

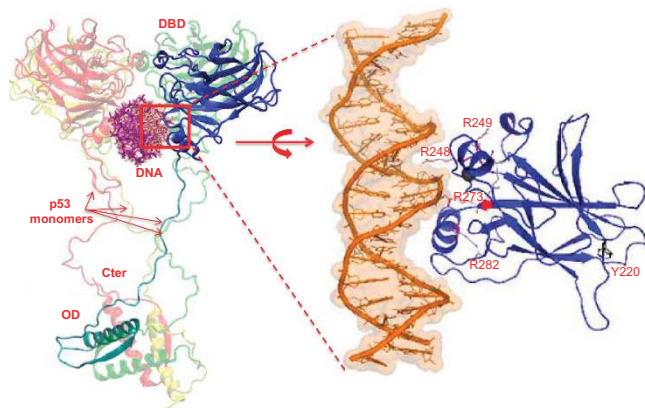
Review

Why are there hotspot mutations in the *TP53* gene in human cancers?

Evan H Baugh¹, Hua Ke², Arnold J Levine^{*3}, Richard A Bonneau^{1,4} and Chang S Chan²

The *p53* gene contains homozygous mutations in ~50–60% of human cancers. About 90% of these mutations encode missense mutant proteins that span ~190 different codons localized in the DNA-binding domain of the gene and protein. These mutations produce a protein with a reduced capacity to bind to a specific DNA sequence that regulates the *p53* transcriptional pathway. Eight of these mutations are localized in codons that account for ~28% of the total *p53* mutations and these alleles appear to be selected for preferentially in human cancers of many tissue types. This article explores the question 'Why are there hotspot mutations in the *p53* gene in human cancers?' Four possible reasons for this are considered; (1) the hotspot mutant alleles produce a protein that has a highly altered structure, (2) environmental mutagens produce allele-specific changes in the *p53* gene, (3) these mutations arise at selected sites in the gene due to a specific DNA sequence, such as a methylated cytosine residue in a CpG dinucleotide, which has a higher mutation rate changing C to T nucleotides, (4) along with the observed change in mutant *p53* proteins, which produce a loss of function (DNA binding and transcription), some mutant proteins have an allele-specific gain of function that promotes cancer. Evidence is presented that demonstrates the first three possibilities all contribute some property to this list of hotspot mutations. The fourth possibility remains to be tested.

Cell Death and Differentiation (2018) 25, 154–160; doi:10.1038/cdd.2017.180; published online 3 November 2017



Most *p53* mutations are at the DNA binding interface: why?

Graphical Abstract

Facts

- *TP53* missense mutations are the most common mutation in human cancers.
- Although missense *TP53* mutations occur at ~190 codons in the gene, eight of these mutations make up ~28% of all *p53* mutations.
- Seven of the eight mutations occur at methylated CpG sites in the gene, which encode arginine residues that contact the DNA and are conserved over evolutionary time scales.

- When all *p53* missense mutations in the DNA-binding domain are examined there is a correlation between the altered structure of the mutant protein and the frequency of *p53* mutant alleles.

Open Questions

- Some or possibly all *p53* missense mutant proteins demonstrate an ability to gain new functions. Do these gain-of-function mutations that favor a cancer phenotype contribute to the selection of some hotspot mutations?
- What properties of this gain of function are selected for in a cancerous cell?
- What is the mechanism that produces a gain-of-function phenotype?
- Not all of the methylated CpG residues in the *TP53* gene contribute equally to the hotspot mutations. The amino acids of codons where hotspot mutations occur are more conserved over evolutionary time scales than the amino acids at other positions of methylated CpG. Is the higher evolutionary conservation constrained by the importance of the amino acid or nucleotide sequence/chromatin at the positions of the hotspot mutations?
- Why is the *TP53* gene observed to be the single most common mutation in human cancers? What properties of the gene and/or protein cause this? Cancers are otherwise very heterogeneous in both the nature and the frequency of mutations found in diverse tumors.

