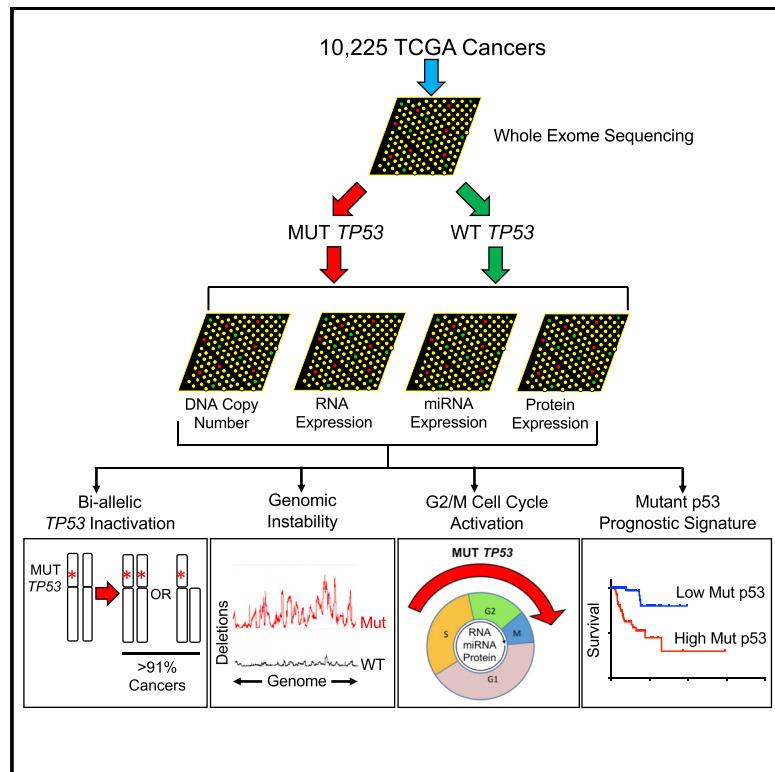


Integrated Analysis of *TP53* Gene and Pathway Alterations in The Cancer Genome Atlas

Graphical Abstract



Authors

Lawrence A. Donehower, Thierry Soussi, Anil Korkut, ..., John N. Weinstein, Rehan Akbani, David A. Wheeler

Correspondence

larryd@bcm.edu

In Brief

Donehower et al. performed a comprehensive analysis of the effects of *TP53* gene mutation in 32 cancer types and 10,225 patients from The Cancer Genome Atlas (TCGA). Data synthesized from five different analysis platforms show how mutant *TP53* increases genomic instability and induces major pathway signaling changes in cancer cells.

Highlights

- *TP53* mutation effects analyzed by five data platforms in 32 cancers/10,225 patients
- More than 91% of cancers with *TP53* mutations show loss of both functional *TP53* alleles
- *TP53* mutation affects genomic stability, global RNA, miRNA, and protein expression
- Mutant p53 RNA expression signature helps prognostic predictions in 11 cancer types



Integrated Analysis of *TP53* Gene and Pathway Alterations in The Cancer Genome Atlas

Lawrence A. Donehower,^{1,2,11,*} Thierry Soussi,^{3,4,5} Anil Korkut,⁶ Yuxin Liu,⁶ Andre Schultz,⁶ Maria Cardenas,¹ Xubin Li,⁶ Ozgun Babur,⁷ Teng-Kuei Hsu,⁸ Olivier Licharge,^{9,10} John N. Weinstein,⁶ Rehan Akbani,⁶ and David A. Wheeler¹

¹Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX 77030, USA

²Department of Molecular Virology and Microbiology, Baylor College of Medicine, Houston, TX 77030, USA

³Sorbonne Université, UPMC University Paris 06, 75005 Paris, France

⁴Department of Oncology-Pathology, Cancer Center Karolinska (CCK), Karolinska Institutet, Stockholm, Sweden

⁵INSERM, U1138, Équipe 11, Centre de Recherche des Cordeliers, Paris, France

⁶Department of Bioinformatics and Computational Biology, Division of Science, M.D. Anderson Cancer Center, Houston, TX 77030, USA

⁷Computational Biology Program, Oregon Health and Science University, Portland, OR 97239, USA

⁸Department of Biochemistry & Molecular Biology, Baylor College of Medicine, Houston, TX 77030, USA

⁹Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA

¹⁰Computational and Integrative Biomedical Research Center, Baylor College of Medicine, Houston, TX 77030, USA

¹¹Lead Contact

*Correspondence: larryd@bcm.edu

<https://doi.org/10.1016/j.celrep.2019.07.001>

SUMMARY

The *TP53* tumor suppressor gene is frequently mutated in human cancers. An analysis of five data platforms in 10,225 patient samples from 32 cancers reported by The Cancer Genome Atlas (TCGA) enables comprehensive assessment of p53 pathway involvement in these cancers. More than 91% of *TP53*-mutant cancers exhibit second allele loss by mutation, chromosomal deletion, or copy-neutral loss of heterozygosity. *TP53* mutations are associated with enhanced chromosomal instability, including increased amplification of oncogenes and deep deletion of tumor suppressor genes. Tumors with *TP53* mutations differ from their non-mutated counterparts in RNA, miRNA, and protein expression patterns, with mutant *TP53* tumors displaying enhanced expression of cell cycle progression genes and proteins. A mutant *TP53* RNA expression signature shows significant correlation with reduced survival in 11 cancer types. Thus, *TP53* mutation has profound effects on tumor cell genomic structure, expression, and clinical outlook.

INTRODUCTION

The p53 tumor suppressor protein is a transcription factor that inhibits cell division or survival in response to various stresses, thus acting as a key fail-safe mechanism of cellular anti-cancer defenses (Kastenhuber and Lowe, 2017; Vousden and Prives, 2009). Frequent mutation of *TP53* in human cancers was initially described by Vogelstein and Minna and colleagues (Baker et al., 1989; Nigro et al., 1989; Takahashi et al., 1989) and was further cataloged in *TP53* mutation databases (Bouaoun et al., 2016; Caron de Fromentel and Soussi, 1992; Hollstein et al., 1991;

Leroy et al., 2014). These compilations established that the pattern of *TP53* mutations in several cancers was linked to carcinogen exposure and provided key insights into key p53 functional domains and the role of p53 in cancer etiology (Caron de Fromentel and Soussi, 1992; Hollstein et al., 1991). Most *TP53* mutations occur in the central DNA binding domain and result in inactivated transcription factor function. Some missense mutations have been associated with dominant-negative inhibition of wild-type p53 and/or oncogenic gain of function in the absence of wild-type p53 in experimental contexts (Muller and Vousden, 2013, 2014; Soussi and Wiman, 2015). Missense mutations often render p53 protein resistant to proteolytic degradation by E3 ubiquitin ligases, such as MDM2, ensuring high levels of stable mutant p53 protein (Frum and Grossman, 2014). The most typical *TP53* mutation configuration is a single *TP53* mutation with loss of the remaining *TP53* allele through a deletion on chromosome band 17p (Baker et al., 1989, 1990). Other less-frequent configurations include mutations of both *TP53* alleles or mutation of one allele and retention of the second wild-type allele. Homozygous *TP53* deletion is a rare event likely because of closely linked cell essential genes (e.g., *POLR2A*) (Liu et al., 2015; Mulligan et al., 1990). Studies in acute myeloid leukemia (AML) have shown that copy-neutral loss of heterozygosity can lead to tumors with the same *TP53* variant in both alleles (Jasek et al., 2010).

The p53 pathway is affected by numerous upstream regulators. In turn, it regulates many targets transcriptionally and non-transcriptionally (Kastenhuber and Lowe, 2017; Mello and Attardi, 2018). Regulators of p53 include binding proteins that stabilize or destabilize it, as well as enzymes that post-translationally modify p53 and activate, deactivate, or modulate its functions (Dai and Gu, 2010; Nguyen et al., 2014). p53 transcriptionally targets hundreds of genes, with targeted specificity depending on cell type and context as well as on the nature and intensity of the cell stressor (Fischer, 2017; Mello and Attardi, 2018; Vousden and Prives, 2009). p53 upregulates genes that encode cell cycle inhibitors, apoptosis inducers, DNA repair



proteins, and metabolic regulators (Vousden and Prives, 2009). The p53 protein also downregulates genes associated with cell cycle progression, although the process involves CDKN1A/p21-dependent indirect mechanisms of gene regulation (Engeland, 2018).

Clinically, *TP53* mutations have been linked to a poorer prognosis for some cancers, but this remains controversial (Robles and Harris, 2010). Many variables, such as cancer type, clinical stage, study size, or quality of *TP53* mutation detection can affect prognostic determinations. *TP53* mutation status is not always indicative of p53 signaling status in a cancer cell (Leroy et al., 2014). For example, p53 signaling can be attenuated through non-mutational mechanisms, such as amplification of p53-negative regulators *MDM2*, *MDM4*, or *PPM1D* (Lu et al., 2008; Matheu et al., 2008; Soussi and Kroemer, 2018; Waslylichen and Lozano, 2016). Thus, a more accurate readout of p53 function in human cancers not based strictly on mutation of the *TP53* gene might lead to more accurate prognostic predictions.

We have capitalized on the integrated The Cancer Genome Atlas Network (TCGA) approach in which many tumors in many cancer types have been simultaneously examined on five independent data platforms, along with extensive clinical annotation, to develop a comprehensive picture of the role of *TP53* mutation. We broadly show how *TP53* mutation confers an array of genomic, transcriptomic, and proteomic changes in cancers that has a major role in regulating clinical outcomes.

RESULTS

TP53 Mutation Profile in TCGA Dataset Is Similar to That of Current TP53 Databases

We analyzed *TP53* mutations in whole exome sequences of 10,225 TCGA patients across 32 different cancer types. We identified 3,786 patients with *TP53* mutations. *TP53* mutation frequencies by cancer type were variable, with ovarian cancer and uterine carcinosarcoma showing greater than 90% incidence, and seven cancer types displaying less than 5% incidence (Figure 1A). The spectrum of TCGA *TP53* mutations obtained by whole-exome sequencing was broadly similar to that detected by conventional Sanger sequencing. Notable hot-spot mutations at codons 175, 248, and 273 in TCGA dataset were also the most frequently mutated *TP53* codons observed in the Sanger dataset from the UMD_TP53 database (36,350 patients with *TP53* mutations, 4,300 different variants) (Leroy et al., 2014) (Figures 1B, 1C, S1A, and S1B). Most TCGA *TP53* mutations occurred in the central DNA-binding domain, encompassing exons 4–8, and aligned with those reported in the Sanger dataset of the UMD_TP53 mutation database. In both groups, missense (~40%), frameshift deletion (~20%), and frameshift insertion (~10%) comprised roughly 70% of all mutations (Figures S1A and S1B). Less frequent were in-frame deletions and insertions, synonymous mutations, and splice-site mutations (Figure S1B). In both datasets, mutation types were enriched in the central domain, although the small in-frame deletions and insertions were surprisingly dispersed (Figure S1A). The TCGA mutation dataset also includes infrequent *TP53* variants in the two newly discovered alternative *TP53* exons, exon beta and

gamma (Figures 1C and S1B). Whether these variants are true somatic pathogenic variants or rare, benign polymorphisms is unknown because this region of the *TP53* gene has not been thoroughly analyzed.

We combined *TP53* single-nucleotide variants (SNVs) from TCGA and UMD_TP53 datasets to assemble a graph showing the fraction of nine possible codon changes for each of the 393 *TP53* triplets encoding the full-length p53 (Figure 1D). The central domain shows at least six of the nine possible codon mutations in almost every location, whereas sites in the N- and C-terminal domains show an absence of any codon changes. The N-terminal mutation-free region correlates closely with the MDM2-binding domain of p53. Because MDM2 is an E3 ubiquitin ligase that mediates p53 degradation, mutation of this binding site should stabilize p53, rather than inactivate it. A mutation-resistant C-terminal domain was also noted, but the mechanism for mutation avoidance is unclear. The absence of variants at codon 114 in loop L1 confirms a study demonstrating that *TP53* variants at this position have little or no functional effect (Zupnick and Prives, 2006).

TCGA effort identified a total of 384 previously unreported variants, most of which were frameshift insertions and deletions, although 69 unreported SNVs were noted (Figures S1C–S1E). These SNVs were more frequently observed in exons 2–4 and 10–11 (Figure 1C, bottom panel), likely because historical *TP53* mutation studies generally sequenced only exons 5–8 or 5–9.

The availability of functional data for most missense *TP53* variants is an unusual feature among cancer genes. In one study, Kato et al. (2003) measured the transcriptional activity of *TP53* variants in yeast assays using eight different *TP53* response elements. A clear correlation between *TP53* variant loss of activity and its frequency in human tumors has been shown (Soussi et al., 2005). Recently, two large-scale analyses of *TP53* variants were performed in mammalian cells using cellular assays measuring variant effects on cell proliferation as a quantitative functional readout (Giacomelli et al., 2018; Kotler et al., 2018). Using the Kato et al. (2003) data, we confirmed the clear correlation between *TP53* loss of activity and the high-frequency *TP53* variants in cancer (Figures S2A and S2B). *TP53* variants from TCGA dataset are also clearly associated with a loss of activity for the eight different *TP53* response elements (see Figure S2B for the *CDKN1A/p21/WAF1* and *AIP1* promoter results). Similarly, we showed that functional analysis data of Kotler et al. (2018) and Giacomelli et al. (2018) correlate strongly with the frequency of *TP53* variants in human cancer and that *TP53* variants from TCGA dataset are generally associated with deleterious *TP53* variants (Figure S2B, bottom graphs). The previously unreported TCGA *TP53* variants showed a relatively high residual p53 activity in all functional assays (Figure S2A and red boxes in Figure S2B).

With respect to the p53 three-dimensional structure, deleterious *TP53* variants were clustered in specific regions of the p53 protein, such as loop L2 and L3, and part of a Helix Loop Helix, including sheet S10 and helix H2 (Figure 1E). These domains are highly enriched for mutations in the UMD_TP53 Mutation Database and TCGA dataset and variants exhibit major functional consequences at virtually every codon (Figure 1E). Loop

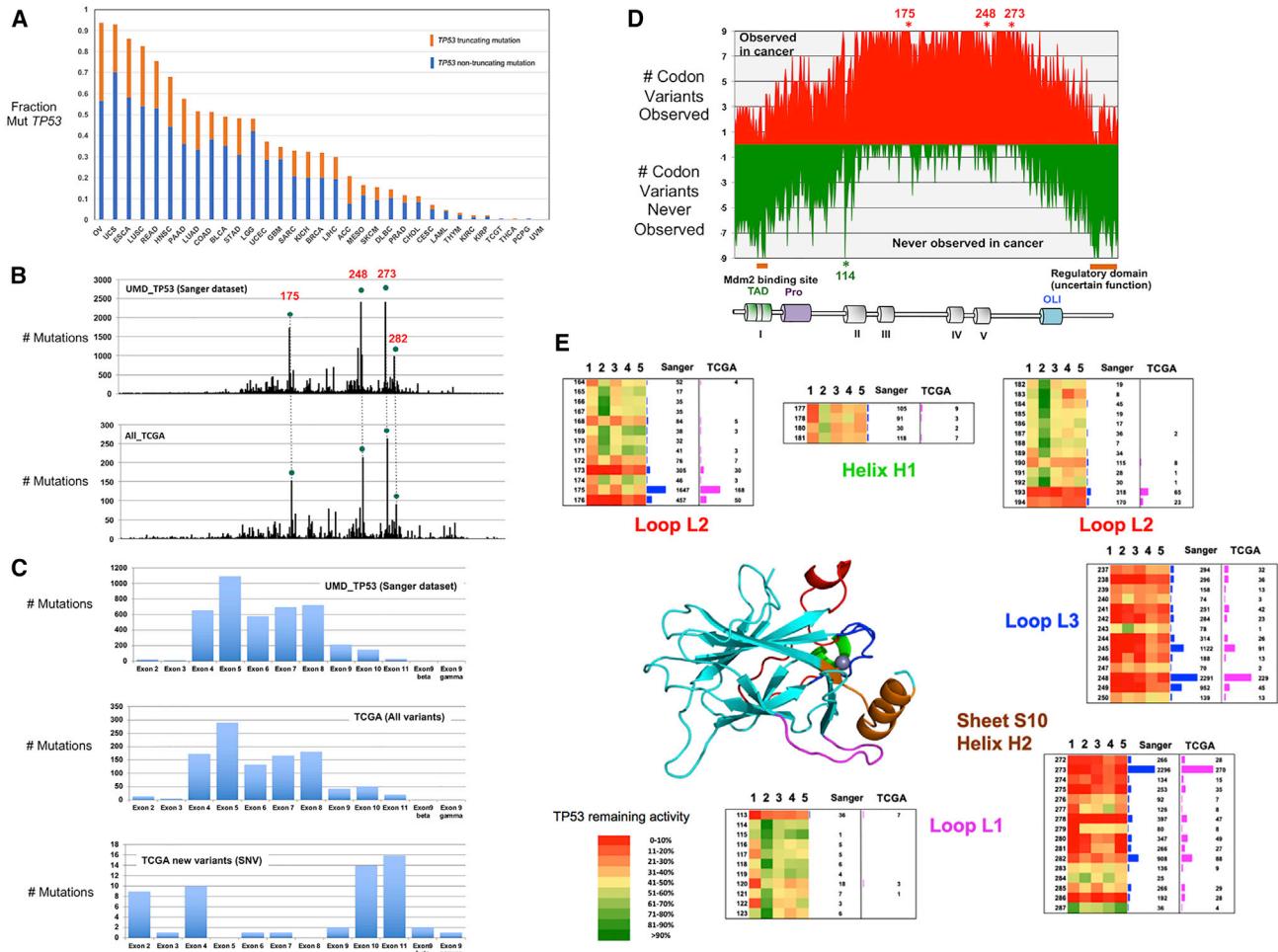


Figure 1. TP53 Mutation Profile in TCGA Dataset Is Similar to That of Current TP53 Databases

(A) TP53 mutation frequency depends on tumor type. For each cancer type, relative fractions of non-truncating and truncating mutations are indicated.

(B) Distribution of global TP53 mutations in TCGA dataset are similar to the Sanger sequenced subset of the UMD_TP53 Mutation Database. TP53 hotspot mutations are indicated.

(C) Exon distribution of TP53 mutations for variants in the Sanger subset of the UMD_TP53 Mutation Database (top panel) are similar to that observed for TCGA dataset TP53 mutations (middle panel). Previously unreported TP53 single-nucleotide variants (SNVs) discovered by TCGA are in the bottom panel.

(D) Distribution of codon SNV sites along the p53 protein recorded to date. For each codon, nine SNVs are possible. Red bars indicate SNVs observed for each codon, whereas green bars indicate SNVs not yet reported. Unmutated residues at codon 114 and hotspot residues at codon 175, 248, and 273 are shown with asterisks. TAD, transactivation domain; Pro, proline-rich domain; OLI, tetramerization domain; I to V, evolutionarily conserved domains.

(E) TP53 variants in functional domains of p53. The DNA-binding surface of p53 protein is composed of two large loops (L2 and L3) stabilized by a zinc ion. Sheet S10 and Helix H2 are components of the LSH (Loop Sheet Helix) domain that includes loop L1. These various domains are essential for DNA recognition. Variants are depicted for each structural domain. Sanger, number of TP53 variants found in the Sanger dataset of the UMD_TP53 database; TCGA, number of TP53 variants in TCGA dataset. Variant frequencies are denoted by bars. The heatmap corresponds to loss of function of the TP53 variants measured by the functional assays of Kato et al. (2003) (1); Kotler et al. (2018) (2), and the three functional assays of Giacomelli et al. (2018) (3–5).

See also Figures S1 and S2.

L3 in particular is critical for DNA binding and exhibits the highest number of mutations.

Inactivation of Both Alleles Occurs in More Than 90% of TCGA Cancers with TP53 Mutations

Tumor suppressors are considered to be recessive at the genetic level; inactivation of both tumor suppressor alleles is generally required for an oncogenic phenotype (Knudson, 1989). While TP53 is a tumor suppressor, there has been evidence that

missense mutations in TP53 result in expression of a stabilized protein of altered conformation that exhibits both dominant-negative activity toward wild-type p53 protein as well as gain of function oncogenic activity (Muller and Vousden, 2013; Soussi and Wiman, 2015). Therefore, mutation of only one TP53 allele could potentially result in a significant oncogenic phenotype.

To assess TP53 allele status in TCGA dataset, we performed an integrated analysis of TP53 mutant tumors using two data platforms: exome sequence data and copy number data.

