Identify non-mutational p53 functional deficiency in human cancers

Qianpeng Li[1,3,4,*], Yang Zhang[1,3,4,*], Sicheng Luo[1,3,4,10], Zhang Zhang[1,3,4], Ann L. Oberg[2], David E. Kozono[5], Hua Lu[6], Jann N. Sarkaria[7], Lina Ma[1,3,4,#], Liguo Wang[2,8,9,#]

[1] China National Center for Bioinformation, Beijing 100101, China
[2] Division of Computational Biology, Mayo Clinic College of Medicine and Science, Rochester, MN 55905, USA
[3] National Genomics Data Center & CAS Key Laboratory of Genome Sciences and Information, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing 100101, China
[4] University of Chinese Academy of Sciences, Beijing 100101, China
[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
[10] China Sino-Danish College, University of Chinese Academy of Sciences, Beijing 100049, China

*Equal contributions

Author email:
Qianpeng Li: liqianpeng2019d@big.ac.cn
Yang Zhang: zhangyang@big.ac.cn
Sicheng Luo: luosicheng@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*'s functional status is critical for cancer genomic medicine. However, there is a significant challenge in identifying tumors with non-mutational p53 inactivations that are not detected through DNA sequencing. These undetected cases are often misclassified as p53-normal, leading to inaccurate prognosis and downstream association analyses. To address this issue, we build the support vector machine (SVM) models to systematically evaluate p53's functional status in multiple TCGA cancer cohorts. Cross-validation demonstrates the excellent performance of the SVM models with mean AUC of 0.9845, precision of 0.9844, and recall of 0.9825. Our model reveals that a significant proportion (87-100%) of wild-type *TP53* (*TP53*^{WT}) tumors actually have compromised p53 function. Further analyses reveal that these genetically normal but functionally deficient tumors (predicted reduced function of p53, *TP53*^{WT}-pRF) share genomic characteristics with tumors carrying p53 mutants. Specifically, they exhibit significantly increased tumor mutation burden, copy number variation burden, aneuploidy score, and hypoxia score. Clinically, patients with *TP53*^{WT}-pRF tumors experience significantly shorter overall survival and disease-free survival compared to those with *TP53*^{WT}-pN (predicted normal function of p53) tumors, and these patients also display heightened 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 through tumor DNA sequencing have significantly advanced our understanding of cancer genetics (Alexandrov et al., 2013; Stephens et al., 2009; Vogelstein et al., 2013) and have revolutionized health care by enabling accurate diagnoses, prognoses, and personalized medicine. An illustrative example is the sensitivity of breast cancers with loss-of-function mutations in *BRCA1*/*BRCA2* to PARP inhibitors (Bryant et al., 2005; Farmer et al., 2005). Genetic alterations resulting in the gain or loss of gene function are major drivers of cancer development, making the identification of causative and targetable genetic abnormalities a vital focus in cancer genomic research.

Tumor suppressor p53, encoded by the *TP53* gene, is a transcription factor that 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, with many have been experimentally validated (Fischer, 2017). Meanwhile, it stands out as the most frequently mutated tumor suppressor gene that undergoes genetic inactivation in 20 different types of cancer (Jia & Zhao, 2019). Somatic mutation frequencies of *TP53* exceed 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). The functional activities of p53 have been frequently associated with the response to 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). Moreover, p53 dysfunction is also linked to immunosuppression and immune evasion (Cortez et al., 2016; Guo, Yu, Xiao, Celis, & Cui, 2017; Maddalena et al., 2021). Therefore, accurately assessing the functional status of p53 is critical for prognosis and personalized medicine.

However, it is crucial to acknowledge that genetic alteration is not the sole mechanism responsible for disrupting or activating protein function. There is a growing recognition that p53 function can be inactivated even in tumors with genetically wild-type *TP53*. While sporadic studies have provided some support for this notion (Bosari et al., 1995; Isaacs, Hardman, Carman, Barrett, & Weissman, 1998; Ishizawa et al., 2018; Jeay et al., 2015; Moll, LaQuaglia, Benard, & Riou, 1995; Moll, Riou, & Levine, 1992; Tominaga et al., 1993), a systematic evaluation using large clinical samples is still lacking. The underlying mechanisms may involve various non-mutational factors such as post-translational modification (including phosphorylation, acetylation, methylation, and E3s mediated ubiquitination) (Chuikov et al., 2004; Kruse & Gu, 2009; E. W. Lee et al., 2009; Pan & Blattner, 2021; Xu, 2003), DNA methylation (Amatya, Naumann, Weller, & Ohgaki, 2005), chromatin states (Hafner, Bulyk, Jambhekar, & Lahav, 2019), and interplaying with microRNAs (Pollutri, Gramantieri, Bolondi, & Fornari, 2016; Suh et al., 2012). These molecular mechanisms are widespread, but their heterogeneous nature presents a formidable challenge in characterizing them. Thus, a comprehensive assessment that goes beyond *TP53* mutation is necessary for a more comprehensive understanding of p53's functional status in tumors.

In this study, our objective is to re-assess functional status of p53 in tumors with wild-type *TP53.* We hypothesized that both mutational and non-mutational inactivation of p53 could be reflected by the altered expression of genes regulated by p53. Toward this end, we first defined cancer type-specific gene sets (encompassing both direct and indirect targets of p53), which can serve as indicators of p53's functional status. This was accomplished through a systematic literature review and differential expression analysis (DEA) of RNA-seq data. We then calculated the composite expression score (CES) derived from these genes using various 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 CESs from non-cancerous normal tissues (referred to as the "NT" group, assumed to have normal p53 function) and tumor samples harboring *TP53* truncating mutations (referred to as the "TM" group, assumed to have lost or reduced p53's tumor suppressor function). These SVM models were applied to tumors with TP53 wildtype (referred to as the “WT” group) and tumors with missense mutations (referred to as the “MM” group).

We evaluated the performance of our SVM models using a testing dataset and compared genome and chromosome instability measurements to examine the prediction results. As anticipated, nearly all samples with missense mutations were predicted to have reduced function (pRF). Interestingly, a significant proportion of $TP53^{WT}$ tumors were also predicted to have reduced p53 function (termed $TP53^{WT}$-pRF). This suggests a functional deficiency in p53 despite the absence of genetic mutations. Further analyses revealed that $TP53^{WT}$-pRF tumors exhibit distinct genomic characteristics compared to other $TP53^{WT}$ tumors predicted to be normal (termed $TP53^{WT}$-pN). $TP53^{WT}$-pRF tumors displayed significantly increased tumor mutation burden (TMB), copy number variation burden, aneuploidy score, and hypoxia score, consistent with p53’s role as a regulator of DNA damage repair and the “guardian of the genome”. Based on our analyses, it has been determined that around 22-25% of $TP53^{WT}$-pRF can be attributed to either false-negative results (where *TP53* mutations are identified through RNA-seq data, but go undetected in the WES assay) or amplification of *MDM2* and *MDM4*.

Clinically, $TP53^{WT}$-pRF patients show significantly reduced overall and disease-free survival rates compared to $TP53^{WT}$-pN patients. Additionally, our data demonstrate that $TP53^{WT}$-pRF tumors exhibited significantly heightened sensitivities to radio- and chemotherapy, offering valuable insights for personalized medicine. The increased sensitivity of $TP53^{WT}$-pRF tumors to radiation therapy is

further supported by the patient-derived xenografts models (PDX) of glioblastomas.

Results

Identification of cancer-specific gene sets reflecting p53's functional status

We identified a set of p53-regulated genes whose expression levels are statistically associated with p53 truncation for each cancer type, enabling their utilization as features for predicting the functional status of p53 status in these cancers. First, we compiled a list of 147 highly reliable target genes, comprising both direct and indirect targets, from a comprehensive study conducted by Fischer (Fischer, 2017). Then, we examined the expression profiles of these 147 genes using RNA-seq from the GTEx datasets, while also assessing p53 bindings using the ReMap2020 ChIP-seq database (Cheneby et al., 2020) (**Figure 1 a**). These genes were found to be reliably expressed across various normal tissues, and their promoters displayed evidence of p53 binding (**Figure 1—Source Data 1**). To further validate these p53-regulated genes, we reanalyzed RNA-seq and ChIP-seq data obtained from MCF-7 breast cancer cells that possess wild-type *TP53*, both before and after p53 activation. Expectedly, minimal or weak p53 binding was observed at the promoters of p53-regulated genes (such as *ATF3*, *BTG2*) prior to p53 activation induced by gamma irradiation. However, upon p53 activation, significant p53 bindings were observed, accompanied by consistent upregulation in the expression levels of these genes (Hafner et al., 2017) (**Figure 1—figure supplement 1**).

The transcriptional programs regulated by p53 vary by tissue (Bouvard et al., 2000; Burns, Bernhard, & El-Deiry, 2001; Karsli Uzunbas, Ahmed, & Sammons, 2019). To define a gene set that can serve as features for predicting p53's functional status in each selected cancer type, we conducted gene expression analyses between NT and TM groups. This analysis was performed using TCGA RNA-seq data for the previously selected 147 genes in each cancer type (**Figure 1 a**). As a result, we identified

a set of genes that were down-regulated (referred to as p53tru-DR genes) and a set of genes that were up-regulated (referred to as p53tru-UR genes) in the TM group for each specific cancer type. It should be noted that p53tru-UR genes primarily represent genes whose expression is negatively regulated by normal p53, and the underlying mechanisms are likely indirect and not yet fully understood (Fischer, 2017). Finally, we identified 47, 55, 36, and 50 p53 regulated genes for BRCA, LUNG, ESCA, and COAD, respectively. Among them, 16 genes (*ABCB1*, *ANLN*, *BIRC5*, *CCNB1*, *CCNB2*, *CDC20*, *CDC25C*, *CDK1*, *CKS2*, *ECT2*, *FAM13C*, *NEK2*, *PCNA*, *PLK1*, *PMAIP1*, *PRC1*) were shared across all four cancer types, and it is likely that 14 of these genes are indirectly regulated by p53. Moreover, these 16 genes demonstrated significant enrichment in well-known p53-associated pathways, including cell cycle control and DNA damage response (**Figure 1—figure supplement 2**).

Composite expression of p53tru-DR and p53tru-UR genes

We analyzed the expression profiles of p53tru-DR and p53tru-UR genes across four distinct groups: NT, WT, MM and TM. As anticipated, p53tru-DR and p53tru-UR genes exhibited the opposite trends; the expressions of p53tru-DR genes were significantly decreased in the TM group, whereas the expression of p53tru-UR genes were significantly increased in the TM group, consistent with the compromised p53 status in this group (**Figure 1 b-d**). Interestingly, similar expression patterns of these p53-regulated genes were observed between the MM and TM groups (**Figure 1 b-d**). This suggests impaired p53 function in the MM samples, and indicates that missense/in-frame mutations and truncating mutations have a comparable impact on p53's cellular activity. This observation aligns with the fact that most missense mutations occur within the DNA binding domain, leading to the disruption of p53's ability to bind to DNA and transactivate its downstream targets (Kato et al., 2003).

In the LUNG and BRCA cohorts, the WT group displayed an intermediate state between the NT and TM groups, suggesting many *TP53*^{WT} tumors have compromised p53 function despite their lack of

genetic mutation (**Figure 1 b-c**). In the COAD cohort, the expression profile of the *TP53*^{WT} tumors was almost indistinguishable from those of the mutant groups (MM + TM) (**Figure 1 d**). These findings collectively indicated a reduced tumor suppressor function of p53 in *TP53*^{WT} samples.

Training and evaluating SVM model based on the expression of p53-regulated genes

Our hypothesis was the inactivation of p53 could be predicted by examining the altered expression of its transcriptional targets, including both directly and indirectly regulated genes, regardless of mutational status of *TP53* gene. To test this hypothesis, we calculated the composite expression score (CES) of p53tru-DR and p53tru-UR genes using four different 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 (**Figure 2—Source Data 1**). Compared to the expression of individual genes, the CES not only provides a combined and stable measure of p53 activity but also reduces the dimensionality of the expression data and helps mitigate potential overfitting of the SVM model.

Using the LUNG dataset as an example, we observed significant inverse correlations between CESs calculated from p53tru-DR and p53tru-UR genes. The Pearson's correlation coefficients were -0.996, -0.998, and -0.749 for GSVA, ssGSEA, and combined Z-score respectively (**Figure 2—figure supplement 1 a-c**). These results indicated that both p53tru-DR and p53tru-UR genes could effectively reflect p53's functional status. In addition, we observed significant positive correlations (Pearson's correlation coefficients ranged from 0.874 to 0.996) among CESs calculated by different algorithms, indicating a high level of concordance among these algorithms (**Figure 2—figure supplement 1 d-i**). When comparing the CESs across the NT, WT, MM, and TM groups, we found CESs of the NT samples tightly congregated within a narrow range. In contrast, CESs of the tumor samples (including WT, MM, and TM groups) showed a higher degree of dispersion, indicating the increased heterogeneity in p53 activity among tumor samples (**Figure 2—figure supplement 2**). Consistent with the gene-level expression data (**Figure 1 b**), CESs of WT samples were intermediate between those of

NT and TM groups, while the CESs of the MM samples resembled those of the TM samples (**Figure 2 a-j**). These results further support the notion that the majority of WT samples exhibit compromised p53 function.