¹New York University, New York, NY, USA; ²Cancer Institute of New Jersey, Rutgers University, New Brunswick, NJ, USA; ³Institute for Advanced Study, Princeton, NJ, USA and ⁴Simons Foundation New York, New York, NY, USA

*Corresponding author: A Levine, Simons Center for Systems Biology, Institute for Advanced Study, Bloomberg Hall, Einstein Drive, Princeton, NJ 08540, USA. Tel: +1 609 734 8005; Fax: +1 609 951 4438. E-mail: alevine@ias.edu

Received 15.5.17; revised 16.8.17; accepted 31.8.17; Edited by F Pentimalli; published online 03.11.17

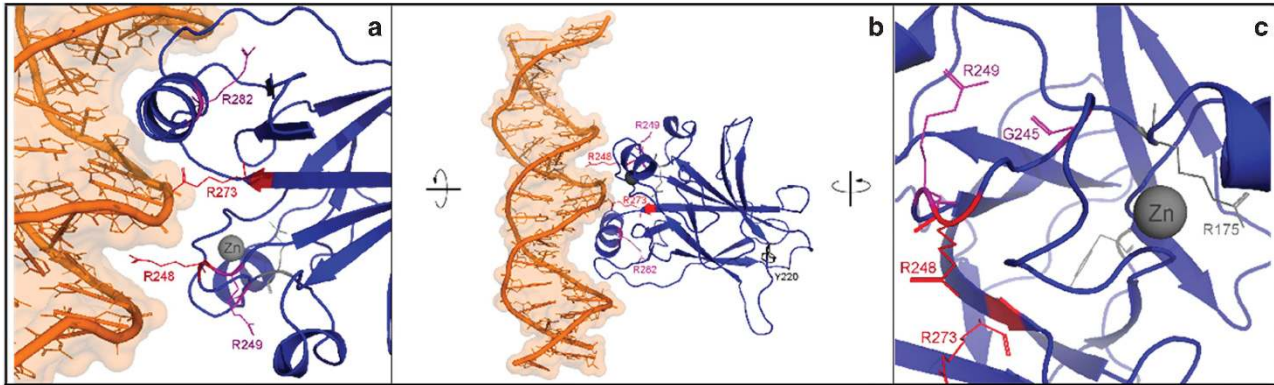


Figure 1 Most hotspot residues are near the TP53 DNA-binding interface: Several 'hotspot' residues that are frequently mutated in human cancers make contact with DNA (PDB code 1TUP, chain B). (a) R248 and R273 make direct contacts with DNA, whereas several other 'hotspot' residues are located near this interface (R249, R282). (b) Other frequently mutated positions occur far from the DNA-binding interface, such as Y220. (c) The Zinc binding site is close to the DNA-binding interface and coordinated by a loop containing R175

The p53 protein is a transcription factor that functions as a suppressor of tumor formation. The TP53 gene is the most commonly mutated gene in a wide variety of human cancers and the functions of the wild-type p53 protein are frequently compromised in many types of cancers.¹ The mutations located in the p53 gene in cancerous cells most commonly occur in its DNA-binding domain between amino-acid residues 102–292 (out of 393 amino acids in the full length protein). About 10% of these are loss of function mutations and produce no protein (through various mechanisms: nonsense or frameshift mutations, deletions), whereas the remainder include missense mutations producing a faulty protein. Loss of function includes mutations that reduce the ability of the mutant protein to bind specifically to DNA sequence motifs (20 base pairs in length) that mediate the transcription of p53 regulated genes. These missense mutations have previously been classified, based upon the structure of the p53 DNA-binding domain, into mutations that disrupt amino acids directly contacting DNA, which provide specificity to DNA binding, and mutations that alter the conformation of the DNA-binding domain (Figure 1). DNA sequencing of TP53 mutations from many diverse cancers has demonstrated that the great majority of the 190 amino acids in the DNA-binding domain are mutated in one or more cancers (and in many cases are homozygous in the tumor) and are therefore thought to contribute to the cancerous phenotypes.¹ Interestingly, these mutant alleles with different missense mutations occur over a 60-fold range of frequencies, with several of the most common p53 mutant alleles occurring at similar high frequencies independent of the tumor tissue type.¹ Table 1 provides a list of 50 of the most common missense mutations ranked by their frequencies in diverse human cancers (derived from human cancers in the IARC database R18).² The top 10 mutations (out of 190 sites) account for ~30% of all missense mutations catalogued in this collection, indicating that there are indeed preferential mutation sites found in different cancerous cells. This preference for some mutant alleles (Table 1) may arise from the selection of mutant TP53 alleles producing a protein whose structure and possibly function(s) maximally contributes to the cancerous phenotype. Several lines of evidence suggest that the different mutant

