nature genetics

LETTERS

https://doi.org/10.1038/s41588-018-0204-y

Mutational processes shape the landscape of *TP53* mutations in human cancer

Andrew O. Giacomelli[1,2], Xiaoping Yang[2], Robert E. Lintner[2], James M. McFarland[2], Marc Duby[2], Jaegil Kim[2], Thomas P. Howard[1,2,3], David Y. Takeda[1,2,3], Seav Huong Ly[1,2], Eejung Kim[2], Hugh S. Gannon[1,2], Brian Hurhula[2], Ted Sharpe[2], Amy Goodale[2], Briana Fritchman[2], Scott Steelman[2], Francisca Vazquez[1,2], Aviad Tsherniak[2], Andrew J. Aguirre[1,2], John G. Doench[2], Federica Piccioni[2], Charles W. M. Roberts[1,3,4], Matthew Meyerson[1,2,3], Gad Getz[2,3,5,6], Cory M. Johannessen[2], David E. Root[2] and William C. Hahn[1,2,3,7*]

Unlike most tumor suppressor genes, the most common genetic alterations in tumor protein p53 (TP53) are missense mutations[1,2]. Mutant p53 protein is often abundantly expressed in cancers and specific allelic variants exhibit dominant-negative or gain-of-function activities in experimental models[3–8]. To gain a systematic view of p53 function, we interrogated loss-of-function screens conducted in hundreds of human cancer cell lines and performed TP53 saturation mutagenesis screens in an isogenic pair of TP53 wild-type and null cell lines. We found that loss or dominant-negative inhibition of wild-type p53 function reliably enhanced cellular fitness. By integrating these data with the Catalog of Somatic Mutations in Cancer (COSMIC) mutational signatures database[9,10], we developed a statistical model that describes the TP53 mutational spectrum as a function of the baseline probability of acquiring each mutation and the fitness advantage conferred by attenuation of p53 activity. Collectively, these observations show that widely-acting and tissue-specific mutational processes combine with phenotypic selection to dictate the frequencies of recurrent TP53 mutations.

Tumor protein p53 (*TP53*) is the most frequently mutated gene in human cancer[1]. By interrogating genome-scale RNA interference (RNAi)[11] and CRISPR–Cas9[12] loss-of-function (LOF) screens, we found that cells harboring wild-type (WT) *TP53* acquired a selective fitness advantage when *TP53* itself or several upstream activators (*TP53BP1*, *CHEK2*, *ATM*, *USP28*) or effector genes (*CDKN1A*, *ZMAT3*) were suppressed or deleted (Fig. 1a–d, Supplementary Fig. 1a, and Supplementary Table 1). We also found that suppression or deletion of *TP53* was well tolerated in cells harboring *TP53* missense mutations, indicating that gain-of-function (GOF) activities associated with these mutations[6–8] did not affect the propagation of cancer cell lines (Fig. 1a and Supplementary Fig. 1a). These observations suggest that loss of canonical p53 function plays a central role in the selective advantage associated with *TP53* mutation. However, because these assays rely on the natural *TP53* variation present in cancer cell lines, they do not provide functional information about every possible *TP53* mutation.

To determine the function of each missense or nonsense *TP53* mutation, we created a comprehensive library of *TP53* mutants using Mutagenesis by Integrated TilEs (MITE)[13–15]. We designed the library so that each allele would contain a single mutation and that each of the 20 natural amino acids and a stop codon would be represented at each codon position. We also included alleles harboring silent mutations, which did not change the amino acid sequence of p53 but allowed us to track WT allele performance. Our mutagenesis approach generated > 99.8% of the expected mutant alleles, as gauged by massively parallel sequencing of the expression plasmid pool (Supplementary Table 2).

To study the function of these alleles in the presence or absence of endogenous p53, we created isogenic WT *TP53* ($p53^{WT}$) and null *TP53* ($p53^{NULL}$) A549 human lung carcinoma cell populations using CRISPR–Cas9-mediated gene editing[16]. These cells display potent p53-dependent drug responses, proliferate robustly in culture, and allow for stable expression of lentivirally delivered *TP53* at near-endogenous levels when present at a single copy per cell (Supplementary Fig. 2a). We capitalized on the differential responses of these isogenic cells to two p53-activating agents, nutlin-3 and etoposide, and performed pooled positive selection screens designed to enrich for dominant-negative (DN), LOF, or WT-like alleles.

Nutlin-3 interrupts the interaction between p53 and E3 ubiquitin-protein ligase Mdm2 (MDM2)[17], a key negative regulator of p53 stability[18] and transcriptional function[19]. Suppression of MDM2 activity is critical for oncogenic stress-induced[20–22] and DNA damage-induced[23] p53 activation. Cancer cell lines harboring WT p53 require MDM2 expression for proliferation (Fig. 1b–d and Supplementary Fig. 1b). As expected, nutlin-3 treatment impaired the proliferation of $p53^{WT}$ A549 cells but had no effect on $p53^{NULL}$ cells (Supplementary Fig. 2b,c). Expression of exogenous WT p53 restored nutlin-3 sensitivity in $p53^{NULL}$ cells but did not alter the nutlin-3 response in $p53^{WT}$ cells. Notably, expression of mutant p53 p.Pro278Ala had no effect on $p53^{NULL}$ cells but rendered $p53^{WT}$ cells partially nutlin-3-resistant, indicating that this allele is deficient for WT function and interferes with endogenously expressed WT p53 in a DN manner. In the heterozygous setting, similar transcriptionally

[1]Dana–Farber Cancer Institute, Boston, MA, USA. [2]Broad Institute of MIT and Harvard, Cambridge, MA, USA. [3]Harvard Medical School, Boston, MA, USA. [4]Department of Oncology, St. Jude Children's Research Hospital, Memphis, TN, USA. [5]Massachusetts General Hospital Center for Cancer Research, Boston, MA, USA. [6]Department of Pathology, Massachusetts General Hospital, Boston, MA, USA. [7]Department of Medicine, Brigham and Women's Hospital, Boston, MA, USA. *e-mail: william_hahn@dfci.harvard.edu

