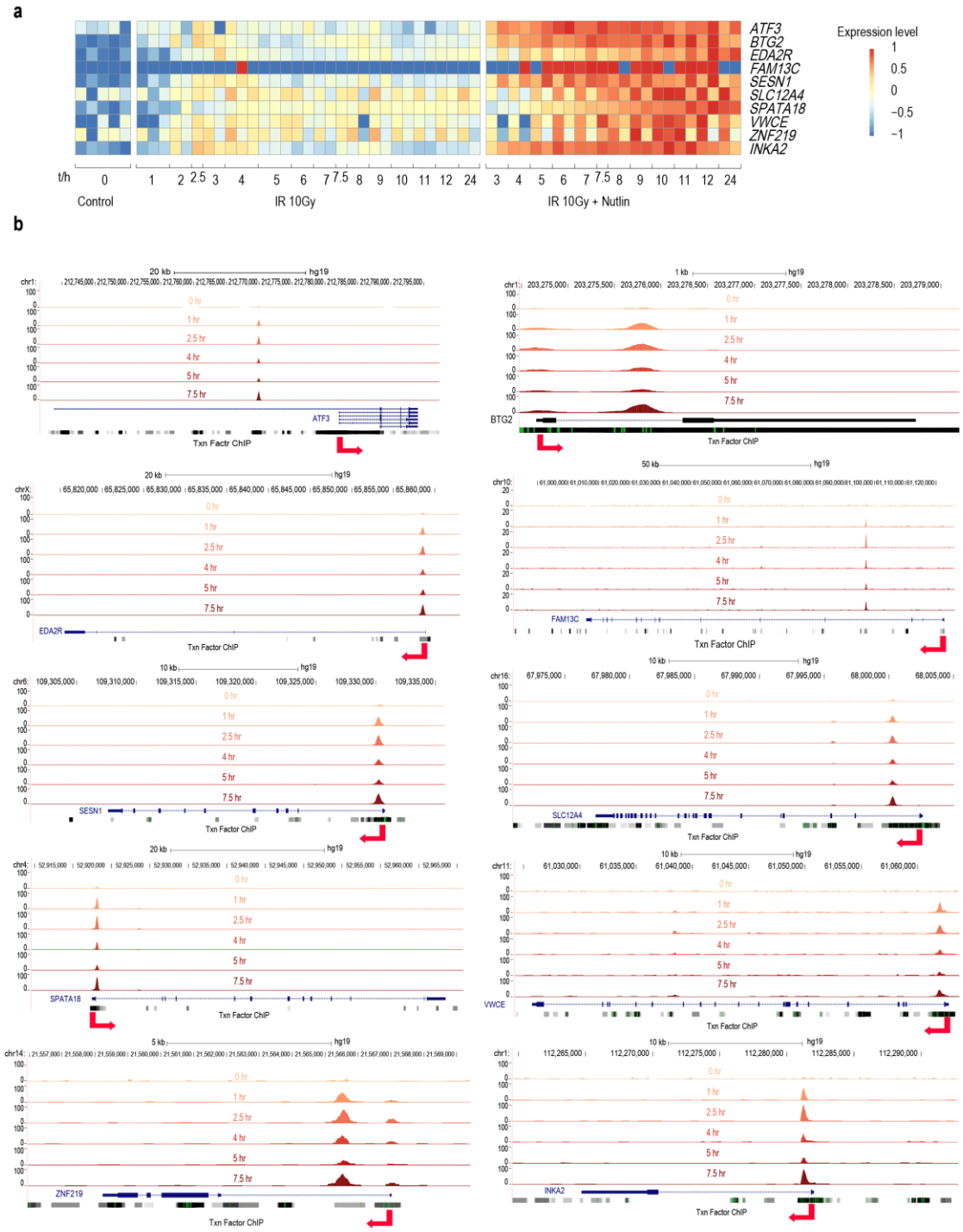


Figure 5. **Dissection of *TP53*^{WT}-pLoF cases.** (**a-b**) Lollipop charts show the position of missense mutations identified using RNA-seq data but missed from the whole-exome sequencing data in TCGA lung adenocarcinoma (LUAD) and breast invasive carcinomas (BRCA), respectively. (**c**) Oncoprint plot shows the *TP53*'s original and predicted status, comparing with the *MDM2* and *MDM4* amplification status. Percentages of samples with *TP53*/*MDM2*/*MDM4* variations among the prediction set (MM and WT samples) are listed on the right side, as well as the numbers of these samples (indicated by bar plots). (**d**) Pie charts show the fraction of *TP53*^{WT}-pLoF cases that can be explained by "mutation detected from RNA-seq" (green) and "*MDM2* or *MDM4*" amplification (red).