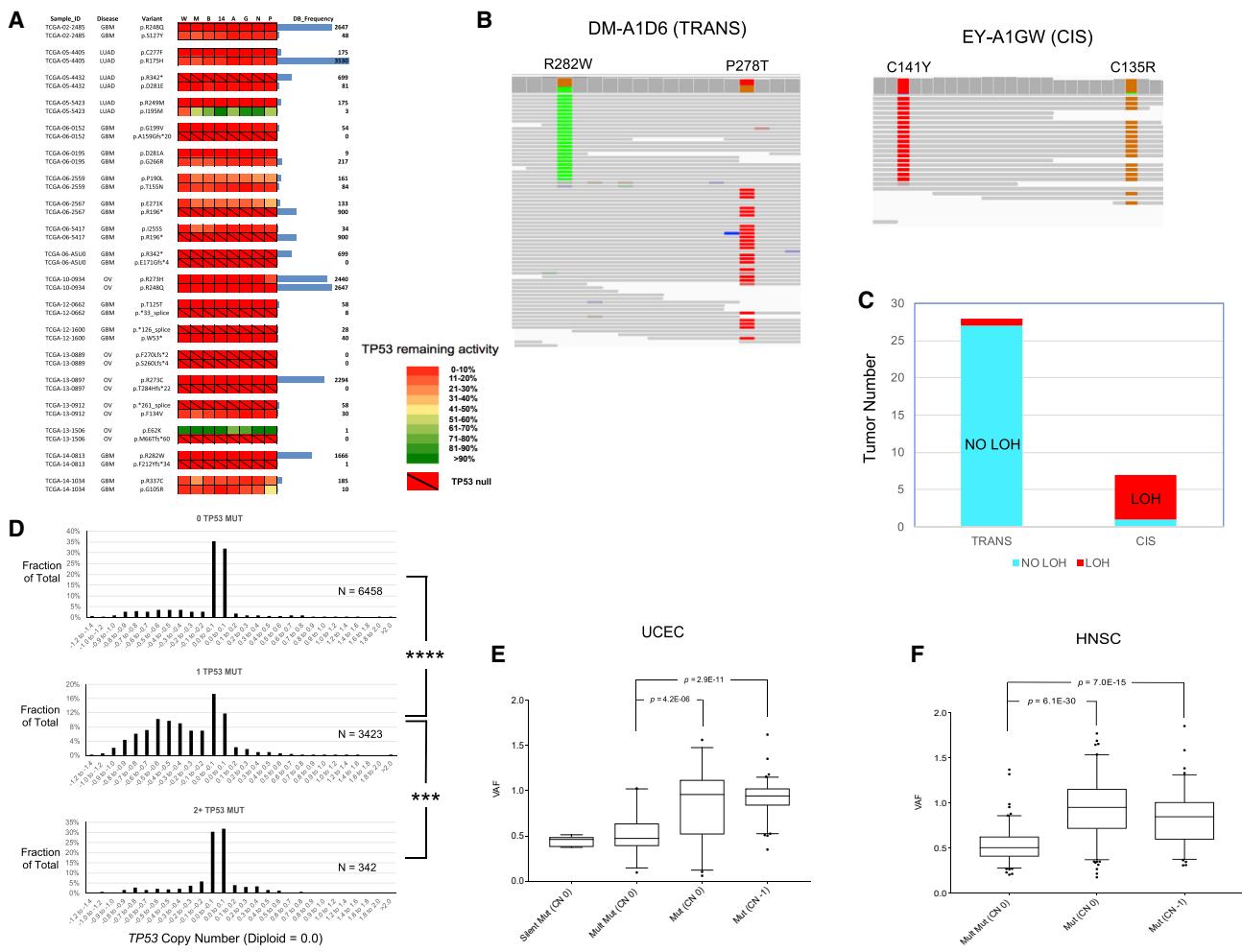


Figure 2. Inactivation of Both Alleles Occurs in Most TCGA Cancers with *TP53* Mutations

(A) p53 functional analyses in some TCGA tumors with two *TP53* mutations. Heatmaps show relative transcriptional activity of *TP53* variants compared with wild-type *TP53* based on data in Kato et al. (2003). Each column shows a p53 transcriptional target, and each row shows a *TP53* variant. W, WAF1 (*CDKN1A*); M, *MDM2*; B, *BAX*; 14, 14-3-3 sigma (*SFN*); A, AIP (*TP53AIP1*); G, *GADD45A*; N, *NOXA* (*PMAIP1*); P, p53R2 (*RRM2B*). Database variant frequency is shown as a blue bar.

(B) Individual DNA sequence reads from tumors with two closely linked *TP53* mutations show *trans* (left) and *cis* (right) mutation configurations. Mutations are at the top. Gray boxes represent individual nucleotides. Gray lines show individual sequence reads, with colored segments indicating mutations.

(C) Most tumors with two closely linked *TP53* mutations have *trans* mutations and retain diploid copy number (NO LOH [no loss of heterozygosity]), whereas a minority have *cis* mutations but show wild-type *TP53* allele loss (LOH [loss of heterozygosity]).

(D) Tumors with one *TP53* mutation exhibit *TP53* copy number loss, whereas tumors with 0 or 2 *TP53* mutations are largely diploid. Copy number values at the *TP53* locus for tumors with 0 (top panel), 1 (middle panel), or 2–3 (lower panel) *TP53* mutations are shown. On the x axis, *TP53* copy-number values are binned in 0.1-value increments, where 0 represents diploidy and values of –0.4 to –0.6 are roughly equivalent to a haploid copy number. Significant differences among each category are indicated. *** $p < 1E-50$, ** $p < 1E-25$.

(E and F) Median variant allele fractions (VAFs) in tumors with one *TP53* mutation approximate 1.0, indicating frequent loss of both wild-type *TP53* alleles. Uterine corpus endometrial carcinoma (UCEC) (E) and head and neck squamous cell carcinoma (HNSC) (F) were stratified by mutation number and copy number. A copy number (CN) of 0 is considered diploid and CN of -1 is considered haploid. "Mult Mut" indicates tumors with 2⁺ *TP53* mutations. VAF distributions are shown, and median values are indicated by the central bar in the box and whiskers plots. Statistical significance was indicated by t tests.

See also Figure S3.

First, we examined the 10% of *TP53* mutant tumors (381 of 3,786 total tumors with *TP53* mutations) with two distinct *TP53* mutations. Functional analysis of the *TP53* mutations revealed that both mutations were usually inactivating (Figure 2A). To determine whether the two *TP53* mutations were on the same allele (*cis*) or on different alleles (*trans*), we examined individual DNA

sequencing reads that were in the same exon and 6–75 nucleotides apart. Those tumors with individual reads consistently displaying both mutations were considered *cis* for *TP53* mutations, and those tumors consistently exhibiting only one of the two mutations were considered trans. Representative read files for *cis* and trans tumors are shown in Figure 2B. For those 35 bi-allelic

TP53 mutation tumors with tightly linked mutations, 28 (80%) were trans and 7 (20%) were *cis* (Figure 2C). Assessment of these 35 tumors for copy number status revealed that 6 of the 7 *cis* tumors exhibited *TP53* copy number loss, while 27 of the 28 trans tumors retained a diploid *TP53* status (Figure 2C). Thus, there is selection for wild-type *TP53* allele loss in the *cis* tumors, whereas trans mutant *TP53* alleles show no selection for allele loss. When we looked at functional effects of individual *TP53* mutations in the *cis* and trans tumors, 5 of the 7 *cis* tumors contained at least one *TP53* mutation with partial wild-type function, whereas almost all of the trans *TP53* mutations resulted in two *TP53* alleles without any wild-type function (Figure S3A).

We also examined *TP53* copy number across all TCGA cancers stratified by 0, 1, or 2+ *TP53* mutation categories. While tumors with no *TP53* mutations were generally diploid at the *TP53* locus, tumors with a single *TP53* mutation showed significant skewing toward copy number values indicating loss of a single *TP53* allele (Figure 2D). Tumors with two or more *TP53* mutations largely retained a diploid copy number.

A closer look at *TP53* DNA copy number status in the single *TP53* mutation tumors showed that 66% of these tumors displayed *TP53* copy number loss, while 34% were apparently diploid in *TP53* copy number. Analysis of individual DNA sequencing reads within multiple cancer types with a single *TP53* mutation revealed that the variant (mutant) allele fraction (VAF) of the *TP53* reads averaged close to 1.0 in tumors with *TP53* allele copy number loss (Figures 2E, 2F, S3B, and S3C). Most tumors with a single *TP53* mutation in a diploid context also displayed a variant allele fraction close to 1.0. This suggests that the mutant *TP53* allele is frequently duplicated through mitotic recombination or another gene duplication mechanism. Consistent with our earlier result (Figure 2C), tumors with two or more *TP53* mutations averaged a variant allele fraction near 0.5 (Figures 2E, 2F, and S2A). In one tumor type (UCEC), tumors with a *TP53* silent mutation showed a variant allele fraction approximating 0.5, indicating little selection for *TP53* allele loss or for mutant allele duplication (Figure 2E). Among tumor types with frequent *TP53* mutations, all showed these patterns (Figures 2E, 2F, S3B, and S3C). Among six tumor types with frequent *TP53* mutation, 91.3% exhibited variant allele fractions (VAFs) consistent with loss of both wild-type alleles, while only 8.7% displayed a likely retention of the wild-type *TP53* allele (Figure S3D). This pattern is repeated among all other cancer types with statistically sufficient numbers of *TP53* mutations (Figure S3D). Thus, the data indicate a strong selection for loss of the second *TP53* allele after mutation of the first allele.

P53 RNA and Protein Expression Are Highly Variable and Dependent on Mutation Type

In the previous section, over 90% of mutant *TP53* tumors displayed loss of the wild-type *TP53* allele. To confirm this phenomenon at the RNA expression level we examined p53 variant allele fractions in the p53 RNA-seq data. After adjusting RNA VAFs for tumor purity, we calculated that over 92% of 799 TCGA tumors with *TP53* missense mutations had p53 VAF values near 1.0 (Figure S3E). Thus, *TP53* generally behaves as a recessive tumor suppressor both at the DNA and RNA level.

Analysis of p53 RNA expression levels in TCGA tumors revealed wide intertumoral variation within both wild-type and mutant *TP53* tumors (Figure S4A). Across all cancer types, p53 RNA in tumors with missense (or in-frame deletion or insertion) *TP53* mutation was modestly increased relative to that in wild-type *TP53* tumors (Figures S4B and S4C). However, p53 RNA expression in tumors with truncating *TP53* mutations (nonsense, frameshift insertion or deletion, splice site) was reduced compared to either wild-type or missense MUT *TP53* tumors (Figures S4B and S4C). This is likely due to nonsense-mediated mRNA decay processes.

The expression of p53 protein in the wild-type and mutant *TP53* tumors as measured by RPPA displayed intertumoral variability as well as some variability based on cancer type (Figure S5A). Cancer types with high fractions of mutant *TP53* tumors generally showed significantly higher levels of p53 protein expression relative to tumors with wild-type p53, though there were a few exceptions (Figure S5B). When the *TP53* mutations in a tumor type were stratified by truncating and non-truncating mutations, both wild-type *TP53* and truncating mutant *TP53* tumors exhibited low p53 protein levels while the non-truncating mutant *TP53* tumors showed significantly higher p53 expression (Figures S5C and S5D). This result may be due to nonsense-mediated decay of p53 truncating mutant RNA and that non-truncating *TP53* mutations often lead to a p53 conformational change with resistance to degradation.

TP53 Mutation Is Significantly Correlated with Increased Chromosomal Instability

P53 has been called the “guardian of the genome,” based on accumulated evidence that it plays a major role in preserving genomic stability (Lane, 1992; Smith and Fornace, 1995; Tappore and Fukasawa, 2002). We examined copy number data for 25,000+ loci in each of the 10,225 TCGA tumors and stratified this data by *TP53* status. We analyzed major copy deviations from diploidy (copy number values of 0). Loci greater than four-fold that of normal copy number (copy number values of greater than 2) were considered indicative of amplification and were counted. Loci with copy number values of less than -1 were considered indicative of deep deletions and these were tallied. The fraction of all TCGA wild-type and mutant *TP53* tumors with amplification and deep deletion were each plotted across the entire genome (Figure 3A). The global genomic profile for deep deletions ($\text{CN} < -1$) shows peaks of enhanced deletions at major tumor suppressors such as *CDKN2A* and *RB1*. Generally, the mutant *TP53* tumors displayed a much greater fraction of tumors with deletions at each frequent deletion site than their wild-type counterparts (Figure 3A, top panel). Likewise, peak regions of amplification (e.g., *MYC*, *CCND1*, *CCNE1*) showed a higher fraction of amplifications in mutant *TP53* tumors relative to wild-type *TP53* tumors (Figure 3A, middle panel). These divergences in amplification frequencies between wild-type and mutant *TP53* were highly significant (Figure 3A, bottom panel). Across the entire genome of all TCGA cancers, mutant *TP53* tumors exhibit roughly 2.5-fold the total number of deletions and amplifications as do wild-type *TP53* tumors ($p = 0$) (Figure 3B). When the relative *TP53*-dependent genomic instability of individual tumor types was compared mutant *TP53* tumors displayed

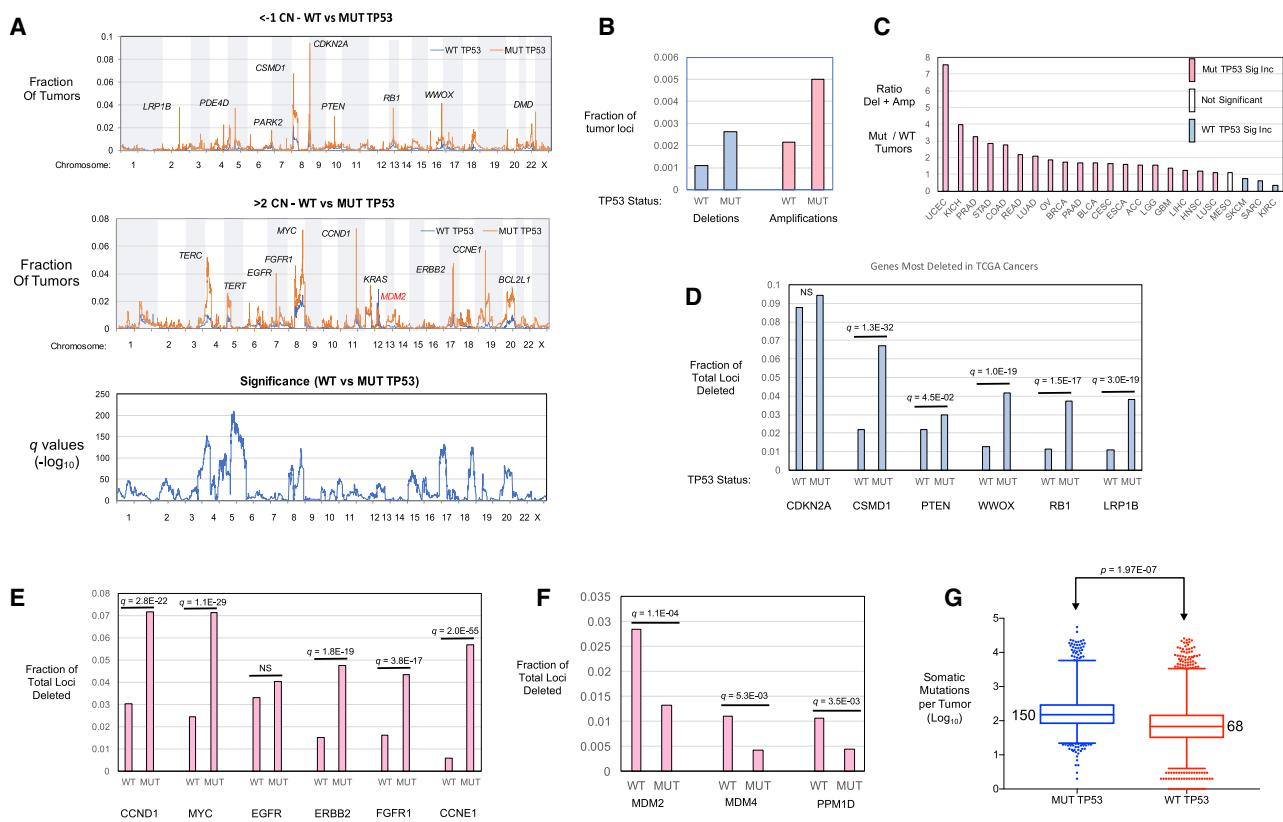


Figure 3. TP53 Mutation Is Correlated with Increased Chromosomal Instability

(A) Genomic profile aggregated from copy number data of all TCGA tumors shows that frequent amplifications and deep deletions occur significantly more in MUT TP53 than they do in wild-type (WT) TP53 tumors. Fraction of copy number losses (CN < -1) (top panel) and gains (CN > 2) (middle panel) for wild-type (blue line) and mutant (orange line) tumors are shown. Peaks correspond to frequent regions of deep deletion (top panel) and amplification (middle panel) and are labeled with the tumor-suppressor gene or oncogene at the epicenter of each peak. Statistical significance in relative frequency of deletion or amplification between WT and MUT TP53 tumors at each gene are indicated in the bottom panel.

(B) Mutant TP53 tumor genomes display roughly 2.5-fold greater rates of amplification and deep deletion relative to their wild-type TP53 counterparts. Gene loci with copy number values greater than 2 (amplification) and less than -1 (deep deletion) were totaled for all wild-type and all mutant TP53 tumors and divided by total loci number in each TP53 category. A t test comparing WT to mutant TP53 tumors was highly significant ($p = 0$).

(C) Most cancer types show significantly increased rates of amplification and deep deletion in the MUT TP53 tumors (Mut TP53 Sig Inc) compared with wild-type TP53 tumors (WT TP53 Sig Inc). The fraction of loci with deep deletion or amplification in MUT TP53 tumors was divided by that in WT TP53 tumors to give a ratio. Significance was determined by t test.

(D) Of the six most-frequent deep deletions, five occurred significantly more frequently in the mutant, compared with the wild-type, TP53 group. For panels (D)–(F), significance between wild-type and mutant TP53 groups was determined by unpaired t test, adjusted for false discovery rate, and presented as q values.

(E) Of the six most-frequent amplifications, five occurred significantly more frequently in the mutant, compared with the wild-type TP53 group.

(F) Three frequently amplified negative regulators of p53 (MDM2, MDM4, and PPM1D) are significantly more amplified in WT, relative to MUT, TP53 tumors.

(G) Nucleotide-level mutation rates are increased in MUT TP53 tumors. Median mutation numbers per tumor for wild-type and mutant TP53 tumors are shown in the box and whiskers plots. An unpaired t test shows that the mutant TP53 tumors have significantly more total mutations per tumor compared with their wild-type counterparts.

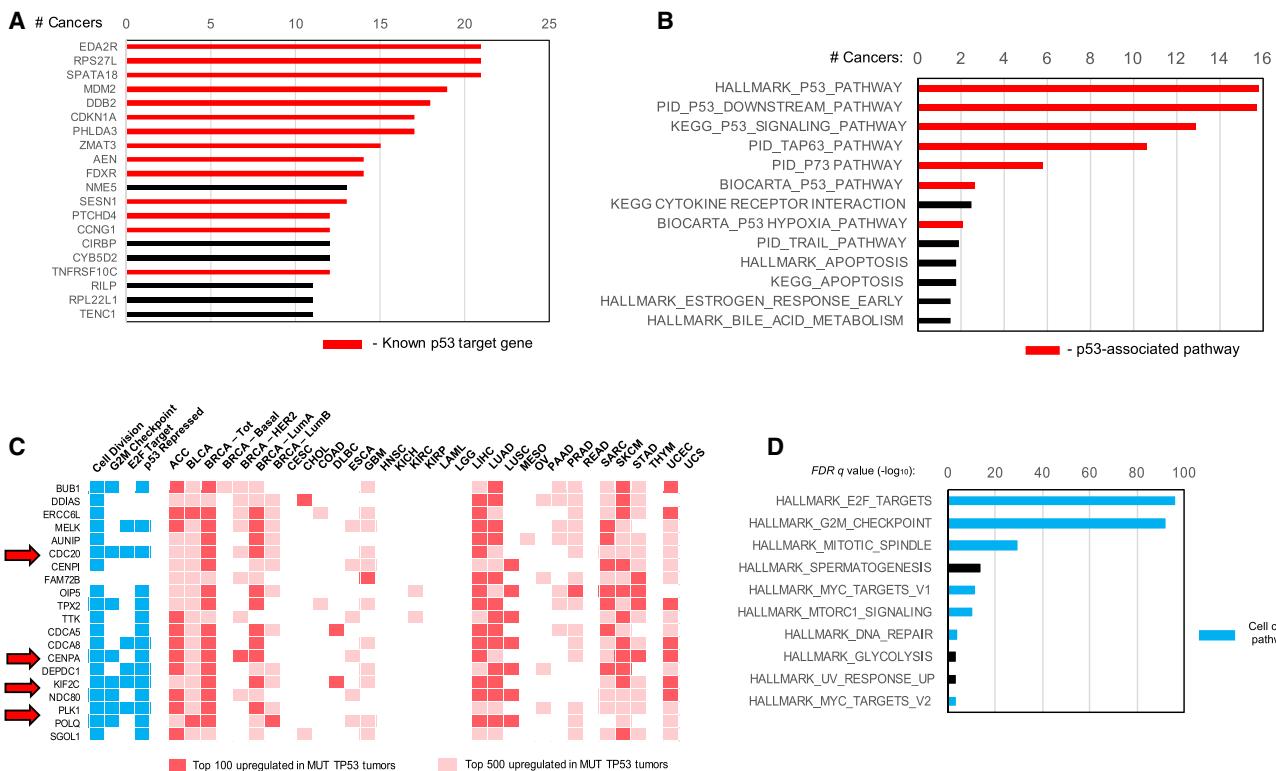
See also Table S1.

significantly more chromosomal instability in 19 of 23 tumor types with sufficient TP53 mutations for comparison (Figure 3C).

When frequent regions of amplification and deletion were examined more closely, major oncogenes and tumor suppressor genes were usually located at the epicenter of each. Again, most of these amplicons and deletions were significantly more frequent in TP53 mutant tumors (Figures 3D and 3E; Tables S1A and S1B). A notable exception were three frequent amplification regions in which MDM2, MDM4, and PPM1D were located at the epicenter, and which were significantly

more amplified in wild-type TP53 tumors (Figure 3F). All three of these genes encode negative regulators of p53 function, illustrating a non-mutational mechanism of inactivating p53 in tumors.