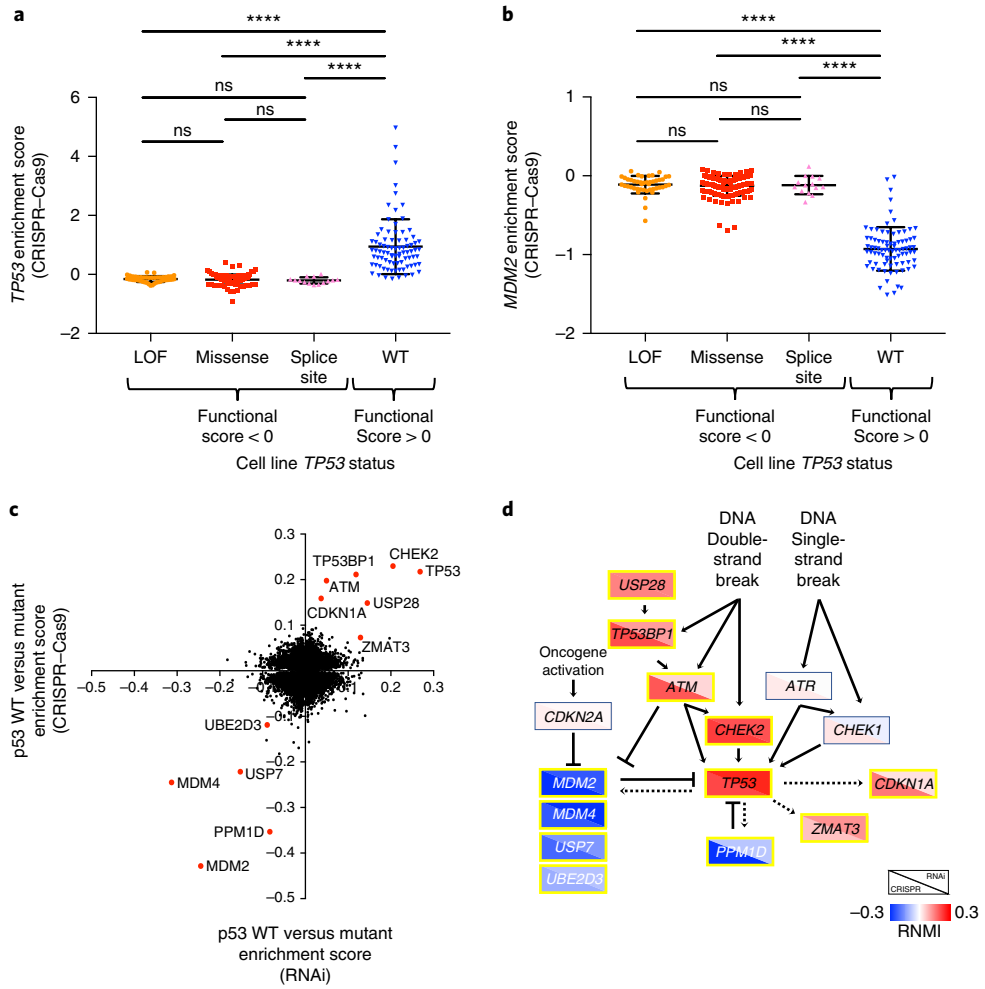

Fig. 1 | Deletion of endogenous WT but not mutant *TP53* impacts fitness in human cancer cells. Comparison of enrichment scores for CRISPR–Cas9 reagents targeting *TP53* (**a**) or *MDM2* (**b**) in cell lines whose *TP53* statuses were defined using *TP53* mutation, copy number, target gene expression, and nutlin-3 sensitivity data (see Supplementary Table 1). Only cell lines with concordant functional and genetic classifications were included in these analyses. Each point represents the gene-level score for a given cell line and the error bars indicate the mean and s.d. of each group. Cell lines in the LOF category harbor homozygous deletion, frameshift, or nonsense mutation of *TP53* and express low levels of p53 protein (see Supplementary Fig. 1c) (****$P < 0.0001$, two-tailed Welch's *t*-test). **c**, PARIS (Probability Analysis by Ranked Information Score), a rescaled normalized mutual information (RNMI)-based statistical analysis was used to nominate genes whose enrichment scores were significantly different between the p53 nonfunctional mutant (LOF, missense, splice site, and indel, with functional score < 0) and p53 functional WT (functional score > 0) cell line classes in genome-scale CRISPR–Cas9[12] and RNAi[11] screens. Reported p53 pathway components that scored as significant (false discovery rate < 0.05) in both analyses are highlighted in red. **d**, Infographic depicting differential enrichment scores of reported p53 pathway members and target genes in genome-scale CRISPR–Cas9 (lower left triangle) and RNAi (upper right triangle) screens. The dashed lines indicate transcriptional regulation. Genes scoring as significant in both analyses are outlined in yellow.

inactive full-length p53 mutants sequester WT p53 from active complexes and thus reduce p53 function more efficiently than mutants that are unstable or prematurely truncated[24,25]. We anticipated that we could identify all such DN *TP53* alleles by introducing the MITE library into $p53^{WT}$ cells and treating them with nutlin-3. Performing the complementary experiment in the $p53^{NULL}$ A549 cells would allow for the selection of LOF alleles and the depletion of alleles with WT-like activity.

Using a competition assay, we found that $p53^{NULL}$ A549 cells were more sensitive than their $p53^{WT}$ counterparts to high doses of DNA double-strand break (DSB)-inducing agents, most notably etoposide and doxorubicin (Supplementary Fig. 3). Although in mouse thymocytes DNA damage-mediated induction of WT p53 leads to apoptosis[26–28], in other contexts, p53 activation induces cell cycle arrest and DNA damage repair and thereby antagonizes p53-independent modes of cell death that result from unresolved DNA damage[29–31]. Our observations indicate that in A549 cells, WT p53 promotes cell survival in response to high doses of DSB-inducing agents.

Indeed, we found that forced expression of WT p53 in $p53^{NULL}$ A549 cells prevented the dramatic cell death induced by 5 μM etoposide treatment, whereas expression of mutant p53 p.Pro278Ala or *Renilla* luciferase, a negative control, had no effect (Supplementary Fig. 2e). We anticipated that by introducing the p53 MITE library into $p53^{NULL}$ cells and treating them with etoposide, we could enrich for alleles bearing WT-like activity and deplete LOF alleles. We noted that although the p.Pro278Ala mutant exhibited LOF in $p53^{NULL}$ cells, it failed to interfere with the etoposide response in $p53^{WT}$ cells, suggesting that its DN activity is limited to a subset of p53 responses (Supplementary Fig. 2d).

To perform the enrichment screens, we introduced the p53 MITE library into $p53^{WT}$ and $p53^{NULL}$ cells at a low multiplicity of infection (MOI), selected cells expressing *TP53* alleles using puromycin,

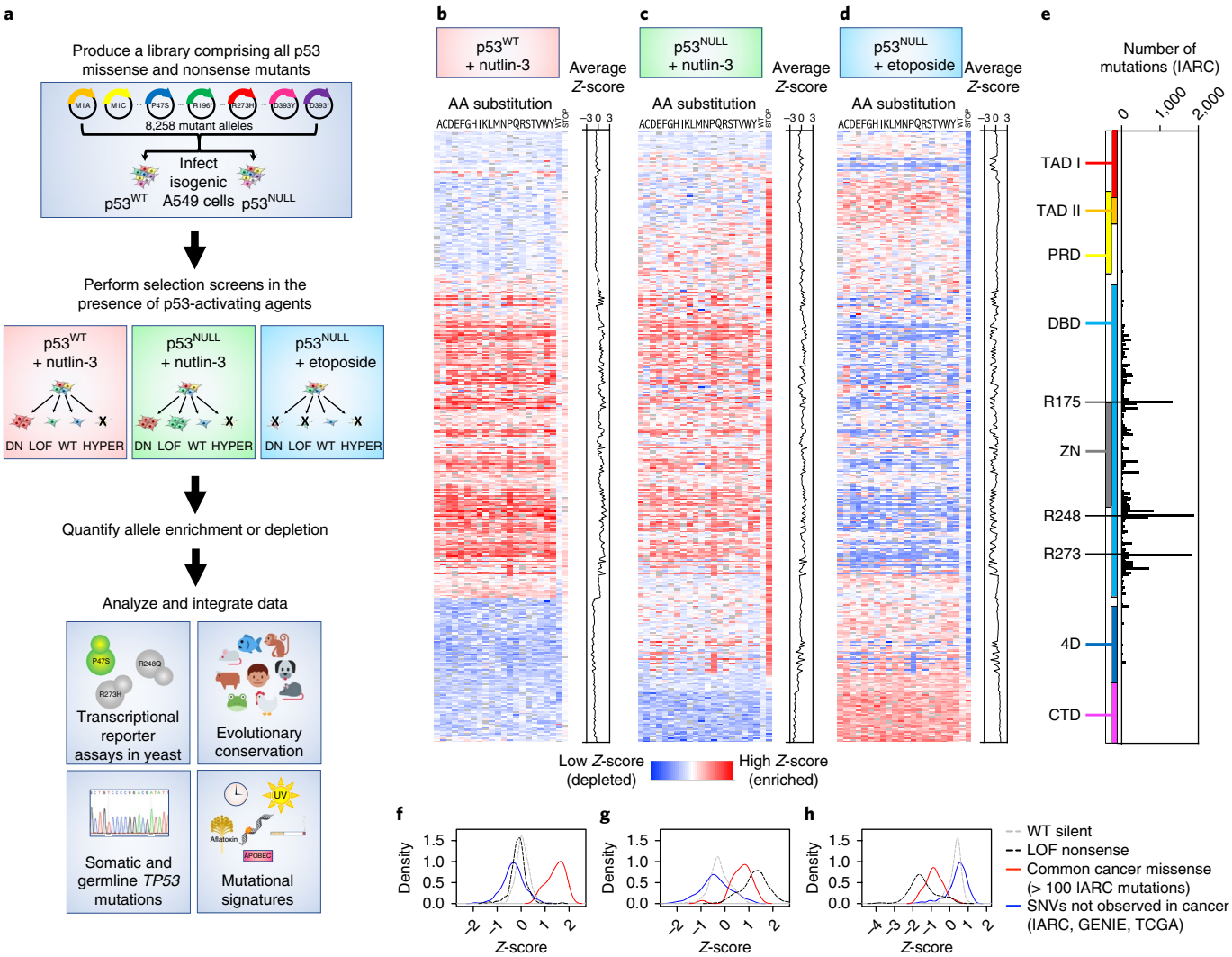

Fig. 2 | Comprehensive mutational scanning of *TP53*. a, A library comprising 8,258 mutant *TP53* alleles was introduced into A549 p53^{WT} and p53^{NULL} cells in a pooled format under conditions that favored the integration of a single vector in each cell. Library-infected p53^{WT} cells were treated with nutlin-3, and library-infected p53^{NULL} cells were treated with either nutlin-3 or etoposide. After 12 days, genomic DNA was collected, PCR-amplified, and subjected to next-generation sequencing. **b–d**, Heat maps of normalized allele enrichment scores (*Z*-scores) with codon-level average *Z*-scores plotted on the right. **e**, Left, The reported domain structure of p53 with residues 175, 248, and 273 highlighted. TAD, transactivation domain; PRD, proline-rich domain; DBD, DNA-binding domain; ZN, zinc-binding domain; 4D, tetramerization domain; CTD, C-terminal domain. Right, total number of missense and nonsense mutations found at each codon in the IARC database[2,32]. **f–h**, Density plots of alleles with silent mutations (WT alleles), nonsense mutations at codons 44–289 (LOF alleles), missense mutations that are common in cancer, and single-nucleotide variant (SNV)-generated missense mutations that have never been observed in cancer. Differences among all groups of alleles were significant in each condition ($P < 0.0001$, Mann-Whitney *U* test). Twitter emoji graphics are licenced under CC-BY 4.0 (see URLs).