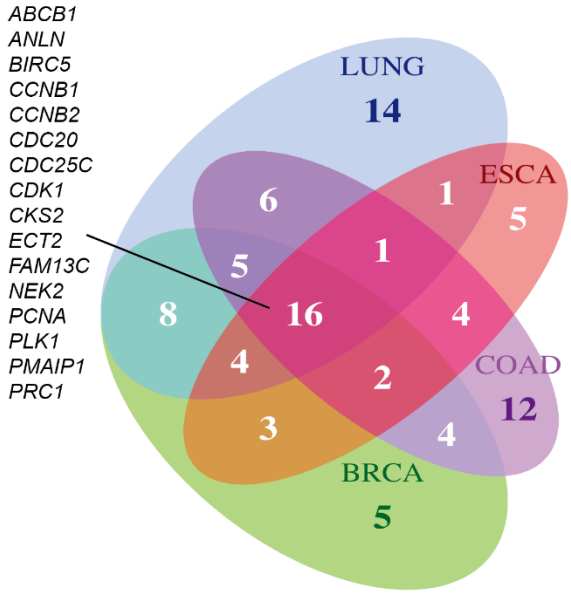
Supplementary figures

Supplementary Fig. 1
Validation of p53 target genes in γ irradiated MCF-7 cells harboring wild type *TP53*.
Supplementary Fig. 2
Core p53 targets among the four cancer cohorts (LUNG, BRCA, COAD and ESCA).
Supplementary Fig. 3
Correlations between different CES scores.
Supplementary Fig. 4
Comparison of CES score distributions among different *TP53* groups.
Supplementary Fig. 5
Pan-cancer analysis.
Supplementary Fig. 6
Distribution of predicted p53 status for $TP53^{WT}$ (WT group) and $TP53^{MM}$ (MM group) tumor samples across the five cohorts.
Supplementary Fig. 7
Comparing progression-free survival (PFS) of $TP53^{MM}$, $TP53^{TM}$, $TP53^{WT}$-pN and $TP53^{WT}$-pLoF patients of TCGA BRCA cohort.
Supplementary Fig. 8
Relationships between p53 status and chemo-and radiation therapy sensitivities after removing the overlapped genes.
Supplementary Fig. 9
Comparison of p53 expression levels between $TP53^{WT}$-pN and $TP53^{WT}$-pLoF samples in LUNG cohort.
Supplementary Fig. 10
Examples of *TP53* somatic mutations identified from RNA-seq data but missed by TCGA WES calls.
Supplementary Fig. 11
p53 mutants identified from RNA-seq data (indicated as the RM group) exhibit similar genomic characteristics as the $TP53^{MM}$ and $TP53^{TM}$ groups.
Supplementary Fig. 12
Genomics alteration status of *TP53* and the other well-known tumor suppressor genes across different TCGA cancer types.

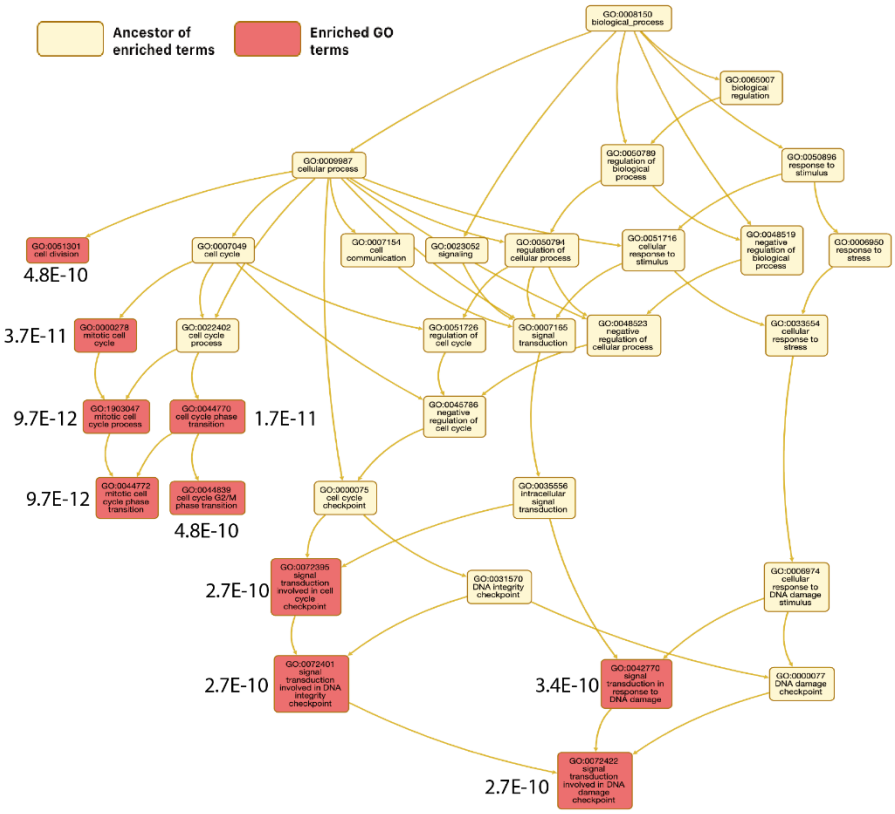


Supplementary Fig. 1. Validation of p53 target genes in γ irradiated MCF-7 cells harboring wild-type *TP53*. (a) Changes of expression levels of p53 targets after γ irradiation, and γ irradiation with nutlin treatment. (b) Visualization of p53 binding sites (p53 ChIP-seq signals) around the promoter regions of genes in (a).