Next, we proceeded to train SVM models using seven CESs (GSVA, ssGSEA, and Z-score for p53tru-DR and p53tru-UR genes respectively, and PCA for all genes) with the purpose of predicting p53's functional status in *TP53*^{WT} and *TP53*^{MM} tumors. To achieve the best performance, these models were independently trained and validated for the four TCGA cancer types including LUNG, BRCA, COAD, and ESCA. To achieve an unbiased validation, we adopted the following strategy. Initially, we randomly partitioned 75% of the NT and TM samples as the training set, while the remaining 25% were reserved as the testing set. Subsequently, within the training set, we identified the p53tru-UR and p53tru-DR genes and constructed the SVM models. The performance of these models was evaluated using the testing set, which notably excluded samples used for feature gene identification and SVM model construction. This validation process, termed hold-out validation, was iterated ten times, as visually depicted in Figure 1a. As exemplified by the LUNG cohort, the mean precision, recall, F1-score, and AUROC were 0.988 ± 0.013, 0.986 ± 0.02, 0.987 ± 0.016, and 0.986 ± 0.02, respectively (**Figure 2 k**). Similar performance was achieved in the BRCA cohort (**Figure 2—Source Data 2**). Due to the smaller sizes of the cohorts and the limited number of truncation cases in other TCGA cancer types, it was impractical to train separate SVM model for these cancer types. Instead, we created a pan-cancer cohort (n = 5,160) and train specific SVM model by combining nine cancer types with relatively high *TP53* mutation rates (**Figure 2—figure supplement 3**). Despite the intrinsic heterogeneity of different cancer types, the pan-cancer SVM model demonstrated outstanding performance (as evaluated using five-fold cross-validation, **Figure 2—figure supplement 3 e, Figure 2—Source Data 2**). However, it should be noted that building a pan-caner SMV model is a compromise when individual cancer types lack sufficient samples to train their own SVM models.

Furthermore, the pan-cancer SVM model may be influenced by a few cancer types with exceptionally larger sample sizes, potentially introducing bias.

Predicting p53 status using SVM models

We examined the predictive power of our SVM models by applying them to *TP53*^{WT} and *TP53*^{MM} tumors, which were deliberately excluded from the training and testing phases. To avoid potential confusion caused by the term "p53 loss of function (LoF)", which typically refers to p53 dysfunction due to mutations, we designated samples predicted to be similar to *TP53*^{TM} samples as "reduced function (RF)" instead. The results revealed that the majority of *TP53*^{WT} samples (87%-100%) and almost all *TP53*^{MM} samples (98%-100%) were predicted to be RF in all five cohorts: LUNG, BRCA, COAD, ESCA, and the pan-cancer cohort (**Figure 3—figure supplement 1**). For instance, in the LUNG cohort, 94 % of the *TP53*^{WT} samples and 100 % of *TP53*^{MM} samples were predicted to be RF (**Figure 3 a)**. These results indicate a prevalent inactivation of p53 irrespective of the *TP53* genetic status, likely occurring through non-mutational mechanisms (Cross et al., 2011; Gogna, Madan, Kuppusamy, & Pati, 2012; Sasaki, Nie, & Maki, 2007) .

To assess the prognostic value of the SVM prediction, we analyzed the LUNG cohort by dividing *TP53*^{WT} tumors into two subgroups: *TP53*^{WT}-pRF, representing samples predicted to be have reduced p53 function, and *TP53*^{WT}-pN, representing samples predicted to have normal p53 function. The comparison of overall survival (OS) in the LUNG cohort revealed the 10-year OS rates for *TP53*^{WT}-pN (n = 20) and *TP53*^{WT}-pRF (n = 310) were 77.0% and 30.2%, respectively. As reference points, the 10-year OS rates for *TP53*^{MM} (n = 419) and *TP53*^{TM} (n = 254) were 24.3% and 16.7%, respectively (**Figure 3 b**). After adjusting for demographic variables such as sex, age, and smoking status using Cox regression, lung cancer patients with *TP53*^{WT}-pRF tumors exhibited a significant 61% decrease in survival compared to those with *TP53*^{WT}-pN tumors (P = 0.016). In contrast, the overall survival of

patients with *TP53*^{WT}-pRF tumors did not differ significantly from that of patients with *TP53*^{MM} or *TP53*^{TM} tumors. In 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. Within this subgroup, the *TP53*^{WT}-pRF patients exhibited significantly reduced prognosis compared to the *TP53*^{WT}-pN group (P = 0.017, log-rank test) (**Figure 3—figure supplement 2 f**). Normal-like tumors were generally considered as "artifacts" resulting from a high percentage of normal specimens or slow-growing basal-like tumors (Parker et al., 2009). However, our data suggested that the PFS is significantly reduced when p53 function is compromised, supporting the classification of normal-like tumors as a distinct subtype rather than mere normal tissue contamination. In summary, in both lung and breast (normal-like) cancers, *TP53*^{WT}-pRF tumors exhibit a significantly worse prognosis compared to the *TP53*^{WT}-pN tumors.

Considering the crucial role of p53 in DNA damage repair (Lane, 1992; Levine, 1997), we hypothesized that p53-defective tumors would accumulate more DNA damage compared to p53 normal tumors. To investigate this hypothesis, we compared measurements of genome and chromosome instability, including TMB, copy number variation burden, and aneuploidy score, between the *TP53*^{WT}-pRF and *TP53*^{WT}-pN groups. The analysis revealed significantly higher levels of genome instabilities in the *TP53*^{WT}-pRF group compared to the *TP53*^{WT}-pN group, indicating deficient DNA damage repair in *TP53*^{WT}-pRF tumors (**Figure 3 c-e**). Similar trends were observed in the other TCGA cohorts (**Figure 3—Source Data 1**). It is known that hypoxia can induce p53 activation (An et al., 1998) and that p53 can 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}-pRF, *TP53*^{MM} and *TP53*^{TM} tumors (**Figure 3 f**). Overall, TP53^{WT}-pRF tumors

exhibited increased genomic instability, worse prognosis, and higher Buffa hypoxia score, resembling the p53 mutant tumors. These data not only functionally reaffirm our SVM predictions but also highlight the necessity to reconsider the determination of *TP53* status based solely on DNA sequencing.

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

The impact of mutant p53 on the response to chemo- and radiation therapy remains controversial. Several studies have associated p53 mutations with reduced sensitivity to these therapies (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 have suggested that p53 inactivation actually increases tumor's sensitivity (Bertheau et al., 2002; Bertheau et al., 2007; Concin et al., 2000). Unfortunately, direct comparisons of therapy efficacy using TCGA samples are not feasible due to the lack of clinical responsiveness data. To overcome this limitation, we assessed the chemo- and radiation sensitivity using previously published gene signatures (**Figure 4—Source Data 1**) and conducted investigations with preclinical animal models to examine the therapeutic effects of radiation. Specifically, we evaluated chemotherapy sensitivity in LUNG and BRCA cohorts using RPS (recombination proficiency score). The RPS correlated positively with DNA recombination proficiency and negatively with the sensitivity of platinum-based chemotherapy, as clinically validated in breast cancer and NSCLC patients (Pitroda et al., 2014). For the assessment of tumor radiation sensitivity, we utilized RSS (radiation sensitivity signature) score and tested in breast cancer. The RSS score was calculated based on the 51-gene signature validated in patients (Speers et al., 2015) (see Methods). To explore the relationship between predicted p53 status and the response to radiation therapy (RT) *in vivo*, we analyzed an independent dataset consisting of 35 patient-derived xenografts (PDX) models that closely mimic the genetic and phenotypic characteristics of glioblastoma (GBM) patients (Vaubel et al., 2020).

When comparing chemo-sensitivity across groups with different p53 statuses in the LUNG cohort, we found that tumors in the NT group exhibited 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}-pRF tumors ($P = 2.8\times10^{-9}$, two-sided Wilcoxon test). Similar results were observed in the BRCA cohort. Notably, 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). While mutations in *BRCA1*/*BRCA2* are commonly known as the primary 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., 2022). Moreover, studies have indicated the involvement of p53 in regulating homologous recombination (Gatz & Wiesmuller, 2006; Sengupta & Harris, 2005). These data suggested that platinum-based chemotherapy may provide greater benefits to *TP53*^{WT}-pRF tumors compared to *TP53*^{WT}-pN tumors.

We assessed radiosensitivity by dividing the RSS gene signature into positive (i.e., genes positively correlated with radiosensitivity) and negative (genes negatively correlated with radiosensitivity) subsets. Our analysis focused on the TCGA BRCA cohort since the RSS signature was derived from breast cancer. When comparing *TP53*^{WT}-pN tumors with *TP53*^{WT}-pRF tumors using positive genes, we found a significant increase in RSS scores for *TP53*^{WT}-pRF tumors (**Figure 4 c**), indicating heightened radiosensitivity. Meanwhile, when measured using negative genes, *TP53*^{WT}-pRF tumors showed significantly decreased RSS scores, also suggesting increased radiosensitivity (**Figure 4 d**). On the other hand, *TP53*^{MM}-pN tumors displayed significantly reduced radiosensitivity compared to *TP53*^{MM}-pRF tumors (**Figure 4 c-d**).

Although there is an overlap of one gene (*RAD51*) from the RPS signature and four genes from the RSS signature with the p53tru-DR and p53tru-UR genes for LUNG or BRCA (**Figure 4—figure supplement 1 a**), we obtained consistent results even after excluding these overlapped genes (**Figure 4—figure supplement 1 b-e**). These findings collectively indicate that tumors predicted to have reduced p53 function, similar to p53 mutant tumors, show increased sensitivity to platinum-based chemotherapy and radiation therapy. This could be attributed to compromised DNA damage repair due to reduced p53 function, making tumor cells more susceptible to the effects of chemo and radiation therapies.

We further confirmed this finding *in vivo* using 35 PDX models. We first defined 17 p53tru-DR and 19 p53tru-UR genes from the TCGA GBM cohort (**Figure 1—Source Data 1**). Due to limited sample sizes of normal tissues and tumors with *TP53* truncating mutations, training an SVM model to predict p53 status was unfeasible. Alternatively, we employed unsupervised clustering to predict the p53 functional status using CESs calculated from RNA-seq data of PDX mouse models (**Figure 4—Source Data 2**). 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, Figure 4—Source Data 2**). Out of 20 PDXs that were predicted to be RF, 15 responded to RT. In contrast, among the 15 PDXs predicted to be p53 normal, only 4 were responders. These results indicate a significant over-representation of RT responders within the pRF group (Fisher's exact test, P = 0.0068) (**Figure 4 f and g**). Although *TP53* somatic mutations (determined from WES) were enriched in the pRF group (**Figure 4 f)**, the association between *TP53* genetic mutation and RT response did not reach statistical significance (Fisher's exact test, P = 0.13). In summary, these findings suggested that our *in silico* predictions effectively uncovered a significant

association between p53 status and RT response, which would have been overlooked if solely examining the *TP53* genetic status.

Dissecting predicted *TP53*^{WT}-pRF

Next, we sought to explore the potential factors and mechanisms that could account for the majority of *TP53*^{WT} tumors being predicted as RF by the SVM model. We investigated the RNA and protein expression, re-evaluated all the *TP53* missense mutations (reported from WES) using RNA-seq data, and examined the alteration status of p53 upstream regulators MDM2 and MDM4. We did not detect significant changes in p53 protein abundance between *TP53*^{WT}-pN and *TP53*^{WT}-pRF tumors. *TP53* RNA expressions were even increased in the *TP53*^{WT}-pRF tumors (**Figure 5—figure supplement 1,** taking LUNG as an example), which aligns with previous research indicating that mutant p53 is associated with increased *TP53* mRNA expression (Donehower et al., 2019). These results also suggest that the impaired p53 function observed in the *TP53*^{WT}-pRF group is unlikely to be attributed to reduced *TP53* mRNA and protein abundances.

We then reassessed all the *TP53* missense mutations, such as R249 and R273, in the *TP53*^{WT}-pRF tumors using the RNA-seq sequence data of the LUNG and BRCA cohorts. Surprisingly, 10.3% (32 out of 310) LUNG and 10.2% (57 out of 561) BRCA tumors in the *TP53*^{WT}-pRF group were indeed genuine p53 mutants. This was evident from the substantial number of RNA-seq reads carrying the mutant alleles (**Figure 5—Source Data 1)**. For example, two lung adenocarcinoma samples (TCGA-55-6987 and TCGA-55-8621), initially classified as *TP53*^{WT} by TCGA, were re-evaluated and found to possess mutant allele fractions (MAF) of 45% (44 out of 98) and 35% (35 out of 100), respectively, at the loci 17: g.7577118C>A and 17: g.7577559G>A. Notably, these two mutations were also detectable in the WES data, albeit with much lower numbers of supporting reads (2 and 5 reads, respectively). This discrepancy explains why they were not identified by the TCGA somatic variant

caller (**Figure 5—figure supplement 2**). The substantial increase in MAFs observed in the RNA-seq data is probably due to the preferential expression of the mutant alleles. Interestingly, when comparing the amino acid locations of missense mutations reported from WES data, these mutations identified through 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**). Missense mutation at the R249 position is generally recognized as a structural mutation that destabilize the p53 protein due to alteration in its 3D structure (Joerger & Fersht, 2016). Furthermore, the 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) as the *TP53*^{MM} and *TP53*^{TM} tumors (**Figure 5—figure supplement 3**).

MDM2 and its homolog MDM4 are well-documented negative regulators of p53, exerting inhibitory effects through various mechanisms such as transcriptional repression and proteasomal degradation (Haupt, Maya, Kazaz, & Oren, 1997; Katz et al., 2018; Wasylishen & Lozano, 2016). In our analysis, we found that amplification of *MDM2* and/or *MDM4* was 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**). Consistent with this observation, all samples displaying amplification of *MDM2* and/or *MDM4* (n = 38 in LUNG, n = 84 in BRCA) were classified as pRF. These samples accounted for 12.3% and 15.0% of all *TP53*^{WT}-pRF samples in the LUNG and BRCA cohorts, respectively (**Figure 5 d**).

Conclusion

In this study, we employed a novel approach to measure p53 activity by defining p53-regulated genes and calculating their CESs as a surrogate. By training and cross-validating SVM models using CES data from the NT (p53-normal) and TM groups (p53-RF), we demonstrated the accuracy and

effectiveness of our *in silico* approach. Our comprehensive analysis revealed the widespread occurrence of non-mutational p53 inactivation in human malignancies. Moreover, our analyses unveiled that the predicted *TP53*^{WT}-pRF tumors exhibited a comparable level of genomic instability to those harboring genetic *TP53* mutation. This included a significant increase in the number of mutations, copy number alterations, and aneuploidy. Importantly, patients with *TP53*^{WT}-pRF tumors showed considerably worse overall survival rates when compared to those with *TP53*^{WT}-pN tumors, highlighting the prognostic value of our prediction. Furthermore, when evaluated using clinically validated signatures, *TP53*^{WT}-pRF tumors demonstrated significantly heightened sensitivity to platinum-based chemotherapy and radiation therapy. This observation was verified in our preclinical animal models of GBM. Additionally, we explored potential factors contributing to the *TP53*^{WT}-pRF classification, such as false-negative mutation calls and amplifications of *MDM2* and *MDM4*.