and treated these populations with nutlin-3 or etoposide for 12 days (Fig. 2a). We then enumerated which *TP53* alleles were present by using massively parallel sequencing. We defined allele enrichments using the $\log_2$-fold change in reads in experimental arms relative to early time point samples and identified alleles that were selected under each of these conditions (Fig. 2b–d). The resulting allele enrichment scores are shown as heat maps juxtaposed to the reported p53 domain structure and the frequency of somatic *TP53* mutations in human tumors[2,32] (International Agency for Research on Cancer (IARC) database R18; see URLs) (Fig. 2e). We found that alleles harboring silent mutations were significantly depleted in p53^{NULL} cells treated with nutlin-3 and enriched in p53^{NULL} cells treated with etoposide relative to alleles bearing premature stop codons, suggesting that these assays effectively discriminated between WT and null alleles (Fig. 2f–h, $P < 0.0001$, Mann-Whitney *U* test). We also found that most alleles, especially those with mutations in the DNA-binding domain (DBD), performed similarly in all screen conditions (Supplementary Fig. 4). Indeed, of the 3,819 DBD missense variants we tested, 1,481 showed evidence of LOF in one or both assays performed in the absence of p53, and 1,219 (82.3%) of these exhibited DN activity (Supplementary Fig. 5). In addition, we identified several mutations in the transactivation (positions 18, 19, 22, 23, 24, 25, 26, 27, 29, and 30), DBD (positions 276, 280, and 281), and tetramerization (positions 330, 332, 337, 338, 341, 344, and 348) domains that exhibited differential responses in the three assays. We also found that not all alleles bearing stop codon mutations behaved as LOF alleles, with different patterns observed for stop codons at positions 1–43, 44–289, 290–362, and 363–393 (Fig. 2b–d, Supplementary Fig. 4–5). We also noted significant associations between the enrichment score of each allele,

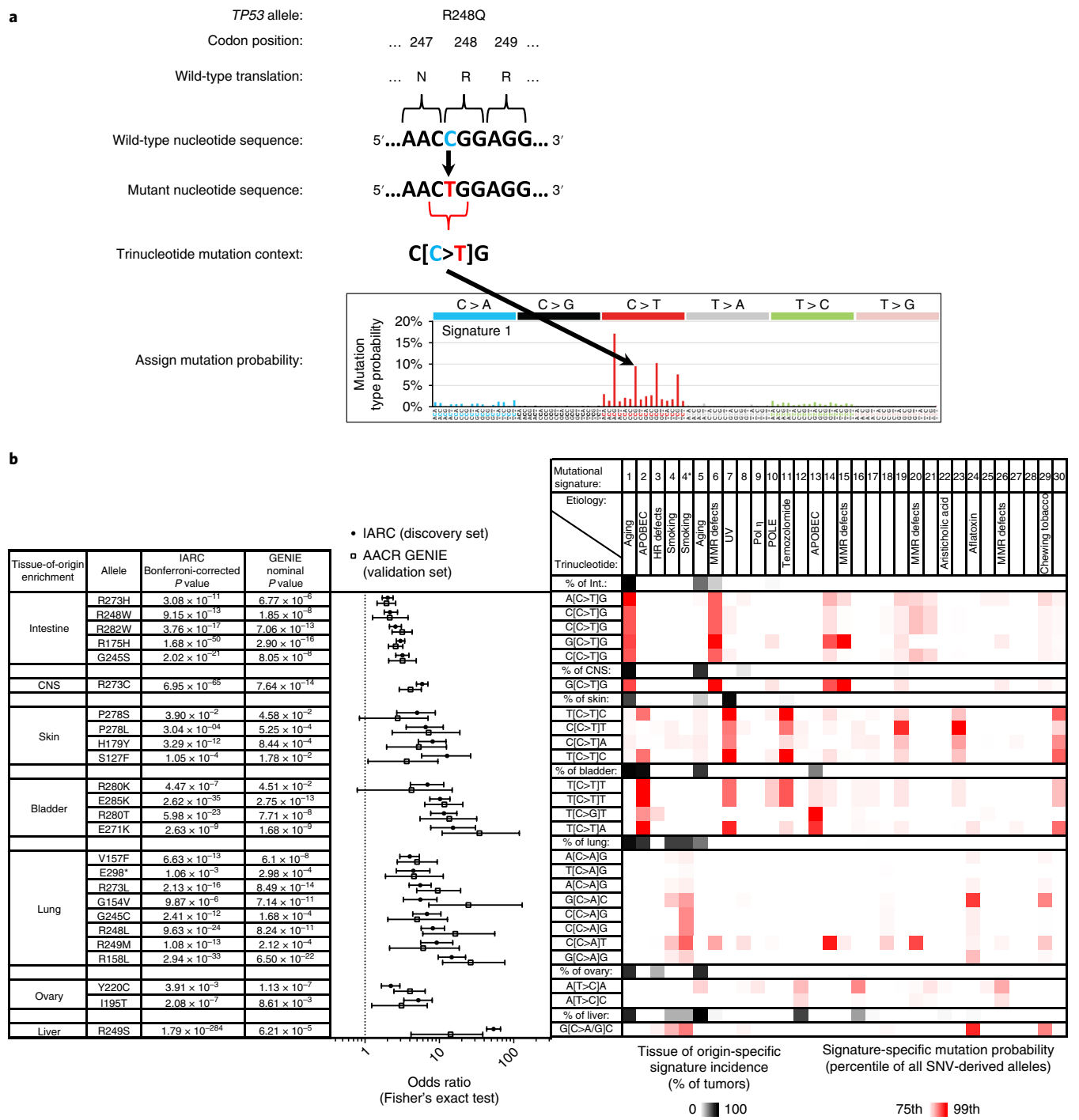

Fig. 3 | Tissue of origin-selective *TP53* mutations are linked to specific mutational processes. a, Each signature in the COSMIC mutational signatures database contains 96 mutation probabilities, one for each trinucleotide mutation type[9,10]. To assign a baseline mutation probability to each *TP53* allele for each signature, we first determined its trinucleotide mutation type and then assigned the corresponding value from the database. Depicted here is the assignment of a mutation probability for Arg248Gln under the influence of signature 1. **b**, Fisher's exact tests were performed to identify *TP53* mutations that occur significantly more frequently in specific tumor types in the IARC[2,32] and GENIE[35] databases. CNS, central nervous system. The heat map shows the relative mutation probabilities for each of the indicated *TP53* mutations under the influence of each signature, depicted graphically on a white-to-red scale as the signature-specific percentile of all single-nucleotide variant (SNV)-derived *TP53* mutations ($n = 2{,}810$). The percentage of tumors of each tissue of origin in which a given signature was found by Alexandrov et al.[10] is depicted graphically on a white-to-black scale.

its transcriptional activity measured in yeast[33], and the degree of evolutionary conservation of the mutated residue, as assessed by the Align-GVGD[34] program (Spearman rank test, $P < 0.0001$; see URLs) (Supplementary Fig. 4j,k). To facilitate interpretation of these data sets and to accommodate new data sets as they become available, we created a website that allows the function of any tested *TP53* allele to be queried (PHenotypic Annotation of TP53 Mutations (PHANTM); see URLs). The raw read counts for each allele in each condition are provided in Supplementary Table 2 and the mean *Z*-scores are reported in Supplementary Table 3. We note that the