Table 1 The 50 most common somatic mutations in TP53 from IARC R18: most of the common mutations in p53 are predicted to disrupt protein structure, having highly deleterious VIPUR scores (> 0.5)

Mutant	Cases	%	VIPUR	Mutant	Cases	%	VIPUR
R175H	1215	5.6	0.742	<i>R280T</i>	108	0.5	0.907
R248Q	949	4.37	0.135	P151S	104	0.48	0.492
R273H	856	3.95	0.655	C141Y	103	0.47	0.953
R248W	765	3.53	0.185	C176Y	103	0.47	0.979
R273C	717	3.31	0.947	R158L	103	0.47	.0942
R282W	613	2.83	0.656	H193R	101	0.47	0.508
G245S	457	2.11	0.407	E286K	99	0.46	0.699
<i>R249S</i>	442	2.04	0.302	C135Y	94	0.43	0.962
Subtotal	6014	27.74		<i>P278S</i>	94	0.43	0.75
Y220C	396	1.83	0.425	V173L	93	0.43	0.62
<i>V157F</i>	213	0.98	0.737	C242F	92	0.42	0.874
C176F	206	0.95	0.982	G245C	92	0.42	0.885
M237I	198	0.91	0.811	G266E	92	0.42	0.895
E285K	186	0.86	0.654	C238Y	91	0.42	0.953
H179R	173	0.8	0.772	Y236C	89	0.41	–
Y163C	168	0.77	0.455	P152L	88	0.41	0.419
G245D	163	0.75	0.857	G245V	87	0.4	0.884
B-R273L	155	0.71	0.83	P278L	86	0.4	0.952
Y234C	145	0.67	0.131	C275Y	85	0.39	0.99
H179Y	133	0.61	0.863	H214R	85	0.39	–
B-R248L	130	0.6	0.319	A161T	83	0.38	0.561
Y205C	122	0.56	0.727	V216M	82	0.38	–
S241F	120	0.55	0.728	G266R	79	0.36	0.995
V272M	114	0.53	0.888	V173M	78	0.36	0.513
R158H	113	0.52	0.553	<i>R280K</i>	77	0.35	0.718
<i>I195T</i>	108	0.5	0.548	Total	11145	51.36	

None of the top 50 missense mutations occurs outside of the DNA-binding region. Among the top eight most common mutations, seven can be explained as positions of methylated CpG transitions due to mutation of 5'-methylated cytosine (highlighted as bold). Several of the other common mutations are likely influenced by environment mutagens (highlighted as italics)

alleles are not equal in their contributions to the production of a cancer *in vivo*. For example Rabadan and his colleagues³ have documented a case of glioblastoma that began with one p53 mutant allele, which occurs at a rare or low frequency and over the evolution of the tumor, the original mutant allele was replaced by a different p53 mutant allele that occurs more commonly in human cancers. This is consistent with some high frequency p53 mutant alleles being selected for more commonly than other low frequency mutant alleles. This

selection likely reflects the loss of function for a high efficiency of DNA binding to a p53-specific DNA sequence, and there is good evidence that different p53 mutant alleles do bind to DNA with different efficiencies; between 0 and 75% of wild type.⁴ When a variety of different p53 mutant alleles were expressed in yeast and tested for their abilities to transcribe a gene regulated by a p53 DNA-binding site, the high frequency p53 mutant alleles transcribed some genes at much lower levels than the lower frequency p53 mutant alleles.⁵ Thus, there could well be a selection pressure for p53 mutant proteins to have an altered structure and/or function that results in a loss of DNA binding to specific p53-regulatory sequences in a cell and/or loss of transcriptional functions.⁴

This review explores known TP53 missense mutant alleles that occur at high and low frequencies in human cancers in the context of high resolution models of these p53 mutant protein structures to determine the structural features of mutations selected at high or low frequencies by cancerous cells and compare these structural features to the mutant allele frequencies. The review then discusses other possible contributions to the selection of hotspot mutant alleles.