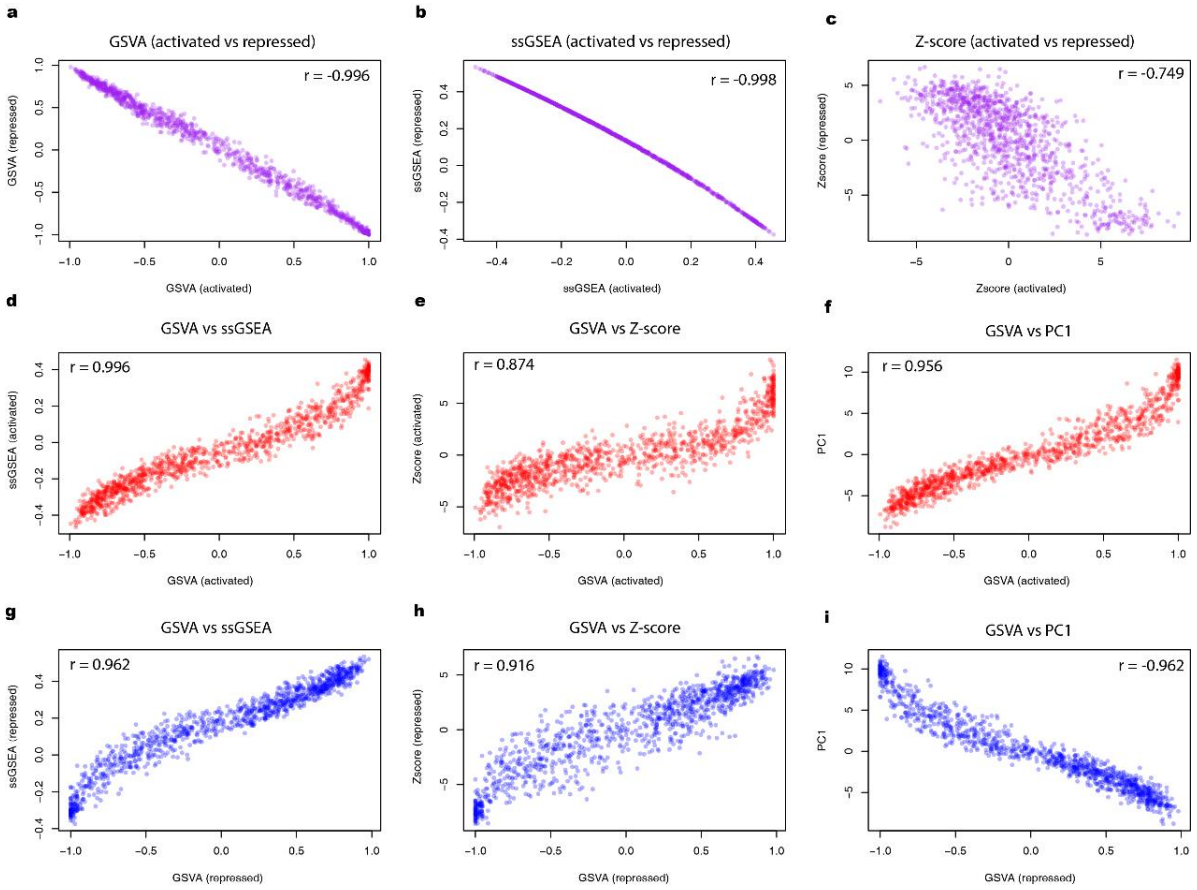
a



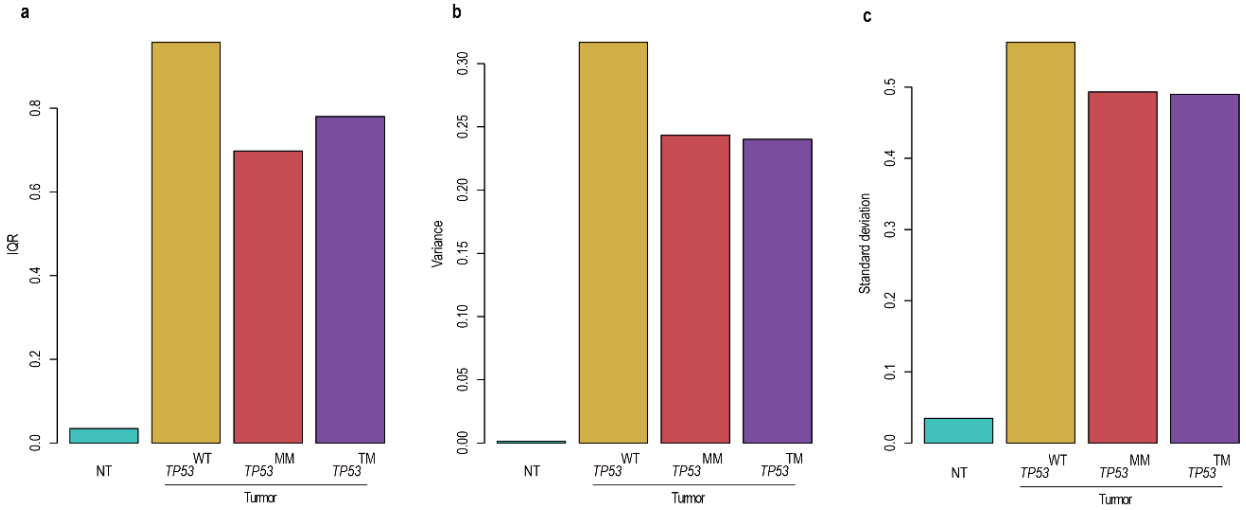
b



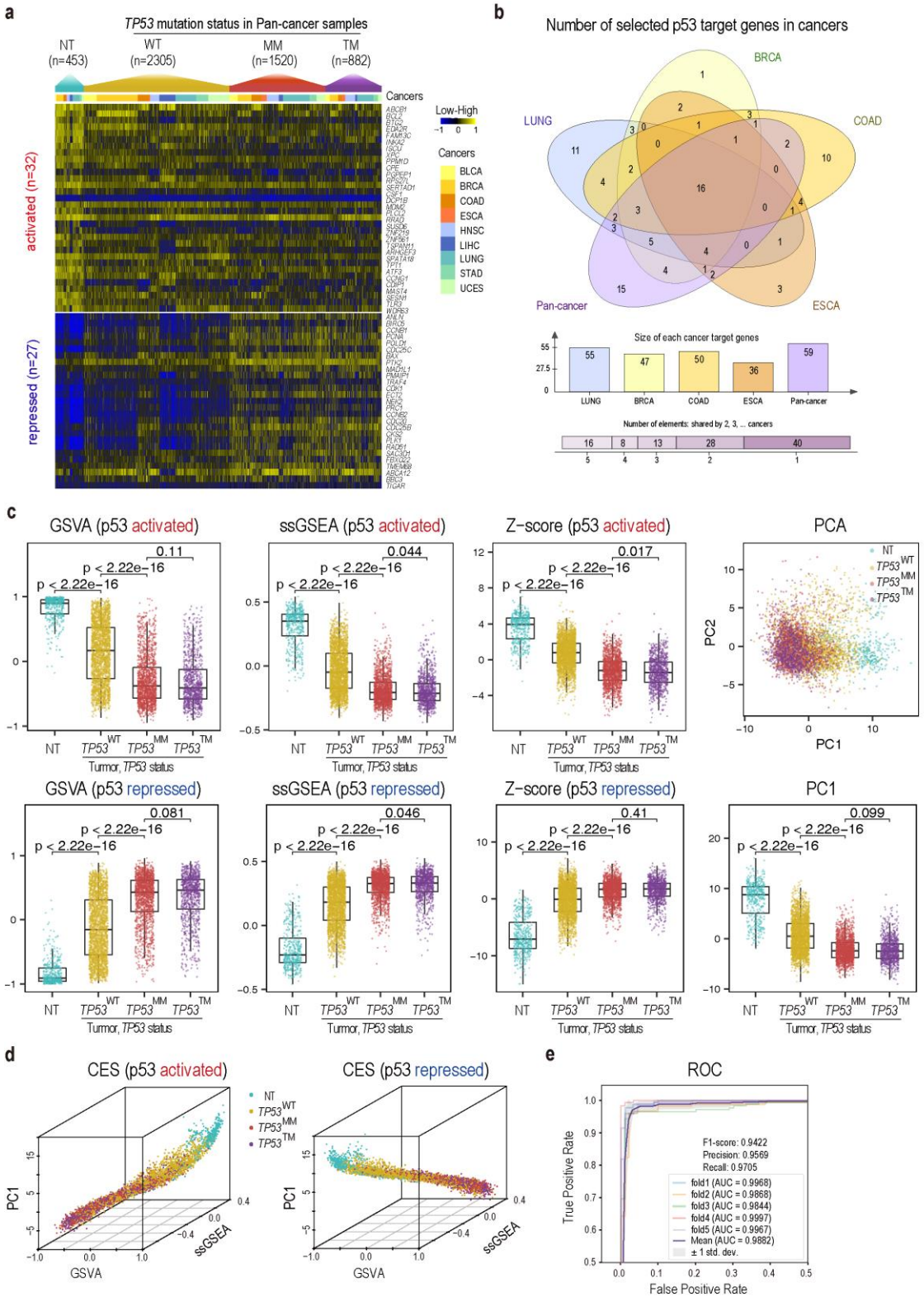
Supplementary Fig. 2. Core p53 targets among the four cancer cohorts (LUNG, BRCA, COAD and ESCA). (a) Venn diagram shows the overlaps of p53 targets among the four cancer types. (b) Gene Ontology (GO) enrichment results of the core p53 targets.



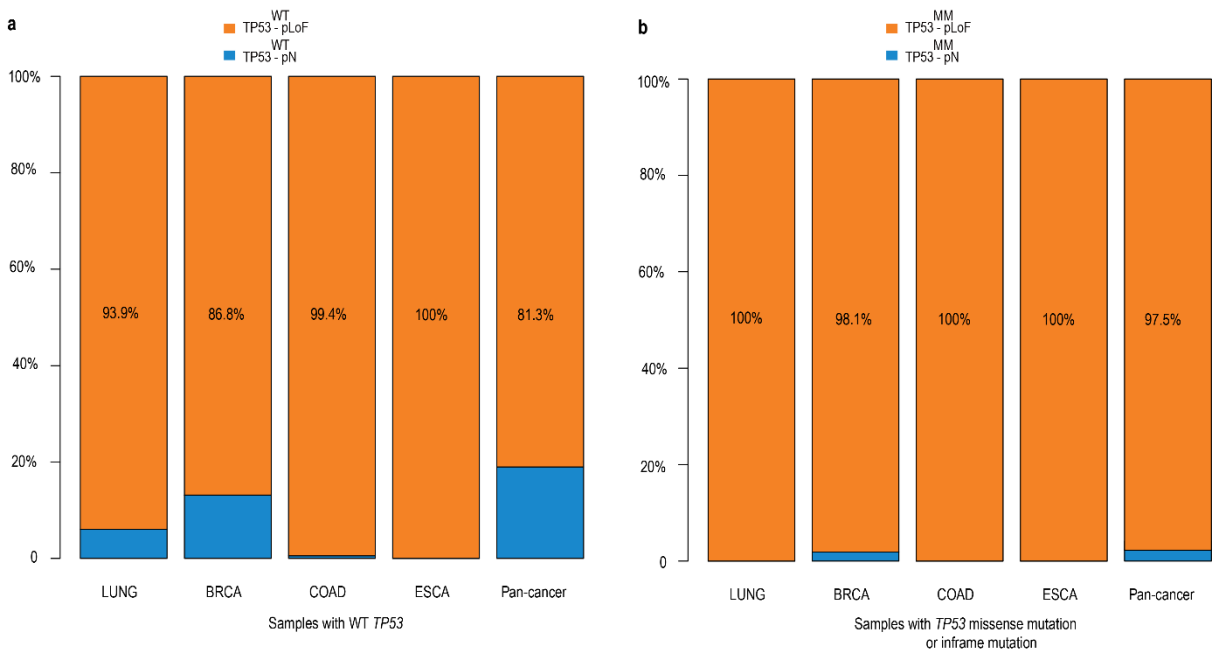
Supplementary Fig. 3. Correlations between different CES scores. (a-c) Significant inverse correlations between CES scores calculated from p53-activated and p53-repressed genes for GSVA, ssGSEA, and Z-score. Significant positive correlations between different CES scores for p53-activated genes (d-f) and between that for p53-repressed genes (g-h). (i) Significant inverse correlation between PC1 and the GSVA of p53-repressed genes.