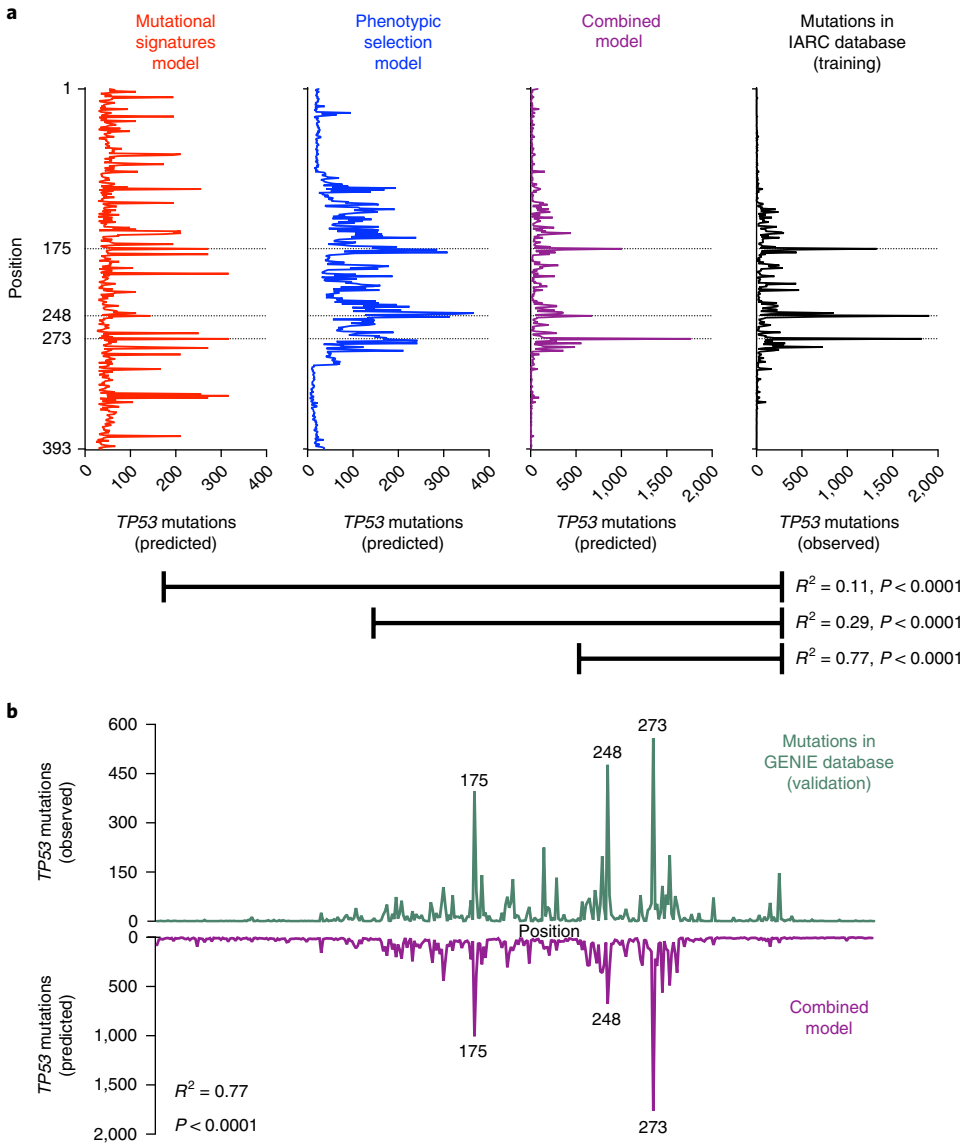

Fig. 4 | The *TP53* mutational spectrum modeled as a function of mutational signatures and phenotypic selection. a, Generalized linear models were trained to predict the mutation frequency of each *TP53* allele in the IARC database[2,32] using mutational signatures from the COSMIC database[9,10], phenotypic selection data from the *TP53* MITE library screens, or both. **b**, Position-level mutation rates predicted by the combined model are plotted downward; observed mutation rates in the GENIE validation database are plotted upward.

effect of missense mutations on splicing cannot be inferred from our data; therefore, mutations at residues adjacent to splice junctions (codons 25, 26, 32, 33, 125, 126, 187, 188, 224, 225, 261, 262, 307, 308, 331, 332, 367, and 368) may also alter *TP53* through their effects on splicing.

The observation that 30% of all p53 missense mutations found in human tumors affect five residues (Arg273, Arg248, Arg175, Arg282, and Gly245)[2,35] suggests that these residues play key roles in p53 function or that these positions are selectively targeted by mutational processes in cancer. When we examined these hotspot missense alleles, we found that while they exhibited LOF and DN activity, they were not the top scoring alleles in the library (Supplementary Fig. 4l). Moreover, we failed to detect *TP53* dependency in cancer cell lines harboring endogenous hotspot mutations (Supplementary Table 1). Thus, although these residues are important for p53 function, these observations argue that mutations may be common at these sites because of the inherent mutability of the DNA sequences that encode them. Indeed, all five of these codons contain methylated CpG dinucleotides, which are highly susceptible to mutation via spontaneous deamination of the cytosine residue[36]; correspondingly, most of the substitutions found at these positions are C>T mutations.

Recently, unbiased approaches to identify and quantify the activity of mutational processes in cancer have been developed using whole-genome and whole-exome data from more than 10,000 human tumor samples[9,10]. These efforts have uncovered 30 distinct mutational signatures that describe the likelihood of acquiring a specific base change, given the activity of an underlying mutational process and the identity of the bases that immediately flank the mutated base—a parameter termed the trinucleotide mutation context (Fig. 3a) (Catalog of Somatic Mutations in Cancer (COSMIC) mutational signatures; see URLs). Of the 30 mutational signatures that have been identified, 17 have been attributed specific etiologies[10], many of which have previously been associated with specific *TP53* mutations[36–43] (Fig. 3b).

We hypothesized that the tissue-selective activity of mutational processes would manifest as tissue-selective enrichment of

certain *TP53* mutations. To identify such mutations, we queried two independent *TP53* mutation databases (IARC[2,32] and GENIE[35]) and found 25 mutations that were significantly overrepresented in a specific tumor type in both databases (Fig. 3b). We next assigned each *TP53* mutation a baseline mutation probability as a function of each mutational signature and found that 23 of the 25 tissue-selective *TP53* mutations were associated with mutational signatures commonly found in the relevant tissue of origin. Among these were mutations enriched in colon, skin, bladder, lung, and liver cancers that are associated with signatures of mismatch repair deficiency (signature 6), ultraviolet (UV) exposure (signature 7), apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like (APOBEC) activity (signatures 2 and 13), tobacco smoke (signature 4*), and aflatoxin exposure (signature 24), respectively (Fig. 3b). We also noted that the most frequent *TP53* mutations found in all tumor types were associated with the most commonly observed mutational signature, signature 1, which reflects the natural degradation of 5-methylcytosine to thymine associated with aging[9,10]. Additionally, we found that several *TP53* alleles, which exhibited WT-like activity in our assays but are recurrently observed in human tumors, exhibit high intrinsic mutability, as shown by signature 1, suggesting that this common mutational process probably generates recurrent *TP53* passenger mutations (Supplementary Figs. 6 and 7). Collectively, these findings suggest that mutational processes play an important role in shaping the landscape of *TP53* somatic driver and passenger mutations.

Therefore, we attempted to model the *TP53* mutational spectrum as a function of (1) gene-agnostic mutational signatures[9,10] and (2) phenotypic selection as assessed by our *TP53* mutagenesis screen data. We included parameters for each of the three screen conditions and projections of mutational signatures that are found across all cancer types (signatures 1 and 5) as well as those that drive tissue-selective *TP53* mutation patterns (signatures 2, 4*, 6, 7, 13, and 24) (Fig. 3b). The models were trained to predict the frequency of somatic *TP53* mutations found in the IARC database[2,32] and were validated in the independent GENIE database[35]. We found that models derived using only mutational signatures or phenotype scores yielded weak but significant correlations with the number of mutations observed at each codon position in p53 in human tumors (Pearson $R^2 = 0.11$ and 0.29 respectively, $P < 0.0001$). However, a combined model that included all parameters recapitulated the observed mutation frequency at each position with high accuracy, capturing both DBD and hotspot enrichment patterns (Pearson's $R^2 = 0.77$, $P < 0.0001$) (Fig. 4a). This combined model was not simply overfitting the training data set, since it performed equally well in the validation database (Pearson's $R^2 = 0.77$, $P < 0.0001$) (Fig. 4b). We also verified that the models showed similar performance when predicting the frequency of mutations at a test set of codon positions not used for model fitting (Supplementary Fig. 6). Ultimately, this model suggests that mutational processes create genetic diversity at the *TP53* locus in somatic tissues using the same rules that govern their activity genome-wide, and that cells acquiring LOF and DN *TP53* mutations have a fitness advantage over those that retain WT p53 function. This model provides a parsimonious explanation for the enrichment of *TP53* missense mutations generally and hotspot mutations, in particular, in human tumors.

It has previously been argued that loss of p53 function is the critical determinant that underlies the selection of *TP53* mutations in cancer[33]. However, the preponderance of missense mutations relative to truncation mutations also argues that full-length mutant p53 actively promotes tumor development. Through our comprehensive screening approach, we found that > 80% of full-length p53 DBD missense mutants that display LOF also display DN activity, suggesting that the ability of mutant p53 to interfere with WT p53 is critical during tumorigenesis. Although other p53 GOF activities may affect different aspects of tumor biology not assessed in this study, the lack of mutant p53 dependency in cancer cell lines coupled with the robust relationships between LOF, DN activity, and mutation frequency indicate that the observed spectrum of *TP53* mutations probably arises due to the selection of LOF and DN alleles that are generated by specific mutational processes.