Discussion

Tumor suppressor genes such as *RB1*, *PTEN*, and *CDKN2A* are generally inactivated through homozygous deletions, resulting in minimal or complete loss of protein expression (https://www.cbioportal.org/). In contrast, *TP53*, unlike other tumor suppressors, predominantly undergoes missense mutations that give rise to altered or dysfunctional protein variants (https://www.cbioportal.org/). In addition, the oncogenic role of mutant p53 has been extensively characterized through multiple lines of evidence (Govindan & Weber, 2014), including the over-expression in p53 mutants, accumulation of mutant p53 protein in the nucleus and the cytoplasm, and 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 objective of this study is to investigate the loss or reduction of p53's tumor suppressor function in tumors that may appear functionally normal due to the absence of *TP53* genetic alterations. First, we observed that p53-regulated genes common among different TCGA cohorts are significantly enriched in pathways

associated with cell division, cell cycle control, DNA integrity checkpoint, and DNA damage repair (**Figure 1—figure supplement 2 b**). Second, we employed SVM models trained on normal tissues (NT samples) and tumor samples with truncating *TP53* mutations (TM samples) that result in shortened protein. Our hypothesis posits that p53's tumor suppressor function remains normal in NT samples, while it is compromised or significantly reduced in TM samples. However, it is essential to recognize that our training dataset is not without imperfections, given that the p53 statuses of these samples is inferred rather than definitively determined. To be more precise, the predicted outcomes should be termed NT-like or TM-like. Looking ahead, datasets that encompass paired RNA- and DNA-sequencing information, coupled with clearly defined p53 statuses are needed to overcome the existing limitations and gain more conclusive insights. Meanwhile, *TP53* missense mutations that were predicted to be functionally reduced ($TP53^{MM}$-pRF) should be interpreted as the "reduced tumor suppressor function of p53". Nevertheless, we could not rule out the possibility that the mutant p53 protein in the $TP53^{MM}$-pRF samples has gained oncogenic functions. Additionally, although p53tru-UR genes are negatively regulated by p53 through indirect and unclear mechanism, we retained these genes for downstream analyses since they can effectively reflect p53's functional status as p53tru-DR genes do. Our strategy allows us to identify tumors with compromised p53 function resulting from multiple mechanisms, and it is plausible that some cases predicted to have reduced p53 function may genuinely represent a loss of function (especially those in the MM group), while others may indicate dysregulation.

According to our analyses, approximately 22-25% of $TP53^{WT}$ tumors predicted as RF ($TP53^{WT}$-pRF) can be attributed to false-negative results (i.e., *TP53* mutations that failed to be detected from the WES assay) or amplification of *MDM2* and *MDM4*. However, the underlining mechanisms for the remaining 75-78% of $TP53^{WT}$-pRF are unknown and warrant further investigation. In order to gain more insights, we analyzed the DNA methylation data generated from the Infinium HumanMethylation450 BeadChip

array. However, we did not observe significant differences in the DNA methylation patterns among all nine CpG sites within the *TP53* gene (cg02087342, cg06365412, cg10792831, cg12041075, cg12041429, cg13468400, cg16397722, cg18198734, cg22949073) when comparing *TP53*^{WT}-pRF and *TP53*^{WT}-pN tumors. This suggests that p53 function may be compromised through other non-mutational mechanisms such as post-translational modification (such as phosphorylation, acetylation, methylation, and E3-mediated ubiquitination) (Chuikov et al., 2004; Kruse & Gu, 2009; E. W. Lee et al., 2009; Pan & Blattner, 2021; Xu, 2003), chromatin states (Hafner, Bulyk, Jambhekar, & Lahav, 2019), and interaction with microRNAs (Pollutri, Gramantieri, Bolondi, & Fornari, 2016; Suh et al., 2012). Unfortunately, due to the unavailability of matched data, we were unable to perform specific analyses in these areas.

In our analysis, we found no significant differences in tumor purities (measured by IHC (Aran, Sirota, & Butte, 2015)) 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). This suggests that the *TP53*^{WT}-pN tumors are not associated with lower tumor cellularity. Additionally, we identified four *TP53*^{MM} samples from the BRCA cohort (TCGA-E2-A1B1, TCGA-LD-A74U, TCGA-BH-A1FE, and TCGA-EW-A1P1) that were predicted to have normal p53 function (**Figure 3—Source Data 2 c**). We confirmed all reported mutations in these four samples through RNA-seq alignments with high mutant allele frequencies, ruling out the possibility of false positives or cross-sample contaminations. The reasons behind these findings remain unclear, and the limited sample size prevents extensive analysis.

While our analytic procedure is robust and can be applied to other datasets, building an SMV model from scratch requires a large number of *TP53* truncating mutations in training data. When obtaining sufficient training data is not practical, alternative unsupervised machine-learning approaches such as K-means clustering, semi-supervised learning, principal component analysis (PCA), and neural

networks could be considered.

TP53 mutations are commonly observed in human cancers. Notably, our analyses have revealed an intriguing aspect: the majority of *TP53*^{WT} tumors are genetically intact but exhibit functional deficiencies in p53. This discovery highlights an inherent limitation in relying solely on DNA markers for patient stratification and segmentation. It emphasizes the need for the development of "ensemble" approaches that incorporate multi-omics data to capture the full complexity of p53 functionality. By considering multiple layers of molecular information, we can gain a more comprehensive understanding of p53 status and its implications in cancer. Such integrated approaches hold great promise for enhancing our ability to characterize tumors and accurately guide personalized treatment strategies.

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 from cBioPortal (https://www.cbioportal.org/). We also obtained from cBioPortal the pre-calculated meta scores including the fraction of genome altered (FGA), mutation count, aneuploidy score, and Buffa hypoxia score. To re-evaluate the p53 mutation status, TCGA WES and RNA-seq BAM files for BRCA and LUNG were downloaded from the GDC data portal (https://portal.gdc.cancer.gov/). Furthermore, we acquired TCGA level-3 RNA-seq data, demographic information, and survival data for selected cancers from the University of California Santa Cruz's Xena web server (https://xena.ucsc.edu/). To identify differentially expressed genes between the NT and TM groups, we used pre-calculated raw

RNA-seq read counts. Log2-transformed FPKM (i.e., Fragments Per Kilobase of exon per Million mapped fragments) was used to calculate CESs, 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-regulated genes across different cancer types

To identify p53-regulated genes that can reflect p53's functional status in each cancer type, we employed a stepwise approach. First, we compiled a set of 147 genes that have been experimentally validated as being regulated by p53. This set includes the 116 genes identified as directly activated targets of p53 (refer to Table 1 in (Fischer, 2017)) and 31 genes that are primarily repressed by p53 and with the evidence of indirect regulation (refer to Table 2 in (Fischer, 2017)). We then analyzed the publicly available ChIP-seq data from ReMap2020 (Cheneby et al., 2020) to identify p53 peak within the gene body or the promotor region, which is defined as +/- 3 kb around the transcription start site (TSS). The basal level expression of p53 regulated genes in normal tissues was evaluated using the RNA-seq data of GTEx (Consortium, 2015). 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. If the mean of median TPMs of each p53 regulated gene was > 1 TPM, the gene was considered to be reliably expressed in GTEx normal tissues and thus was included in our downstream analyses.

Since p53 might regulate a specific set of genes in different cancer types (Fischer, 2017), and to determine genes whose expression could reflect p53's functional status in each cancer type, we performed differential expression analysis for the 147 genes in each cancer by comparing the NT group (considered to be p53 functionally normal) to the TM group (considered to be p53 functionally

impaired). We employed the R package DESeq2 to perform differential analyses on the raw RNA-seq reads counts (Love, Huber, & Anders, 2014). The significance threshold was set at an adjusted *p*-value ≤ 0.05 and |fold change| ≥ 2. By applying these criteria, we were able to identify p53-regulated genes specific to each TCGA cancer type. We defined "p53tru-DR" genes as those showing downregulation in p53-truncated samples compared to p53-normal samples. Conversely, "p53tru-UR" genes referred to genes showing upregulation in p53-truncated samples.

It should be noted that this approach aims to identify genes whose expression alterations are significantly associated with p53 functional status, encompassing both direct and indirect regulation. Interestingly, we found that the indirectly regulated targets display a higher sensitivity to p53 truncation, with 25 among the 31 indirect targets demonstrating significant differential expression upon p53 truncation in at least one cancer type. In comparison, 71 out of the 116 direct targets exhibit significant differential expression (Figure 1—Source Data 1). Thus, from a computational point of view, we incorporated both direct and indirect targets to construct cancer-type-specific SVM models.

Computation of composite expression scores

The CES is a single enrichment score that provides intuitive and stable measurement of p53 activity in each cancer. We employed four algorithms to calculate the CES, namely: (1) Gene Set Variation Analysis (GSVA), a gene set enrichment method that estimates the 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 $\gamma = \{1,\cdots,k\}$ 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 separately for p53tru-DR and p53tru-UR genes, while PC1 was computed using all p53-regulated genes, which resulted in a total of seven CESs for each sample. The GSVA package (https://www.bioconductor.org/packages/release/bioc/html/GSVA.html) was used to calculate 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 CESs of NT (normal tissue, coded as "0") and TM (truncating mutations, coded as "1") groups. GridSearch was employed 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 varied based on the specific cancer type. Specifically, 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 (**Figure 3—Source Data 2 a**). A total of seven features (GSVA, ssGSEA and Zscore for p53tru-DR and p53tru-UR genes respectively, and PC1 for both p53tru-DR and p53tru-UR genes) 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.

To better evaluate the performance of the SVM model of individual cancer types (e.g. LUNG SVM model), our approach encompassed several steps. Firstly, we employed the "train_test_split" function from the "sklearn.model_selection" class to partition the samples—comprising both NT and TM samples—into a training set (constituting 75% of the data) and a separate testing set (comprising the remaining 25%). Subsequently, we performed the differential analyses for feature gene selection and trained the SVM models with samples exclusively from the training set. Finally, we used the testing set to evaluate the performance. We repeated this process ten times to mitigate the potential variability in the outcomes of cross-validation. In each iteration, a confusion matrix was made, and performance measurements (i.e., sensitivity/recall, precision, accuracy) were calculated. The Pan-cancer model was validated using five-fold cross-validation. 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, ten-time hold-out validation, and five-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 **Figure 2—Source Data 3**. All TCGA barcodes (i.e., sample

identifiers) for training and prediction used in this study are listed in **Figure 3—Source Data 2 b-f.** Python source code is available from https://github.com/liguowang/epage.

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

The genomic positions 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'(Li et al., 2009) pileup engine and an in-house Python script. Samples with fewer than 10 mapped reads at a given genome site were excluded (n = 41), and a cut-off of minor allele frequency = 0.1 was employed to determine the genotype. A sample was considered *TP53* mutated if it had 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 inspect the variants and visualize the alignments manually.

Investigation of two independent treatment-related signatures

Tumor chemotherapy sensitivity (i.e., RPS score) was estimated based on 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 based on 51-gene panel reported by Speers *et al.* (Speers et al., 2015), and the genes were divided into "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. Notably, 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 our study, RPS was represented by minus GSVA score calculated from 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 into groups of 5–10 mice and treated with placebo and radiation therapy (RT). To determine the RT responsiveness of PDX, we calculated 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 (**Figure 4—Source Data 2)**.

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 numbers 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 CESs, 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 (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 (Source Data).

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].

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
p53tru-DR genes, genes down-regulated upon p53 truncation
p53tru-UR genes, genes up-regulated upon p53 truncation
PCA, principal component analysis
PDX, patient derived xenografts
ROC, receiver operating characteristic
RF, reduced function
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

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

Amatya, V. J., Naumann, U., Weller, M., & Ohgaki, H. (2005). TP53 promoter methylation in human gliomas. *Acta Neuropathol, 110*(2), 178-184. doi:10.1007/s00401-005-1041-5

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

Aran, D., Sirota, M., & Butte, A. J. (2015). Systematic pan-cancer analysis of tumour purity. *Nat Commun, 6*, 8971. doi:10.1038/ncomms9971

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

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. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/7677190

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

Chuikov, S., Kurash, J. K., Wilson, J. R., Xiao, B., Justin, N., Ivanov, G. S., . . . Reinberg, D. (2004). Regulation of p53 activity through lysine methylation. *Nature, 432*(7015), 353-360. doi:10.1038/nature03117

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

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

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

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., Bulyk, M. L., Jambhekar, A., & Lahav, G. (2019). The multiple mechanisms that regulate p53 activity and cell fate. *Nat Rev Mol Cell Biol, 20*(4), 199-210. doi:10.1038/s41580-019-0110-x

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

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

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. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/9690622

Ishizawa, J., Nakamaru, K., Seki, T., Tazaki, K., Kojima, K., Chachad, D., . . . Andreeff, M. (2018). Predictive Gene Signatures Determine Tumor Sensitivity to MDM2 Inhibition. *Cancer Res, 78*(10), 2721-2731. doi:10.1158/0008-5472.CAN-17-0949

Jeay, S., Gaulis, S., Ferretti, S., Bitter, H., Ito, M., Valat, T., . . . Graus Porta, D. (2015). A distinct p53 target gene set predicts for response to the selective p53-HDM2 inhibitor NVP-CGM097. *Elife, 4*. doi:10.7554/eLife.06498

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

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

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

Kruse, J. P., & Gu, W. (2009). MSL2 promotes Mdm2-independent cytoplasmic localization of p53. *J Biol Chem, 284*(5), 3250-3263. doi:10.1074/jbc.M805658200

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, E. W., Lee, M. S., Camus, S., Ghim, J., Yang, M. R., Oh, W., . . . Song, J. (2009). Differential regulation of p53 and p21 by MKRN1 E3 ligase controls cell cycle arrest and apoptosis. *EMBO J, 28*(14), 2100-2113. doi:10.1038/emboj.2009.164

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

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

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

Pan, M., & Blattner, C. (2021). Regulation of p53 by E3s. *Cancers (Basel), 13*(4). doi:10.3390/cancers13040745