Protein Structural Alterations

Several frameworks for protein structure prediction have recently shown great success in tasks that range from *de novo* and homology-based (template-based) protein structure prediction to the design of new folds and functions.⁶ The algorithms have been tested in the context of double-blind predictions (where predictions are made before protein structure is determined)⁶ and in the context of annotating new function and novel protein variants.⁷ More recently one of these algorithms, VIPUR, which extends the Rosetta code for protein structure prediction and design, has been employed to model the structure of missense mutant proteins based upon models of the structure of the wild-type protein (initially informed by X-ray crystallography or homology modeling with the modeling of the mutation allowing small changes from the wild-type model).⁸ VIPUR has been employed successfully in the study, a large number of spontaneous missense mutations that arise in a diverse group of children with autism,⁸ demonstrating a positive correlation between a score predicting the degree of a deleterious protein structure due to mutations and the enrichment for *de novo* proband mutations. High VIPUR scores (0.8–1.0) are associated with poorly folded (or non-functional) proteins while lower VIPUR scores (0.1–0.5) are closer to a native protein structure.

Specific alleles for TP53 missense mutations may arise spontaneously and then be selected for a protein structure/function that optimizes for the development of a cancer, giving rise to different allele-specific frequencies of cancers. Higher cancer penetrance may be selected for by mutant p53 proteins with disrupted structure and function. This hypothesis provides the opportunity to test whether the structural integrity of a p53 protein with a missense mutation, predicted by its VIPUR score, is related to the frequency of different p53 mutant alleles in cancer databases? Herein, the predicted protein structures or VIPUR scores of the top 10 most frequent mutant alleles are compared with the remainder (180) of p53 mutant alleles observed at lower frequencies in cancers. This helps determine if the most common

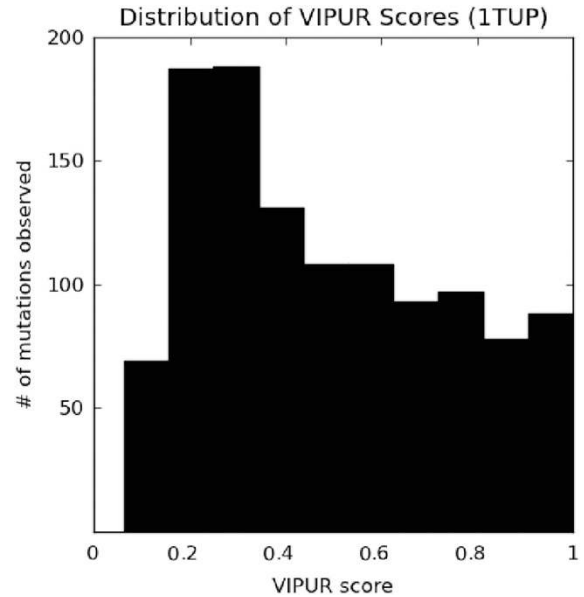


Figure 2 Most mutations in the TP53 DNA-binding domain appear neutral-like: the distribution of VIPUR scores is skewed toward neutral (<0.5) scores, suggesting that many mutations in TP53 are more wildtype-like