URLs. IARC, http://p53.iarc.fr/TP53SomaticMutations.aspx. AlignGVGD, http://agvgd.hci.utah.edu/. Twitter emojis, https://github.com/twitter/twemoji. PHANTM, http://mutantp53.broadinstitute.org. COSMIC mutational signatures, http://cancer.sanger.ac.uk/cosmic/signatures. Cancer Cell Line Encyclopedia, http://portals.broadinstitute.org/ccle. Genomics of Drug Sensitivity in Cancer, http://www.cancerrxgene.org. Cancer Target Discovery and Development, https://ocg.cancer.gov/programs/ctd2/data-portal. cBioPortal, http://cbioportal.org. GenePattern, https://genepattern.broadinstitute.org. TIDE: Tracking of Indels by DEcomposition, https://tide.deskgen.com. Genetic Perturbation Platform, https://portals.broadinstitute.org/gpp/public/resources/protocols. ORFcall, https://github.com/broadinstitute/ORFCall/releases/tag/v1.0. AACR GENIE, http://www.cbioportal.org/genie.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at https://doi.org/10.1038/s41588-018-0204-y.

Received: 22 November 2017; Accepted: 26 July 2018;
Published online: 17 September 2018

References

1. Hainaut, P. & Pfeifer, G. P. Somatic *TP53* mutations in the era of genome sequencing. *Cold Spring Harb. Perspect. Med.* **6**, a026179 (2016).
2. Petitjean, A. et al. Impact of mutant p53 functional properties on *TP53* mutation patterns and tumor phenotype: lessons from recent developments in the IARC TP53 database. *Hum. Mutat.* **28**, 622–629 (2007).
3. Milner, J. & Medcalf, E. A. Cotranslation of activated mutant p53 with wild type drives the wild-type p53 protein into the mutant conformation. *Cell* **65**, 765–774 (1991).
4. Harvey, M. et al. A mutant p53 transgene accelerates tumour development in heterozygous but not nullizygous p53-deficient mice. *Nat. Genet.* **9**, 305–311 (1995).
5. de Vries, A. et al. Targeted point mutations of p53 lead to dominant-negative inhibition of wild-type p53 function. *Proc. Natl. Acad. Sci. USA* **99**, 2948–2953 (2002).
6. Olive, K. P. et al. Mutant p53 gain of function in two mouse models of Li–Fraumeni syndrome. *Cell* **119**, 847–860 (2004).
7. Lang, G. A. et al. Gain of function of a p53 hot spot mutation in a mouse model of Li–Fraumeni syndrome. *Cell* **119**, 861–872 (2004).
8. Freed-Pastor, W. A. & Prives, C. Mutant p53: one name, many proteins. *Genes Dev.* **26**, 1268–1286 (2012).
9. Alexandrov, L. B. et al. Signatures of mutational processes in human cancer. *Nature* **500**, 415–421 (2013).
10. Alexandrov, L. B. et al. Clock-like mutational processes in human somatic cells. *Nat. Genet.* **47**, 1402–1407 (2015).
11. Tsherniak, A. et al. Defining a cancer dependency map. *Cell* **170**, 564–576.e16 (2017).
12. Meyers, R. M. et al. Computational correction of copy number effect improves specificity of CRISPR–Cas9 essentiality screens in cancer cells. *Nat. Genet.* **49**, 1779–1784 (2017).
13. Melnikov, A., Rogov, P., Wang, L., Gnirke, A. & Mikkelsen, T. S. Comprehensive mutational scanning of a kinase in vivo reveals substrate-dependent fitness landscapes. *Nucleic Acids Res.* **42**, e112 (2014).
14. Brenan, L. et al. Phenotypic characterization of a comprehensive set of MAPK1/ERK2 missense mutants. *Cell Rep.* **17**, 1171–1183 (2016).
15. Majithia, A. R. et al. Prospective functional classification of all possible missense variants in *PPARG*. *Nat. Genet.* **48**, 1570–1575 (2016).
16. Drost, J. et al. Sequential cancer mutations in cultured human intestinal stem cells. *Nature* **521**, 43–47 (2015).
17. Vassilev, L. T. et al. In vivo activation of the p53 pathway by small-molecule antagonists of MDM2. *Science* **303**, 844–848 (2004).
18. Fang, S., Jensen, J. P., Ludwig, R. L., Vousden, K. H. & Weissman, A. M. Mdm2 is a RING finger-dependent ubiquitin protein ligase for itself and p53. *J. Biol. Chem.* **275**, 8945–8951 (2000).

19. Momand, J., Zambetti, G. P., Olson, D. C., George, D. & Levine, A. J. The mdm-2 oncogene product forms a complex with the p53 protein and inhibits p53-mediated transactivation. *Cell* **69**, 1237–1245 (1992).
20. Zhang, Y., Xiong, Y. & Yarbrough, W. G. ARF promotes MDM2 degradation and stabilizes p53: *ARF-INK4a* locus deletion impairs both the Rb and p53 tumor suppression pathways. *Cell* **92**, 725–734 (1998).
21. Lin, A. W. & Lowe, S. W. Oncogenic *ras* activates the ARF-p53 pathway to suppress epithelial cell transformation. *Proc. Natl. Acad. Sci. USA* **98**, 5025–5030 (2001).
22. Eischen, C. M., Weber, J. D., Roussel, M. F., Sherr, C. J. & Cleveland, J. L. Disruption of the ARF–Mdm2–p53 tumor suppressor pathway in Myc-induced lymphomagenesis. *Genes Dev.* **13**, 2658–2669 (1999).
23. Khosravi, R. et al. Rapid ATM-dependent phosphorylation of MDM2 precedes p53 accumulation in response to DNA damage. *Proc. Natl. Acad. Sci. USA* **96**, 14973–14977 (1999).
24. Friedman, P. N., Chen, X., Bargonetti, J. & Prives, C. The p53 protein is an unusually shaped tetramer that binds directly to DNA. *Proc. Natl. Acad. Sci. USA* **90**, 3319–3323 (1993).
25. Wang, P. et al. p53 domains: structure, oligomerization, and transformation. *Mol. Cell. Biol.* **14**, 5182–5191 (1994).
26. Clarke, A. R. et al. Thymocyte apoptosis induced by p53-dependent and independent pathways. *Nature* **362**, 849–852 (1993).
27. Lowe, S. W., Schmitt, E. M., Smith, S. W., Osborne, B. A. & Jacks, T. p53 is required for radiation-induced apoptosis in mouse thymocytes. *Nature* **362**, 847–849 (1993).
28. Lowe, S. W., Ruley, H. E., Jacks, T. & Housman, D. E. p53-dependent apoptosis modulates the cytotoxicity of anticancer agents. *Cell* **74**, 957–967 (1993).
29. Lukin, D. J., Carvajal, L. A., Liu, W. J., Resnick-Silverman, L. & Manfredi, J. J. p53 promotes cell survival due to the reversibility of its cell-cycle checkpoints. *Mol. Cancer Res.* **13**, 16–28 (2015).
30. Bunz, F. et al. Requirement for p53 and p21 to sustain G2 arrest after DNA damage. *Science* **282**, 1497–1501 (1998).
31. Bunz, F. et al. Disruption of p53 in human cancer cells alters the responses to therapeutic agents. *J. Clin. Invest.* **104**, 263–269 (1999).
32. Bouaoun, L. et al. *TP53* variations in human cancers: new lessons from the IARC TP53 database and genomics data. *Hum. Mutat.* **37**, 865–876 (2016).
33. Kato, S. et al. 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 (2003).
34. Mathe, E. et al. Computational approaches for predicting the biological effect of p53 missense mutations: a comparison of three sequence analysis based methods. *Nucleic Acids Res.* **34**, 1317–1325 (2006).
35. André, F. et al. AACR Project GENIE: powering precision medicine through an international consortium. *Cancer Discov.* **7**, 818–831 (2017).
36. Rideout, W. M. 3rd, Coetzee, G. A., Olumi, A. F. & Jones, P. A. 5-Methylcytosine as an endogenous mutagen in the human LDL receptor and p53 genes. *Science* **249**, 1288–1290 (1990).
37. Tornaletti, S., Rozek, D. & Pfeifer, G. P. The distribution of UV photoproducts along the human p53 gene and its relation to mutations in skin cancer. *Oncogene* **8**, 2051–2057 (1993).
38. Aguilar, F., Hussain, S. P. & Cerutti, P. Aflatoxin B1 induces the transversion of G– > T in codon 249 of the p53 tumor suppressor gene in human hepatocytes. *Proc. Natl. Acad. Sci. USA* **90**, 8586–8590 (1993).
39. Puisieux, A., Lim, S., Groopman, J. & Ozturk, M. Selective targeting of *p53* gene mutational hotspots in human cancers by etiologically defined carcinogens. *Cancer Res.* **51**, 6185–6189 (1991).
40. Bressac, B., Kew, M., Wands, J. & Ozturk, M. Selective G to T mutations of p53 gene in hepatocellular carcinoma from southern Africa. *Nature* **350**, 429–431 (1991).
41. Hsu, I. C. et al. Mutational hotspot in the p53 gene in human hepatocellular carcinomas. *Nature* **350**, 427–428 (1991).
42. Brash, D. E. et al. A role for sunlight in skin cancer: UV-induced p53 mutations in squamous cell carcinoma. *Proc. Natl. Acad. Sci. USA* **88**, 10124–10128 (1991).
43. Pierceall, W. E., Mukhopadhyay, T., Goldberg, L. H. & Ananthaswamy, H. N. Mutations in the p53 tumor suppressor gene in human cutaneous squamous cell carcinomas. *Mol. Carcinog.* **4**, 445–449 (1991).