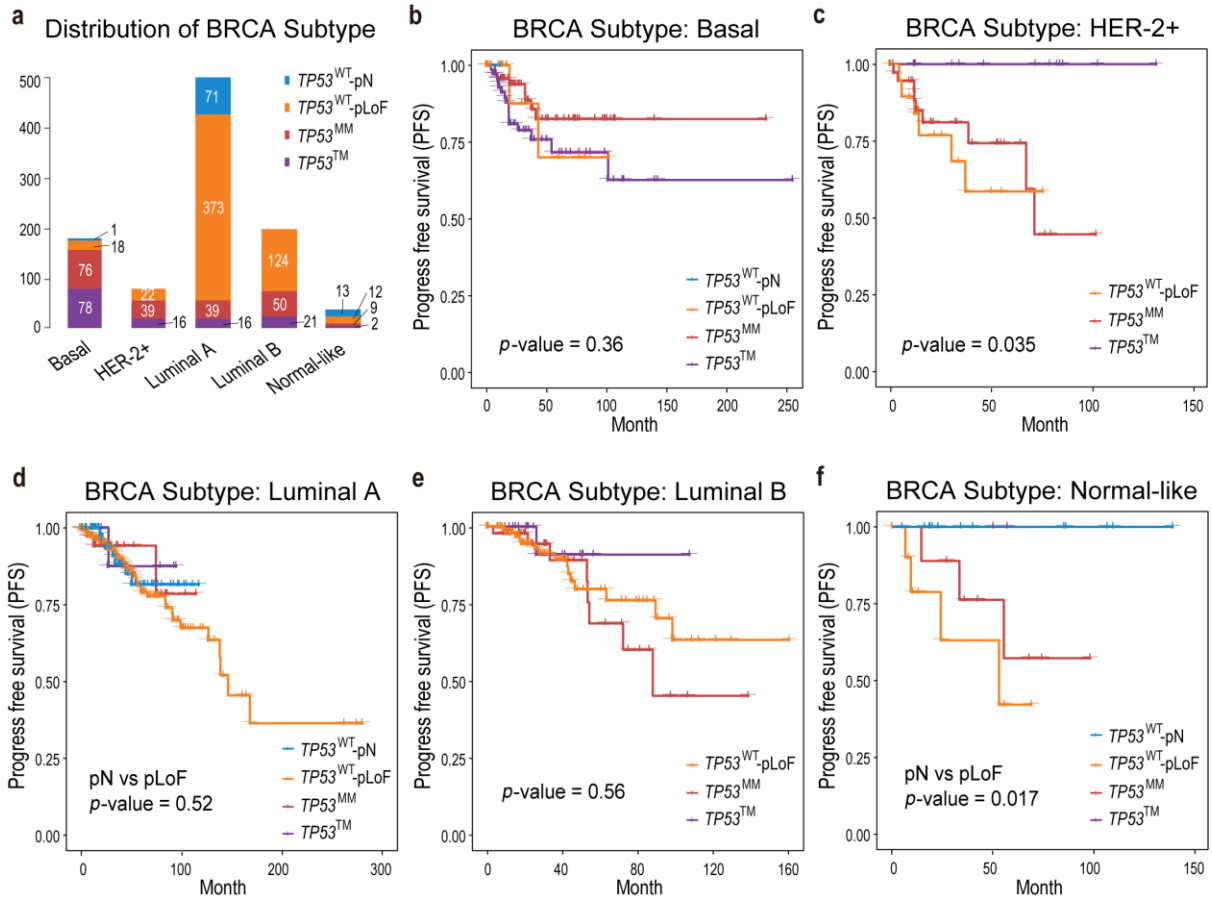
Supplementary Fig. 4. Comparison of CES score distributions among different *TP53* groups. Bar plots show the Interquartile Range (a), Variance (b), and Standard Deviation (c) of CES scores in NT, WT, MM, and TM groups.



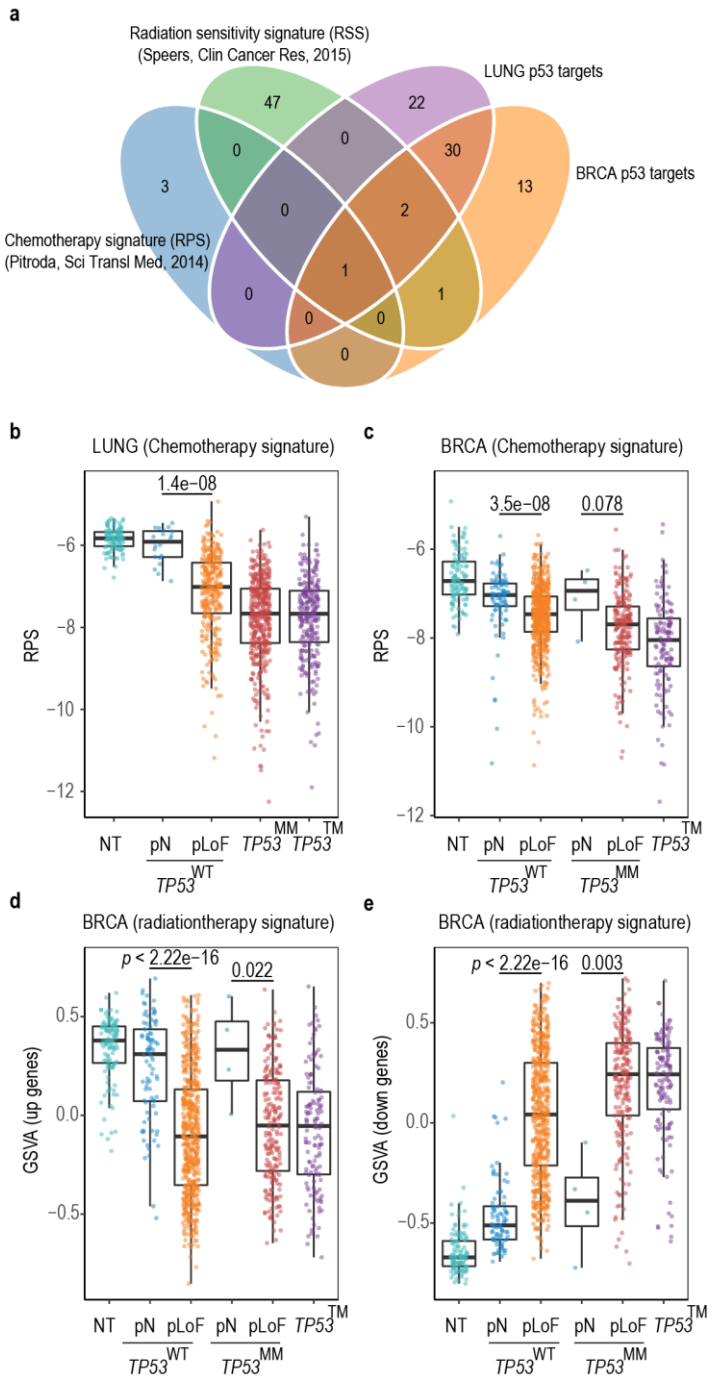
Supplementary Fig. 5. Pan-cancer analysis. (a) Heatmap shows the expression profiles of p53-activated genes and repressed genes across 9 cancer types. (b) Overlaps of p53 targets among pan-cancer cohort and four previously selected cancer cohorts. (c-d) CES scores calculated from p53 targets by different algorithms in pan-cancer cohort (similar to Figure 2). (e) ROC curve of the SVM model built from the pan-cancer cohort.