We also assessed the effects of TP53 mutation on genomic instability at the individual nucleotide level by comparing whole exome mutation rates in individual wild-type and mutant TP53 tumors. Across all TCGA tumors we found a moderately increased rate of whole exome nucleotide level mutations in the individual mutant TP53 tumors. The median number of whole

**Figure 4. Comparison of Global RNA Expression Reveals p53-Dependent Pathways in Cancer**

(A) The most significantly upregulated gene RNAs in wild-type *TP53* cancers are mostly known p53 target genes, and the number of cancer types in which they were upregulated is indicated. Direct p53 target genes are indicated by red bars. See also Tables S2 and S3.

(B) Pathway analyses based on genes expressed at significantly higher rates in wild-type *TP53* cancers show that p53-related pathways (red bars) are highly enriched. See also Table S4A.

(C) The most significantly upregulated RNAs in mutant *TP53* cancers are cell-division promoters. For each cancer, the top 100 and 500 gene RNAs most highly expressed in mutant, relative to wild-type, *TP53* tumors were identified, and the top 20 upregulated genes across all mutant *TP53* cancers are shown. Roles for cell division, G2/M checkpoint control, E2F target genes, and documented repression by p53 are indicated by blue boxes. Significant upregulation in individual cancer types are indicated by red and pink boxes. Red arrows indicate four genes comprising the mutant p53 signature discussed later (Figure 7).

(D) Pathway GSEA analysis based on mutant *TP53* upregulated genes in TCGA cancers confirms the importance of cell cycle regulation (blue bars). See also Table S4B.

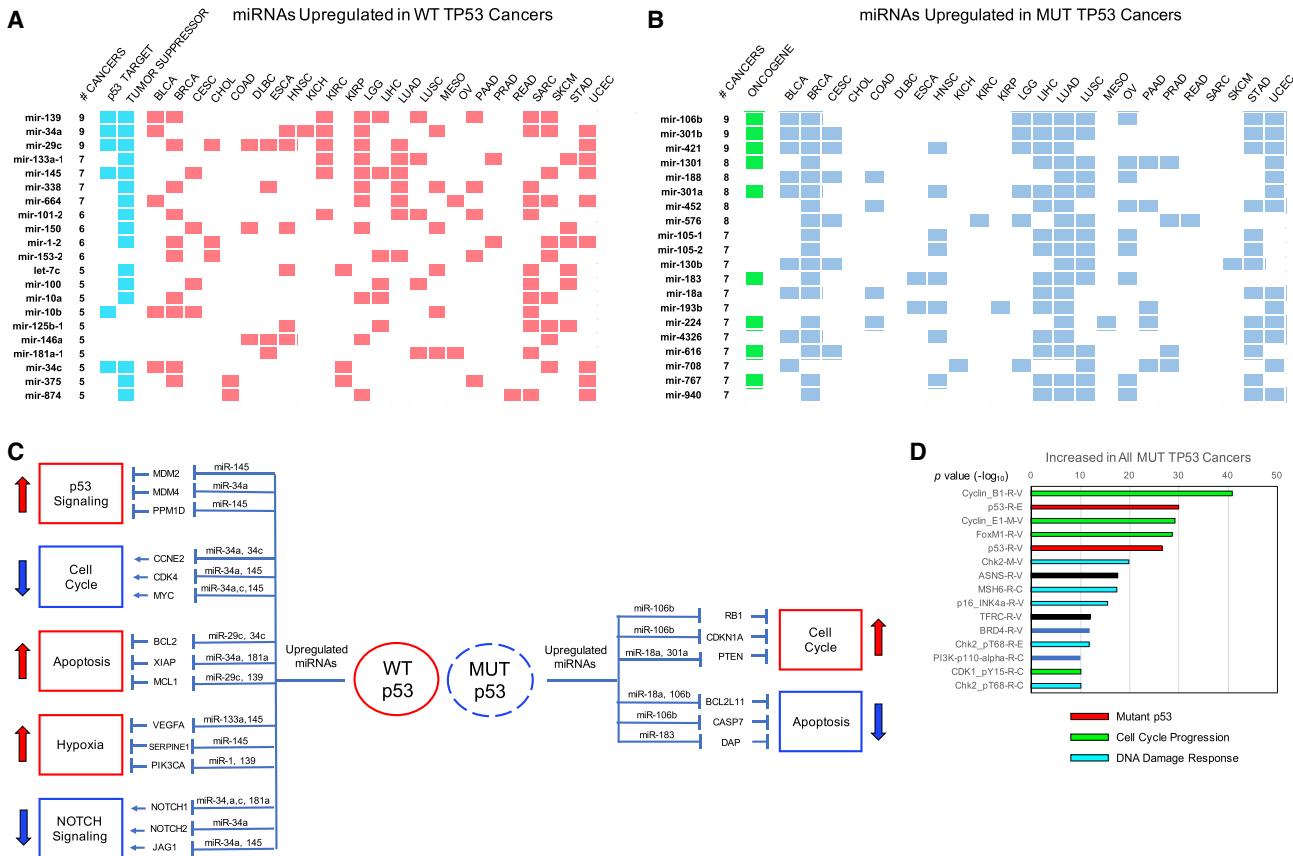
exome mutations across all wild-type *TP53* tumors was 68 compared to 150 in mutant *TP53* tumors (Figure 3G). There did not appear to be significant differences in the types or patterns of mutations based on *TP53* status. Despite the apparent increase in nucleotide level mutation rates in mutant *TP53* tumors, it is difficult to determine whether *TP53* mutation plays a causative role in nucleotide level instability or whether increased *TP53* mutations are correlated with other processes that globally affect cellular mutation rates and patterns. Consistent with the latter idea, a recent analysis of cancer mutational signatures indicated that *TP53* mutation patterns were often dependent on mutational signatures commonly found in the relevant tissue of origin (Giacomelli et al., 2018).

Comparison of Global RNA, MicroRNA, and Protein Expression Reveal p53-Dependent Pathways in Cancer

Because p53 is a transcriptional regulator, we compared gene expression patterns in the wild-type and mutant *TP53* tumors for each cancer type (Figures 4A–4D; Table S2). We directly

compared RNA expression levels for each gene and ranked by t test the most significant up- and downregulated genes in wild-type versus mutant *TP53* tumors. Of the top 20 most upregulated genes in wild-type *TP53* tumors, 14 were established p53 upregulated target genes (Figure 4A; Table S2 and S3). P53 target genes *EDA2R*, *RPS27L*, and *SPATA18* were each significantly upregulated in 21 different wild-type *TP53* cancer types (Figure 4A; Table S3). To perform pathway analyses based on differences in individual gene expression, we identified the top 500 most consistently upregulated genes in wild-type *TP53* tumors and performed gene set enrichment analysis (GSEA) (Subramanian et al., 2005) on them (Table S4A). Not surprisingly, as shown in Figure 4B, most of the significantly enhanced pathways in wild-type *TP53* cancers involved p53-related signaling pathways. Even those enriched pathways not directly related to p53 signaling are functionally associated with p53, and included a number of apoptosis pathways.

Using similar approaches as described above, we also identified the most consistently upregulated genes in mutant *TP53*

**Figure 5. Comparison of Global MicroRNA and Protein Expression Reveals p53-Dependent Pathways in Cancer**

(A) Tumors with wild-type *TP53* show upregulation of a subset of miRNAs relative to tumors with mutant *TP53*. For each tumor type, those miRNAs most significantly upregulated for expression in wild-type, relative to mutant, *TP53* cancers were determined. The top 20 most significantly upregulated miRNAs are shown. Blue rectangles indicate whether each miRNA is a direct p53 transcriptional target or has been shown to exhibit tumor-suppressor activity. Pink rectangles indicate the miRNA is significantly upregulated in specific cancers. See also Table S5A.

(B) Tumors with mutant *TP53* show significant upregulation of a subset of miRNAs relative to tumors with wild-type *TP53*. For each tumor type, the top 20 miRNAs most significantly upregulated for expression in mutant *TP53* cancers are shown. Green rectangles indicate miRNA oncogenic activity, and blue rectangles indicate significant upregulation in specific cancers. See also Table S5B.

(C) Schematic diagram outlining miRNAs significantly upregulated in wild-type *TP53* cancers (left) and mutant *TP53* cancers (right), their key target genes, and their proposed pathway impacts.

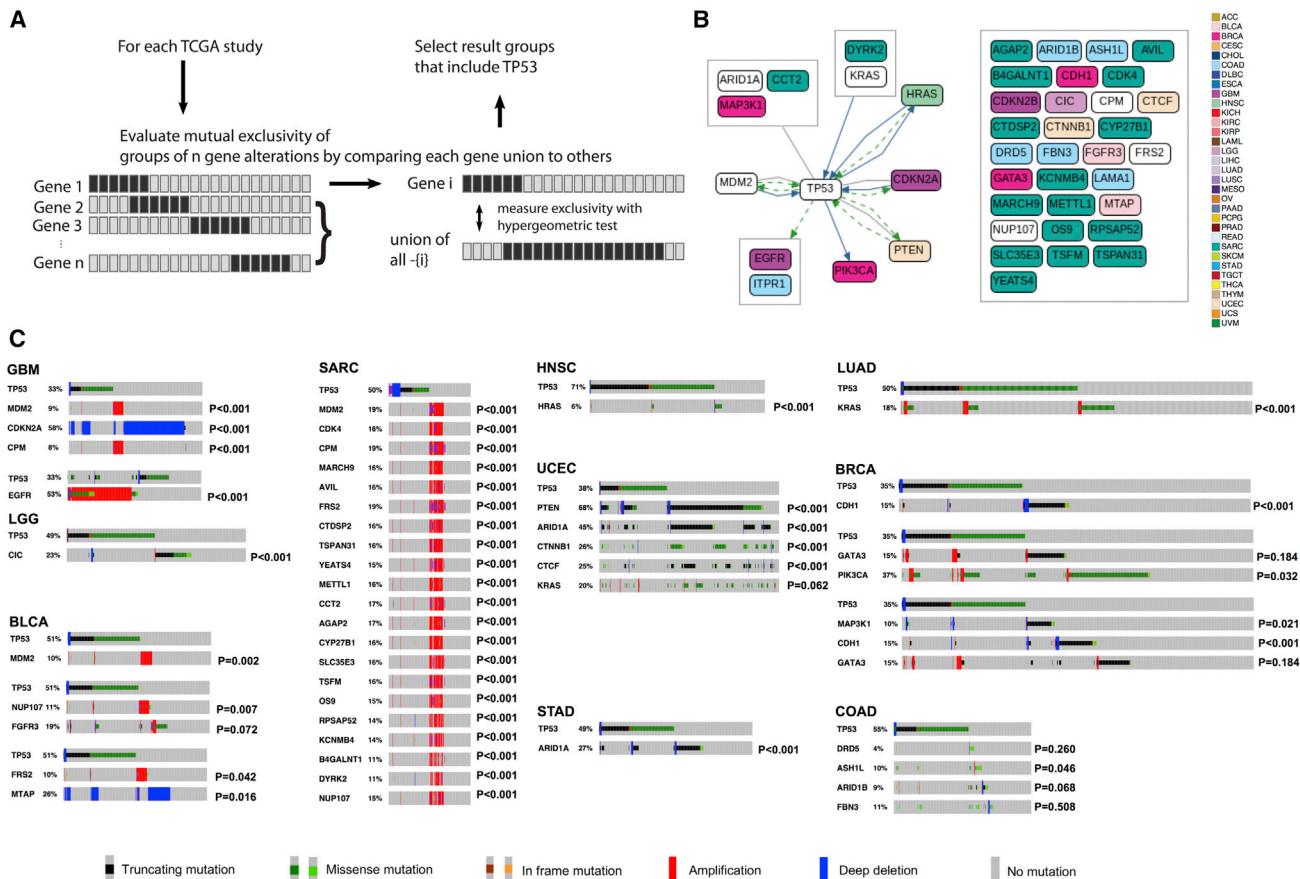
(D) GSEA analysis of RPPA data indicates that proteins most upregulated in TCGA mutant *TP53* tumors are enriched for cell-cycle progression (green bars) and DNA-damage response (blue bars). See Table S6B.

cancers. All but one of the 20 most consistently upregulated genes were directly related to cell division regulation (Figure 4C). Many of these top 20 genes also played roles in cell cycle promotion (particularly G2/M checkpoint control), were E2F targets (E2F is a major S phase/G2 promoting transcription factor), and were documented to be repressed by wild-type p53 protein (Figure 4C). GSEA performed on the top 500 most upregulated genes in mutant *TP53* tumors revealed highly significant enrichment of pathways directly involved in promoting cell cycle progression (Figure 4D; Table S4B).

Differential microRNA expression was also compared, as p53 is known to transcriptionally target a number of specific miRNAs (Hermeking, 2012; Lujambio and Lowe, 2012). Of the top 20 microRNAs consistently upregulated in wild-type *TP53* cancers, 6 of these had previously been shown to be direct transcriptional

targets of p53 (Figure 5A; Table S5A). A literature search of these top 20 miRNAs indicated that 16 of them had previously been associated with tumor suppressor activity in other experimental contexts. Among the 20 miRNAs consistently upregulated in mutant *TP53* cancers, 9 had been experimentally associated with oncogenic function (Figure 5B; Table S5B). An examination of the gene targets of the miRNAs upregulated in mutant and wild-type *TP53* cancers suggested that in wild-type *TP53* cancers there appeared to be enrichment for miRNAs that enhance apoptosis, p53 signaling, and hypoxia as well as suppressing cell cycle progression and NOTCH signaling (Figure 5C). In contrast, miRNAs enriched in mutant *TP53* cancers appeared to promote cell cycle progression and reduce apoptosis (Figure 5C).

We also compared TCGA RPPA data to determine whether *TP53* mutation status affected protein expression

**Figure 6. Genomic Alterations Mutually Exclusive with TP53 Mutations**

(A) Schematic overview of Mutex algorithm for determining gene mutual exclusivity (see Method Details).

(B) Pathway representation of mutually exclusive partners of TP53 across all cancer types. The color code represents the cancer type in which mutual exclusivity is observed. White indicates mutual exclusivity in more than one cancer. Genes are grouped based on their topology in the network. The disconnected gene groups on the right do not have connections to TP53 in the Pathway Commons database.

(C) The oncrint representation of TP53 mutation mutual exclusivity genomic modules. Note the partners in SARC are all amplified on the same samples, which is a chromosomal event at chromosome 12q. The p values represent significance of the observed mutual exclusivity between each gene and TP53 as calculated by a Fisher's Exact Test.

See also Figure S6.

(Tables S6A and S6B). In TP53 mutant tumors, we observed significant upregulation of proteins associated with cell cycle progression, such as cyclin B1, cyclin E1, FOXM1, and CDK1 (Figure 5D; Table S6B). As expected, mutant TP53 cancers contained enhanced expression of p53, presumably the mutant form which is known to be stabilized following a non-truncating mutation. Proteins associated with DNA damage response were also shown to upregulate in mutant TP53 cancers.

Genomic Alterations Correlated with TP53 Mutation Status Vary by Cancer Type

Driver oncogenic aberrations in the same pathway are rare since only one alteration is generally necessary to induce pathway signaling alteration. Thus, frequent mutually exclusive alterations within a cancer type may reveal multiple members of one signaling pathway, whereas frequent co-occurring mutations may be indicative of distinct but cooperating pathways. To

catalog the aberrations that share similar oncogenic endpoints with TP53 mutations, we identified the genomic alterations that are mutually exclusive with TP53. For this purpose, we used the Mutex algorithm, which performs a greedy search to identify modules of genetic alterations with maximized mutual exclusivity score (see STAR Methods) (Figure 6A). We have performed the analysis in 32 cancer types using both mutation and copy number data and detected all the mutual exclusivity modules that involve TP53. Searches in the Pathway Commons database identified known pathway interactions between TP53 and a fraction of the mutually exclusive genes (Figure 6B). MDM2, KRAS, and ARID1A are mutually exclusive with TP53 in more than one cancer type and their functional relation to TP53 has been characterized. In addition, a group of genes demonstrated mutual exclusivity with TP53 in at least one cancer type but searches in signaling databases did not capture a known functional relationship to TP53. Those genes are grouped without a connection to

TP53 in the pathway diagram on Figure 6B. Among those, *NUP107*, *FRS2*, and *CPM* were mutually exclusive with *TP53* in multiple cancer types. It is not certain whether these genes are oncogenic drivers and have a true functional relation with *TP53* as they co-locate with *MDM2* on the chromosome 12q, suggesting co-amplification. Mutations and copy number alterations showing mutual exclusivity with *TP53* are shown in Figure 6C. In most cases, *TP53* was mutually exclusive with known driver genes. In glioblastoma multiforme (GBM), *TP53* mutations were mutually exclusive with alterations in *MDM2*, *CDKN2A*, and *EGFR*. In breast cancers, we observed mutually exclusivity between alterations in *TP53* and *GATA3* in two separate modules with contributions from *PI3KCA* in module and *MAPK3* and *CDH1* in the other. Interestingly, *TP53* mutations were mutually exclusive with mutations in *KRAS* in LUAD and with *HRAS* in HNSC. *ARID1A* mutations were mutually exclusive with *TP53* mutations in UCEC and STAD, while *ARID1B* mutations were detected in COAD. *MDM2* amplifications were mutually exclusive with *TP53* in three cancer types (GBM, bladder urothelial carcinoma [BLCA], and sarcoma [SARC]).

With respect to *TP53* co-occurring mutations, we noted that *IDH1* and *ATRX* were highly significant in LGG and GBM (Figure S6A), as has been previously reported (Kannan et al., 2012; Liu et al., 2012). In UCEC, *TP53* and *PPP2R1A*, a negative growth regulatory subunit of protein phosphatase 2, have highly significant co-occurring mutations, as previously noted (McConechy et al., 2012). *TP53* mutations also significantly occur with certain copy number alterations, particularly those that regulate cell cycle progression and cell division (Figure S6B). These include amplifications of *CCNE1* in UCEC and STAD, *CCND1* in HNSC, *CCND2* in LGG, and *MYC* in PAAD, UCEC, and BRCA, as well as *TERT* in LUAD and BRCA.

Development of a Prognostically Useful Mutant *TP53* Signature

There are numerous studies examining whether *TP53* mutation affects survival and in some cases *TP53* mutation has been associated with poorer prognosis (Robles and Harris, 2010). In our initial stratification of TCGA cancers into wild-type and mutant *TP53* categories, we examined the effect of *TP53* mutation on overall survival. For most TCGA cancer types, no statistically significant differences in overall survival were observed for mutant versus wild-type *TP53* cancers (Table S7A) but that may be due to the limited clinical follow-up for some of TCGA patient population. We showed that *TP53* mutations across all TCGA cancers were significantly more frequent in patients with less than 1 year survival after diagnosis (Figure S7A). We assessed age at initial diagnosis for each cancer type and found significant differences in only a few cancer types (Figure S7B). Notably, *TP53* mutations in cervical squamous cell carcinoma (CESC) associated with later diagnosis, suggesting human papillomavirus infection may be more carcinogenic than *TP53* mutation (because these two events tend to be mutually exclusive) (Crook et al., 1992).

Using *TP53* mutation as a prognostic marker, although useful in some contexts, may not be useful in others. For example, p53 protein function can be inactivated not only by mutation but also by overexpression of key p53 regulatory proteins, such as

MDM2, *MDM4*, or *PPM1D*. Thus, we sought to develop an expression signature correlated with *TP53* mutation that might be more prognostically predictive. We tested a mutant p53 expression signature based on the expression of four cell-cycle regulatory genes consistently upregulated in mutant *TP53* cancers, *CDC20*, *CENPA*, *KIF2C*, and *PLK1* (Figure 4D, arrows). We ranked all TCGA tumors of each type for relative RNA expression of these four signature genes and averaged their rankings for a final signature score. High signature scores correlated well with *TP53* mutation status. For each cancer type with sufficient *TP53* mutation numbers, we then compared overall survival of patients displaying signature scores within the bottom quartile to those within the top quartile of mutant *TP53* signature expression. We found that 11 of 24 cancer types with high p53 signatures resulted in significantly poorer overall survival relative to their low-signature counterparts (Figures 7A–7D and S7C–S7F; Table S7B). For 8 of those 11 cancer types, the mutant *TP53* signature was distinctly better at predicting poorer outcomes than *TP53* mutation status was (Figures 7B, 7D, and S7C–S7F). Two representative cancers, lower-grade glioma (LGG) and skin cutaneous melanoma (SKCM), showed strong correlation of *TP53* mutation and p53 signature (Figures 7A and 7C), but the p53 signature was much more prognostic than *TP53* mutation status was (Figures 7B and 7D). We also applied the web-based software program Kaplan-Meier Plotter (Nagy et al., 2018) on 16 TCGA cancer types using the four gene signature and found strong correlations between our prognosis calculations and the Kaplan-Meier Plotter-based calculations (Table S7B).

To facilitate clinical use of the mutant p53 four-gene signature, we normalized RNA expression values of each of the signature genes to that of a linked control gene, resulting in a four gene-normalization set. Each normalization gene was chosen by having (1) a median expression almost identical to that of its linked signature gene, (2) an extremely low average deviation from median expression across all tumors of that type, and (3) no correlation with expression of the signature gene across all tumors. Thus, ratios of the four signature genes to the four normalization genes would be dependent only on expression of the signature genes (Figures S7G and S7H). Increasing ratios of signature to control gene expression above 1.0 would represent tumors with predicted poorer prognosis, whereas ratios below 1.0 would predict better prognosis (Figure S7H). Testing of this normalization approach on all TCGA cancer types that were prognostically dependent on the four gene signature (Table S7B) showed almost identical overall survival results between the two methodological approaches (Table S7C).