Acknowledgements

This work was funded in part by grants from the US National Cancer Institute (U01 CA176058, U01 CA199253). T.P.H. is the recipient of training grants from the US National Institutes of Health (T32GM007753 and T32GM007226). G.G. was partially funded by the Paul C. Zamecnik Chair in Oncology from the Massachusetts General Hospital Cancer Center.

Author contributions

A.O.G. and W.C.H. designed the study. A.O.G., X.Y., and R.E.L. performed the experiments with help from T.P.H., D.Y.T., S.H.L., E.K., H.S.G., B.H., A.G., and B.F. A.O.G., X.Y., R.E.L., J.M.M., M.D., J.K., and D.E.R. analyzed the data with help from T.S., S.S., F.V., A.T., A.J.A., J.G.D., F.P., and G.G. A.O.G. and W.C.H. wrote the manuscript. D.E.R., C.M.J., M.M., and C.W.M.R. revised the manuscript.

Competing interests

M.M. is a consultant for OrigiMed and receives research support from Bayer. W.C.H. is a consultant for KSQ Therapeutics.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41588-018-0204-y.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to W.C.H.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Methods

***TP53* annotation of human cancer cell lines.** The functional and genetic *TP53* status of 966 cell lines was determined using the Cancer Cell Line Encyclopedia (CCLE; see URLs)[44], Genomics of Drug Sensitivity in Cancer (GDSC; see URLs)[45], Cancer Target Discovery and Development (CTD^2; see URLs)[46], and The Cancer Genome Atlas (TCGA accessed via the cBioPortal; see URLs)[47] databases. Cell lines were first separated into two functional classes by considering nutlin-3 sensitivity data from GDSC and CTD^2 and a p53 target gene expression signature[48] computed using CCLE data. Each cell line was provisionally considered as p53 functional if the functional score—calculated as the target genes CCLE *Z*-score – nutlin-3 CTD^2 *Z*-score – nutlin-3 Sanger *Z*-score—was > 0, and provisionally considered as p53 nonfunctional if this value was < 0. Cell lines in the p53 functional class were declared p53 WT if no *TP53* alterations were detected by CCLE, GDSC, or TCGA ($n = 252$), and discarded as ambiguous if any *TP53* alterations were found ($n = 104$). Cell lines in the p53 nonfunctional class were declared p53 mutant if any genetic *TP53* alteration was found ($n = 528$), and discarded as ambiguous if no *TP53* alterations were found ($n = 82$). The p53 mutant class was further divided into four subclasses: an LOF subclass, comprising cell lines with nonsense mutations, frameshift mutations, or homozygous deletions; a missense subclass; a splice site subclass; and an in-frame insertion/deletion subclass. Cell lines with multiple *TP53* alterations were classified using the following precedence order: missense > in-frame > splice site > LOF. Refer to Supplementary Table 1 for the full classification matrix.

Analyses of Project Achilles screening data. Pooled genome-scale LOF screens have been described[11,12]. The p53 classification scheme allowed us to annotate 348 cell lines in the RNAi data set[11] (LOF, $n = 67$; missense, $n = 150$; splice site, $n = 21$; in-frame, $n = 7$; WT, $n = 103$) and 276 cell lines in the CRISPR–Cas9 data set (LOF, $n = 45$; missense, $n = 130$; splice site, $n = 14$; in-frame, $n = 4$; WT, $n = 83$). Gene-level enrichment scores for RNAi screens were computed using DEMETER[11], an algorithm that maximizes the contributions of on-target reagents and controls for microRNA-like seed effects to reduce the contributions of off-target reagents. CERES[12], an algorithm that minimizes the contributions of copy number effects was used to compute gene-level scores for the CRISPR–Cas9 screens. PARIS (Probability Analysis by Ranked Information Score)[49], a permutation-based analytical tool, was used to identify differential gene-level enrichment scores between cell lines in the p53 WT class and cell lines in the p53 mutant class (GenePattern; see URLs). A p53 Pathway Score was generated for each cell line using either DEMETER or CERES scores using the following formula: (*MDM2* + *MDM4* + *PPM1D* + *USP7* + *UBE2D3*) – (*TP53* + *TP53BP1* + *CHEK2* + *ATM* + *CDKN1A*).

Deletion of *TP53* in A549 cells using CRISPR–Cas9. A549 cells were obtained from the American Type Culture Collection and were authenticated using Fluidigm SNP fingerprinting. Cells were not tested for Mycoplasma contamination. 2×10^5 cells were seeded in Costar 6-well dishes (Corning) in normal culture media (DMEM containing 10% fetal bovine serum, 1% penicillin-streptomycin, 2× L-glutamine) and transiently transfected with a Cas9 expression vector (pLX_TRC311) along with one of several single-guide RNA (sgRNA) expression vectors (pXPR003) at a 10:1 (w/w) ratio using the *Trans*IT-LT1 transfection reagent (Mirus) (See Supplementary Table 4); 48 h after transfection, cells were trypsinized and replated in media containing 2.5 μM nutlin-3 (Cayman Chemical). Cells were expanded under nutlin-3 selection for four weeks to enrich for cells that had deleted WT p53. Thereafter, cells were maintained in media lacking nutlin-3. Stable $p53^{NULL}$ populations were derived from three independent sgRNAs, one of which was chosen for subsequent rescue experiments and genetic screens (sgTP53-4, 5′-CCCCGGACGATATTGAACAA-3′). Using the Tracking of Indels by DEcomposition (TIDE) algorithm[49] (see URLs), we confirmed the presence of single base insertions (57.5% of sequences, $P = 0.0$), and deletions of one base (10.2%, $P = 3.5 \times 10^{-59}$), two bases (9.8%, $P = 4.4 \times 10^{-55}$), or four bases (18.3%, $P = 3.6 \times 10^{-180}$), all of which lead to premature termination codons. Loss of endogenous p53 protein expression was determined in all populations by immunoblot. Although we could have selected a single clone for these studies, the use of cell populations decreases the likelihood of studying clone-specific effects.

Immunoblots. Cells were seeded into 6-well dishes at 2×10^5 cells per well in standard culture media. Twenty-four hours later, cells were treated with nutlin-3 (10 μM) or dimethylsulfoxide (DMSO) vehicle (0.1%) (Sigma-Aldrich) and incubated for a further 24 h. Cells were then washed in ice-cold PBS buffer solution (Corning) and protein was isolated using a radioimmunoprecipitation assay buffer (Cell Signaling Technology) containing cOmplete protease inhibitors (Roche) and PhosSTOP phosphatase inhibitors (Roche). Protein was quantified by bicinchoninic acid assay (Thermo Fisher Scientific) and equal amounts of protein were heated to 95 °C in lithium dodecyl sulfate buffer (Thermo Fisher Scientific) containing 1% 2-mercaptoethanol (Sigma-Aldrich). Protein (20 μg) was loaded onto a 4–12% Bis-Tris gel (Thermo Fisher Scientific), resolved by gel electrophoresis, and transferred onto a nitrocellulose membrane using an iBlot apparatus (Thermo Fisher Scientific). Membranes were blocked in PBS and 0.05% Tween containing 10% milk (LabScientific) or 10% Bovine Serum Albumin Fraction V (Thermo Fisher Scientific) and probed with primary antibodies targeting p53 (DO-1, Santa Cruz Biotechnology), p21 (12D1, Cell Signaling Technology), or β-actin (C4, horseradish peroxidase (HRP)-conjugated, Santa Cruz Biotechnology). HRP-conjugated secondary antibodies targeting mouse or rabbit immunoglobulin G (Thermo Fisher Scientific) were used in conjunction with Western Lightning Plus-ECL reagent (PerkinElmer) to visualize protein bands.

Expression of WT or mutant p53. A construct encoding mutant p53 p.Pro278Ala was obtained from the Dana–Farber Cancer Institute (CCSB ORFeome in pDONR223). To produce a plasmid encoding WT p53, this plasmid was subjected to site-directed mutagenesis using a QuikChange Lightning II site-directed mutagenesis kit (Agilent). The open reading frames (ORFs) of both p53 variants or a *Renilla* luciferase control were cloned into the pLX_TRC313 lentiviral destination vector (Broad Institute Genetic Perturbation Platform (GPP)) using the Gateway cloning system (Thermo Fisher Scientific) and lentivirus was produced according to the Broad Institute GPP protocol (GPP; see URLs). A549 $p53^{WT}$ and $p53^{NULL}$ cells were then stably infected at an MOI < 1 and selected in hygromycin B gold (1 mg ml^{-1}) (InvivoGen). p53 protein expression and activity (p21 induction) were assessed by immunoblot.