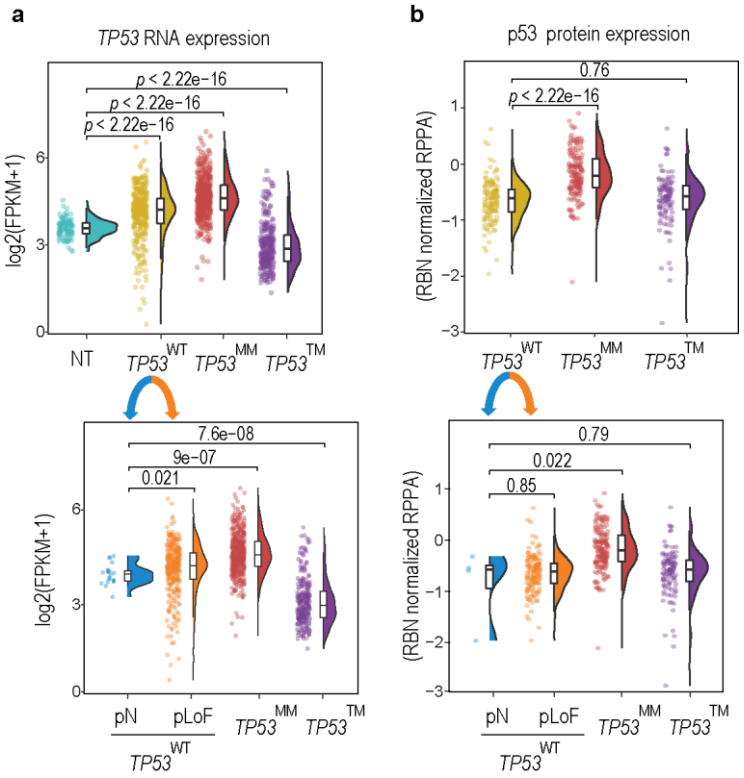
Supplementary Fig. 6. Distribution of predicted p53 status for *TP53*^{WT} (WT group) and *TP53*^{MM} (MM group) tumor samples across the five cohorts. Most tumor samples in WT (a) and MM (b) groups are predicted as LoF.