DISCUSSION

TCGA-sponsored, large-scale analyses of 32 different cancer types using five high-throughput data platforms integrated with relevant clinical data has provided a rich dataset from which to derive important insights (Weinstein et al., 2013). Despite more than 90,000 papers that have been published on p53, previous studies have generally used tumor samples of one cancer type and perhaps one or two experimental methodologies to examine p53 effects. TCGA

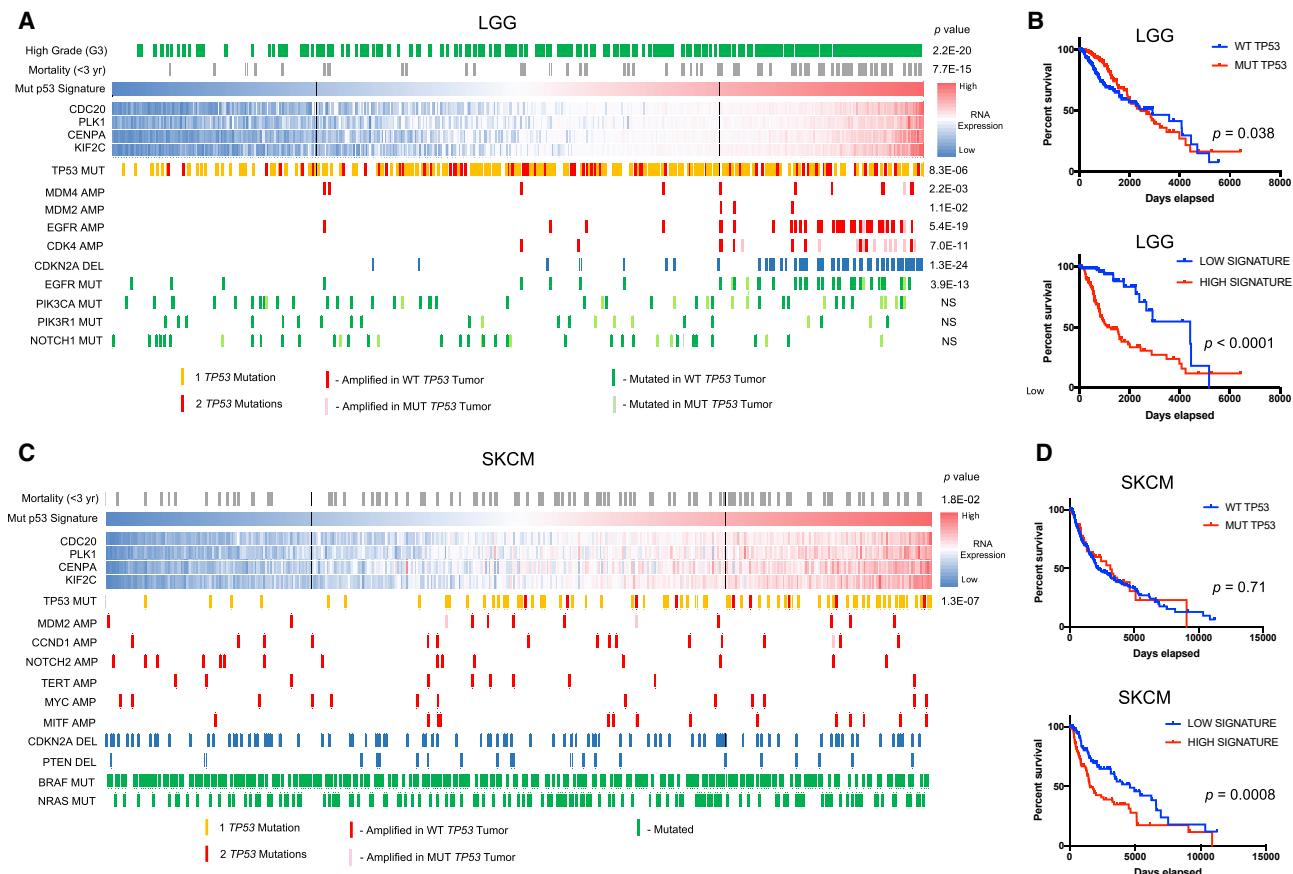


Figure 7. Development of a Prognostic Mutant TP53 Signature

(A) Lower-grade glioma (LGG) stratified by mutant p53 expression signature correlates with a number of clinical and molecular parameters. For all LGG, RNA expression of the four genes comprising the mutant p53 signature (*CDC20*, *PLK1*, *CENPA*, and *KIF2C*) were ranked from low to high expression and combined ("Mut p53 Signature"). The bottom and top signature-expression quartiles are demarcated with black vertical lines. Also indicated are tumor grade and mortality. Below the signature, *TP53* mutation status and copy number and mutation status of a number of cancer driver genes relevant to LGG development are shown. p values to the right indicate the significance of the difference of the low- and high-signature quartiles for each parameter.

(B) *p53* mutant signature status is more prognostically predictive than *TP53* mutation status for overall survival in LGG. Log rank analysis was performed on the LGG overall survival data based on *TP53* mutation status (top panel) or on mutant p53 signature status (bottom panel) in a top-versus-bottom quartile analysis. To the right of each survival graph in (B) and (D), p values indicate the significance of differences in survival of mutant versus wild-type *TP53* cancers (top graph) and low versus high p53 signature cancers (bottom graph).

(C) Skin cutaneous melanoma (SKCM) stratified by p53 expression signature shows that it correlates with *TP53* mutation and mortality. The heatmaps shown here are similar to those described for (A), although driver genes relevant to SKCM are shown.

(D) *p53* mutant signature status is more prognostically predictive than *TP53* mutation status for overall survival in SKCM. Overall survival in SKCM is compared based on *TP53* mutation status (top panel) or on mutant p53 signature status (bottom panel) as in (B).

See also Figure S7 and Table S7.

large-scale, integrated, multi-data platform approach facilitates extraction of statistically significant patterns that might be obscured by background noise associated with smaller, less-diverse datasets.

The integration of two data platforms, exome sequencing and DNA copy number, was particularly useful in the analysis of individual *TP53* alleles in both wild-type and mutant *TP53* tumors. First, we found that roughly 10% of tumors with *TP53* mutations had two distinct *TP53* mutations, and both alleles were affected in 80% of cases. Moreover, about two-thirds of tumors with a single *TP53* mutation exhibited loss of the wild-type *TP53* allele and a high fraction of the remaining third with diploid *TP53* copy

number exhibited copy-neutral loss of heterozygosity. This type of wild-type *TP53* loss has been previously reported (Jasek et al., 2010; Parikh et al., 2014; Saeki et al., 2011; Svobodova et al., 2016) and is likely the result of mitotic recombination or gene conversion events in the emerging tumor (Kumar et al., 2015; Stewart et al., 2012). Overall, by our assay methods, more than 91% of tumors with *TP53* mutations had structural losses of both *TP53* alleles, which was further corroborated by analyses of p53 RNA sequencing (RNA-seq) data showing that more than 92% of single-mutation tumors exhibited RNA variant-allele fractions close to 1.0. Although cell culture studies indicate the mutant p53 protein can behave in a dominant-negative fashion

to inactivate wild-type p53 protein (Muller and Vousden, 2013, 2014; Giacomelli et al., 2018; Soussi and Wiman, 2015), the data presented here argue that there is still a strong selection for inactivation of the wild-type *TP53* allele in tumors with a single *TP53* mutation. Thus, *TP53* usually behaves like a classic, recessive tumor suppressor in the requirement for inactivation of both alleles (Knudson, 1996).

Genomic instability is a central characteristic of most cancers (Negrini et al., 2010). *TP53*, since its designation as “guardian of the genome” by David Lane (1992) has been known to prevent this instability. Studies in cell culture, experimental animal models, and human cancers have shown that mutations in the *TP53* gene are associated with enhanced chromosomal instability largely because of a loss of the cell cycle checkpoint control (Donehower, 1997; Negrini et al., 2010; Smith and Fornace, 1995; Tainsky et al., 1995; Tomasini et al., 2008). However, few large-scale, systematic studies on genomic instability across multiple cancers have been published. Our examination of TCGA cancer types (those with sufficient numbers of tumors with *TP53* mutations) revealed that 19 of 23 examined cancer types had significantly enhanced copy number instability in the MUT *TP53* cohort relative to their wild-type *TP53* counterparts. This global copy-number instability was closely associated with increased amplification of known oncogenes (e.g., *CCND1*, *CCNE1*, *ERBB2*, and *MYC*) and deep deletion of known tumor suppressors (*RB1*, *PTEN*, and *WWOX*). An exception to that trend was the enhanced amplification of *MDM2*, *MDM4*, and *PPM1D* loci in wild-type *TP53* tumors. These three genes all encode negative regulators of p53 and, thus, might undergo selection in nascent wild-type *TP53* cancer cells (Lu et al., 2008; Oliner et al., 2016; Wasylissen and Lozano, 2016). Our genomic instability results are consistent with recent experimental data indicating that p53 is directly involved in suppression of aneuploidy by engaging a ploidy-sensing checkpoint that blocks proliferation of tetraploid and aneuploid cells (Dalton et al., 2010; Ganem et al., 2007; Hanel and Moll, 2012; Talos and Moll, 2010). Moreover, amplifications have been shown experimentally to be enhanced by mutant *TP53* in part because of defective double-strand break repair and absence of p53-mediated apoptosis in response to proliferation of cells with double-strand DNA breaks (Hanel and Moll, 2012; Lengauer et al., 1998; Livingstone et al., 1992).

Analysis of the MUT *TP53* cancers across the RNA, microRNA (miRNA) and protein expression data platforms consistently showed strong enhancement of pathways regulating cell-cycle progression. MUT *TP53* cancers showed enhanced expression of cell-cycle progression genes and S-phase-promoting E2F target genes. The E2F results are consistent with the findings that wild-type p53 may indirectly suppress E2F1/2 through *CDKN1A* (p21^{CIP1}) activation, which, in turn, results in suppression of RB1 phosphorylation and cell cycle inhibition (Polager and Ginsberg, 2009). p53 may also have suppressive effects on a number of E2F target genes, including *FOXM1*, one of the highest differentially upregulated proteins in the MUT *TP53* RPPA dataset (Barsovi and Prives, 2009).

Studies on *TP53* mutations as a prognostic marker have been historically mixed, and many variables may contribute to that (Robles and Harris, 2010). One problem in these clinical studies

is that there are mutationally independent mechanisms to inactivate the p53 signaling pathway. To circumvent that, we searched for downstream transcriptional signatures based on RNA expression data of four genes highly and significantly upregulated across almost all MUT *TP53* TCGA tumors relative to wild-type *TP53* tumors. In addition, these genes (*CDC20*, *CENPA*, *KIF2C*, and *PLK1*) were cell-cycle-promoting genes and established p53-repression targets. Eleven of 24 TCGA cancers showed significantly poorer survival with the high-p53 signature, and significance values were usually more robust than those observed in the previous *TP53* mutation prognostic tests. Moreover, no cancer types showed poorer survival with a low-p53 signature. Finally, we developed a normalization approach for each of 11 cancer types that would facilitate prognostic predictions on samples from individual patients entering the clinic. Thus, we believe this four-gene RNA-expression signature could serve as an improved prognostic marker and a better indicator of the absence of p53 functionality in some cancer types.

In conclusion, the large-scale multi-data platform approach pioneered by TCGA effort has provided an unparalleled opportunity to better understand structural mechanisms of p53 pathway inactivation and the resulting effect on the genetics and biology of many of the cancers examined. We believe that this article may facilitate the development of diagnostic and therapeutic tools based on a more robust knowledge of the p53 signaling pathway in cancer.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCE TABLE
- LEAD CONTACT AND MATERIALS AVAILABILITY
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
 - TP53 Mutation Analyses
 - Assessment of TP53 Allele Status through integration of TP53 mutation and copy number data
 - Analysis of p53 RNaseq Data to Determine Relative Expression of Mutant and Wild-type p53 in Tumors with TP53 Missense Mutations
 - TP53 Status and Global Copy Number Alterations
 - TP53 Status and RNA Expression-based Pathway Analysis
 - TP53 Status and Protein Expression-based Pathway Analysis
 - TP53 Status and miRNA Expression-based Pathway Analysis
 - Analyses of mutually exclusive and co-occurring genomic alterations
 - TP53 Status and Clinical Parameter Correlations
 - Development and Testing of a Mutant p53 RNA Expression Signature
- QUANTIFICATION AND STATISTICAL ANALYSIS
 - Assessment of TP53 allele status through integration of TP53 mutation and copy number
 - TP53 Status and Global Copy Number Alterations

- TP53 Status and RNA Expression-based Pathway Analysis
- TP53 Status and Protein Expression-based Pathway Analysis
- Analyses of mutually exclusive and co-occurring genomic alterations
- TP53 Status and Clinical Parameter Correlations
- Development and Testing of a Mutant p53 RNA Expression Signature

● DATA AND CODE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at <https://doi.org/10.1016/j.celrep.2019.07.001>.

ACKNOWLEDGMENTS

We thank Tajhal Dayaram for editorial assistance on the manuscript. We are also grateful for funding from multiple sources: NIH/NCI U24 CA210950, U24 CA210949, U24 CA199461, UM1HG008898, and P30 CA016672; and DoD/CDMRP W81XWH-16-1-0237. Personal grants for T.S. to support the development of the UMD_TP53 database were received from Radiumhemmets Forskningsfondet.

AUTHOR CONTRIBUTIONS

The Cancer Genome Atlas Research (TCGA) Network contributed collectively to the acquisition of patient samples and the generation of data underlying this study. Data generation and analyses were performed by Genome Sequencing Centers, Genome Characterization Centers, and Genome Data Analysis Centers. The efforts reported here were also assisted by TCGA Pan-Cancer Analysis Project. The authors in this p53 pathway-specific analysis group included L.A.D., who analyzed much of the data presented and wrote the manuscript; D.A.W., who led and organized the study, assisted in the analysis of *TP53* sequencing data, and helped edit the manuscript; T.S., who performed analysis of the *TP53* mutations, contributed to the generation of Figures 1, 2, and S1–S3, and helped in the editing of the paper; A.K., who provided data analyses for Figures 4 and 6, and edited the statistical analyses throughout the paper; Y.L., who contributed the reverse-phase protein array (RPPA) analysis that is the basis for Figures 5 and S7; R.A., who performed analyses on p53 pathway activity and PanCanAtlas expression data and provided computational expertise for some of the analyses; A.S., who performed TP53 pathway and mRNA expression computational analyses; J.N.W., who led the computational scientists assisting this effort and provided editing assistance on the paper; M.C., who extracted and analyzed individual *TP53* sequencing reads from the Genomic Data Commons (GDC) portal and who assisted in the analyses shown in Figure 2; X.L. and O.B., who performed the mutual exclusivity analyses in Figure 6; and T.-K.H. and O.L., who assisted on the role of *TP53* mutations in patient survival shown in Figure 7.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: November 29, 2018

Revised: May 9, 2019

Accepted: June 27, 2019

Published: July 30, 2019; corrected online: September 4, 2019

REFERENCES

- Babur, Ö., Dogrusoz, U., Demir, E., and Sander, C. (2010). ChiBE: interactive visualization and manipulation of BioPAX pathway models. *Bioinformatics* 26, 429–431.
- Babur, Ö., Aksoy, B.A., Rodchenkov, I., Sümer, S.O., Sander, C., and Demir, E. (2014a). Pattern search in BioPAX models. *Bioinformatics* 30, 139–140.
- Babur, Ö., Dogrusoz, U., Çakır, M., Aksoy, B.A., Schultz, N., Sander, C., and Demir, E. (2014b). Integrating biological pathways and genomic profiles with ChiBE 2. *BMC Genomics* 15, 642.
- Babur, Ö., Gönen, M., Aksoy, B.A., Schultz, N., Ciriello, G., Sander, C., and Demir, E. (2015). Systematic identification of cancer driving signaling pathways based on mutual exclusivity of genomic alterations. *Genome Biol.* 16, 45.
- Baker, S.J., Fearon, E.R., Nigro, J.M., Hamilton, S.R., Preisinger, A.C., Jessup, J.M., vanTuinen, P., Ledbetter, D.H., Barker, D.F., Nakamura, Y., et al. (1989). Chromosome 17 deletions and p53 gene mutations in colorectal carcinomas. *Science* 244, 217–221.
- Baker, S.J., Preisinger, A.C., Jessup, J.M., Paraskeva, C., Markowitz, S., Willson, J.K., Hamilton, S., and Vogelstein, B. (1990). p53 gene mutations occur in combination with 17p allelic deletions as late events in colorectal tumorigenesis. *Cancer Res.* 50, 7717–7722.
- Barsotti, A.M., and Prives, C. (2009). Pro-proliferative FoxM1 is a target of p53-mediated repression. *Oncogene* 28, 4295–4305.
- Bouaoun, L., Sonkin, D., Ardin, M., Hollstein, M., Byrnes, G., Zavadil, J., and Olivier, M. (2016). TP53 variations in human cancers: new lessons from the IARC TP53 database and genomics data. *Hum. Mutat.* 37, 865–876.
- Caron de Fromentel, C., and Soussi, T. (1992). TP53 tumor suppressor gene: a model for investigating human mutagenesis. *Genes Chromosomes Cancer* 4, 1–15.
- Cerami, E.G., Gross, B.E., Demir, E., Rodchenkov, I., Babur, Ö., Anwar, N., Schultz, N., Bader, G.D., and Sander, C. (2011). Pathway Commons, a web resource for biological pathway data. *Nucleic Acids Res.* 39, D685–D690.
- Cerami, E., Gao, J., Dogrusoz, U., Gross, B.E., Sumer, S.O., Aksoy, B.A., Jacobsen, A., Byrne, C.J., Heuer, M.L., Larsson, E., et al. (2012). The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* 2, 401–404.
- Crook, T., Wrede, D., Tidy, J.A., Mason, W.P., Evans, D.J., and Vousden, K.H. (1992). Clonal p53 mutation in primary cervical cancer: association with human-papillomavirus-negative tumours. *Lancet* 339, 1070–1073.
- Dai, C., and Gu, W. (2010). p53 post-translational modification: deregulated in tumorigenesis. *Trends Mol. Med.* 16, 528–536.
- Dalton, W.B., Yu, B., and Yang, V.W. (2010). p53 suppresses structural chromosome instability after mitotic arrest in human cells. *Oncogene* 29, 1929–1940.
- Donehower, L.A. (1997). Genetic instability in animal tumorigenesis models. *Cancer Surv.* 29, 329–352.
- Engeland, K. (2018). Cell cycle arrest through indirect transcriptional repression by p53: I have a DREAM. *Cell Death Differ.* 25, 114–132.
- Fischer, M. (2017). Census and evaluation of p53 target genes. *Oncogene* 36, 3943–3956.
- Frum, R.A., and Grossman, S.R. (2014). Mechanisms of mutant p53 stabilization in cancer. *Subcell. Biochem.* 85, 187–197.
- Ganem, N.J., Storchova, Z., and Pellman, D. (2007). Tetraploidy, aneuploidy and cancer. *Curr. Opin. Genet. Dev.* 17, 157–162.
- Gao, J., Aksoy, B.A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S.O., Sun, Y., Jacobsen, A., Sinha, R., Larsson, E., et al. (2013). Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal.* 6, pl1.
- Giacomelli, A.O., Yang, X., Lintner, R.E., McFarland, J.M., Duby, M., Kim, J., Howard, T.P., Takeda, D.Y., Ly, S.H., Kim, E., et al. (2018). Mutational processes shape the landscape of TP53 mutations in human cancer. *Nat. Genet.* 50, 1381–1387.
- Hanel, W., and Moll, U.M. (2012). Links between mutant p53 and genomic instability. *J. Cell. Biochem.* 113, 433–439.
- Hermeking, H. (2012). MicroRNAs in the p53 network: micromanagement of tumour suppression. *Nat. Rev. Cancer* 12, 613–626.