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

Pollutri, D., Gramantieri, L., Bolondi, L., & Fornari, F. (2016). TP53/MicroRNA Interplay in Hepatocellular Carcinoma. *Int J Mol Sci, 17*(12). doi:10.3390/ijms17122029

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. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/10640274

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

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

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

Suh, S. S., Yoo, J. Y., Nuovo, G. J., Jeon, Y. J., Kim, S., Lee, T. J., . . . Croce, C. M. (2012). MicroRNAs/TP53 feedback circuitry in glioblastoma multiforme. *Proc Natl Acad Sci U S A, 109*(14), 5316-5321. doi:10.1073/pnas.1202465109

Takamatsu, S., Brown, J. B., Yamaguchi, K., Hamanishi, J., Yamanoi, K., Takaya, H., . . . Matsumura, N. (2022). Utility of Homologous Recombination Deficiency Biomarkers Across Cancer Types. *JCO Precis Oncol, 6*, e2200085. doi:10.1200/PO.22.00085

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

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. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/8378077

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

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

Xu, Y. (2003). Regulation of p53 responses by post-translational modifications. *Cell Death Differ, 10*(4), 400-403. doi:10.1038/sj.cdd.4401182

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

Figures

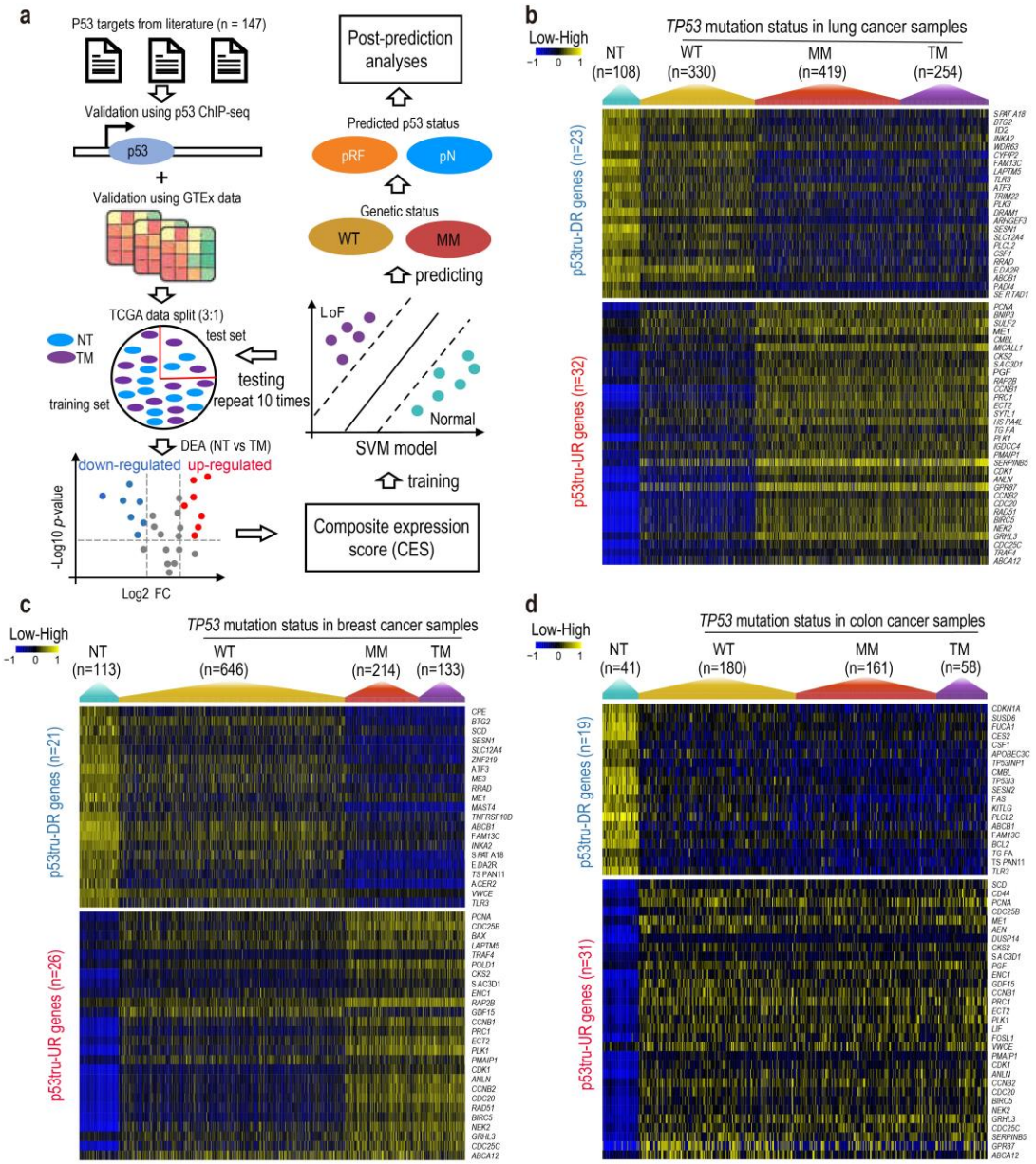


Figure 1. Identification of p53-regulated genes and their expression profiles in TCGA lung, breast, and colon cancers. (**a**) Analytic workflow depicting the process to identify p53tru-DR genes (genes down-regulated upon p53 truncation) and p53tru-UR genes (genes up-regulated upon p53 truncation), train and validate the SVM models. (**b-d**) Heatmaps illustrating the expression profiles of p53tru-DR and p53tru-UR genes in TCGA lung, breast, and colon cancers. The samples within each cohort were categorized into four groups based on TCGA designation: 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; pRF, predicted reduced function; pN, predicted normal; SVM, support vector machine.

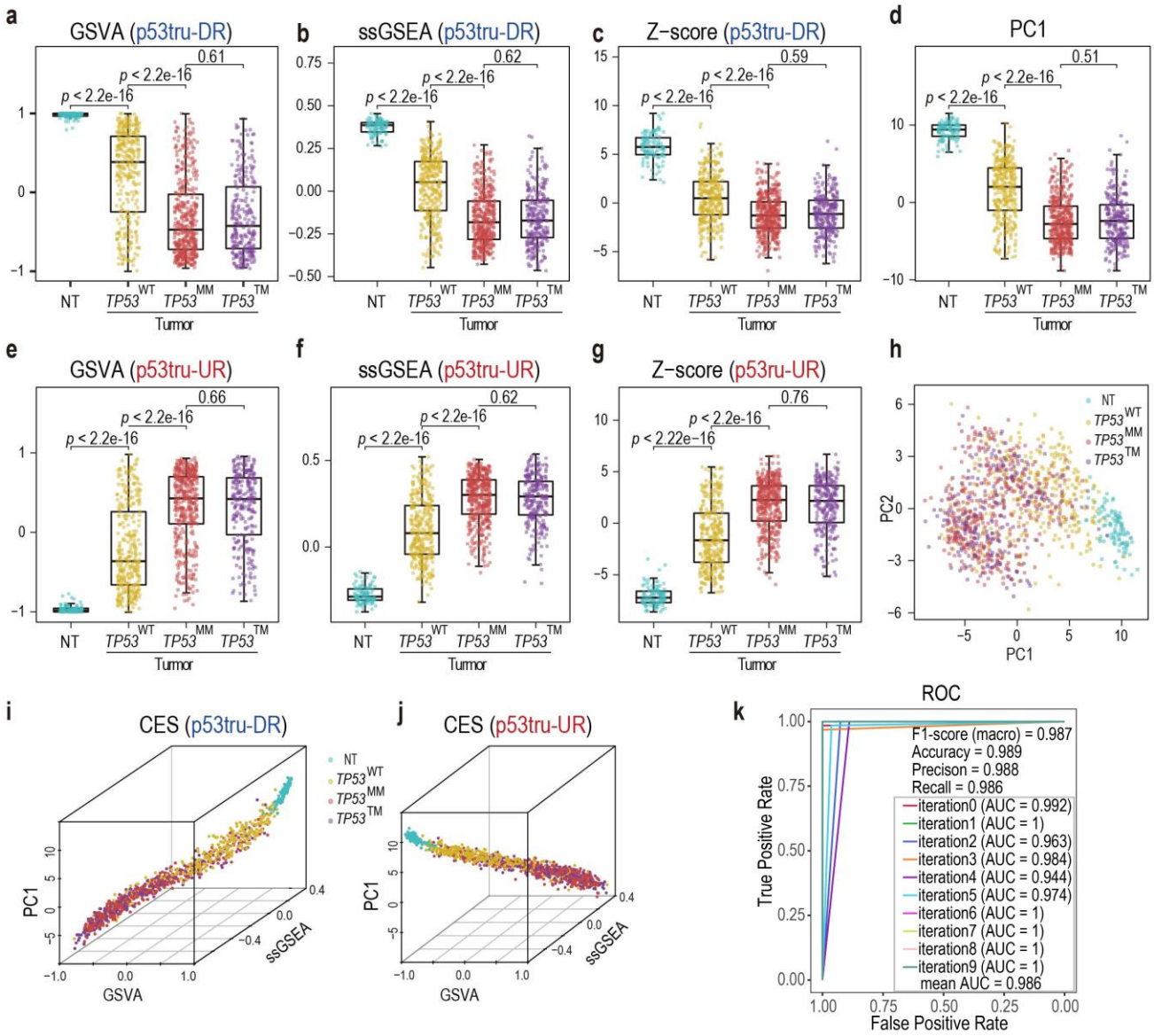


Figure 2. **Composite expression scores (CESs) and the ROC curve of the SVM model built from the TCGA lung cancer cohort.** (**a-c**) CESs of p53tru-DR genes calculated by GSVA, ssGSEA, and Z-score methods. In each panel, the blue, yellow, red, and purple box-and-whisker plots indicate the distribution of CESs in NT, WT, MM, and TM groups, respectively. (**d**) CES calculated by the first principal component scores (PC1) combining p53tru-DR and p53tru-UR genes. (**e-g**) CESs of p53tru-UR genes calculated by GSVA, ssGSEA, and Z-score methods. (**h**) Two-dimensional PCA plot demonstrating the clusters of NT, WT, MM, and TM samples. (**i-j**) Three-dimensional scatter plots illustrating the combinatorial effects of CES calculated from p53tru-DR genes and p53tru-UR genes, respectively. (**k**) Performance evaluation of the SVM model using ten-time hold-out 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, adjacent 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; p53tru-DR genes, genes down-regulated upon p53 truncation; p53tru-UR genes, genes up-regulated upon p53 truncation

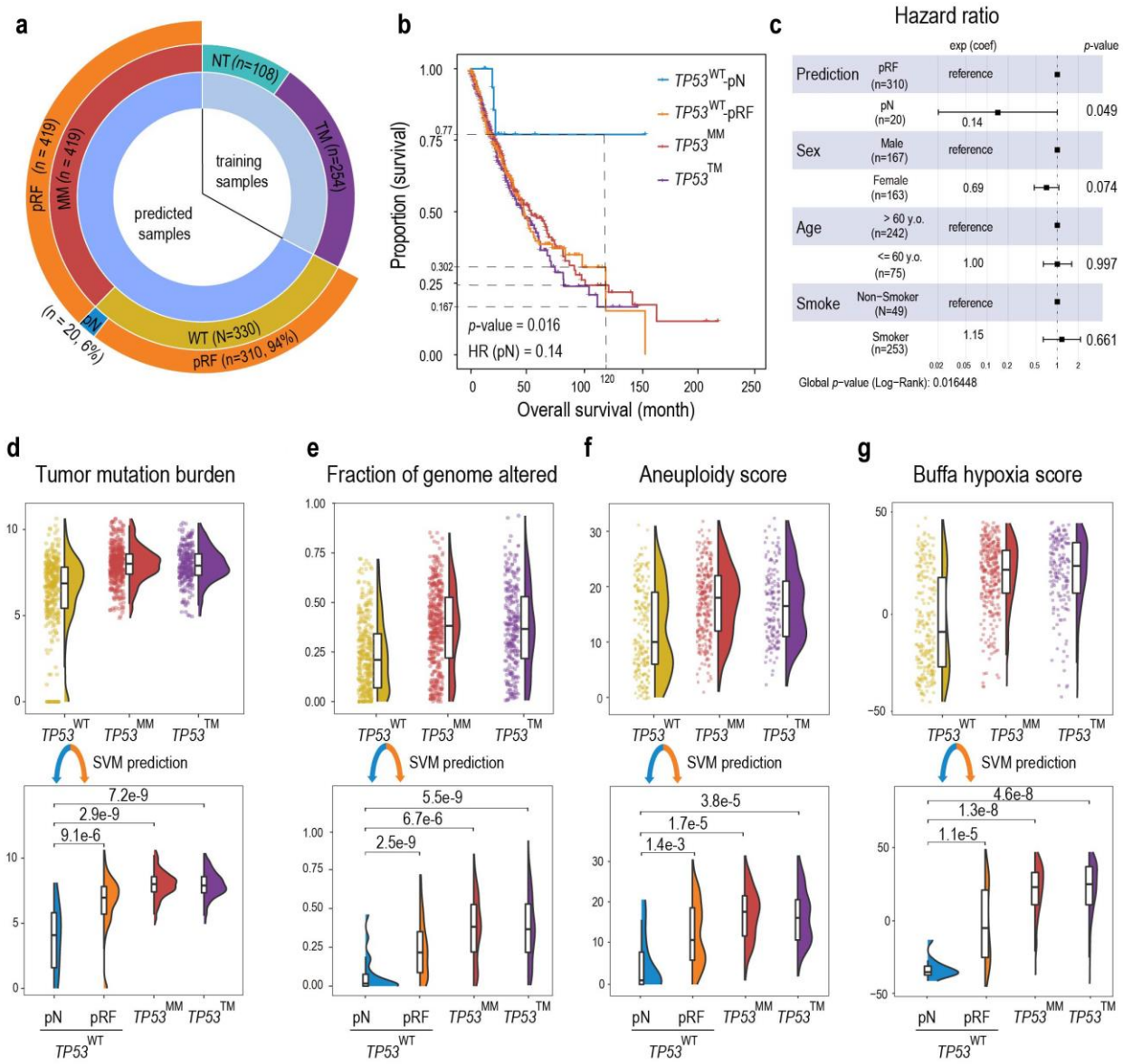


Figure 3. **Evaluation of the SVM predictions in the TCGA lung cancer cohort.** (**a**) A sunburst chart showing the breakdown of the lung cancer samples before (inner and middle layer circles) and after (outer layer circle) SVM prediction. Training samples consist of NT (adjacent normal tissue) and *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 as normal (pN) or RF (pRF) by the SVM model. (**b**) Overall survival of patients with *TP53*^{MM}, *TP53*^{TM}, *TP53*^{WT}-pN (wild-type *TP53* predicted as normal) and *TP53*^{WT}-pRF (wild-type *TP53* predicted as reduced function) tumors. (**c**) Forest plot of a Cox Proportional-Hazards model showing *p*-values, hazard ratios, and the 95% confidence intervals for covariates. (**d**) 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}-pRF) based on SVM prediction. (**e**) Similar to (d), comparison of the copy number variation burden measured by "Fraction of genome altered". (**f**) Similar to (d), comparison of the aneuploidy score. (**g**) Similar to (d), comparison of hypoxia score.