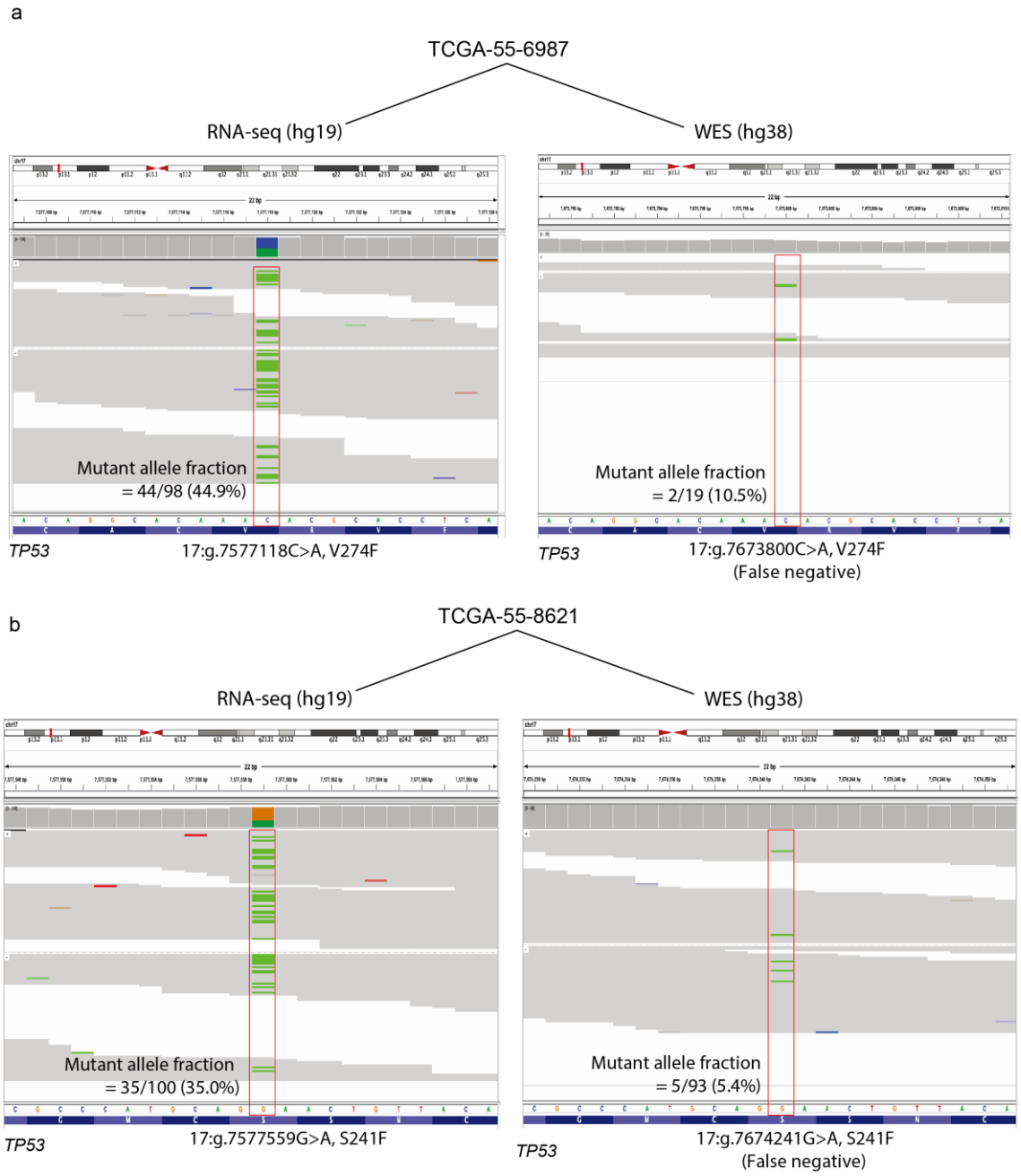
Supplementary Fig. 7. Comparing progression-free survival (PFS) of *TP53*^{MM}, *TP53*^{TM}, *TP53*^{WT}-pN and *TP53*^{WT}-pLoF patients of TCGA BRCA cohort. (a) Bar plot shows the breakdown of BRCA patients into five subtypes, including "Basal/triple negative", "HER2^{+}", "Luminal A", "Luminal B" and "Normal-like". Each subtype is further divided into four subgroups based on *TP53* status. (b-f) Comparison of PFS for Basal, HER2^{+}, Luminal A, Luminal B, and Normal-like subtype, respectively.



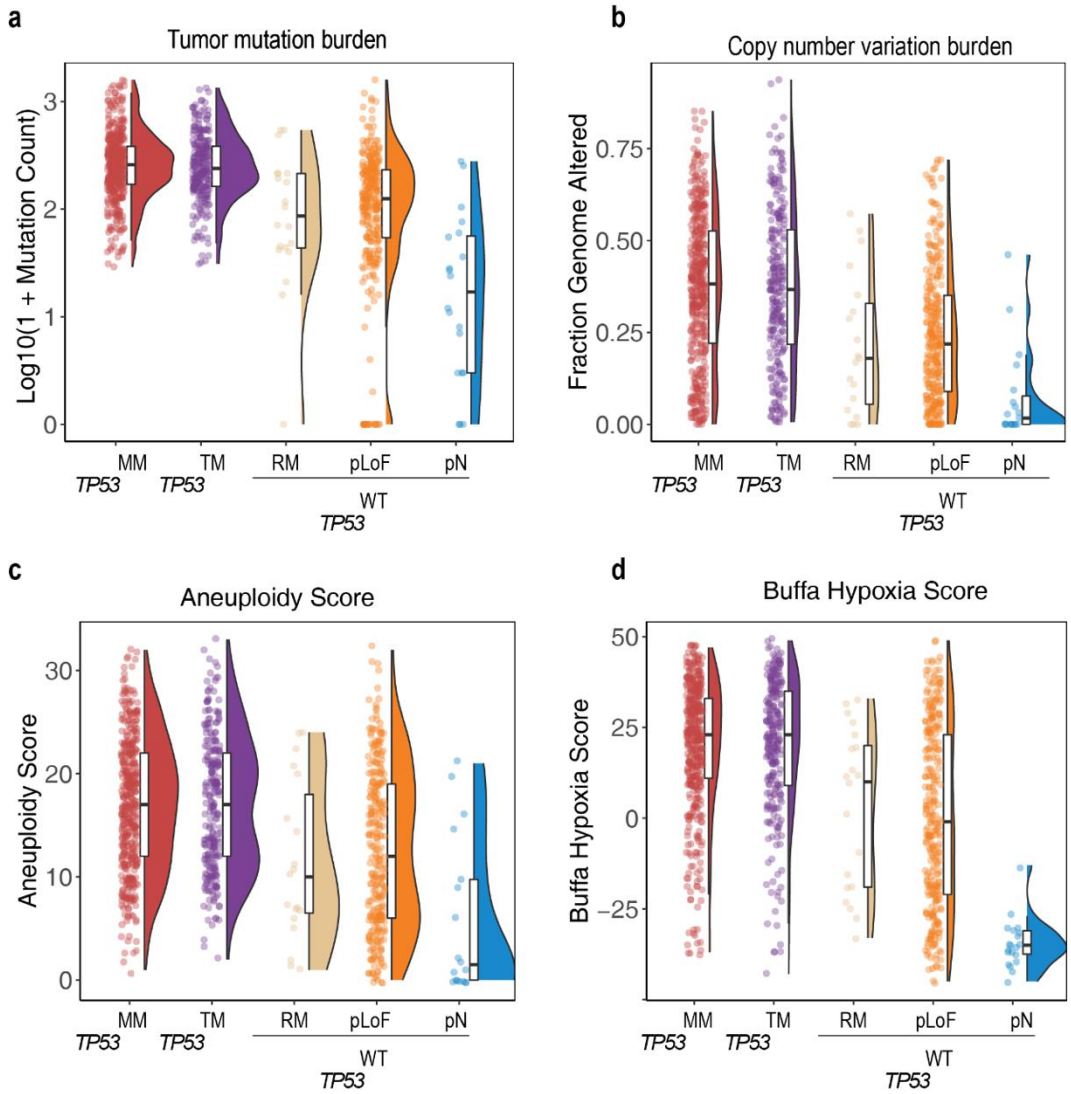
Supplementary Fig. 8. Relationships between p53 status and chemo-and radiation therapy sensitivities after removing the overlapped genes. (a) Venn diagram shows the overlap between p53 targets used in the LUNG SVM model, p53 targets used in the BRCA SVM model, the chemotherapy gene signature (recombination proficiency score, RPS), and the radiation sensitivity signature (RSS). (b-c) Comparison of the RPS scores (represented by -GSVA here, see Methods) amongst NT, *TP53*^{WT}-pN, *TP53*^{WT}-pLoF, *TP53*^{MM}, and *TP53*^{TM} in the TCGA LUNG and BRCA cohort, respectively (similar to Figure 4 a-b). (d-e) Comparison of the RSS scores amongst NT, *TP53*^{WT}-pN, *TP53*^{WT}-pLoF, *TP53*^{MM}-pN, *TP53*^{MM}-pLoF and *TP53*^{TM} in the TCGA BRCA cohort (similar to Figure 4 c-d). GSVA was used to calculate the PRS and RSS scores. Genes overlapped with p53 targets of either LUNG or BRCA were removed when calculating the RPS and RSS scores.