Cell viability assays. Cells were seeded at 200 cells per well in clear-bottom, opaque-walled 96-well dishes (Costar) in 200 μl of normal culture media and allowed to adhere overnight. Cells were then treated with nutlin-3 or etoposide (Sigma-Aldrich) at 10 concentrations ranging from 20 μM to 39 nM (twofold dilutions), and incubated for 7 days with periodic visual inspection. At the end of each experiment, media was aspirated from each well and 50 μl of a 1:1 mixture of CellTiter-Glo reagent (Promega) and PBS (Corning) was added. Plates were protected from light and incubated for 20 min at room temperature before being read on a Wallac EnVision plate reader (PerkinElmer). Readings from drug-treated wells were normalized to DMSO vehicle-treated wells (set to 100% luminescence) and wells containing only media (set to 0% luminescence).

LucifeRace competition assay. A549 $p53^{WT}$ and $p53^{NULL}$ cells were stably infected at an MOI < 1 with lentivirus encoding firefly or *Renilla* luciferase under control of the human EF1α promoter (pLX_TRC313) and selected in hygromycin B gold (1 mg ml^{-1}) (InvivoGen). Cells were mixed at a 1:1 ratio and seeded at 2,500 cells per well in two replicate 96-well dishes in 200 μl of normal culture media and allowed to adhere overnight. Cells were then treated with compounds at six doses and incubated for 2 days. One plate was then subjected to a dual luciferase assay[50] and luminescence readings were obtained using a Wallac EnVision plate reader (PerkinElmer). Readings from drug-treated wells were normalized to DMSO vehicle-treated wells (set to 100% luminescence) and wells containing only media (set to 0% luminescence). Normalized luminescence values were then expressed as a ratio. The second replicate plate was passaged at a 1:4 dilution to continue the assay. Briefly, media was aspirated, 25 μl of trypsin 0.25% (Thermo Fisher Scientific) was added, and plates were incubated for 5 min. To quench the trypsin, 175 μl of culture media was added and cells were lifted and mixed by repeated pipetting; 50 μl aliquots of these suspensions were then transferred to two new replicate plates, each containing 150 μl of media and fresh compound. The process of reading and replating was repeated every two days.

Gene synthesis and cloning. The EM7 promoter and *TP53* coding sequence bearing an amber stop codon were synthesized (GenScript) in-frame with the *Sh ble* gene, which confers Zeocin resistance[51]. This construct was sequence-confirmed and cloned into pUC57-KAN (EcoRV/HindIII) (New England Biolabs) to create the entry vector backbone template pUCTP53EV.

Oligonucleotide synthesis and tile amplification. The *TP53* ORF was tiled using 150-base oligonucleotides each having 30 bases of complementary sequences flanking a 90-base variable region[13]. Oligonucleotides were synthesized (CustomArray) with all possible amino acid substitutions at each position. Adjacent tile oligonucleotides were manufactured on separate medium-density (12 K) array chips. Raw oligonucleotide tile pools were amplified by emulsion PCR (Micellula DNA Emulsion & Purification Kit, Roboklon) using primers designed to the 30-base constant sequence regions and PfuUltra II DNA Polymerase & PCR Master Mix (Agilent). Emulsions were split into seven 50 μl reactions, thermal-cycled, cleaned according to the Micellula kit instructions, and then purified on a 2% E-Gel (Thermo Fisher Scientific).

Entry vector assembly. Primers designed to anneal to the constant region for each tile and Phusion high-fidelity DNA polymerase (New England Biolabs) were used to linearize pUCTP53EV. The PCR reactions were purified on a 1% E-Gel. The DpnI-treated linear plasmid backbone was mixed with the relevant tile and assembled via multiplexed in vitro recombination[13] with the NEBuilder HiFi DNA Assembly Master Mix (New England Biolabs). The assembly reactions were purified, electroporated into TG1 *Escherichia coli* cells (Lucigen), and recovered for 1 h at 37 °C in Recovery Medium (Lucigen). Aliquots from the transformations were used to inoculate overnight cultures of lysogeny broth (LB) containing 25 μg ml^{-1} of Zeocin (Thermo Fisher Scientific) to minimize constructs harboring

frameshift mutations. Cells were collected by centrifugation and plasmid DNA (pDNA) was isolated using the *Quick*-DNA Miniprep Plus Kit (QIAGEN). Entry libraries were verified by sequencing each tile plasmid pool using Nextera XT technology (Illumina).

Expression library construction. The lentiviral vector pMT_BRD025 was developed by the Broad Institute GPP. ORFs can be cloned into this expression vector through restriction/ligation cloning. pMT_BRD025 allows for ORF expression under control of the human EF1α promoter. A *pac* gene, which confers puromycin resistance, is driven by the simian vacuolating virus 40 (SV40) promoter. To clone the *TP53*-MITE entry library into this vector, the entry library DNA was digested with restriction enzymes NheI/MluI (New England Biolabs) and ligated with the pMT_BRD025 vector that had been opened with the same two enzymes. To avoid bottlenecks in clone distribution, we sought to obtain 1,000 bacterial colonies per variant, or 8 million colonies for the entire *TP53*-MITE expression library. pDNA was extracted from the collected colonies using QIAGEN Plasmid Plus Maxi Kits. The resulting pDNA library was sequenced using Illumina Nextera XT platform to determine the distribution of variants.

Lentivirus production. Lentivirus was produced and titered by the Broad Institute GPP. The detailed protocols are available (GPP; see URLs). Briefly, 293T viral packaging cells were transfected using the *Trans*IT-LT1 transfection reagent with the pDNA library, a packaging plasmid containing the *gag*, *pol*, and *rev* genes (psPAX2, Addgene), and an envelope plasmid containing vesicular stomatitis virus G protein (pMD2.G, Addgene). Media was changed 6–8 h after transfection and the virus was collected 24 h thereafter. To measure virus titer of the library, appropriately diluted virus was used to infect A549 cells. The infected cells were then selected in puromycin and surviving cells were quantified using the alamarBlue cell viability assay reagent. The lentiviral titer was determined to be 4.4×10^5 particles ml^{-1}.

***TP53*-MITE library screen.** A549 $p53^{WT}$ and $p53^{NULL}$ cells were infected with the *TP53*-MITE library using the polybrene infection reagent (5 μg ml^{-1}) (Santa Cruz Biotechnology) in two independent experiments (4×10^7 cells mixed with 1.2×10^7 viral particles) and selected in puromycin (2 μg ml^{-1}) (InvivoGen). Cells were then split into five equal fractions: one aliquot was saved as an early time point reference; one was treated with DMSO vehicle at 0.1% v/v; one with etoposide at 5 μM; one with nutlin-3 at 2.5 μM; and one with nutlin-3 at 5 μM. Nutlin-3- and DMSO-treated cells were trypsinized, counted, and reseeded every 3 days for 12 days, whereas etoposide-treated cells were collected and counted only on day 12. Aliquots comprising approximately 8×10^6 cells were frozen in PBS (Corning) at each collection. The following samples were subjected to genomic DNA (gDNA) isolation and subsequent analysis: $p53^{WT}$early time point; $p53^{WT}$ + nutlin-3 (2.5 μM); $p53^{NULL}$early time point; $p53^{NULL}$ + nutlin-3 (5 μM); and $p53^{NULL}$ + etoposide (5 μM). gDNA was isolated using QIAamp DNA Blood Midi kits (QIAGEN). Spectrophotometric analysis of the gDNA indicated that between 2.2×10^6 and 5.5×10^6 genome equivalents were recovered from each condition.

ORF purification from gDNA. Twelve PCR reactions were performed for each gDNA sample. The volume of each PCR reaction was 100 μl and contained ~3 μg of gDNA. Herculase II (Agilent) was used as the DNA polymerase. All 12 PCR reactions for each gDNA sample were pooled, concentrated with a PCR cleanup kit (QIAGEN), loaded onto a 1% agarose gel, and separated by gel electrophoresis. Bands of the expected size were excised and DNA was purified first using a QIAquick Gel Extraction Kit (QIAGEN) then an Agencourt AMPure XP kit (Beckman Coulter).

Nextera sequencing. Sequencing sample preparation reactions were performed according to the Illumina Nextera XT protocol. For each ORF sample, we set up six Nextera reactions, each with 1 ng of purified ORF DNA. Each reaction was indexed with unique i7/i5 index pairs. After the limited-cycle PCR step, the Nextera reactions were purified with the Agencourt AMPure XP kit. All samples were then pooled and detected using an Hiseq4000 platform (Illumina) using two reads, each 150 bases in length.