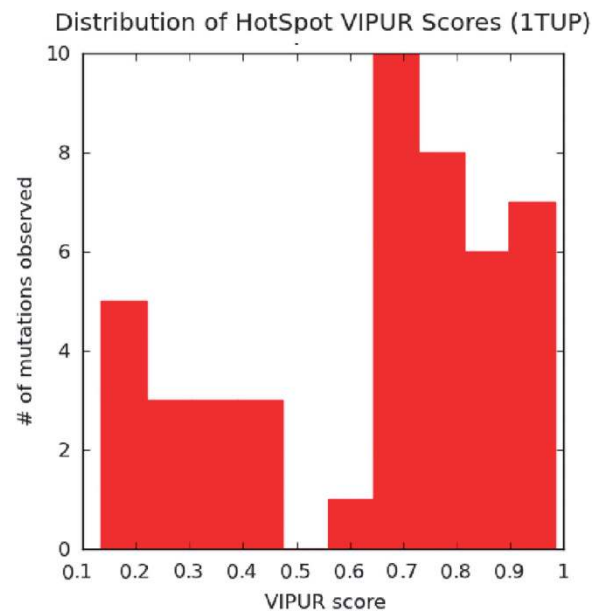


Figure 3 'Hotspot' mutations are mostly deleterious: Mutations occurring at designated 'hotspot' residues are commonly found in tumors and two-thirds of these mutants achieve highly deleterious scores in VIPUR. Eight positions were considered hotspots, with 46 mutations accessible from a single-nucleotide change to the TP53 DNA sequence. VIPUR suggests that most of these mutations above 0.7 score will be highly damaging to TP53 function

p53 mutant alleles produced proteins with significantly different structures than the rest of the p53 mutant alleles.

The COSMIC database of tumors from cancer patients contains 3686 mutants in the p53 protein with a single amino-acid change in the DNA-binding domain of the protein, all of which can be modeled to produce VIPUR scores.

The structure of the wild-type protein (PDB 1TUP, chain B) DNA-binding domain of 190 amino-acid residues was employed to calculate the changes in structure owing to these mutations. Figure 2 provides a distribution of VIPUR scores of all the TP53 missense mutations observed in the DNA-binding domain (190 codons) of the p53 protein.

This distribution of VIPUR scores for all p53 protein missense mutations in the DNA-binding domain has a bias towards lower scores (0–0.5), indicating that most p53 mutant proteins are predicted to have a more wild-type-like structure. The distribution also has a flat tail, with examples of mutant p53 proteins that are structurally disrupted to a greater extent (VIPUR scores 0.8–1.0). By contrast, Figure 3 presents the distribution of VIPUR scores for mutant missense proteins that occur at TP53 gene sites producing p53 proteins that are observed with the 10 highest frequencies in human cancers. These 'hotspot' positions form a VIPUR score distribution (Figure 3), that is, significantly different from the overall distribution of p53 mutant VIPUR scores (Figure 2, P -value of $3.32e-6$, KS test). The more common p53 mutants split into two groups, one with a p53 protein structure more similar to the wild-type molecule and another with a decidedly more destabilized structure. About two-thirds of the TP53 mutants in the top 10 frequencies are in the group with high VIPUR scores and a more disrupted structure.

Clearly, the mutant alleles that occur in cancer have a very different distribution when we subset out the positions where mutations occur at the highest frequencies from the total set of mutants. There is not a simple correlation between the poorest structural p53 proteins with high VIPUR scores and the top 10 TP53 missense mutations, based upon the frequencies of their appearance.

The biphasic nature of the VIPUR scores as a function of the COSMIC data set counts of each individual mutation in the data set (Figure 4) suggest that the mutants form two groups with very different structural effects. We therefore focused our attention on mutant alleles that are found in each of these two groups and attempted to explain why some p53 mutant alleles have structures more like wild-type molecules, whereas others were more denatured in their structure.

Figure 4 presents the distribution of all the mutations in the database (each mutation is a black dot) and the numbers of the mutants that occur at the top 10 frequencies are shown in dots at the highest COSMIC counts. All of the mutant TP53 alleles not in the top 10 frequencies (black dots) are distributed such that increasing VIPUR scores correlate with increasing log₁₀ COSMIC counts, with a Pearson correlation coefficient, $r=0.41$ and a P -value $< 1e-20$. A. Fersht and his colleagues⁴ have measured the thermal stability of the p53 wild-type protein and several mutant proteins and their ability to bind to DNA. The relative thermal stabilities and DNA-binding capabilities of different p53 mutant proteins largely agree with the structural assessment and stabilities predicted by VIPUR with the wild-type protein being the most stable, perhaps with notable exceptions at R273H and R273C. The reason for this is that VIPUR incorporates predictions of energetic stability and inter-species conservation to identify deleterious variants. Both R273C and R273H are predicted to be less stable than the wild-type protein, each obtaining VIPUR structural scores close to 0.7, an intermediate deleterious prediction. However,