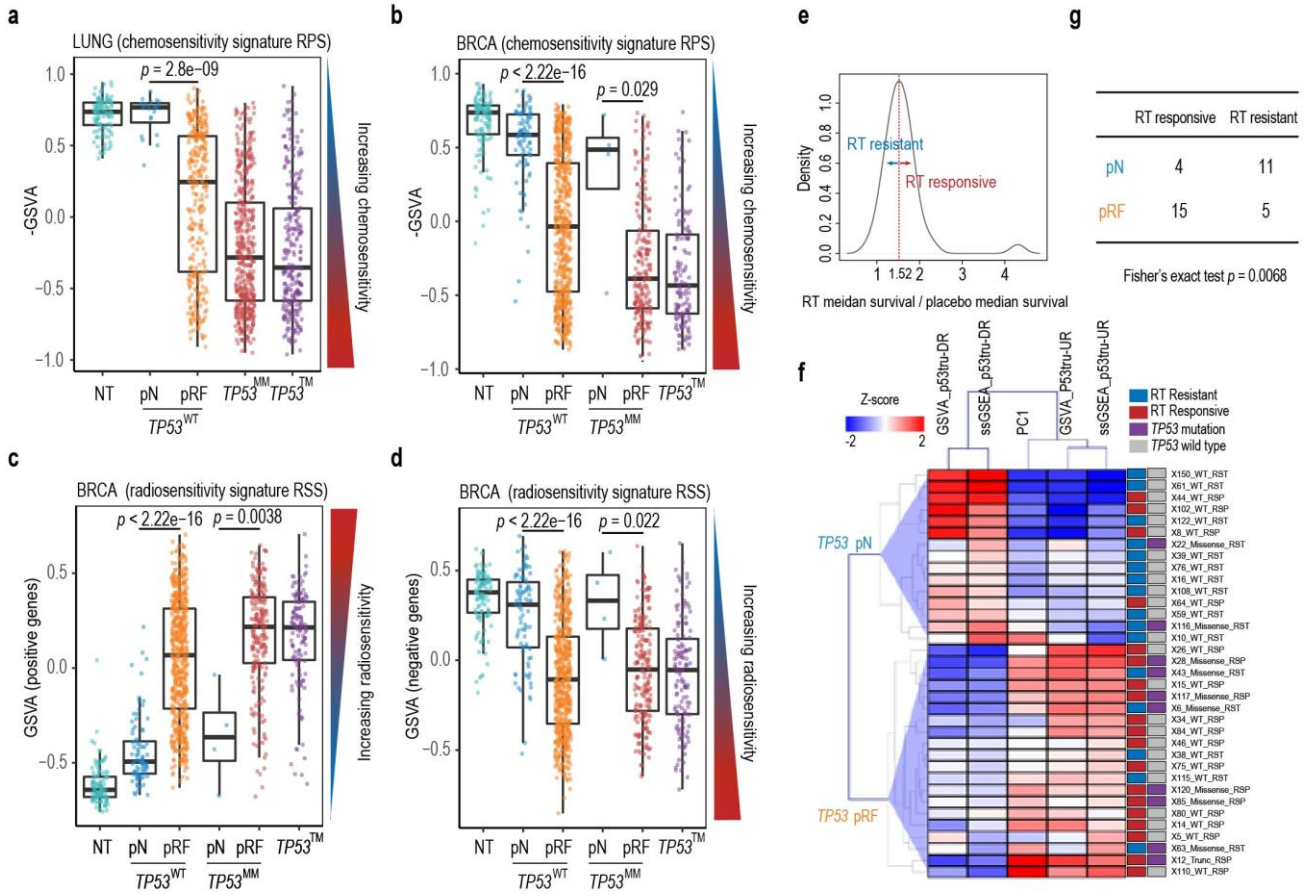


Figure 4. Relationships between p53 status and chemo-and radiation therapy sensitivities. (**a-b**) Comparison of the RPS scores (chemosensitivity signature, represented by negative GSVA score, see Methods) amongst the NT, *TP53*^{WT}-pN, *TP53*^{WT}-pRF, *TP53*^{MM}, and *TP53*^{TM} groups in the TCGA LUNG and BRCA cohorts, respectively. (**c-d**) Comparison of the RSS scores (radiosensitivity signature, represented by GSVA score, see Methods) of positive and negative genes amongst NT, *TP53*^{WT}-pN, *TP53*^{WT}-pRF, *TP53*^{MM}-pN, *TP53*^{MM}-pRF, and *TP53*^{TM} groups in the BRCA cohort, respectively. (**e**) Distribution of the ratio of median survival days between the RT (radiation therapy) group and the placebo group. PDX models with ratios >= 1.52 were considered as RT-responsive. (**f**) A heatmap showing the p53 functional status predicted by unsupervised clustering using composite expression (GSVA, ssGSEA, and PC1) of GBM-specific p53tru-DR and p53tru-UR genes in PDX data. RT response and *TP53* mutation status of each sample are color coded on the right side of the heatmap. (**g**) A contingency table shows that pRF (predicted reduced function) samples are significantly associated with RT responsiveness.

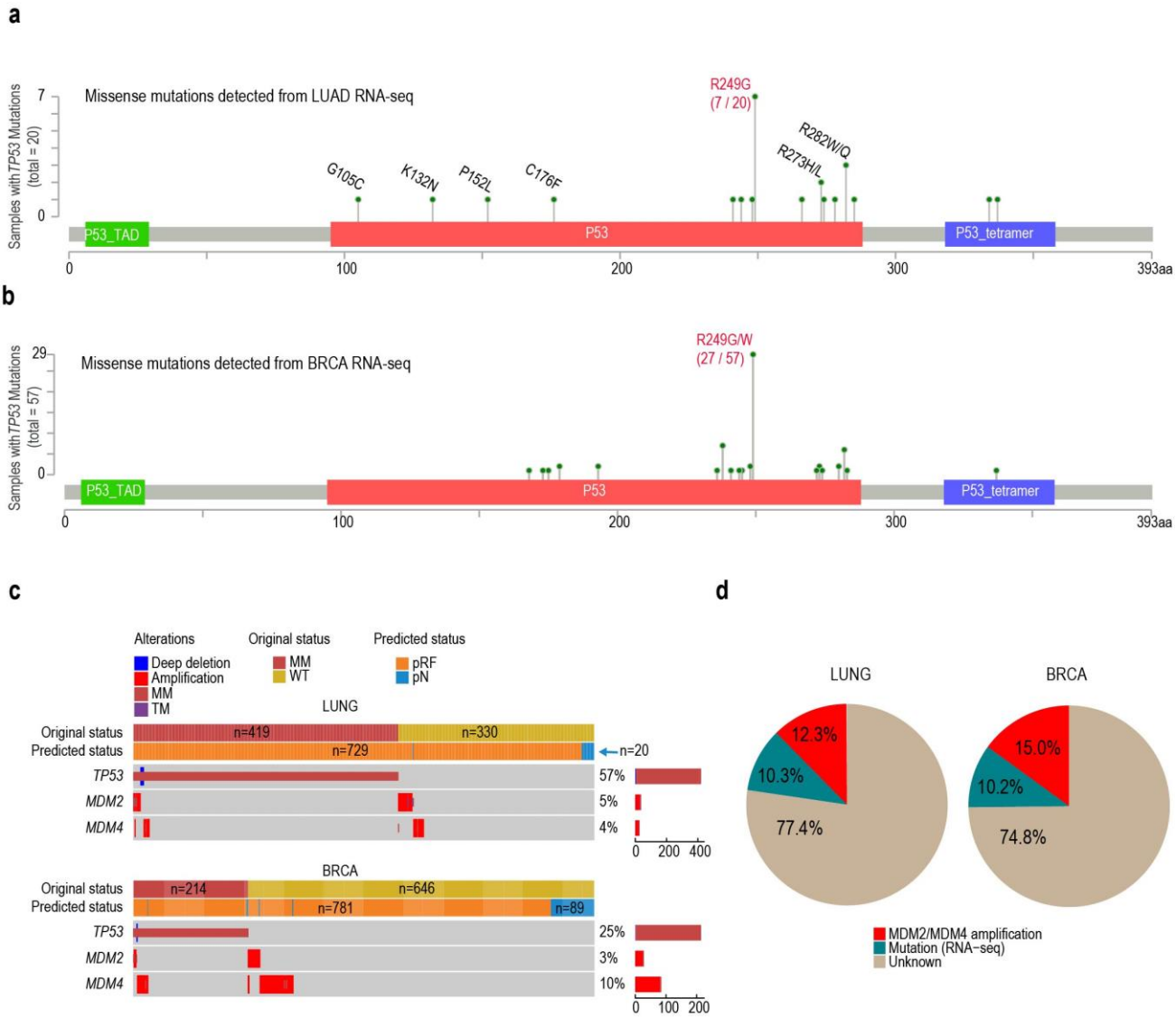


Figure 5. **Dissection of *TP53*^{WT}-pRF cases.** (**a-b**) Lollipop charts showing the positions of missense mutations identified using RNA-seq data but missed from the whole-exome sequencing data in TCGA lung adenocarcinoma (LUAD) and breast invasive carcinomas (BRCA), respectively. (**c**) Oncoprint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

Supplementary Figures

Figure 1—figure supplement 1. Validation of p53 target genes in γ irradiated MCF-7 cells harboring wild-type *TP53*.

Figure 1—figure supplement 2. Shared p53tru-DR and p53tru-UR genes among the four cancer cohorts (LUNG, BRCA, COAD and ESCA).

Figure 2—figure supplement 1. Correlations between different CESs. (a-c) Significant inverse correlations between CESs calculated from p53tru-DR and p53tru-UR genes for GSVA, ssGSEA, and Z-score.

Figure 2—figure supplement 2. Comparison of CES distributions among different *TP53* groups.

Figure 2—figure supplement 3. Pan-cancer analysis.

Figure 3—figure supplement 1. Distribution of predicted p53 status for $TP53^{WT}$ (WT group) and $TP53^{MM}$ (MM group) tumor samples across the five cohorts.

Figure 3—figure supplement 2. Progression-free survival (PFS) comparison among $TP53^{MM}$, $TP53^{TM}$, $TP53^{WT}$-pN and $TP53^{WT}$-pRF patients of TCGA BRCA cohort.

Figure 4—figure supplement 1. Relationships between p53 status and chemo-and radiation therapy sensitivities after removing the overlapped genes.

Figure 5—figure supplement 1. Comparison of p53 expression levels between $TP53^{WT}$-pRF and $TP53^{WT}$-pN samples in LUNG cohort.

Figure 5—figure supplement 2. Examples of *TP53* somatic mutations identified from RNA-seq data but missed by TCGA WES calls.

Figure 5—figure supplement 3. p53 mutants identified from RNA-seq data (indicated as the RM group) exhibit similar genomic characteristics as the $TP53^{MM}$ and $TP53^{TM}$ groups.