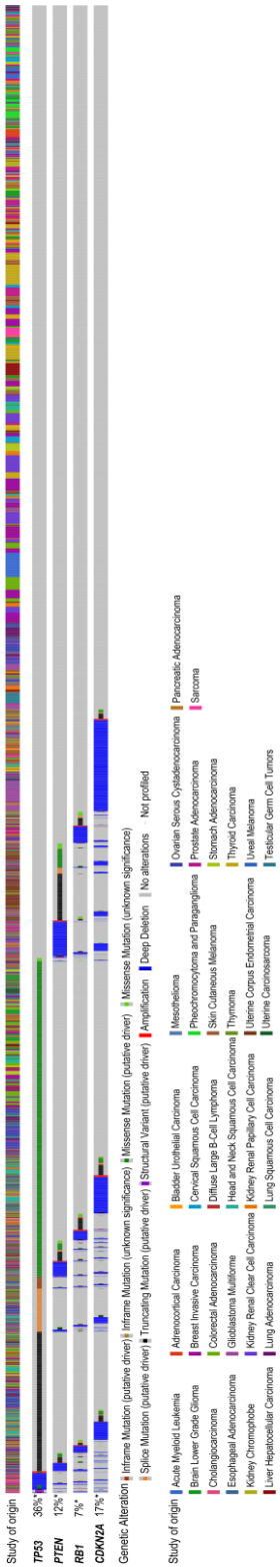
Supplementary Fig. 9. Comparison of p53 expression levels between *TP53*^{WT}-pLoF and *TP53*^{WT}-pN samples in LUNG cohort. (a) Comparison of the *TP53* RNA expression. (b) Comparison of the p53 abundance.



Supplementary Fig. 10. Examples of *TP53* somatic mutations identified from RNA-seq data but missed by TCGA WES calls. Mutant allele fraction (MAF) is measured by the ratio between the "number of reads supporting the mutant allele" and the "total number of reads" covering the mutation site.



Supplementary Fig. 11. p53 mutants identified from RNA-seq data (indicated as the RM group) exhibit similar genomic characteristics as the *TP53*^{MM} and *TP53*^{TM} groups. Compared to the *TP53*^{WT}-pN tumors, tumors of the RM group show increased tumor mutation burden (a), copy number variation burden (b), aneuploidy score (c), and Buffa hypoxia score (d).



Supplementary Fig. 12. Genomics alteration status of *TP53* and the other well-known tumor suppressor genes across different TCGA cancer types.

Supplementary tables

Supplementary tables are included in file "SupplementaryData1.xlsx" (table s1-s7 and s9) and "SupplementaryData2.xlsx" (table s8), which contain our main results and numeric data used to generate figures.

Supplementary Table 1
List of p53 target genes used in this study.

Supplementary Table 2
CES scores of p53 target genes in TCGA LUNG, BRCA, COAD, ESCA, Pan-cancer (including 12 types or subtypes) cohorts.

Supplementary Table 3
The performance metrics of the SVM model in TCGA lung and breast cohorts.

Supplementary Table 4
Genomic and clinical characteristics of TCGA LUNG, BRCA, COAD, ESCA, and Pan-cancer cohorts.

Supplementary Table 5
Gene signatures reflecting chemo- or radiosensitivity.

Supplementary Table 6
Median survival days and CES values of PDX models treated with placebo and radiation therapy (RT).

Supplementary Table 7
List of lung and breast cancer samples that have TP53 missense mutations detected from RNA-seq.

Supplementary Table 8
List of SVM training samples, prediction samples, TP53 original genetic states, SVM-predicted states and probabilities of TCGA cancers analyzed in our study.

Supplementary Table 9
Summary of SVM models according to the DOME (Data, Optimization, Model and Evaluation) recommendations.