- Hollstein, M., Sidransky, D., Vogelstein, B., and Harris, C.C. (1991). p53 mutations in human cancers. *Science* 253, 49–53.
- Jasek, M., Gondek, L.P., Bejanyan, N., Tiu, R., Huh, J., Theil, K.S., O'Keefe, C., McDevitt, M.A., and Maciejewski, J.P. (2010). TP53 mutations in myeloid malignancies are either homozygous or hemizygous due to copy number-neutral loss of heterozygosity or deletion of 17p. *Leukemia* 24, 216–219.
- Kannan, K., Inagaki, A., Silber, J., Gorovets, D., Zhang, J., Kastenhuber, E.R., Heguy, A., Petrini, J.H., Chan, T.A., and Huse, J.T. (2012). Whole-exome sequencing identifies ATRX mutation as a key molecular determinant in lower-grade glioma. *Oncotarget* 3, 1194–1203.
- Kastenhuber, E.R., and Lowe, S.W. (2017). Putting p53 in Context. *Cell* 170, 1062–1078.
- Kato, S., Han, S.Y., Liu, W., Otsuka, K., Shibata, H., Kanamaru, R., and 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. USA* 100, 8424–8429.
- Knudson, A.G., Jr. (1989). Nakahara memorial lecture. Hereditary cancer, oncogenes, and anti-oncogenes. *Int. Symp. Princess Takamatsu Cancer Res. Fund* 20, 15–29.
- Knudson, A.G. (1996). Hereditary cancer: two hits revisited. *J. Cancer Res. Clin. Oncol.* 122, 135–140.
- Kotler, E., Shani, O., Goldfeld, G., Lotan-Pompan, M., Tarcic, O., Gershoni, A., Hopf, T.A., Marks, D.S., Oren, M., and Segal, E. (2018). A systematic p53 mutation library links differential functional impact to cancer mutation pattern and evolutionary conservation. *Mol. Cell* 71, 178–190 e178.
- Kumar, Y., Yang, J., Hu, T., Chen, L., Xu, Z., Xu, L., Hu, X.X., Tang, G., Wang, J.M., Li, Y., et al. (2015). Massive interstitial copy-neutral loss-of-heterozygosity as evidence for cancer being a disease of the DNA-damage response. *BMC Med. Genomics* 8, 42.
- Lane, D.P. (1992). Cancer. p53, guardian of the genome. *Nature* 358, 15–16.
- Lek, M., Karczewski, K.J., Minikel, E.V., Samocha, K.E., Banks, E., Fennell, T., O'Donnell-Luria, A.H., Ware, J.S., Hill, A.J., Cummings, B.B., et al.; Exome Aggregation Consortium (2016). Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 536, 285–291.
- Lengauer, C., Kinzler, K.W., and Vogelstein, B. (1998). Genetic instabilities in human cancers. *Nature* 396, 643–649.
- Leroy, B., Anderson, M., and Soussi, T. (2014). TP53 mutations in human cancer: database reassessment and prospects for the next decade. *Hum. Mutat.* 35, 672–688.
- Liu, X.Y., Gerges, N., Korshunov, A., Sabha, N., Khuong-Quang, D.A., Fontebasso, A.M., Fleming, A., Hadjadj, D., Schwartzentruber, J., Majewski, J., et al. (2012). Frequent ATRX mutations and loss of expression in adult diffuse astrocytic tumors carrying IDH1/IDH2 and TP53 mutations. *Acta Neuropathol.* 124, 615–625.
- Liu, Y., Zhang, X., Han, C., Wan, G., Huang, X., Ivan, C., Jiang, D., Rodriguez-Aguayo, C., Lopez-Berestein, G., Rao, P.H., et al. (2015). TP53 loss creates therapeutic vulnerability in colorectal cancer. *Nature* 520, 697–701.
- Livingstone, L.R., White, A., Sprouse, J., Livanos, E., Jacks, T., and Tlsty, T.D. (1992). Altered cell cycle arrest and gene amplification potential accompany loss of wild-type p53. *Cell* 70, 923–935.
- Lu, X., Nguyen, T.A., Moon, S.H., Darlington, Y., Sommer, M., and Donehower, L.A. (2008). The type 2C phosphatase Wip1: an oncogenic regulator of tumor suppressor and DNA damage response pathways. *Cancer Metastasis Rev.* 27, 123–135.
- Lujambio, A., and Lowe, S.W. (2012). The microcosmos of cancer. *Nature* 482, 347–355.
- Matheu, A., Maraver, A., and Serrano, M. (2008). The Arf/p53 pathway in cancer and aging. *Cancer Res.* 68, 6031–6034.
- McConechy, M.K., Ding, J., Cheang, M.C., Wiegand, K., Senz, J., Tone, A., Yang, W., Prentice, L., Tse, K., Zeng, T., et al. (2012). Use of mutation profiles to refine the classification of endometrial carcinomas. *J. Pathol.* 228, 20–30.
- Mello, S.S., and Attardi, L.D. (2018). Deciphering p53 signaling in tumor suppression. *Curr. Opin. Cell Biol.* 51, 65–72.
- Mootha, V.K., Lindgren, C.M., Eriksson, K.F., Subramanian, A., Sihag, S., Lehar, J., Puigserver, P., Carlsson, E., Ridderstråle, M., Laurila, E., et al. (2003). PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat. Genet.* 34, 267–273.
- Muller, P.A., and Vousden, K.H. (2013). p53 mutations in cancer. *Nat. Cell Biol.* 15, 2–8.
- Muller, P.A., and Vousden, K.H. (2014). Mutant p53 in cancer: new functions and therapeutic opportunities. *Cancer Cell* 25, 304–317.
- Mulligan, L.M., Matlashewski, G.J., Sorbara, H.J., and Cavenee, W.K. (1990). Mechanisms of p53 loss in human sarcomas. *Proc. Natl. Acad. Sci. USA* 87, 5863–5867.
- Nagy, Á., Lánczky, A., Menyhárt, O., and Györffy, B. (2018). Validation of miRNA prognostic power in hepatocellular carcinoma using expression data of independent datasets. *Sci. Rep.* 8, 9227.
- Negrini, S., Gorgoulis, V.G., and Halazonetis, T.D. (2010). Genomic instability—an evolving hallmark of cancer. *Nat. Rev. Mol. Cell Biol.* 11, 220–228.
- Nguyen, T.A., Menendez, D., Resnick, M.A., and Anderson, C.W. (2014). Mutant TP53 posttranslational modifications: challenges and opportunities. *Hum. Mutat.* 35, 738–755.
- Nigro, J.M., Baker, S.J., Preisinger, A.C., Jessup, J.M., Hostetter, R., Cleary, K., Bigner, S.H., Davidson, N., Baylin, S., Devilee, P., et al. (1989). Mutations in the p53 gene occur in diverse human tumour types. *Nature* 342, 705–708.
- Oliner, J.D., Saiki, A.Y., and Caenepeel, S. (2016). The role of MDM2 amplification and overexpression in tumorigenesis. *Cold Spring Harb. Perspect. Med.* 6, a026336.
- Parikh, N., Hilsenbeck, S., Creighton, C.J., Dayaram, T., Shuck, R., Shinbrot, E., Xi, L., Gibbs, R.A., Wheeler, D.A., and Donehower, L.A. (2014). Effects of TP53 mutational status on gene expression patterns across 10 human cancer types. *J. Pathol.* 232, 522–533.
- Polager, S., and Ginsberg, D. (2009). p53 and E2f: partners in life and death. *Nat. Rev. Cancer* 9, 738–748.
- Robles, A.I., and Harris, C.C. (2010). Clinical outcomes and correlates of TP53 mutations and cancer. *Cold Spring Harb. Perspect. Biol.* 2, a001016.
- Saeki, H., Kitao, H., Yoshinaga, K., Nakanoko, T., Kubo, N., Kakeji, Y., Morita, M., and Maehara, Y. (2011). Copy-neutral loss of heterozygosity at the p53 locus in carcinogenesis of esophageal squamous cell carcinomas associated with p53 mutations. *Clin. Cancer Res.* 17, 1731–1740.
- Smith, M.L., and Fornace, A.J., Jr. (1995). Genomic instability and the role of p53 mutations in cancer cells. *Curr. Opin. Oncol.* 7, 69–75.
- Soussi, T., and Kroemer, G. (2018). MDM2-TP53 crossregulation: an underestimated target to promote loss of TP53 function and cell survival. *Trends Cancer* 4, 602–605.
- Soussi, T., and Wiman, K.G. (2015). TP53: an oncogene in disguise. *Cell Death Differ.* 22, 1239–1249.
- Soussi, T., Kato, S., Levy, P.P., and Ishioka, C. (2005). Reassessment of the TP53 mutation database in human disease by data mining with a library of TP53 missense mutations. *Hum. Mutat.* 25, 6–17.
- Soussi, T., Leroy, B., Devir, M., and Rosenberg, S. (2019). High prevalence of cancer-associated TP53 variants in the gnomAD database: a word of caution concerning the use of variant filtering. *Hum. Mutat.* 40, 516–524.
- Stewart, D.R., Pemov, A., Van Loo, P., Beert, E., Brems, H., Sciot, R., Claes, K., Pak, E., Dutra, A., Lee, C.C., and Legius, E. (2012). Mitotic recombination of chromosome arm 17q as a cause of loss of heterozygosity of NF1 in neurofibromatosis type 1-associated glomus tumors. *Genes Chromosomes Cancer* 51, 429–437.
- Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., and Mesirov, J.P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* 102, 15545–15550.

- Svobodova, K., Zemanova, Z., Lhotska, H., Novakova, M., Podskalska, L., Belickova, M., Brezinova, J., Sarova, I., Izakova, S., Lizcova, L., et al. (2016). Copy number neutral loss of heterozygosity at 17p and homozygous mutations of TP53 are associated with complex chromosomal aberrations in patients newly diagnosed with myelodysplastic syndromes. *Leuk. Res.* **42**, 7–12.
- Tainsky, M.A., Bischoff, F.Z., and Strong, L.C. (1995). Genomic instability due to germline p53 mutations drives preneoplastic progression toward cancer in human cells. *Cancer Metastasis Rev.* **14**, 43–48.
- Takahashi, T., Nau, M.M., Chiba, I., Birrer, M.J., Rosenberg, R.K., Vinocour, M., Levitt, M., Pass, H., Gazdar, A.F., and Minna, J.D. (1989). p53: a frequent target for genetic abnormalities in lung cancer. *Science* **246**, 491–494.
- Talos, F., and Moll, U.M. (2010). Role of the p53 family in stabilizing the genome and preventing polyploidization. *Adv. Exp. Med. Biol.* **676**, 73–91.
- Tapapore, P., and Fukasawa, K. (2002). Loss of p53 and centrosome hyperamplification. *Oncogene* **21**, 6234–6240.
- Tomasini, R., Mak, T.W., and Melino, G. (2008). The impact of p53 and p73 on aneuploidy and cancer. *Trends Cell Biol.* **18**, 244–252.
- Vousden, K.H., and Prives, C. (2009). Blinded by the Light: The Growing Complexity of p53. *Cell* **137**, 413–431.
- Wasylissen, A.R., and Lozano, G. (2016). Attenuating the p53 pathway in human cancers: many means to the same end. *Cold Spring Harb. Perspect. Med.* **6**, a026211.
- Weinstein, J.N., Collisson, E.A., Mills, G.B., Shaw, K.R., Ozenberger, B.A., Ellrott, K., Shmulevich, I., Sander, C., and Stuart, J.M.; Cancer Genome Atlas Research Network (2013). The Cancer Genome Atlas pan-cancer analysis project. *Nat. Genet.* **45**, 1113–1120.
- Wildeman, M., van Ophuijen, E., den Dunnen, J.T., and Taschner, P.E. (2008). Improving sequence variant descriptions in mutation databases and literature using the Mutalyzer sequence variation nomenclature checker. *Hum. Mutat.* **29**, 6–13.
- Wilks, C., Cline, M.S., Weiler, E., Diekhans, M., Craft, B., Martin, C., Murphy, D., Pierce, H., Black, J., Nelson, D., et al. (2014). The Cancer Genomics Hub (CGHub): overcoming cancer through the power of torrential data. *Database (Oxford)*. **2014**, bau093.
- Zupnick, A., and Prives, C. (2006). Mutational analysis of the p53 core domain L1 loop. *J. Biol. Chem.* **281**, 20464–20473.

STAR★METHODS**KEY RESOURCE TABLE**

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited Data		
cBioPortal for Cancer Genomics	Gao et al., 2013; Cerami et al., 2012	https://www.cbioportal.org
TCGA Pan-Cancer Atlas Data Portal	TCGA, Weinstein et al., 2013	https://www.synapse.org
CGHub BAM files	Wilks et al., 2014	https://gdc.cancer.gov/
UMD_TP53 mutation database	Leroy et al., 2014	http://p53.fr/tp53-database
The Genome Aggregation Database (gnomAD)	Soussi et al., 2019	https://gnomad.broadinstitute.org
Software and Algorithms		
Gene Set Enrichment Analysis (GSEA)	Subramanian et al., 2005; Mootha et al., 2003	http://software.broadinstitute.org/gsea/index.jsp
Mutalyzer	Wildeman et al., 2008	https://mutalyzer.nl/
Mutex	Babur et al., 2015	Babur et al., 2015
Kaplan-Meier plotter	Nagy et al., 2018	http://kmplot.com/analysis/index.php?p=service

LEAD CONTACT AND MATERIALS AVAILABILITY

Further information and requests for data analysis results should be directed to the lead contact and corresponding author, Lawrence A. Donehower (larry@bcm.edu). The analyses in this report were solely computational and no physical experiments were performed, so no materials, reagents, or resources other than data analyses and data tables are available for sharing.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

All analyzes carried out in this paper are derived from The Cancer Genome Atlas Research (TCGA) Network effort. TCGA collected tumor and non-tumor biospecimens from 10,225 human patients. All data files have been deposited in the TCGA Pan-Cancer Atlas Data portal in the [Synapse.org](https://synapse.org) Website. All data used in the analysis reported here is from Data Freeze 1.3.1. Exome sequencing data (syn4924181) from this portal was used to identify TP53 and other relevant gene mutations. For individual TP53 DNAseq reads and RNAseq reads we accessed CGHub BAM files (<https://gdc.cancer.gov/>). RNA expression data was obtained from RNAseq files (syn4557678.9 and syn4874822.6). Copy number data was downloaded from the GISTIC copy number files (Syn5049514.1). MicroRNA expression data was obtained from two files (syn45577894.9 and syn4557787.8). RPPA data (syn4557674.9) and clinical data (syn4983466.1) was also downloaded. All publicly available TCGA tumor data complies with U.S. law protecting patient confidentiality and other ethical standards.

For comparison purposes, we also utilized the UMD_TP53 mutation database. Version 2017_R1 was used for all studies. This release includes 80,406 TP53 mutations identified in tumors, cell lines (somatic mutations) or in patients with hereditary cancer (germline mutations) (database freeze Oct 2017). These mutations can be grouped into 6,874 different TP53 variants and are from studies using conventional Sanger sequencing, NGS or both. For comparison with TP53 variants from the TCGA dataset, a specific dataset (Sanger_dataset) was created by selecting only studies using Sanger sequencing. The Sanger dataset includes 37,299 TP53 mutations (4,299 variants). The database also includes functional data for most missense mutations. Residual transactivating activity for WAF (CDKN1A), MDM2, BAX, 14-3-3-sigma, AIP, GADD45A, NOXA and p53R2 promoters was originally published by Kato et al. and later used to assess TP53 variant deleteriousness ([Soussi et al., 2005](#)). The residual transcriptional activity of mutant p53 was assayed in yeast and always compared to wild-type p53 for the same promoter. Two large-scale analysis of the functional activities of TP53 variants in mammalian cells have been released recently and are now included in the UMD TP53 database ([Giacomelli et al., 2018; Kotler et al., 2018](#)). [Kotler et al. \(2018\)](#) have defined the growth arrest potential of TP53 variants in H1299, a TP53 null cell line whereas Giacomelli et al. have analyzed the growth suppressive effect of TP53 in three different cellular settings including cells either WT TP53 or TP53KO. Data from both studies have been included in the UMD_TP53 database. For the purpose of all analysis, all datasets have been normalized from 0 to 1 with the lowest value corresponding to the most detrimental activity for TP53. Only missense variants issued from single nucleotide variations have been used for all comparison.

The Genome Aggregation Database (gnomAD) is a repository of SNP data from 125,748 exome sequences and 15,708 whole-genome sequences from unrelated individuals sequenced as part of various population genetic studies. It contains predominately

frequent and rare non-pathogenic SNP from the normal population although a few pathogenic germline variants have been identified (Lek et al., 2016; Soussi et al., 2019).

METHOD DETAILS

TP53 Mutation Analyses

All *TP53* mutations were downloaded from the TCGA PanCanAtlas portal in Synapse (synapse.org). Exome sequencing data from Freeze 1.3.1 (syn4924181) was utilized for these analyses. The following types of *TP53* mutations were scored as bona fide *TP53* mutations for analysis purposes: non-synonymous missense mutations, indels (in frame insertions and deletions and out of frame insertions and deletions, nonsense mutations, and splice-site mutations. Synonymous missense mutations were not scored as bona fide mutations, except for NM_000546.5:c.375G > A (NP_000537.3:p.(Thr125 =)) and NM_000546.5:c.672G > A (NP_000537.3:p.(Glu224 =)) two synonymous variants which are known to be a frequent *TP53* splice site mutation. T125T is known to be a frequent *TP53* splice site mutation. Other mutations not considered bona fide mutations were non-exonic mutations in the 5' UTR, the 3' UTR, and introns.

In using the UMD_TP53 database for comparison purposes, minimal genomic information such as genomic coordinates and genetic events were extracted from each dataset to define a correct annotation using HGVS recommendations. In a second step the variant annotation were validated using the Name Checker tool developed by Mutalyzer (<https://mutalyzer.nl/>) (Wildeman et al., 2008). Mutalyzer handles all types of variations that can target the *TP53* gene, such as substitutions, insertions, duplications, deletions, or more complex insertion or deletion. The current version of Mutalyzer (Mutalyzer 2.0.26) uses the stable NCBI sequence NG_017013.2 as a reference for TP53.

Assessment of TP53 Allele Status through integration of TP53 mutation and copy number data

For six cancer types with high numbers of *TP53* mutations (UCEC, LGG, OV, LUAD, LUSC, HNSC) we downloaded all *TP53* mutations for each *TP53* locus and stratified each tumor into one of two categories: (a) tumors with two or more distinct *TP53* mutations, or (b) tumors with one *TP53* mutation (tumors with no *TP53* mutations were not further analyzed). *TP53* copy number data for each tumor with *TP53* mutations was then downloaded. Copy number values for the *TP53* alleles downloaded from the TCGA PanCanAtlas portal in each tumor were categorized by GISTIC scores (0 = diploid, -1 = haploid, -2 = nullizygous) obtained from the Memorial Sloan Kettering Cancer Center cBioPortal for Cancer Genomics (<https://www.cbioperl.org/>) and each GISTIC score aligned with its *TP53* mutation status. As indicated in Figure 2C, virtually all tumors with two or more *TP53* mutations coincided with diploid copy number values. However, two thirds of tumors with one *TP53* mutation showed copy number loss and had GISTIC scores of -1, while about one third displayed diploid 0 copy number scores. To determine the variant allele fraction (VAF) in each of these one *TP53* mutation tumors, we obtained the total number of wild-type *TP53* allele reads and mutant *TP53* reads in each tumor from the exome sequencing data. We then adjusted the number of wild-type *TP53* allele reads by multiplying this number by the purity fraction of that tumor. The VAF was then determined by dividing the total number of mutant alleles by the purity-adjusted total wild-type allele number. Those diploid (0) one mutation *TP53* tumors with VAF less than 0.75 were considered to retain a wild-type *TP53* allele (no LOH), while those with VAF greater than 0.75 were considered to be copy neutral LOH tumors (CN LOH). All haploid (-1) one mutation *TP53* tumors were considered to have lost the second wild-type *TP53* allele (LOH). VAF for each *TP53* category (2 *TP53* mutations, 1 *TP53* mutation – diploid, and 1 *TP53* mutation – haploid) were compared by box and whisker plots (GraphPad Prism 7) for all six cancer types as shown in Figures 2E, 2F, S3B, and S3C. Statistical differences in VAF for tumors with 2 *TP53* mutations relative to those with one *TP53* mutation were determined by two sided t test.

Analysis of p53 RNaseq Data to Determine Relative Expression of Mutant and Wild-type p53 in Tumors with TP53 Missense Mutations

To determine relative mutant and wild-type p53 RNA expression in individual TCGA tumors, BAM files containing individual p53 reads were downloaded from the GDC portal. Individual mapped reads were obtained using the SAMtools mpileup utility. For each tumor with a documented non-synonymous *TP53* missense mutation, total numbers of p53 wild-type and mutant RNA sequence reads were quantified. For each tumor, tumor purity fractions were determined by downloading tumor purity data from the TCGA PanCanAtlas Data portal in the [Synapse.org](https://synapse.org) Website. Purity adjusted p53 variant allele fractions (VAFs) were then determined for each tumor. Tumors that retain one expressed copy of WT *TP53* and one equally expressed copy of mutant *TP53* would be expected to display a VAF value of 0.5. Tumors that lose their wild-type *TP53* allele through copy number loss or copy neutral loss should have values approximating 1.0 (or greater if tumors express p53 RNA at higher levels than adjacent normal cells in the tumor sample).

TP53 Status and Global Copy Number Alterations

GISTIC copy number data for 25,129 individual genetic loci in each of 10,225 tumors was downloaded from the Synapse TCGA PanCanAtlas data portal. For each tumor, the copy number data was stratified by cancer type and by *TP53* mutation status (either wild-type or mutant for *TP53*). Then, every locus was categorized by the following GISTIC copy number filters: (a) CN > +2.0 was considered indicative of amplification; (b) CN > +1.0 was considered copy number gain; (c) CN < +1.0 and > -0.5 was considered diploidy/near diploidy; (d) CN < -0.5 was considered copy number loss; and (e) CN < -1.0 was considered deep

deletion. For Figure 3A the fraction of those tumors at each locus that were in the amplification ($CN > 2$) and deep deletion ($CN < -1$) categories was determined. The fraction of copy number gains at each locus for each *TP53* category were then graphed across the entire genome (Figure 3A). The fractions of all loci in all wild-type and mutant *TP53* tumors with a GISTIC copy number less than -1 or greater than $+2$ are shown in Figure 3B. Wild-type and mutant *TP53* tumor differences in amplification/deletion frequencies were statistically determined to be significant by chi-square test.