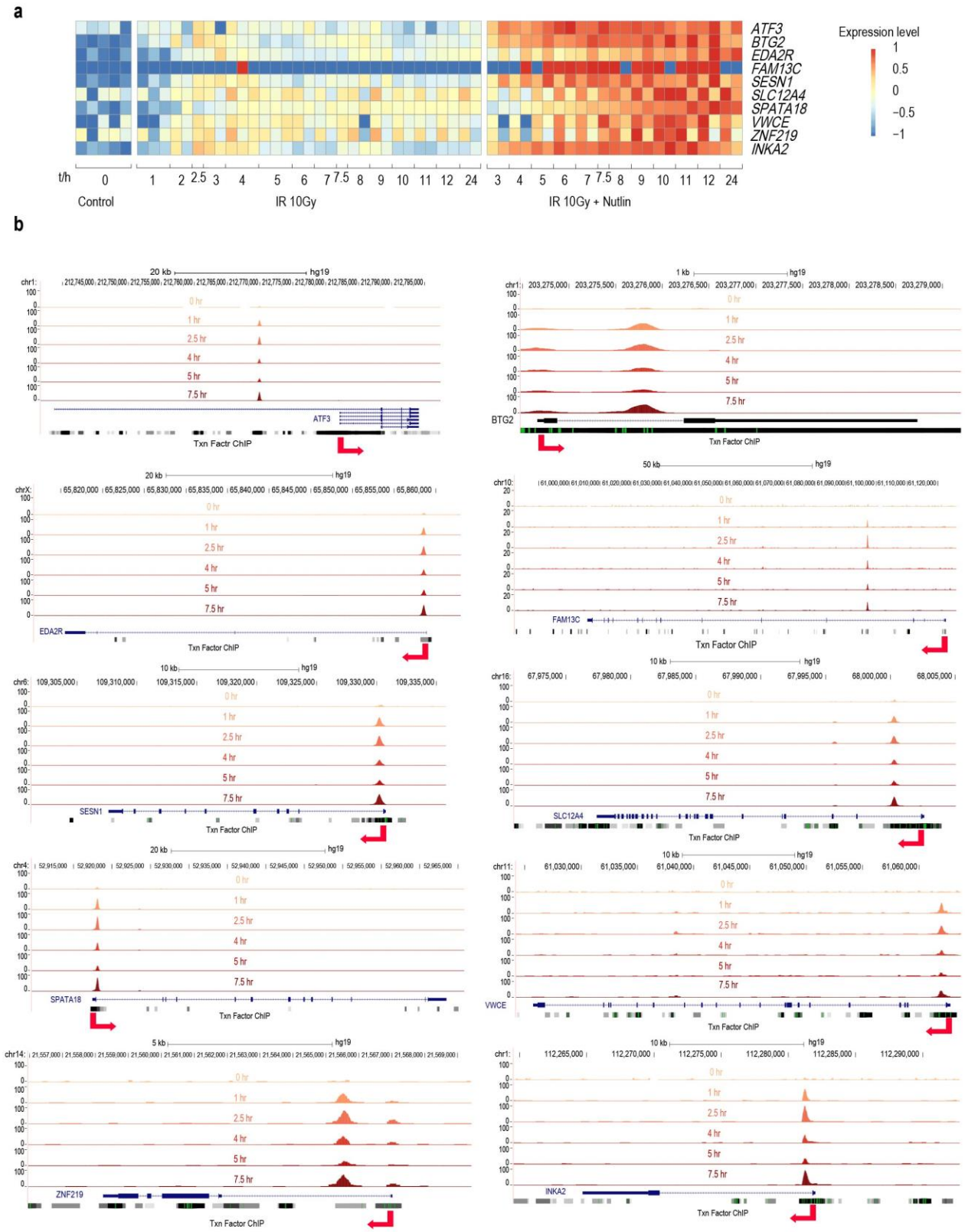
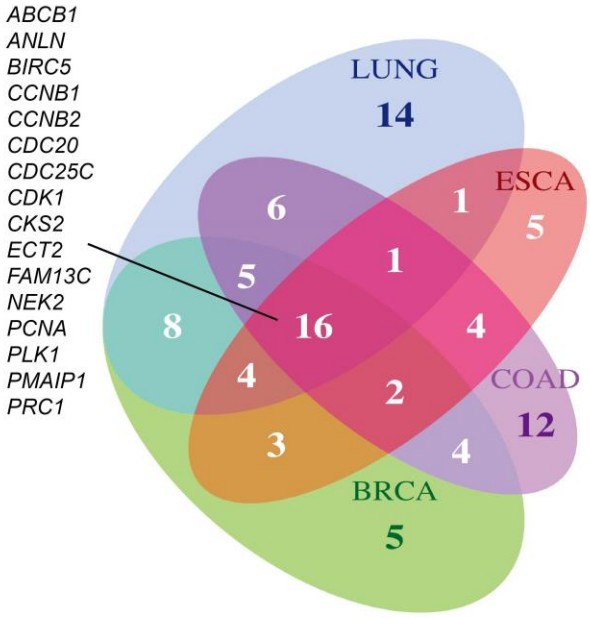


Figure 1—figure supplement 1. Validation of p53 target genes in γ irradiated MCF-7 cells harboring wild-type *TP53*. (a) Changes of expression levels of p53 targets after γ irradiation, and γ irradiation with nutlin treatment. (b) Visualization of p53 binding sites (p53 ChIP-seq signals) around the promoter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

a



b

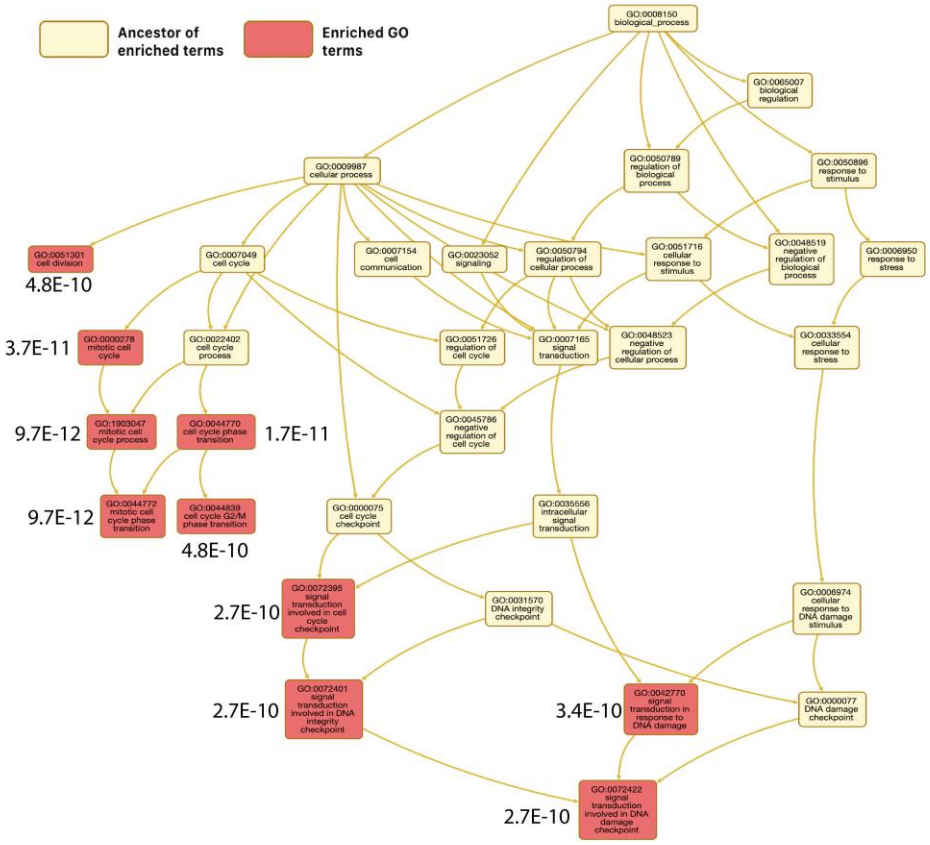


Figure 1—figure supplement 2. Shared p53tru-DR and p53tru-UR genes among the four cancer cohorts (LUNG, BRCA, COAD and ESCA). (a) Venn diagram shows the overlaps of p53 regulated genes among the four cancer types. (b) Gene Ontology (GO) enrichment results of the shared p53 regulated genes.

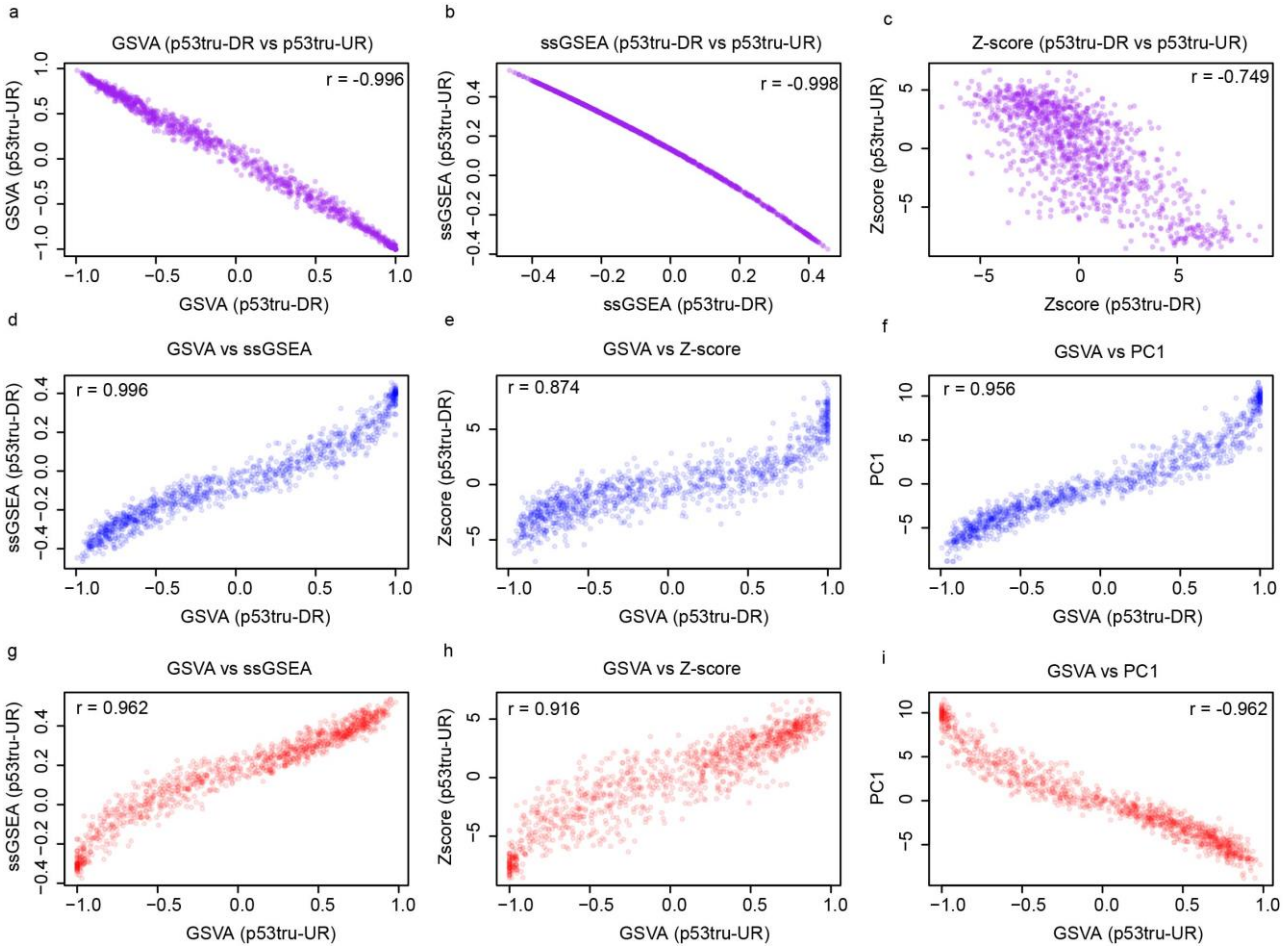


Figure 2—figure supplement 1. Correlations between different CESs. (a-c) Significant inverse correlations between CESs calculated from p53tru-DR and p53tru-UR genes for GSVA, ssGSEA, and Z-score. Significant positive correlations between different CESs for p53tru-DR genes (d-f) and between that for p53tru-UR genes (g-h). (i) Significant inverse correlation between PC1 and the GSVA of p53tru-UR genes.

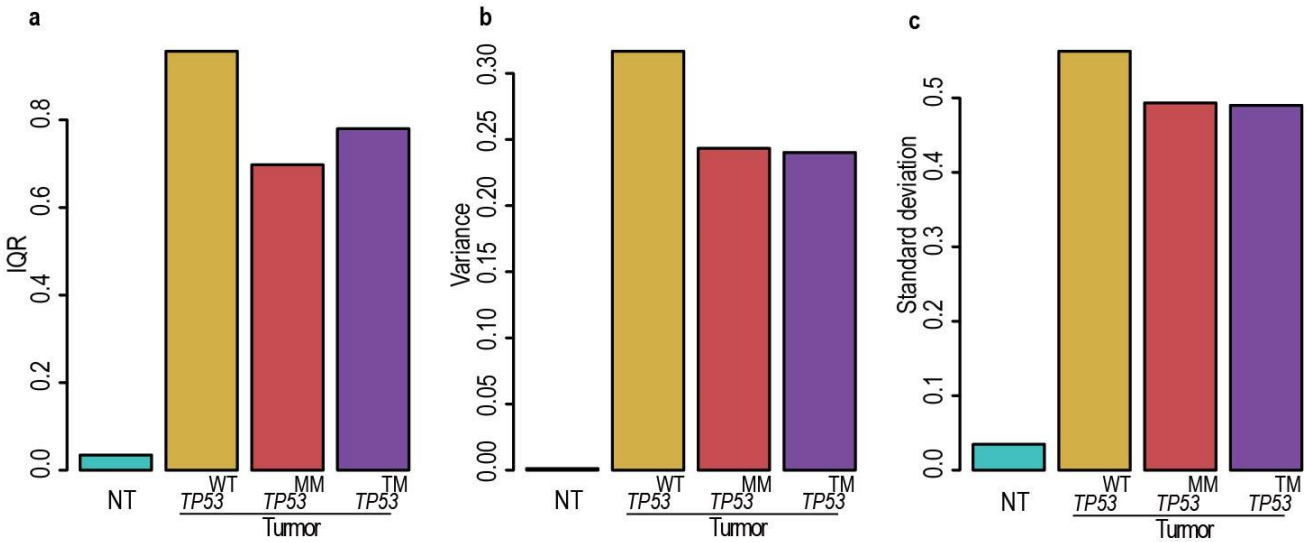


Figure 2—figure supplement 2. Comparison of CES distributions among different *TP53* groups. Bar plots show the Interquartile Range (a), Variance (b), and Standard Deviation (c) of CESs in NT, WT, MM, and TM groups.