***TP53*-MITE library screen data analysis.** Hiseq4000 data were processed with the ORFcall software developed by the Broad Institute[15] (See URLs and Supplementary Note) and aligned to the *TP53* reference sequence. The number of reads corresponding to each variant were then tallied. At each codon position, there were counts for all 20 amino acids and a stop codon. At 378 codon positions, there were counts for synonymous codons (silent mutations) (Supplementary Table 2). The raw read counts were normalized to the fraction of counts at each codon position. Experimental replicates yielded highly correlated read counts for each allele (Supplementary Fig. 4a,d,g). To assess the level of enrichment or depletion of each amino acid variant, we calculated the $\log_2$-fold change in fractional read counts relative to the early time point samples for all non-WT residues and synonymous variants; then, we averaged and standardized these values to produce *Z*-scores (Supplementary Table 3, Fig. 2, and Supplementary Fig. 4b,c,e,f,h,i). In the attached data files, synonymous variants are represented by the amino acid letter 'B' and stop codons are represented by the letter 'Z'. A combined phenotype score was calculated as: $(Z\text{-score}_{(p53^{WT}+\text{nutlin-3})}+Z\text{-score}_{(p53^{NULL}+\text{nutlin-3})}-Z\text{-score}_{(p53^{NULL}+\text{etoposide})})/3 \pm \text{s.d.}(+Z\text{-score}_{(p53^{WT}+\text{nutlin-3})}+Z\text{-score}_{(p53^{NULL}+\text{nutlin-3})}-Z\text{-score}_{(p53^{NULL}+\text{etoposide})})/\sqrt{3}$.

Assignment of mutation probability based on trinucleotide mutation context. Each base in the WT *TP53* complementary DNA sequence was systematically changed to every other base in silico, and the resulting sequences were translated using the standard genetic code. By considering the nucleotide before and after the altered base, each translated mutant p53 protein was assigned to its trinucleotide mutation type. The 192 trinucleotide mutation types were collapsed to 96 types by considering base changes from the perspective of the complementary strand read in the opposite direction. For example, 5′-A[G > T]C-3′ mutations were considered equivalent to 5′-G[C > A]T-3′ mutations. Mutation probabilities were then assigned to each variant using all 30 signatures in the COSMIC mutational signatures database, as well as a transcription strand-specific version of signature 4 (signature 4*, Supplementary Table 5). Mutation probabilities for nucleotide changes that yielded the same amino acid change were then added together as these events were assumed to be rare and independent in a population.

Systematic identification of tissue of origin-enriched *TP53* mutations. Enrichment analyses were performed using the IARC R18 database (see URLs)[2,32]. Fisher's exact tests were performed on each mutant allele in each tissue type that was represented by > 300 samples, and a Bonferroni correction was applied to correct for multiple hypothesis testing. Enrichments were considered significant at an α level of 0.05. These mutation type–tumor type relationships were then validated in an independent data set (AACR Project GENIE; see URLs)[35] using a nominal *P* value cutoff of 0.05.

Modeling the *TP53* mutational spectrum. We created a generalized linear model in R (3.3.3 GUI 1.69 Mavericks build) using the 'glm' function, assuming a Poisson noise distribution and a logarithmic link function. This function was used to predict the mutation count for each *TP53* allele in the IARC database using inputs for all three phenotypic screens and projections of mutational signatures 1 and 5, which are associated with aging and have been found in all tumor types, as well as signatures 2 (APOBEC), 4* (smoking), 6 (mismatch repair deficiency), 7 (UV), 13 (APOBEC), and 24 (aflatoxin), which were identified as contributing tissue of origin-selective *TP53* mutations.

To derive the models, the following three calls were used: (1) *modelSigglm* <– *IARC ~ glm(Sig1 + Sig2 + Sig4star + Sig5 + Sig6 + Sig7 + Sig13 + Sig24, family = poisson)*; (2) *modelFuncglm* <– *IARC ~ glm(Z_DN + Z_LOF + Z_WT, family = poisson)*; and (3) *modelFuncSigglm* <– *IARC ~ glm(Z_DN + Z_LOF + Z_WT + Sig1 + Sig2 + Sig4star + Sig5 + Sig6 + Sig7 + Sig13 + Sig24, family = poisson)*.

SigX represents the mutation probability for each allele under the influence of signature *X*, as defined in the COSMIC mutational signatures database, and *Z_DN*, *Z_LOF*, and *Z_WT* represent the *Z*-score for each allele in the $p53^{WT}$ + nutlin-3, $p53^{NULL}$ + nutlin-3, and $p53^{NULL}$ + etoposide screens, respectively. As outlined in Supplementary Fig. 6, each of these parameters significantly contributed to the combined model (*P* values ranged from 4.74×10^{-5} to $< 2.2\times10^{-16}$), and all parameters except *Z_WT* produced positive coefficients; this was expected given that LOF alleles were depleted in the etoposide screen. Model inputs and predictions for each amino acid variant are reported in Supplementary Table 3.

In addition to testing the predictive accuracy of the combined model on a separate database of mutations than the one used for training (GENIE)[35], we also controlled for the possibility of the models being overfitted to correlated noise in the two mutation databases by using tenfold cross-validation. Briefly, we created ten complementary data partitions (folds) using the mutation counts from the IARC R18 database. For each fold, we used 90% of the alleles to train a mutational signatures model, a phenotypic selection model, and a combined model. We then used the outputs of these models to predict the mutation counts for the remaining 10% of alleles we had set aside as a test set. We found that the contributions made by each of the input variables to the combined model were similar across all folds, as well as in the final model, and that the descriptive accuracy was similar to the predictive accuracy, suggesting that the coefficients had not been overfitted (Supplementary Fig. 6).

Reporting Summary. Further details on research design can be found in the Nature Research Reporting Summary linked to this article.

Data availability

All data sets analyzed in the current study are included in this published manuscript and its supplementary information files or can be found in the published works cited herein.

References

44. Barretina, J. et al. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **483**, 603–607 (2012).
45. Garnett, M. J. et al. Systematic identification of genomic markers of drug sensitivity in cancer cells. *Nature* **483**, 570–575 (2012).

46. Seashore-Ludlow, B. et al. Harnessing connectivity in a large-scale small-molecule sensitivity dataset. *Cancer Discov.* **5**, 1210–1223 (2015).
47. Gao, J. et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal.* **6**, pl1 (2013).
48. Jeay, S. et al. A distinct p53 target gene set predicts for response to the selective p53-HDM2 inhibitor NVP-CGM097. *eLife* **4**, e06498 (2015).
49. Cowley, G. S. et al. Parallel genome-scale loss of function screens in 216 cancer cell lines for the identification of context-specific genetic dependencies. *Sci. Data* **1**, 140035 (2014).
50. Brinkman, E. K., Chen, T., Amendola, M. & van Steensel, B. Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Res.* **42**, e168 (2014).
51. Dyer, B. W., Ferrer, F. A., Klinedinst, D. K. & Rodriguez, R. A noncommercial dual luciferase enzyme assay system for reporter gene analysis. *Anal. Biochem.* **282**, 158–161 (2000).
52. Bennett, R. P., Cox, C. A. & Hoeffler, J. P. Fusion of green fluorescent protein with the Zeocin-resistance marker allows visual screening and drug selection of transfected eukaryotic cells. *Biotechniques* **24**, 478–482 (1998).

nature research

Corresponding author(s): William C. Hahn

☐ Initial submission ☐ Revised version ☒ Final submission

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

Experimental design

1. Sample size

Describe how sample size was determined. N/A

2. Data exclusions

Describe any data exclusions. See Methods: TP53 annotation of human cancer cell lines

3. Replication

Describe whether the experimental findings were reliably reproduced. N/A

4. Randomization

Describe how samples/organisms/participants were allocated into experimental groups. N/A

5. Blinding

Describe whether the investigators were blinded to group allocation during data collection and/or analysis. N/A

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
☐	☒ A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ A statement indicating how many times each experiment was replicated
☐	☒ The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
☐	☒ A description of any assumptions or corrections, such as an adjustment for multiple comparisons
☐	☒ The test results (e.g. *P* values) given as exact values whenever possible and with confidence intervals noted
☐	☒ A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
☐	☒ Clearly defined error bars

See the web collection on statistics for biologists for further resources and guidance.

Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this Commercial and open source: R (v3.3.3 GUI 1.69 Mavericks build), GraphPad Prism

study.	(v7.0), and Microsoft Excel (v15.32) Internally developed by Broad Institute: PARIS, DEMETER, CERES, and ORFcall (see methods section for references)

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* guidance for providing algorithms and software for publication provides further information on this topic.

Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.	No unique materials were used

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).	See Methods: Immunoblots

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.	A549 cells were obtained from the ATCC
b. Describe the method of cell line authentication used.	Fluidigm SNP fingerprinting was used to authenticate A549 cells
c. Report whether the cell lines were tested for mycoplasma contamination.	Cell lines were not tested for mycoplasma contamination
d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.	No commonly misidentified cell lines were used

Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.	N/A

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.	N/A