TP53 Status and RNA Expression-based Pathway Analysis

Level 3 normalized global RNA expression data files for each tumor were downloaded from the TCGA PanCancerAtlas SYNPASE portal. For each cancer type, tumors were stratified by *TP53* mutation status and the RNA expression of each gene averaged for the wild-type and mutant *TP53* tumor cohorts. Then, for each gene 2-sided t tests comparing RNA expression of wild-type versus mutant *TP53* tumors were used to determine whether each gene was differentially expressed to a significant degree. Several cancer types had less than five patients with *TP53* mutations and these were excluded from analysis. The 500 most differentially upregulated genes (determined by lowest t test p values) in wild-type and mutant *TP53* tumors of each cancer type were then ranked (Table S3). Genes highly upregulated in wild-type *TP53* tumors were enriched for known p53 target genes (Figure 4A; Tables S2 and S3). These were identified from a manually curated list of known p53 target genes from the literature (see Table S3). This p53 target gene list contained genes identified from individual gene experimental studies as well as genes from global expression genomic studies shown to be consistently upregulated in multiple datasets (Fischer, 2017). Pathway analysis was performed on both the 500 most upregulated genes in wild-type and mutant *TP53* tumors of each cancer type by Gene Set Enrichment Analysis (GSEA) (Mootha et al., 2003; Subramanian et al., 2005). Uploading of the differentially regulated 500 gene sets into the “canonical pathways” search function of GSEA resulted in ranked lists of the most significantly enriched pathways for each cancer type. The number of cancer types with wild-type *TP53* showing enrichment for a particular pathway are shown in Figure 4B and Table S4A. In the case of mutant *TP53* tumors, the 500 most frequently observed upregulated genes across all mutant *TP53* cancers were used to search the GSEA “canonical pathways” function (Figure 4D; Table S4B).

TP53 Status and Protein Expression-based Pathway Analysis

RPPA data was downloaded from the TCGA PanCancerAtlas SYNPASE portal (syn4557674.9). Similar to methods used for RNA expression data, RPPA values for each protein was stratified according to the *TP53* mutation status of the corresponding patient tumor. Mean protein expression levels were determined for proteins in the wild-type and mutant *TP53* tumor cohorts (in cancer types with sufficient numbers of *TP53* mutations) and significant upregulation of expression in wild-type and mutant *TP53* cancers were statistically analyzed by two sided t tests. Wild-type and mutant *TP53* tumors were compared in individual cancer types and all cancer types combined. P values for statistical differences in proteins upregulated in the combined mutant *TP53* cancer types are shown in Figure 5D and Table S6B. P values for statistical differences in proteins upregulated in wild-type *TP53* cancer types are shown in Table S6A.

TP53 Status and miRNA Expression-based Pathway Analysis

MicroRNA data was downloaded from the TCGA PanCancerAtlas SYNPASE portal (syn4557787.8). Similar to methods used for RNA expression data, expression values for each miRNA was stratified according to the *TP53* mutation status of the corresponding patient tumor. Mean expression levels were determined for miRNAs in the wild-type and mutant *TP53* tumor cohorts (in cancer types with sufficient numbers of *TP53* mutations) and significant upregulation of expression in wild-type and mutant *TP53* cancers were statistically analyzed by two sided t tests in each cancer type. Those miRNAs significantly upregulated in wild-type or mutant *TP53* tumors were noted (Table S5; Figures 5A and 5B). The most consistently upregulated miRNAs in wild-type and mutant *TP53* tumors are indicated in Figures 5A and 5B, respectively. For each of the top 20 most consistently upregulated genes in wild-type *TP53* cancers, literature searches were used to identify those that have been shown to be direct transcriptional targets of p53 and exhibit tumor suppressor functions in experimental contexts (Figure 5C; Table S5A). For each of the top 20 most consistently upregulated genes in mutant *TP53* cancers, literature searches were used to identify those that have been shown to exhibit oncogenic or cell growth promoting functions in experimental contexts (Figure 5C; Table S5B).

Analyses of mutually exclusive and co-occurring genomic alterations

To understand if *TP53* genomic alterations exhibit mutual exclusivity with other gene alterations in cancer patients, we used the Mutex algorithm (Babur et al., 2015) on 32 TCGA studies iteratively. For each study, we compiled an alteration matrix from detected gene mutations and copy number alterations that are also confirmed by altered gene expression, as described in the original Mutex manuscript. Mutex finds groups of genes with less overlap in gene alterations than expected by random, and provides a score for the group based on the worst fitting gene, a significant score making sure every member of the group contributes significantly to the pattern. We did not restrict the search space using pathways, used 0.01 as Mutex score cutoff, and limited maximum group size to 5. Mutex identified 46 distinct groups in 10 cancer studies containing *TP53* and other 43 genes. We used ChiBE (Babur et al., 2014b; Babur et al., 2010) to retrieve known pathway relations in Pathway Commons (Cerami et al., 2011) between the genes in the result groups and filtered-in the relations of *TP53*. For the query, we selected relation types “controls-state-change-of,” “controls-expression-of,” “in-complex-with” and “interacts-with” and merged the last two types (Figure 6). These binary interactions were derived from detailed processes in Pathway Commons resources using a pattern detection algorithm (Babur et al., 2014a).

To identify co-occurring genomic alterations with *TP53* mutation, we utilized the cBioPortal for Cancer Genomics “Enrichments” function (<https://www.cbioportal.org/>) (Cerami et al., 2012; Gao et al., 2013). By entering “*TP53*” and “Mutations” for each TCGA PanCanAtlas tumor type in the “Query” box, and then entering “Enrichments,” statistically prioritized lists of altered genes significantly co-occurring with *TP53* mutation could be obtained, using both the “Mutations” and “Copy Number” subfunctions. For those cancer types exhibiting highly significant co-occurring gene alterations, these were entered along with *TP53* into the “Oncoprint” function to obtain the oncoprints shown in [Figure S6](#). Statistical significance of each co-occurring alteration was accompanied by q values derived from the Benjamini-Hochberg procedure.

TP53 Status and Clinical Parameter Correlations

Clinical data (overall survival, age at first diagnosis) were downloaded from the patient data section of the TCGA PanCancerAtlas SYNAPSE portal (syn4983466.1). Overall survival data of all cancer types (with sufficient numbers of *TP53* mutations) was combined and stratified into three groups: (a) dead within one year post diagnosis; (b) still alive at four or more years post-diagnosis; and (c) dead after one year post-diagnosis or still alive less than four years post-diagnosis. These three groups were then sub-stratified by type of *TP53* mutation ([Figure S7A](#)). The significance of the difference in *TP53* mutation frequency patterns in the three groups was determined by chi square test. Finally, age at first diagnosis was also compared according to *TP53* mutation status for each cancer type and mean values for age at first diagnosis for each cancer type and *TP53* mutation category is shown in [Figure S7B](#). For each cancer type, two sided t tests were used to determine whether *TP53* mutation status had a significant effect on age of first diagnosis ([Figure S7B](#); [Table S7A](#)). Overall survival (still incomplete due to limited years of followup for a number of patients) was stratified by *TP53* mutation status for cancer types with sufficient numbers of *TP53* mutations ([Table S7](#); [Figures 7B, 7D, and S7C–S7F](#)). Overall survival for patients with wild-type *TP53* tumors was compared to patients with mutant *TP53* tumors for each cancer type by log rank test using GraphPad Prism 7 software. Results for each of these comparisons is indicated in [Table S7A](#).

Development and Testing of a Mutant p53 RNA Expression Signature

The mutant p53 RNA expression signature was based on the aggregated expression of four genes, *CDC20*, *PLK1*, *CENPA*, and *KIF2C*, in the TCGA PanCancer dataset. These four genes were almost invariably significantly overexpressed in all cancer types with mutant *TP53*. Moreover, these genes have been established as (a) targets of wild-type p53 repression, (b) promoters of cell cycle progression, (c) components of the G2/M checkpoint, and (d) established E2F targets (with the exception of *CENPA*). For each individual cancer type (with sufficient numbers of *TP53* mutations) each of the four signature genes was ranked by RNA expression levels. These rankings were then added together to give a combined ranking of the four genes across all tumors of a given cancer type to give relative mutant p53 signature values. Those patient tumors in the lowest quartile for mutant p53 signature expression were then compared to those patients in the highest quartile for signature expression for each cancer type. First, each cancer type was tested for how well the signature correlated with *TP53* mutation status. For most cancers, tumors with *TP53* mutation correlated significantly with a high mutant p53 expression signature. For each cancer type, tumors of the low and high mutant p53 signature quartiles were then compared for overall survival by log rank test. The resulting survival curves are shown in [Figures 7B, 7D, and S7C–S7F](#) and detailed in [Table S7B](#).

Many of the TCGA cancer types were further analyzed for overall survival based on mutant p53 signature by the Web-based tool Kaplan-Meier plotter (Nagy et al., 2018) by both top and bottom quartile splits and by median splits (comparison of tumors with signature values above and below the median signature values) ([Table S7B](#), last column). Values from these analyses correlated well with our earlier analyses. Likewise, we performed Kaplan-Meier plotter analyses on non-TCGA tumor RNA expression datasets, including breast, ovarian, lung, gastric, and two liver carcinoma datasets and found that these survival analyses closely matched our results with TCGA RNA expression datasets.

The mutant p53 signature studies were developed from TCGA datasets subjected to fairly uniform data collection methods for each cancer type. So, application of the signature analysis to any newly accrued patient cancers might be subject to data collection methods and conditions quite different from the original TCGA data collection and analysis methods. Thus, absolute signature values derived from newly collected patient data might not correlate well with the values obtained from the original TCGA-derived analyses. To internally normalize signature values within each patient tumor sample we matched expression of each of the four mutant p53 signature genes (*CDC20*, *PLK1*, *KIF2C*, and *CENPA*) with expression of a normal control gene in that individual tumor. To qualify as a matched control gene, the gene had to meet the following criteria: (1) it had to have a nearly identical median expression as the matched signature gene; (2) it had to exhibit a very low average deviation across all tumors within a cancer type; and (3) it had to display no evidence of correlated expression with the matched signature gene (see example in [Figure S7G](#)). For each of the 11 TCGA cancer types that showed prognostic dependence on the mutant p53 signature, we chose four normalization genes based on the above criteria and these are indicated in [Table S7C](#). For each tumor of a given type, we then totaled the expression of the four signature genes and the four normalization control genes and determined ratios of total signature expression/total control expression for all of the tumors. [Figure S7H](#) shows a typical plot of these ratios for TCGA lung adenocarcinomas. When we compared the quartile of tumors with the highest signature/control ratios to the lowest for overall survival by log-rank test, we saw significantly poorer survival for the high quartiles versus the low quartiles ([Table S7C](#)), correlating well with the original mutant p53 signature analyses reported in [Table S7B](#). For most cancers, we found that signature/control ratios above 1.8 put the tumor sample in the high risk quartile, while ratios below 0.5 put the sample in the low risk quartile.

To determine whether this approach could be used for a non-TCGA dataset, we examined relative expression of the four LIHC signature and normal genes in an RNA expression dataset of 243 hepatocellular carcinomas from China and obtained very similar results to those of the 366 TCGA LIHC. Nevertheless, if the appropriate tumor RNA expression data is available, we believe it is preferable to develop normalization genes from a locally available dataset using the three criteria described above for choosing normalization genes.

QUANTIFICATION AND STATISTICAL ANALYSIS

Assessment of TP53 allele status through integration of TP53 mutation and copy number

Statistical differences in variant allele fractions for tumors with 2 *TP53* mutations relative to those with one *TP53* mutation were determined by two sided t test.

TP53 Status and Global Copy Number Alterations

Wild-type and mutant *TP53* tumor differences in amplification/deletion frequencies were statistically determined to be significant by chi-square test.

TP53 Status and RNA Expression-based Pathway Analysis

Pathway analysis was performed on both the 500 most upregulated genes in wild-type and mutant *TP53* tumors of each cancer type by Gene Set Enrichment Analysis (GSEA) (Mootha et al., 2003; Subramanian et al., 2005).

TP53 Status and Protein Expression-based Pathway Analysis

Mean protein expression levels were determined for proteins in the wild-type and mutant *TP53* tumor cohorts (in cancer types with sufficient numbers of *TP53* mutations) and significant upregulation of expression in wild-type and mutant *TP53* cancers were statistically analyzed by two sided t tests. Wild-type and mutant *TP53* tumors were compared in individual cancer types and all cancer types combined. P values for statistical differences in proteins upregulated in the combined mutant *TP53* cancer types are shown in Figure 4D and Table S6B. P values for statistical differences in proteins upregulated in wild-type *TP53* cancer types are shown in Table S6A.

Analyses of mutually exclusive and co-occurring genomic alterations

To understand if *TP53* genomic alterations exhibit mutually exclusivity with other gene alterations in cancer patients, we used the Mutex algorithm (Babur et al., 2015) on 33 TCGA studies iteratively. We used ChiBE (Babur et al., 2014b; Babur et al., 2010) to retrieve known pathway relations in Pathway Commons (Cerami et al., 2011) between the genes in the result groups and filtered-in the relations of *TP53*. To identify co-occurring genomic alterations with *TP53* mutation, we utilized the cBioPortal for Cancer Genomics “Enrichments” function (<https://www.cbioportal.org/>) (Cerami et al., 2012; Gao et al., 2013).

TP53 Status and Clinical Parameter Correlations

The significance of the difference in *TP53* mutation frequency patterns in overall survival of all TCGA patients for the three groups (dead < 1 year, alive > 4 years, all other patients) was determined by chi square test (Figure 7A). For each cancer type, two sided t tests were used to determine whether *TP53* mutation status had a significant effect on age of first diagnosis (Figure S7B; Table S7A). Overall survival for patients with wild-type *TP53* tumors was compared to patients with mutant *TP53* tumors for each cancer type by log rank test using GraphPad Prism 7 software. Results for each of these comparisons is indicated in Table S7A.

Development and Testing of a Mutant p53 RNA Expression Signature

For each cancer type, tumors of the low and high mutant p53 signature quartiles were compared for overall survival by log rank test. Survival curves are shown in Figures 7B, 7D, and S7C–S7F and detailed in Table S7B. Many of the TCGA cancer types were further analyzed for overall survival based on mutant p53 signature by the Web-based tool Kaplan-Meier plotter (Nagy et al., 2018) by both top and bottom quartile splits and by median splits (comparison of tumors with signature values above and below the median signature values) (Table S7B, last column).

DATA AND CODE AVAILABILITY

The published article includes all datasets generated or analyzed during this study. The supplementary (Tables S1, S2, S3, S4, S5, S6, and S7) contain these datasets.

Supplemental Information

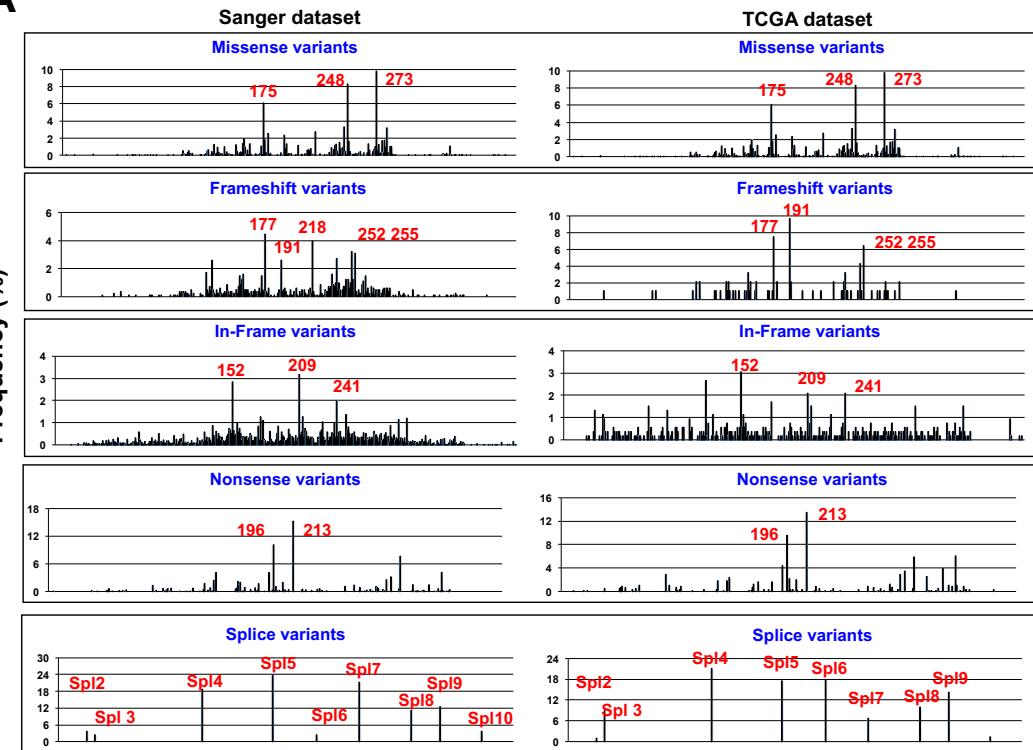
Integrated Analysis of *TP53* Gene and Pathway

Alterations in The Cancer Genome Atlas

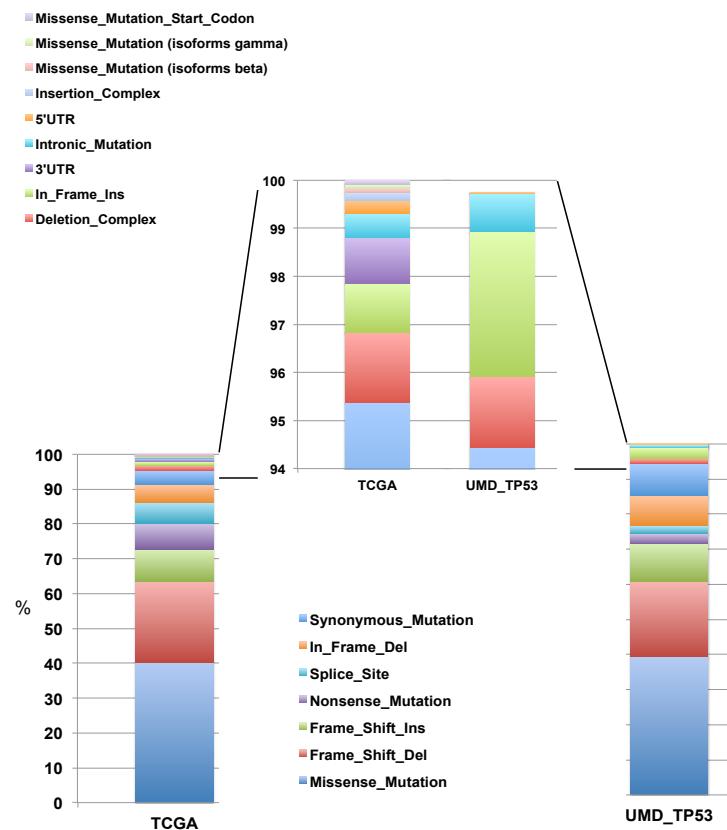
Lawrence A. Donehower, Thierry Soussi, Anil Korkut, Yuexin Liu, Andre Schultz, Maria Cardenas, Xubin Li, Ozgun Babur, Teng-Kuei Hsu, Olivier Lichtarge, John N. Weinstein, Rehan Akbani, and David A. Wheeler

Supplemental Figure 1

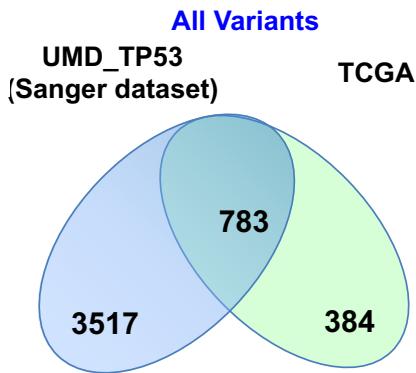
A



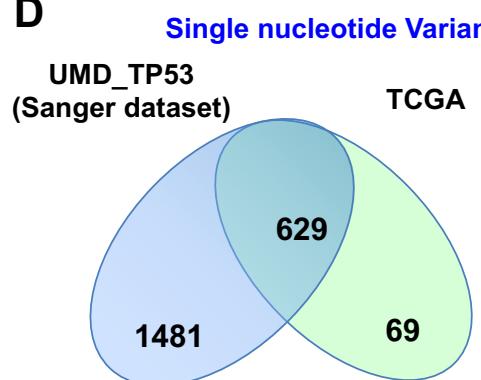
B



C



D



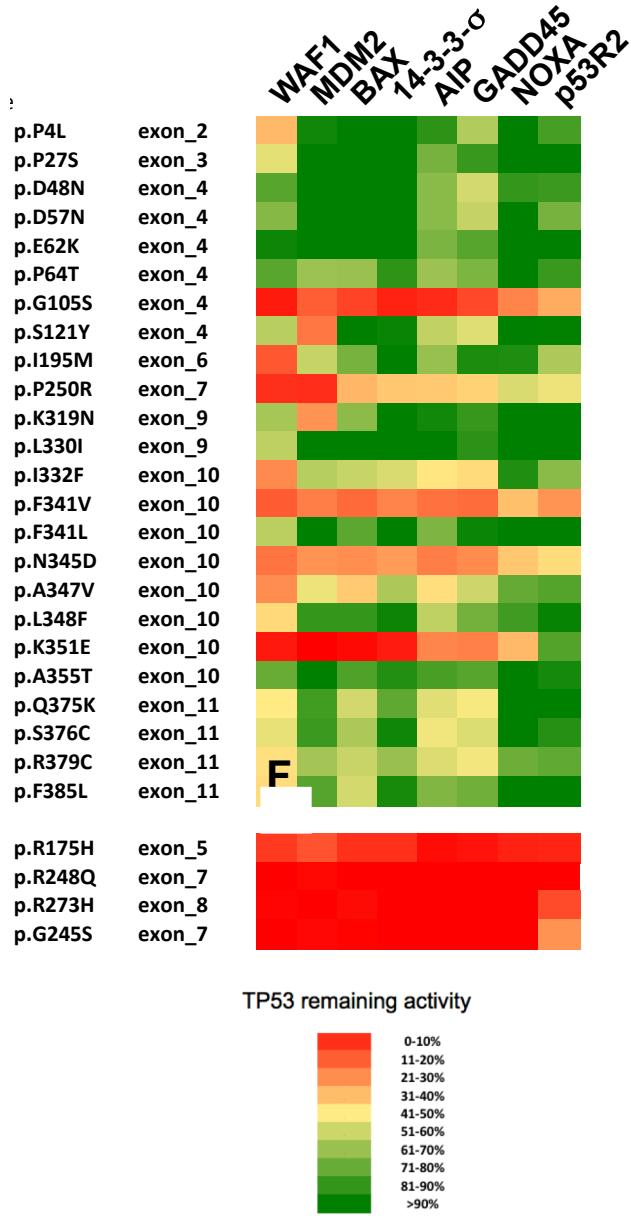
E

Event	Number
3'UTR	11
5'UTR	3
Intron	8
Missense_Mutation	24
Nonsense_Mutation	6
Splice_Site	7
Synonymous_mutation	9
Translation_Start_Site	1

Figure S1. *TP53* mutation profile in TCGA dataset (related to Fig. 1). (A) Distribution of *TP53* mutations in the TCGA dataset according to specific mutation type is similar to the Sanger sequenced subset of the UMD_ TP53 Mutation Database. Bar heights indicate mutation frequency at individual codons across the protein. Major *TP53* hotspot mutations are indicated. (B) Similar overall fractions of mutation types are observed in the TCGA (left) and UMD_ TP53 Mutation (right) databases. Infrequently observed mutation types are magnified at the top. (C,D) Venn diagrams showing the overlap between *TP53* variants from the Sanger subset of the UMD_ TP53 database (blue) and the TCGA dataset (green) for all *TP53* variants (C) or for Single Nucleotide Variants (D). Most *TP53* variants unique to the TCGA dataset are nucleotide indels (315 out of 384). (E) Unique Single Nucleotide Variants (SNVs) of *TP53* in the TCGA dataset (69) are dispersed in different parts of the *TP53* gene.