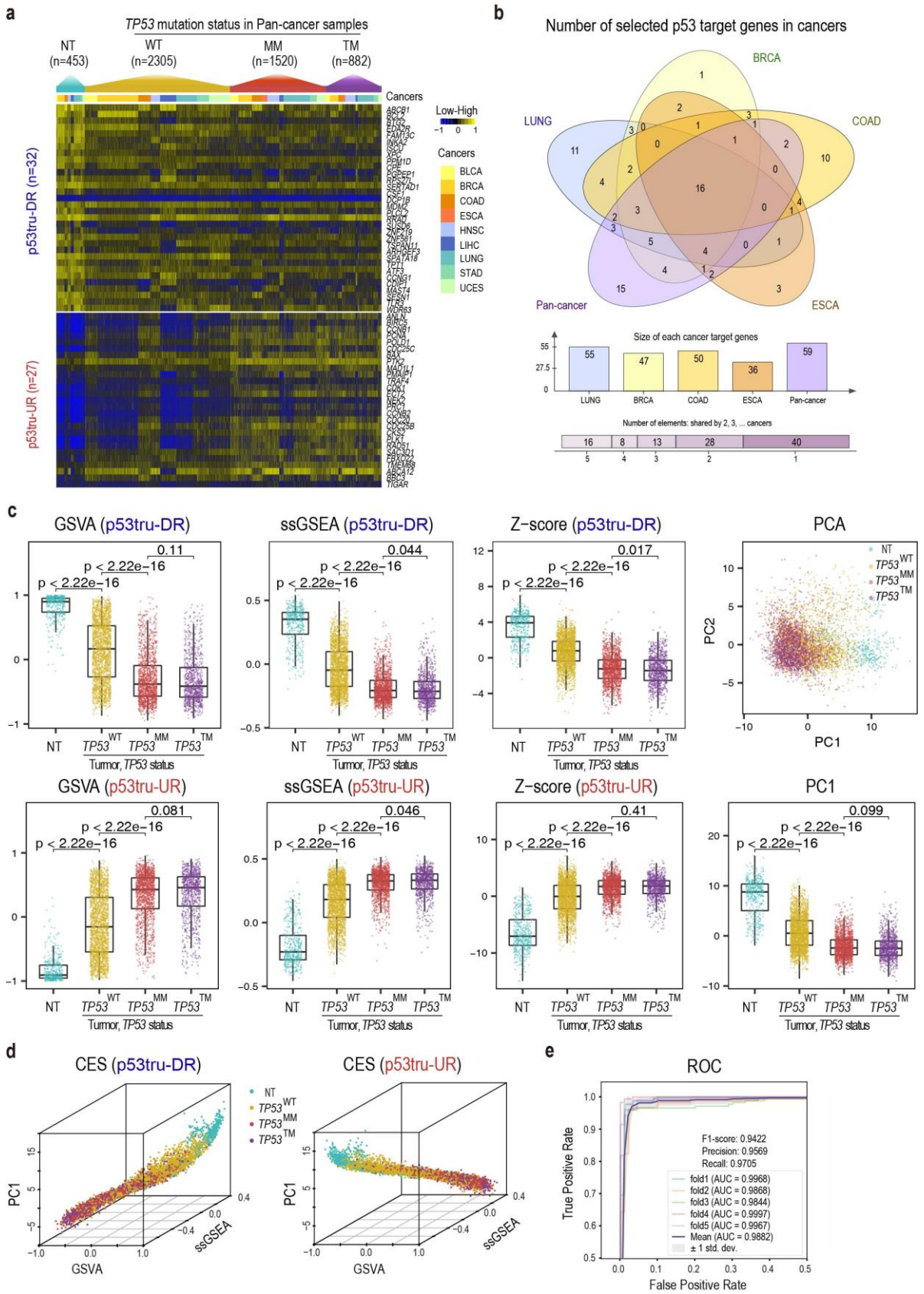


Figure 2—figure supplement 3. Pan-cancer analysis. (a) Heatmap shows the expression profiles of p53tru-DR genes and p53tru-UR genes across 9 cancer types. (b) Overlaps of p53 regulated genes among pan-cancer cohort and four previously selected cancer cohorts. (c-d) CESs calculated from p53tru-DR genes and p53tru-UR genes by different algorithms in pan-cancer cohort (similar to Figure 2). (e) ROC curve of the SVM model built from the pan-cancer cohort.

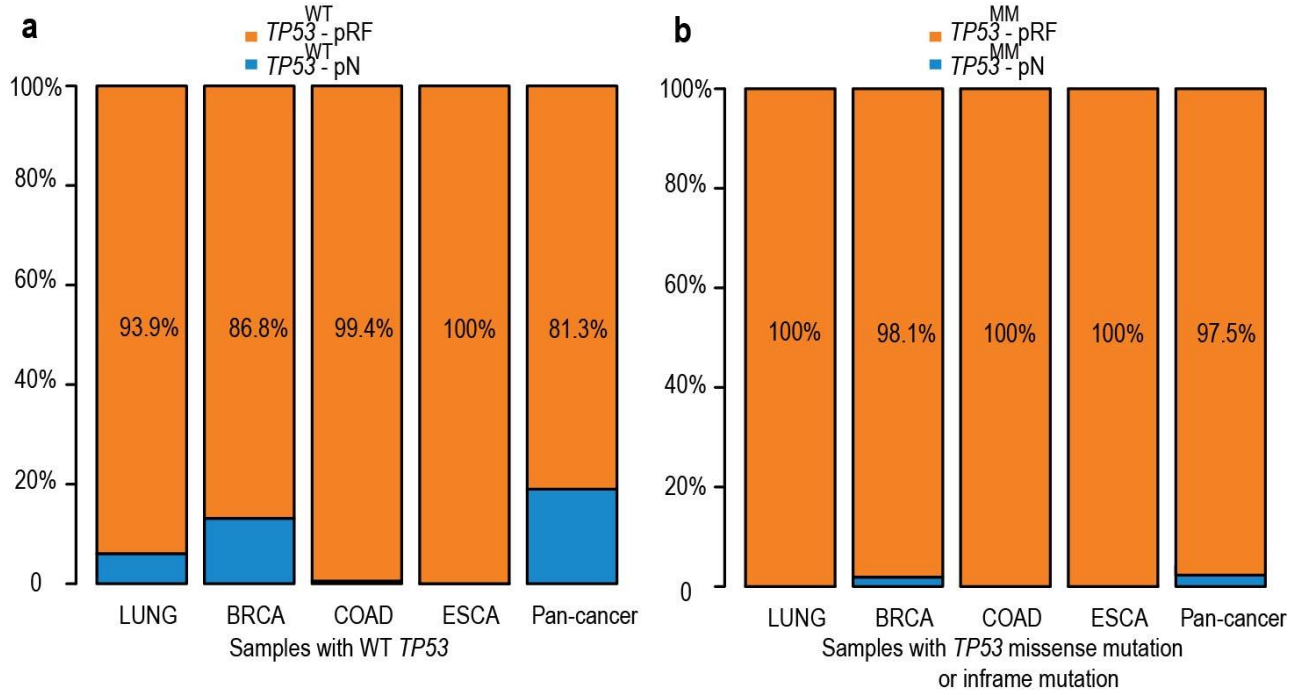


Figure 3—figure supplement 1. Distribution of predicted p53 status for $TP53^{WT}$ (WT group) and $TP53^{MM}$ (MM group) tumor samples across the five cohorts. Most tumor samples in WT (a) and MM (b) groups are predicted as RF.

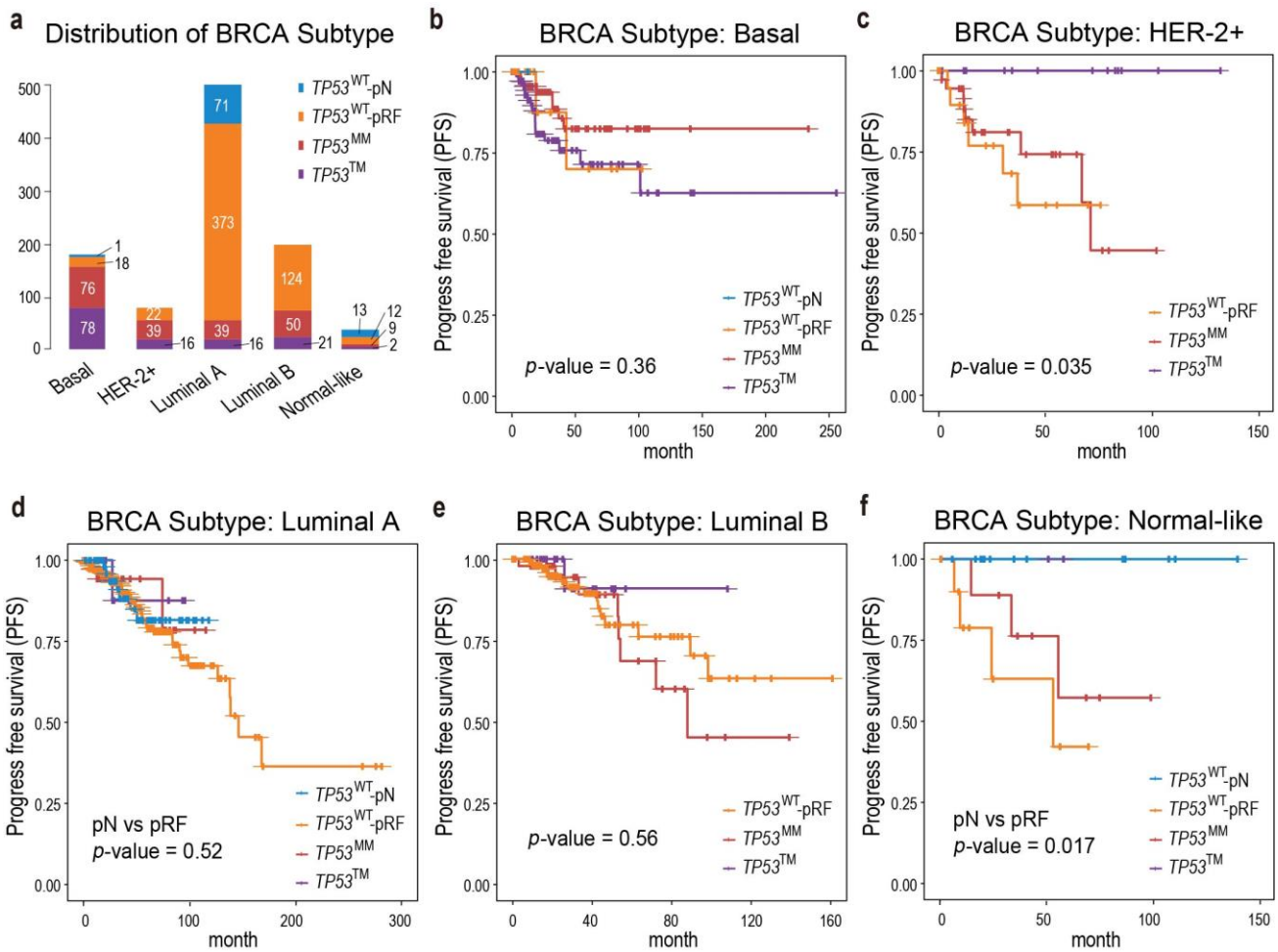


Figure 3—figure supplement 2. Progression-free survival (PFS) comparison among *TP53*^{MM}, *TP53*^{TM}, *TP53*^{WT}-pN and *TP53*^{WT}-pRF patients of TCGA BRCA cohort. (a) Bar plot shows the breakdown of BRCA patients into five subtypes, including "Basal/triple negative", "HER2^{+}", "Luminal A", "Luminal 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.

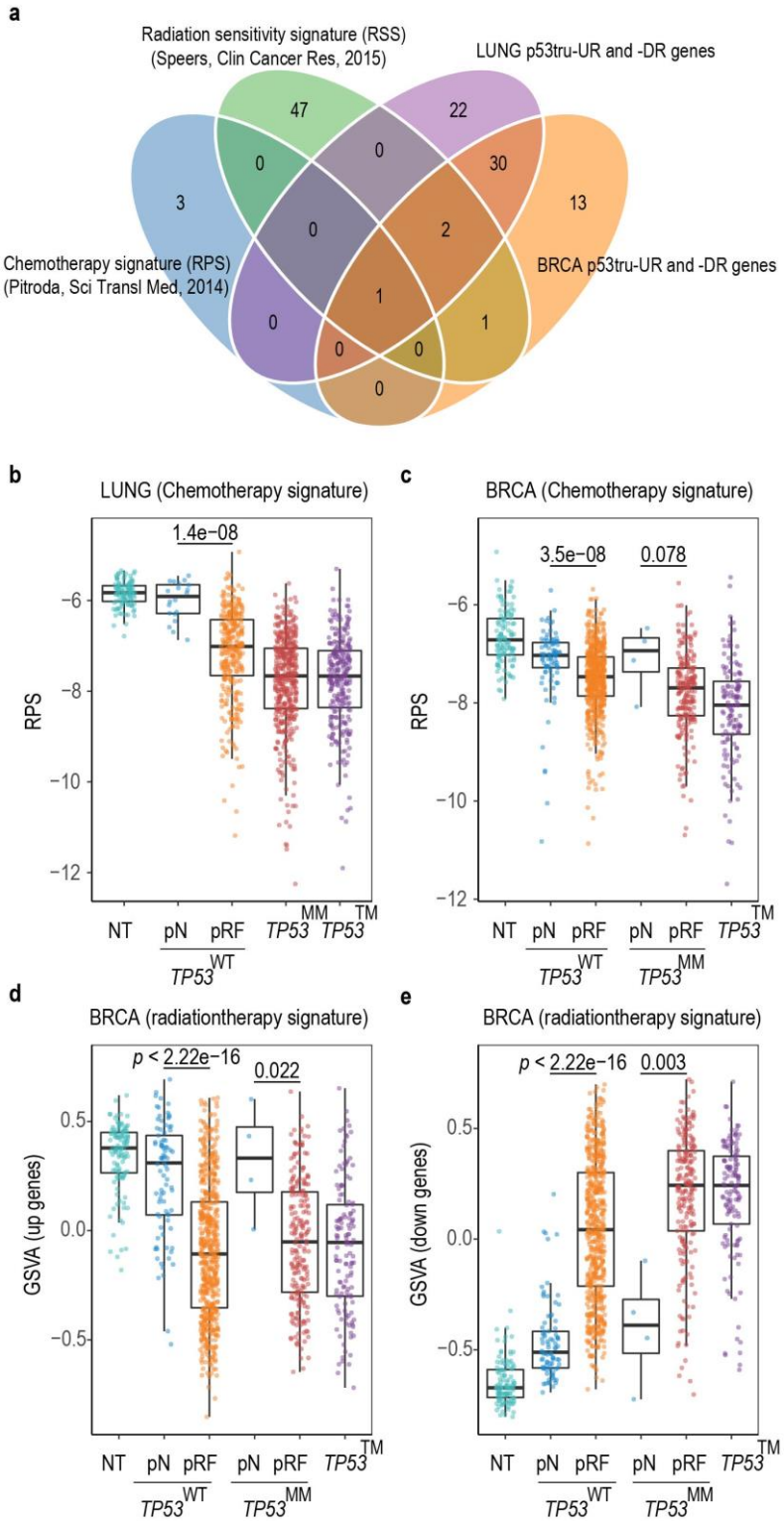


Figure 4—figure supplement 1. Relationships between p53 status and chemo-and radiation therapy sensitivities after removing the overlapped genes. (a) Venn diagram shows the overlap between p53tru-UR and p53tru-DR genes used in the LUNG and the BRCA SVM models, the chemotherapy gene signature (recombination 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}-pRF, *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}-pRF, *TP53*^{MM}-pN, *TP53*^{MM}-pRF 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 p53tru-DR or p53tru-UR genes of either LUNG or BRCA were removed when calculating the RPS and RSS scores.

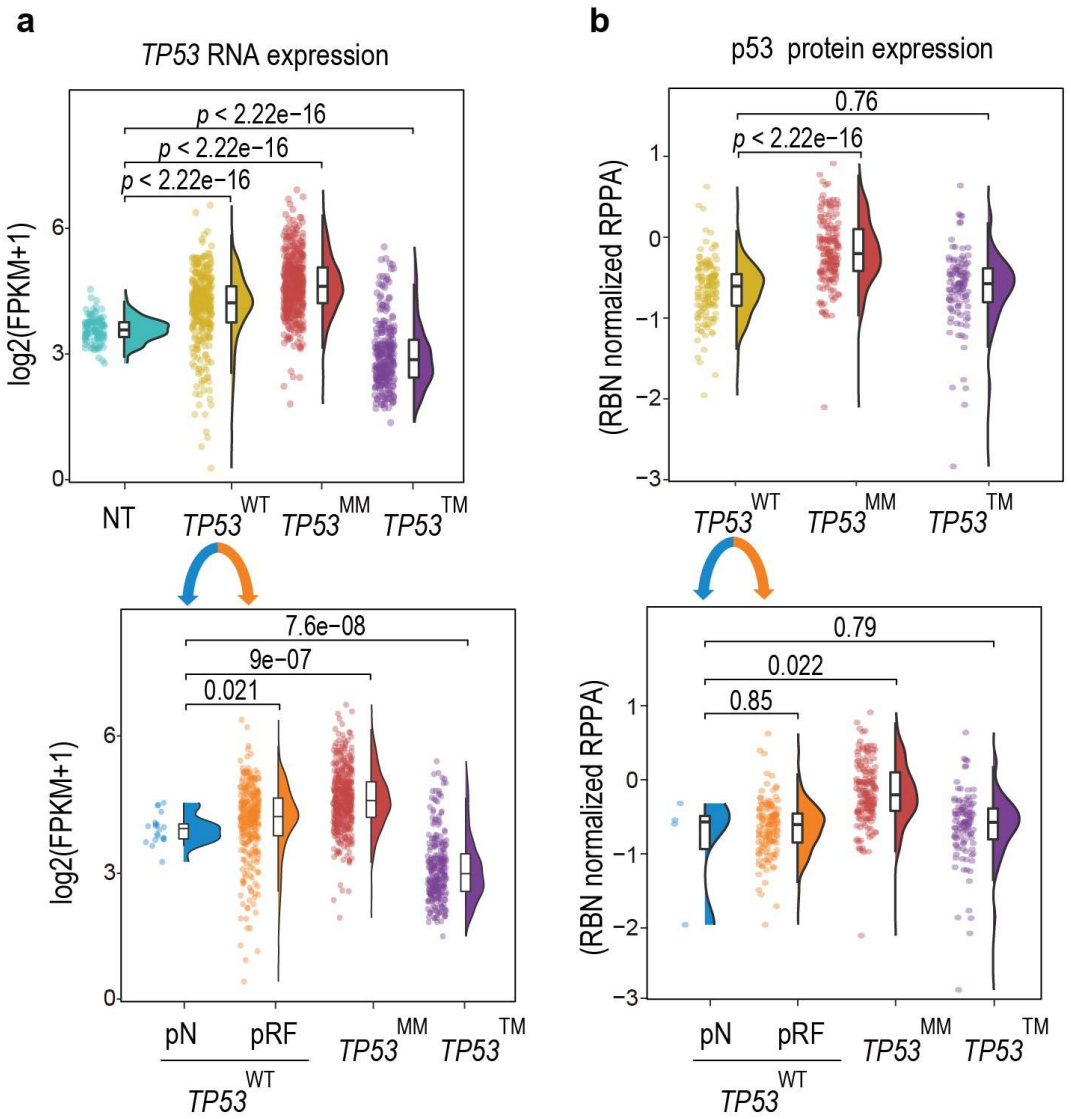


Figure 5—figure supplement 1. Comparison of p53 expression levels between *TP53*^{WT}-pRF and *TP53*^{WT}-pN samples in LUNG cohort. (a) Comparison of the *TP53* RNA expression. (b) Comparison of the p53 protein abundance.

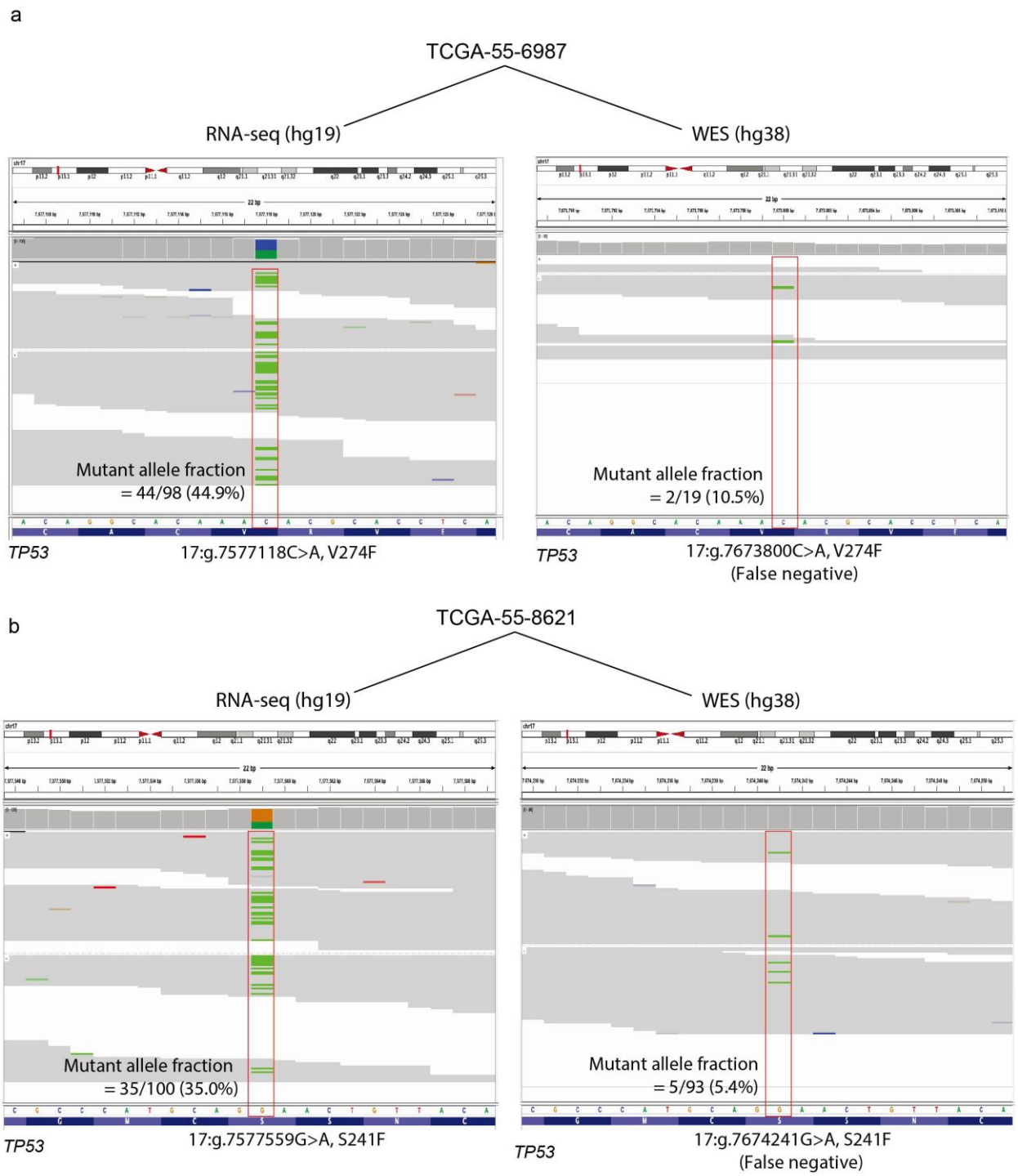


Figure 5—figure supplement 2. Examples of *TP53* somatic mutations identified from RNA-seq data but missed by TCGA WES calls. Mutant allele fraction (MAF) is measured by the ratio between the "number of reads supporting the mutant allele" and the "total number of reads" covering the mutation site.

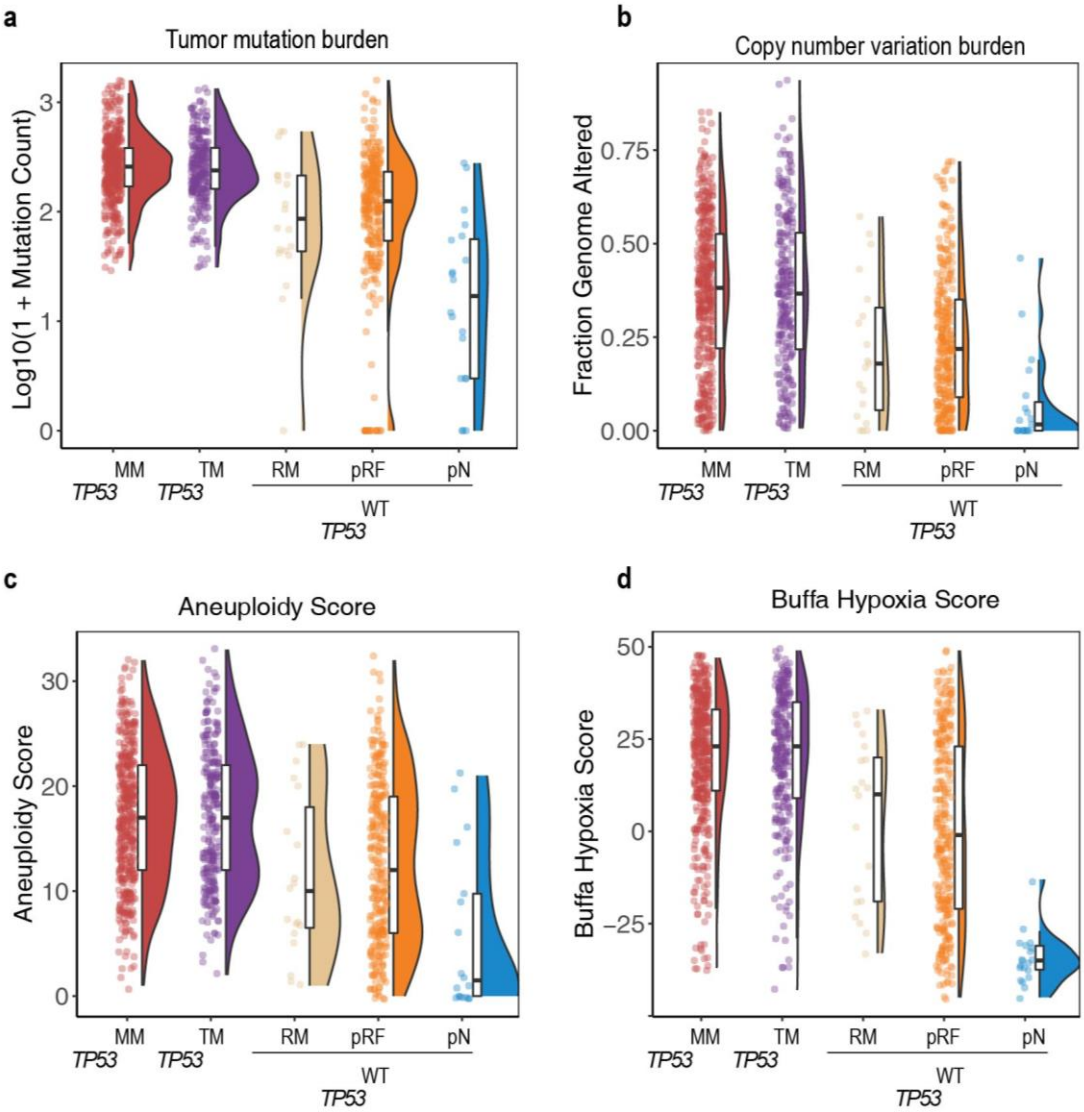


Figure 5—figure supplement 3. p53 mutants identified from RNA-seq data (indicated as the RM group) exhibit similar genomic characteristics as the *TP53*^{MM} and *TP53*^{TM} groups. Compared to the *TP53*^{WT}-pN tumors, tumors of the RM group show increased tumor mutation burden (a), copy number variation burden (b), aneuploidy score (c), and Buffa hypoxia score (d)

Source data

Figure 1—Source Data 1
List of p53tru-DR and p53tru-UR genes used in this study.
Figure 2—Source Data 1
CESs of p53tru-DR and p53tru-UR genes in TCGA LUNG, BRCA, COAD, ESCA, Pan-cancer (including 12 types or subtypes) cohorts.
Figure 2—Source Data 2
The performance metrics of the SVM model in TCGA LUNG and BRCA cohorts.
Figure 2—Source Data 3
Summary of SVM models according to the DOME (Data, Optimization, Model and Evaluation) recommendations.
Figure 3—Source Data 1
Genomic and clinical characteristics of TCGA LUNG, BRCA, COAD, ESCA, and Pan-cancer cohorts.
Figure 3—Source Data 2
List of SVM training samples, prediction samples, *TP53* original genetic states, SVM-predicted states and probabilities of TCGA cancers analyzed in our study.
Figure 4—Source Data 1
Gene signatures reflecting chemo- or radiosensitivity.
Figure 4—Source Data 2
Median survival days and CES values of PDX models treated with placebo and radiation therapy (RT).
Figure 5—Source Data 1
List of lung and breast cancer samples that have *TP53* missense mutations detected from RNA-seq.