Supplemental Figure 2

A



B

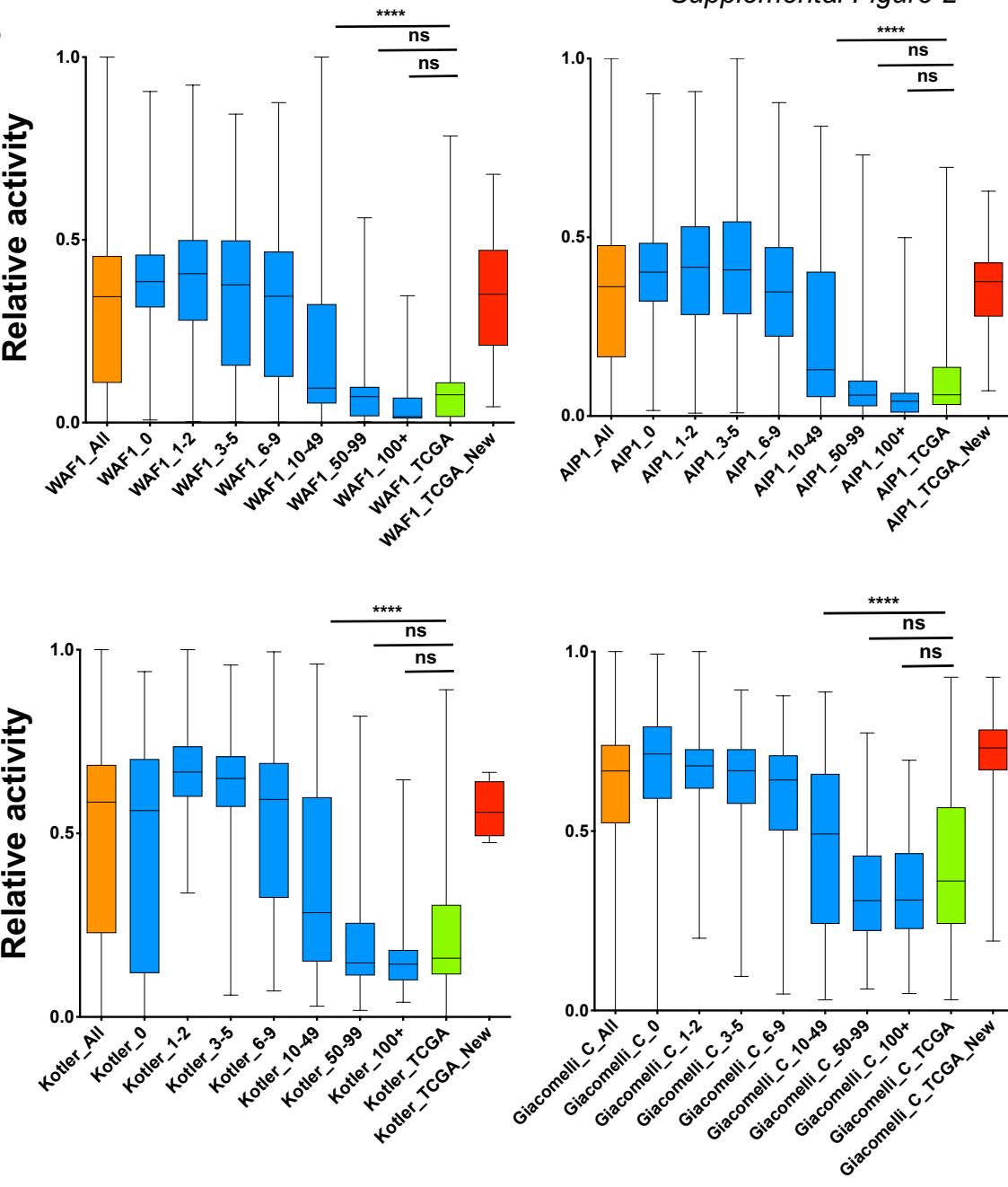


Figure S2. P53 Functionality in tumors with *TP53* mutations (related to Fig. 1). (A) Analysis of previously unreported TCGA-specific *TP53* variant loss of function for the 24 SNVs localized in coding exons. Residual transactivating activity based on *TP53* variant functional assays reported by Kato et al. (2003) for *WAF* (*CDKN1A*), *MDM2*, *BAX*, *I4-3-3- σ* , *AIP*, *GADD45A*, *NOXA* and *P53R2* promoters ranges from 0 (red) to 100% (green) as indicated in the legend. The remaining transcriptional activity for four *TP53* hotspot variants is shown at the bottom of the figure. (B) Analysis of functionality of all UMD_ TP53 Mutation Database variants (orange box and whiskers plots), eight classes of mutant *TP53* alleles based on their frequency of occurrence (indicated by frequency numbers on the X axis) in the UMD_ TP53 mutation database (blue box plots), all TCGA *TP53* variants (green box plots), as well as previously unreported identified TCGA variants (red box plots). Upper graphs are based on functional data analysis of transcriptional activity of mutant p53 proteins on a *CDKN1A* (WAF1) p53 response element (left) or an AIP1 p53 response element (right) in yeast-based assays from Kato et al. (2003). The lower left graph is based on functional data of growth suppressive activity of p53 variants in cell-based assays performed by Kotler et al. (2018). The lower right graph is based on tests for dominant-negative activity of variants in cellular functional assays performed by Giacomelli et al. (2018). The Y axis shows mean variant functionality relative to wildtype p53 based on the parameters of each of the functional assays employed.

Supplemental Figure 3

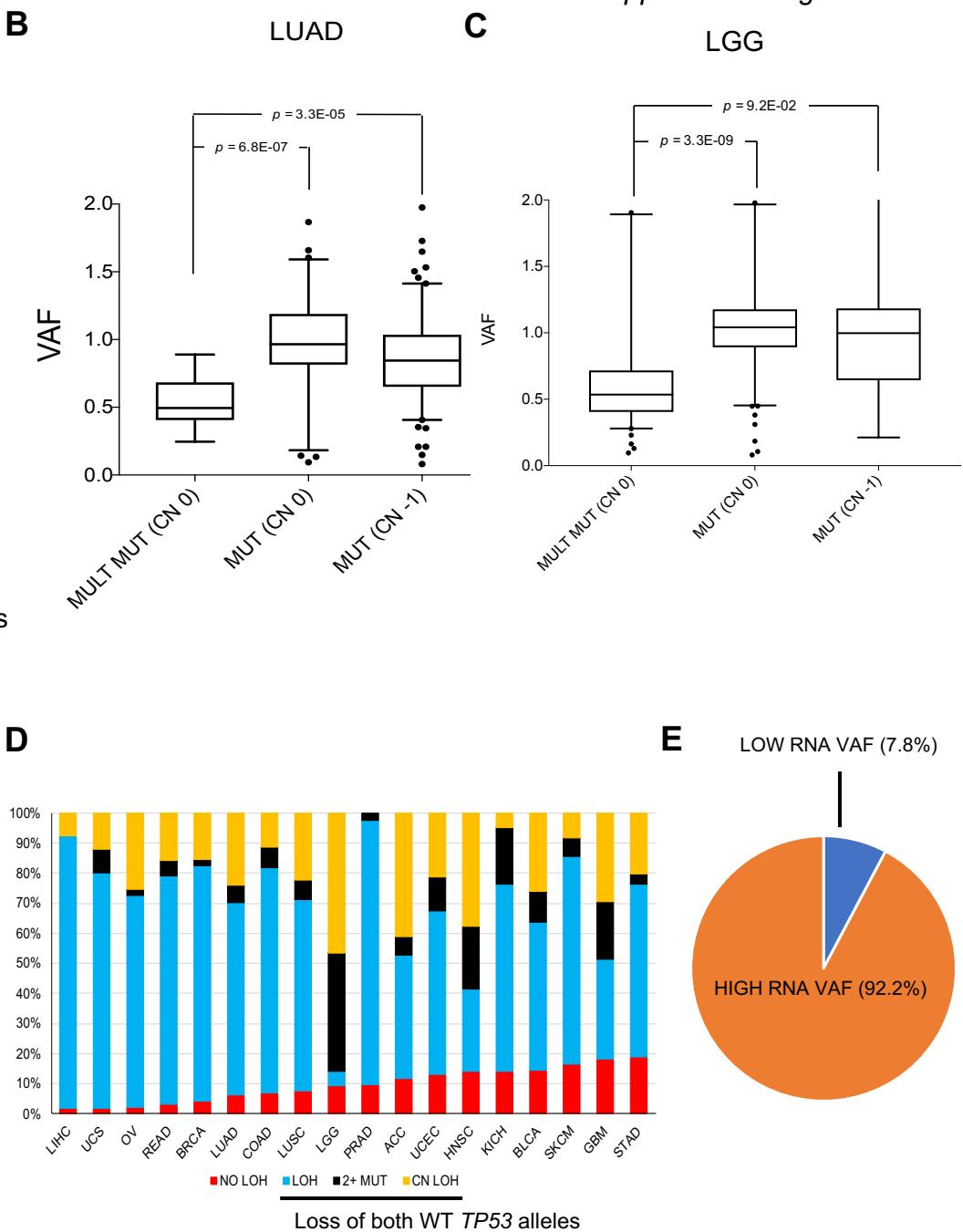
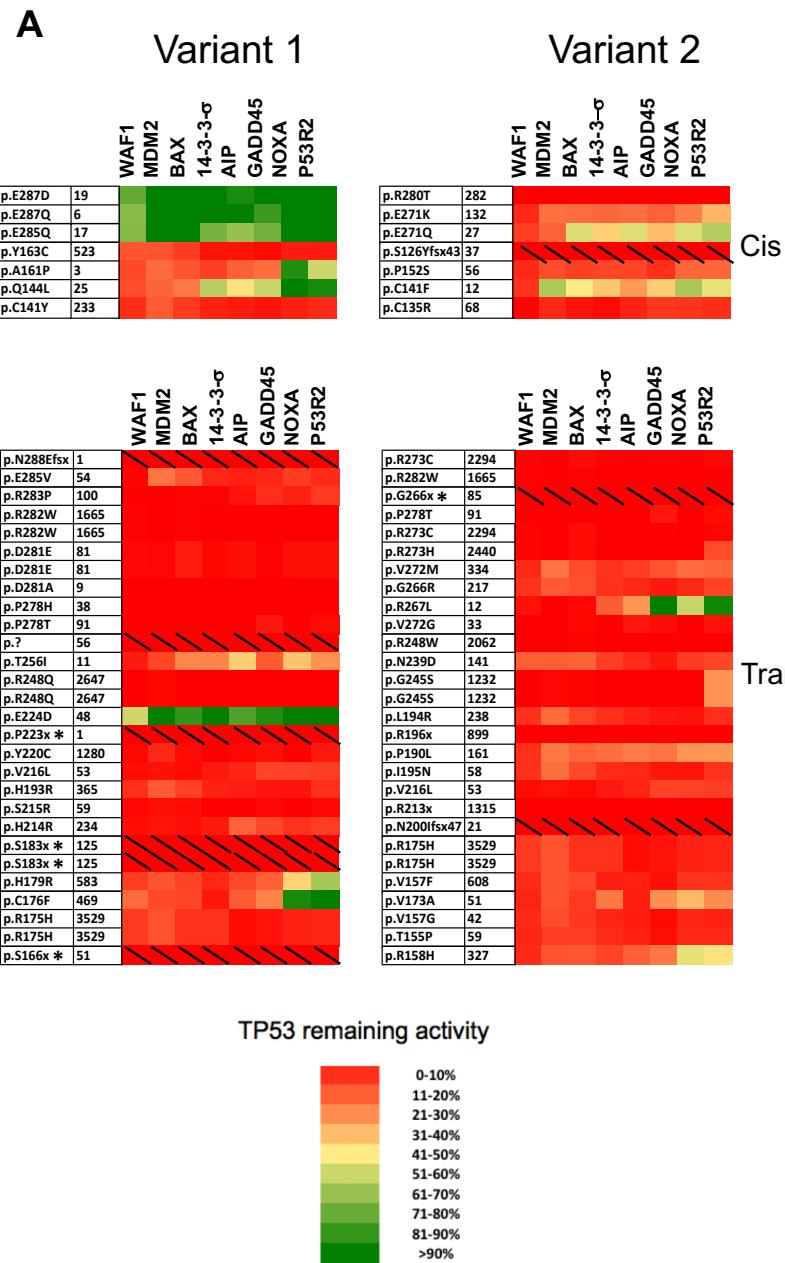
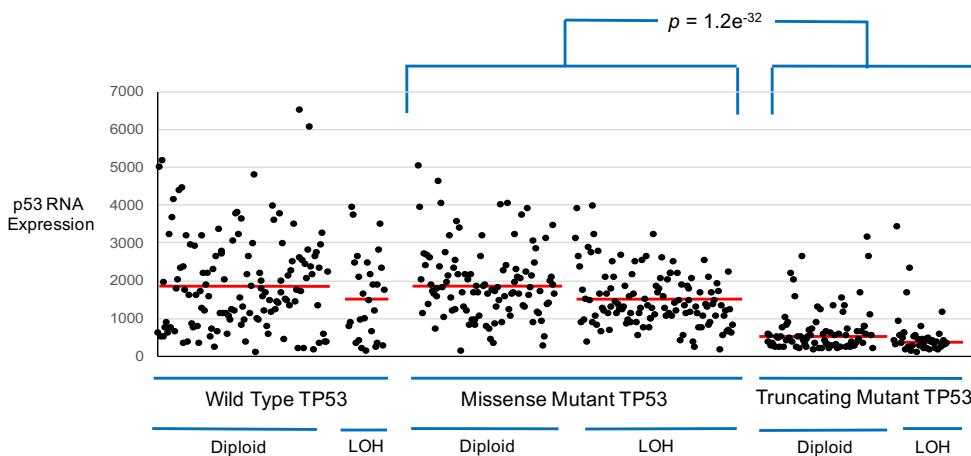


Figure S3. Structural and functional status of both *TP53* alleles in TCGA cancers with mutant *TP53* (related to Fig. 2) (A) TCGA tumors with two *TP53* mutations located within 75 nucleotides on the same exon were determined to be either on the same allele (Cis) or on separate alleles (Trans) by analysis of individual DNA sequence reads. Both *TP53* mutations and frequencies of occurrence of each variant in the UMD p53 database in each tumor are indicated to the left of the two heat maps using p53 protein (NP_000537.3) as a reference. While each horizontal bar of the heat map corresponds to a single tumor, the columns represent transcriptional function assays for each mutant p53 on known p53-responsive gene promoters (indicated above each column) as described in the legend of Fig. S2. The legend at the bottom indicates relative transcriptional activity of each mutant p53 in the transcriptional assay. Red boxes with diagonal bars indicate *TP53* truncating mutations that are assumed to have less than 10% functional activity. (B,C) Median variant allele fractions (VAF) in lung adenocarcinomas (B) and low grade gliomas (C) with one *TP53* mutation are near 1.0, indicating frequent loss of both wildtype *TP53* alleles. Tumors with *TP53* mutations were stratified by mutation number and copy number. A copy number (CN) of 0 is considered diploid and copy number of -1 is considered haploid. Individual DNA sequence reads for wild and variant (mutant) *TP53* were totaled, adjusted for tumor purity fraction, and variant allele fraction (VAF) determined. VAF distributions for each mutation category are shown and median values are indicated by the central bar in the box and whiskers plots. Statistical significance by t tests compare tumors with two or more (MULT) *TP53* mutations versus tumors with a single *TP53* mutation that are either diploid (CN 0) or haploid (CN -1) for *TP53*. (D) Most TCGA tumors display loss of both wildtype *TP53* alleles. Based on *TP53* VAF and copy number analyses as described for Fig. 2E,F and Fig. S3B,C, bar graphs were generated showing fractions of various types of *TP53* mutant configurations for all cancer types. *TP53* mutations in the “MUT (CN 0)” category were classified as CN LOH (copy neutral *TP53* mutation) if they display VAF values greater than 0.75. Those MUT *TP53* tumors with VAF values less than 0.75 were classified as “NO LOH” (red segments of bar graph) and are the only mutant *TP53* category not showing loss of both *TP53* alleles. (E) The vast majority of TCGA tumors with missense *TP53* mutations show little or no residual expression of WT p53 RNA. The individual tumor purity-adjusted *TP53* variant allele fractions based on RNAseq were quantified for each of 799 TCGA tumors with non-synonymous *TP53* mutations. VAF values above 0.75 were considered “high” and VAF values below 0.75 were considered “low”.

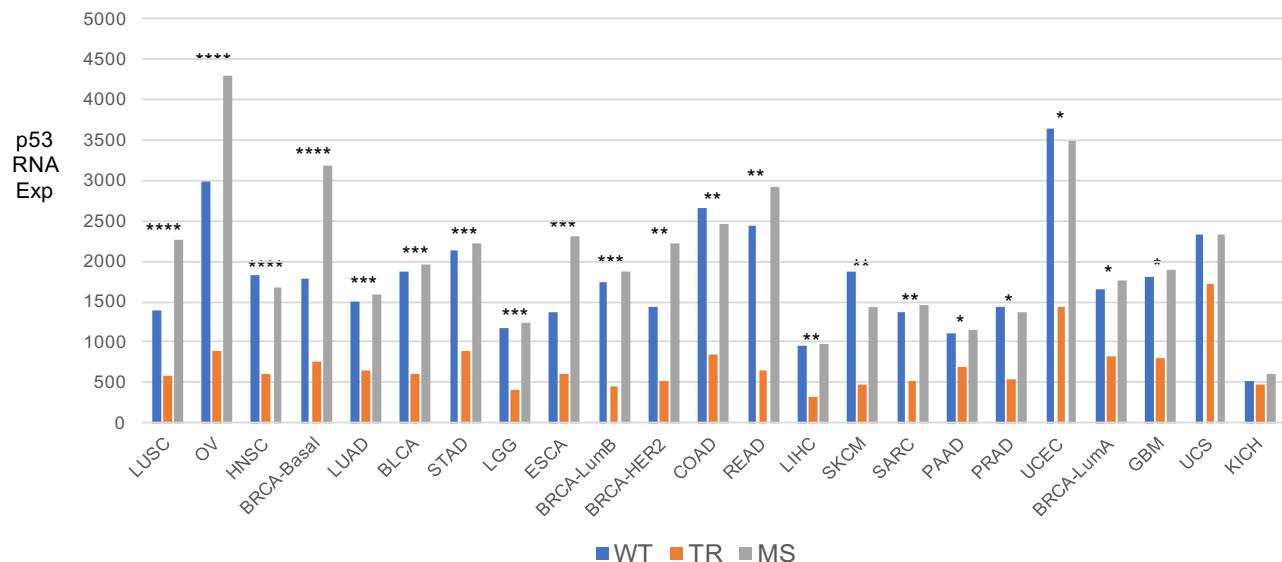
Supplemental Figure 4

A



B

RNA Expression in Tumors with Wild Type TP53 or with Missense or Truncating Mutations



**** $p < 10E-30$
 *** $p < 10E-20$
 ** $p < 10E-10$
 * $p < 10E-03$

C

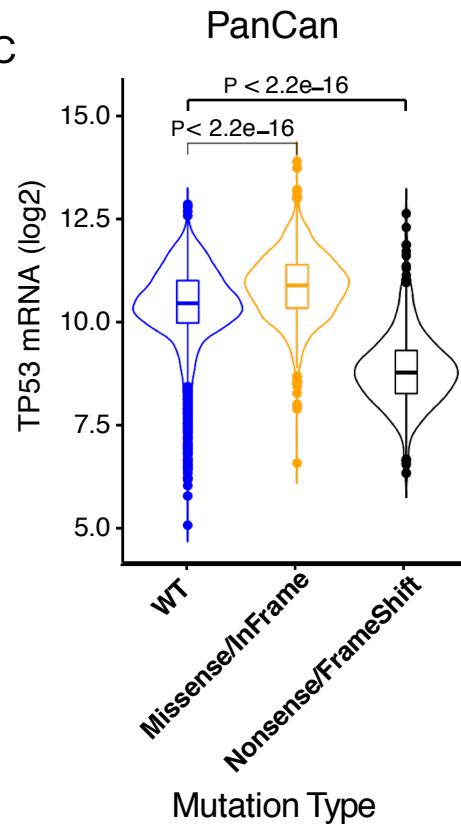
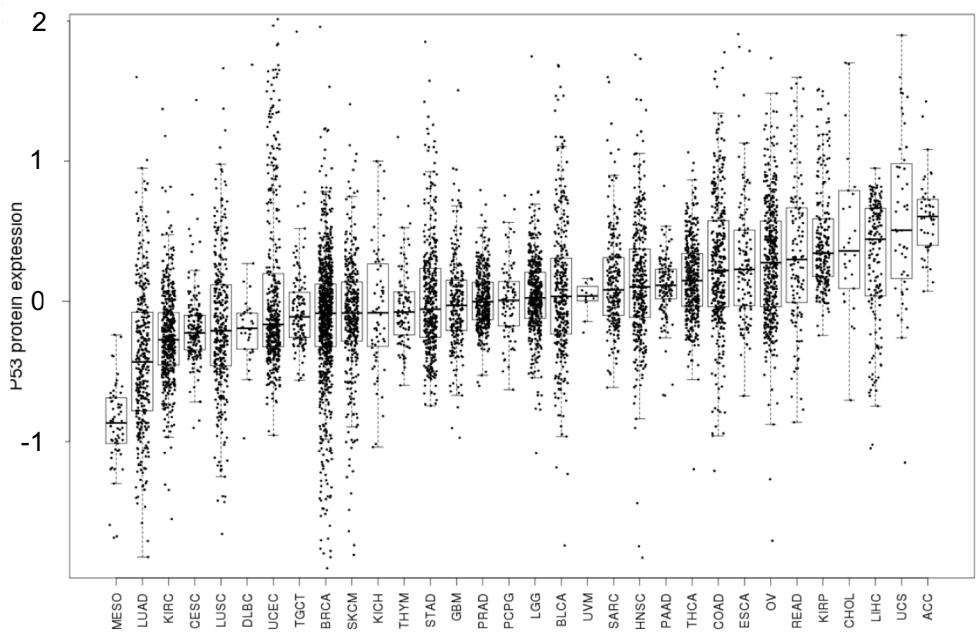


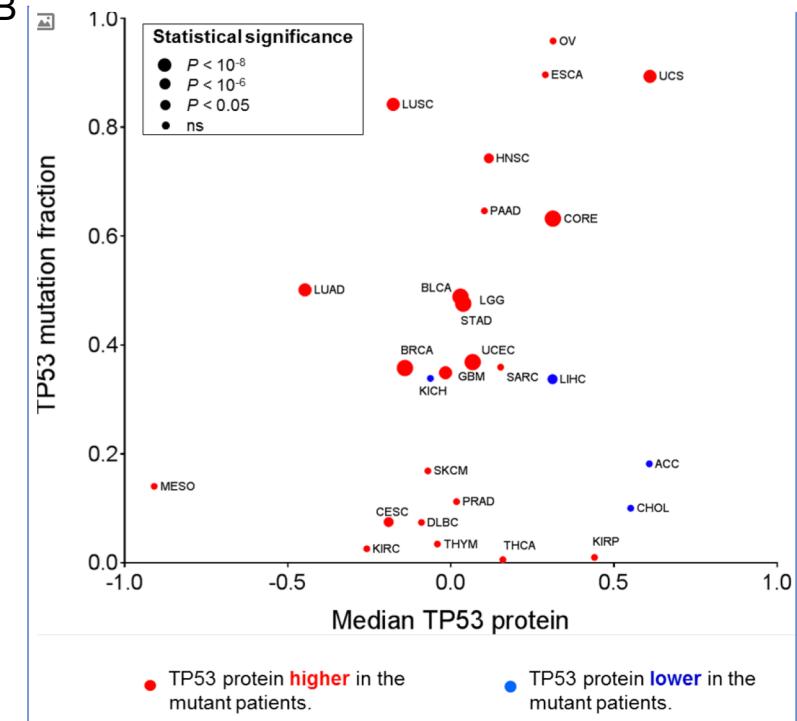
Figure S4. p53 RNA expression differs in tumors depending on *TP53* mutation status (related to Fig. 4). (A) Individual tumor p53 RNA expression shows dramatic reduction of p53 RNA in those tumors with truncating *TP53* mutations. RNA expression values in all breast cancers are stratified according to *TP53* mutation status (wildtype, non-truncating mutation, or missense mutation) and *TP53* copy number status (diploid or haploid). Mean expression levels for each *TP53* category are indicated by horizontal red lines. The significant p53 RNA expression differences between tumors with non-truncating and truncating *TP53* mutations (by t test) is indicated. (B) Mean p53 RNA expression in each of the three major *TP53* mutation categories for each cancer type are indicated by bars. WT (blue bar) = Wildtype *TP53*; TR (orange bar) = Truncating *TP53* mutation; MS (gray bar) = missense *TP53* mutation. Asterisks indicate the relative significance of the difference in p53 RNA expression in tumors with truncating and non-truncating *TP53* mutations. (C) Violin plots showing p53 mRNA levels in all TCGA cancers were aggregated and stratified into three groups: (a) WT *TP53* tumors (blue); (b) MUT *TP53* tumors containing missense/in frame *TP53* mutations (yellow); and (c) MUT *TP53* tumors with nonsense/frameshift/splice site *TP53* mutations (black).

Supplemental Figure 5

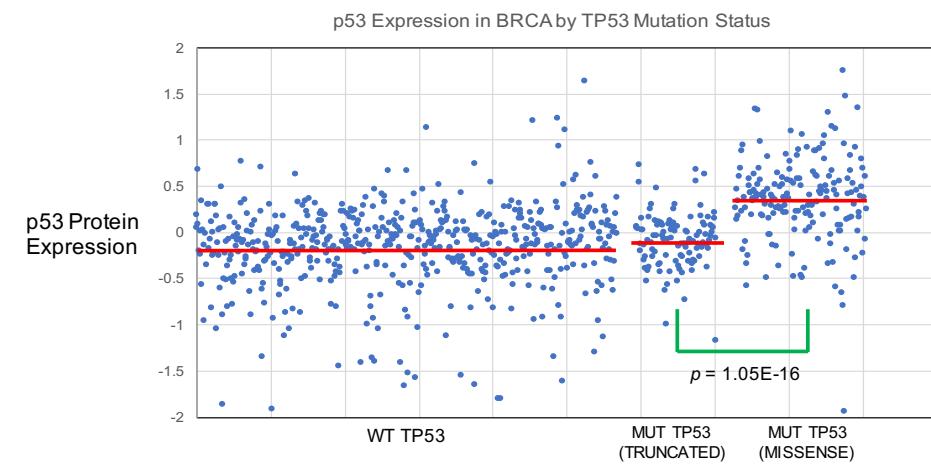
A



B



C



D

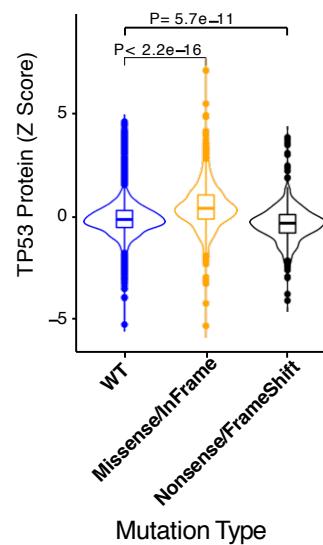


Figure S5. P53 protein expression varies in tumors depending on *TP53* mutation status (related to Fig. 4). (A) Range of relative p53 protein expression in each of the TCGA tumor types. (B) Correlation of *TP53* mutation fraction and median p53 protein expression levels in individual tumors. Red data points indicate cancers in which p53 protein levels are higher in the mutant *TP53* tumors of a given type are higher than their WT *TP53* counterparts. (C) TCGA breast cancers with non-truncating *TP53* mutations express significantly more p53 protein than those with no *TP53* mutations or truncating *TP53* mutations. Mean protein expression levels for each *TP53* category are indicated by horizontal red lines. An unpaired t test was used to show significance of the differences in p53 protein expression between tumors with truncating and non-truncating mutations in *TP53*. (D) Violin plots showing p53 protein levels in all TCGA cancers (RPPA data) were aggregated and stratified into three groups: WT *TP53* tumors (blue); MUT *TP53* tumors containing missense/in frame *TP53* mutations (yellow); and MUT *TP53* tumors with nonsense/frameshift/splice site *TP53* mutations (black).

Supplemental Figure 6

A



READ



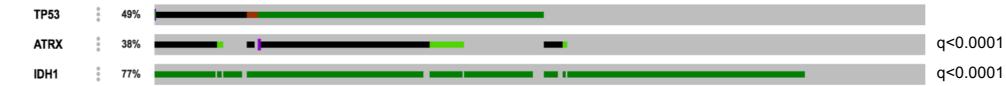
UCEC



PAAD



LGG



GBM



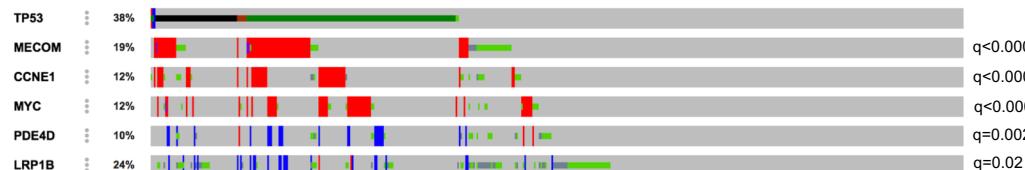
B



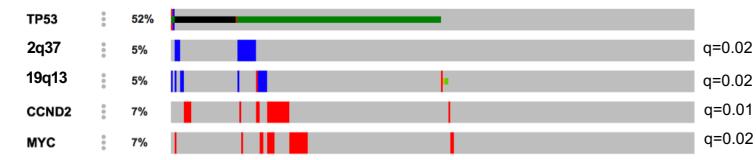
PAAD



UCEC



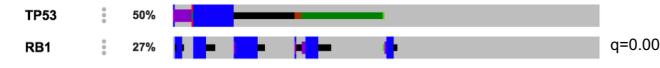
LGG



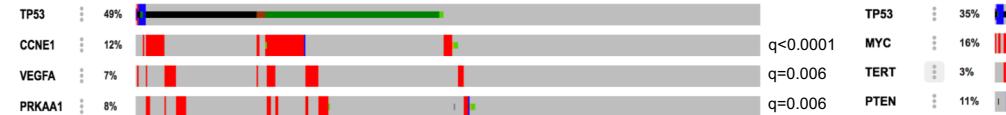
HNSC



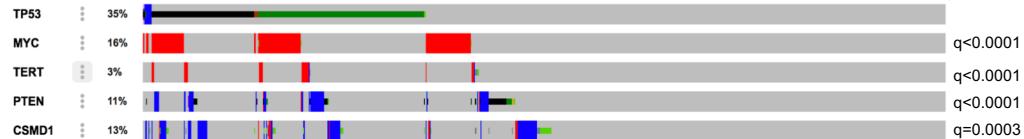
SARC



STAD



BRCA



Truncating mutation

Missense mutation

In frame mutation

Amplification

Deep deletion

No mutation

Figure S6. Genomic alterations that significantly co-occur with *TP53* mutations in individual cancer types (related to Fig. 6). (A) The cBioPortal for Cancer Genomics algorithms were used to generate oncoprints showing gene mutations that significantly co-occur with *TP53* mutations across all samples in six TCGA cancer types. Types of mutations arising in *TP53* and co-occurring genes are designated in the legend at the bottom. For missense mutations and in frame mutations the darker squares indicate likely driver mutations and the lighter squares indicate mutations of unknown driver potential. Significance of the co-occurrences are indicated by q values to the right. (B) Deep deletions and amplifications that significantly co-occur with *TP53* mutations in eight cancer types. Types of gene copy number alterations are designated in the legend. Significance of the co-occurrences are indicated by q values to the right.

Supplemental Figure 7

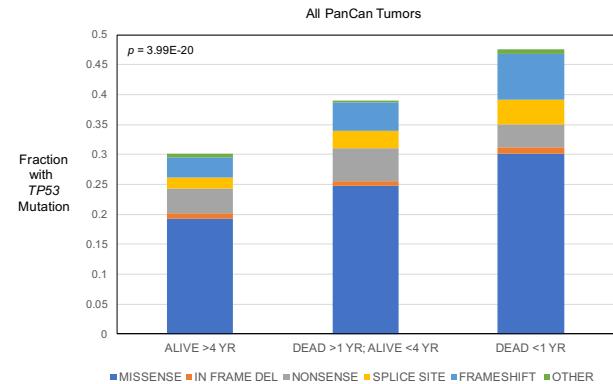
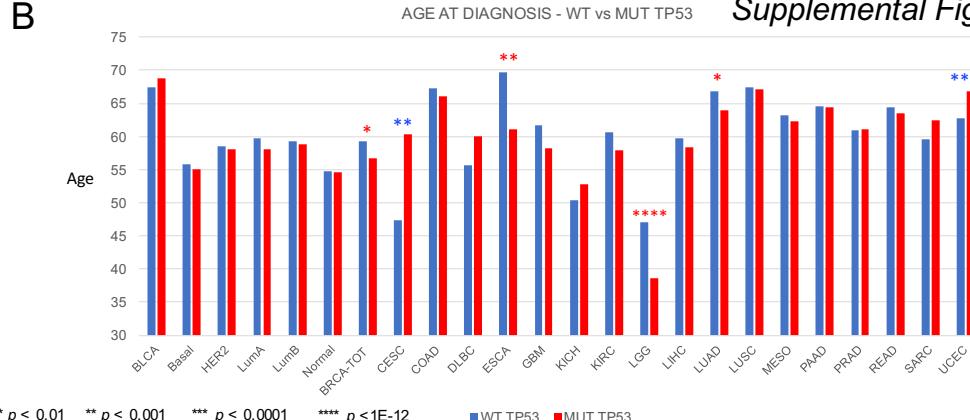
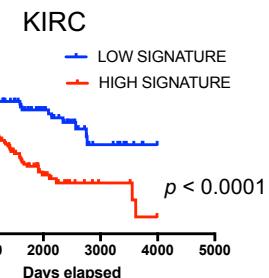
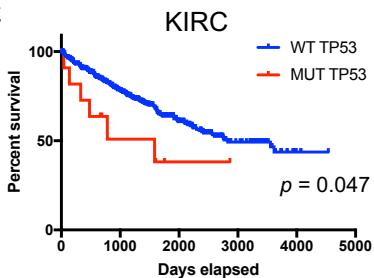
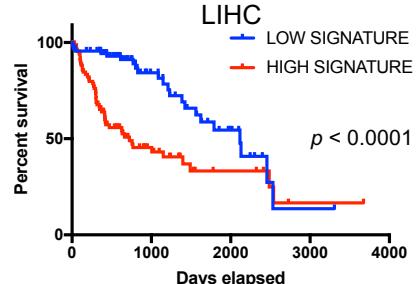
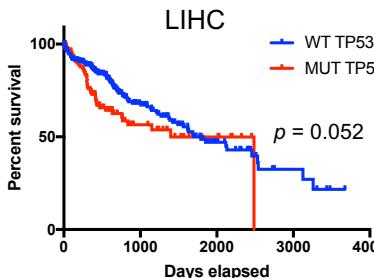
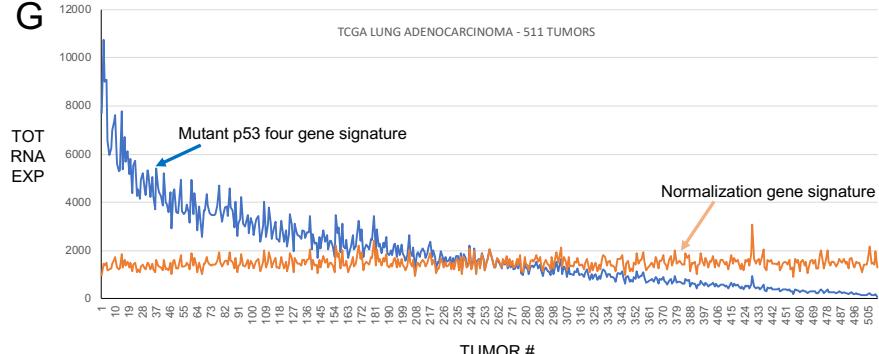
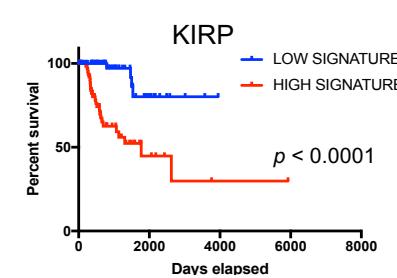
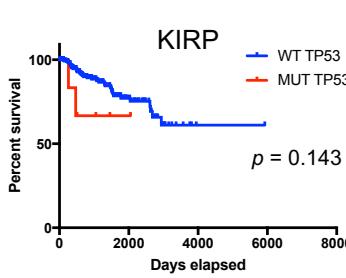
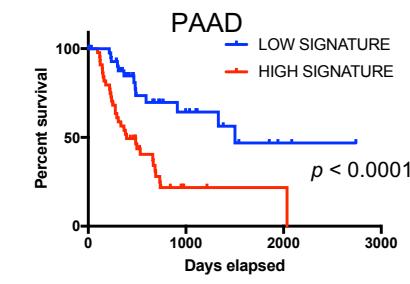
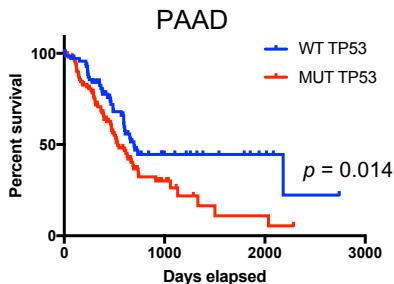
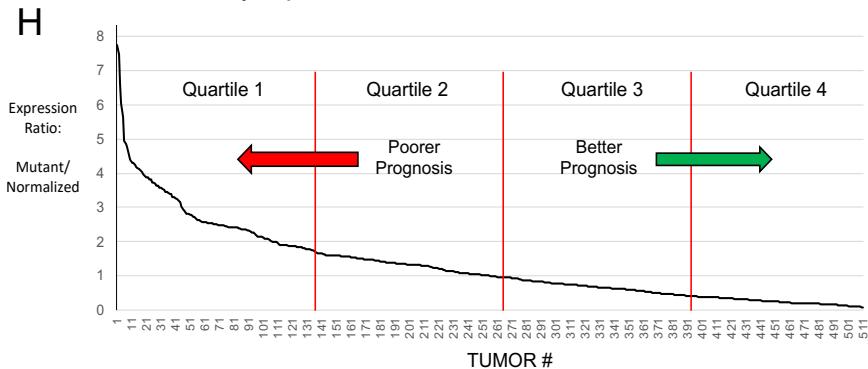
A**B****C****E****G****D****F****H**

Figure S7. Effects of TP53 mutation status and mutant p53 signature on overall survival (related to Fig. 7). (A) Comparative effects of *TP53* mutation status on short and long term cancer survivors. The clinical data for 20 TCGA cancer types with sufficient fractions of *TP53* mutations were aggregated and classified into three survival categories: (i) patients who died within one year of initial diagnosis, (ii) patients who were alive 4 years or more after initial diagnosis, and (iii) all other patients. The fraction of each patient survival category that showed *TP53* mutation was then determined. Types of *TP53* mutation for each patient category were also determined and these are indicated by different colors in each bar (legend at the bottom of panel A). (B) Effects of *TP53* mutation status on age of initial cancer diagnosis. The TCGA clinical data was examined for all cancer types with sufficient *TP53* mutation numbers and the average age at initial diagnosis of cancer determined. This data was then stratified by *TP53* mutation status. Significant *TP53* mutation-dependent differences in age at initial diagnosis for a particular cancer type are noted by asterisks (legend). (C-F) For some TCGA cancers, stratification by mutant p53 signature is more prognostically predictive of overall survival than stratification by *TP53* mutation status. For each cancer, overall survival is compared by log rank analysis using stratification by *TP53* mutation status (left) or by mutant p53 signature status (right). Log rank *p* values are given for each comparison. (C) KIRC (kidney renal clear cell carcinoma), (D) KIRP (kidney renal papillary carcinoma), (E) LIHC (liver hepatocellular carcinoma), (F) PAAD (pancreatic adenocarcinoma). (G) Normalization of the mutant p53 signature values to expression of four normal control genes within each tumor facilitates evaluation of individual patient tumors. Total RNA expression values in each of 511 lung adenocarcinomas for the four genes of the mutant p53 signature (*CDC20*, *PLK1*, *KIF2C*, *CENPA*, blue line) are compared to expression of the normalization genes (*UBIAD1*, *ZMYM6*, *RPP30*, *ZNF17*, brown line). (H) Ratios of expression in lung adenocarcinomas of four mutant p53 signature genes and four normalization genes from panel G facilitate stratification of high prognosis tumors and low prognosis tumors. Log rank survival analysis of high and low quartiles in this graph gave *p* values <0.0001 with a high/low hazard ratio of 2.372. See also Table S7C.