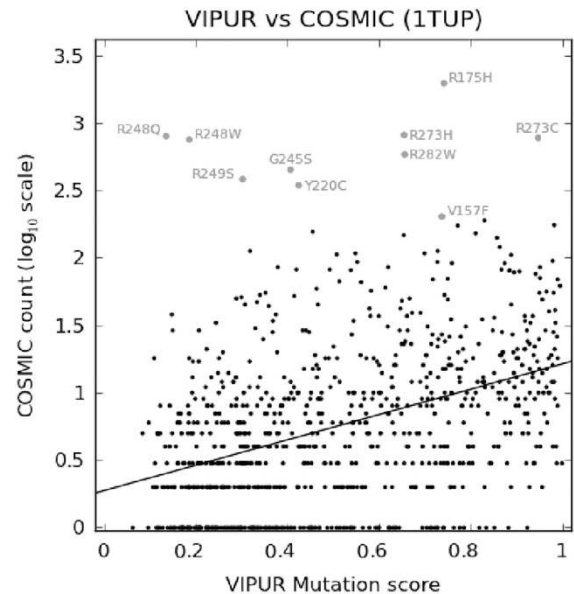


Figure 4 VIPUR deleterious scores correlate with COSMIC Counts but do not explain the outliers: A significant positive correlation exists between VIPUR scores and tumor prevalence in the COSMIC database ($r=0.41$, P -value ~ 0 , Pearson correlation with log₁₀ COSMIC counts). Although most mutations in TP53 can be described using VIPUR as structural loss of function mutations, many mutations occur much more commonly than would be expected. Curiously, the 10 most common mutations are distributed over the entire VIPUR score range. Although frequently occurring mutations with high VIPUR scores may simply be more damaging, the frequency of mutations with neutral scores (< 0.5) is not explained by VIPUR. The prevalence of these mutations may be explained by other factors

the inter-species conservation term rejects R273C specifically (This is a highly conserved region of the protein), raising its VIPUR score to 0.95, a highly deleterious prediction, suggesting that R273C will not be able to properly bind p53 DNA targets. VIPUR's predictions do not disagree so much with the measured thermal stabilities of A. Fersht and his colleagues,⁴ but demonstrate the difficulties in translating energetic stability scores into predictions of specific functional effects (like DNA binding). VIPUR as currently formulated only considers the energetic impact of a mutation in the context of the protein monomer and future improvements will allow us to include DNA–protein-binding interfaces and help to precisely identify mutations disrupting inter-molecular interactions.

Although the distribution of most low frequency p53 mutations in the COSMIC data set shows a reasonable proportionality between higher frequencies of mutant alleles in cancers and a more unfolded protein structure ($r=0.41$), the top 10 mutant alleles are distributed quite differently, with five alleles like wild-type and five alleles more unfolded. The p53 mutant alleles at residues 245, 248, 273 and 282 are all known to make contacts at or near the DNA and tend to be poor DNA-binding proteins (see Figure 1). The 157, 220 and 175 mutant alleles are farther from the DNA contact sites and tend to unfold the protein preventing proper folding and thus DNA binding. This helps to explain the biphasic character of the top frequency mutant alleles, where proteins that are folded but fail to bind DNA can have neutral (folded) VIPUR scores and proteins with variants away from the DNA-binding site are

likely to disrupt global structure (with a corresponding higher VIPUR score). The 175 mutant protein fails to bind zinc, whereas the wild-type protein has one mole of zinc per monomer protein and this helps to fold the DNA-binding domain of the protein.^{9,10} Thus, at least two classes of p53 mutant proteins are identified based upon the overall structure of the DNA-binding domain and at least two different mechanisms (DNA contact mutants and structurally unfolded mutants) contribute to the loss of function phenotype, poor or no DNA binding and no transcription.⁴ Figure 4 also demonstrates that many mutant p53 alleles occur 1–10 times in the database (see the horizontal lines of black dots with increasing numbers, 1–10, of COSMIC counts). The error rates of DNA sequencing are high and these sequences are contributed by many different laboratories; so that p53 mutant alleles observed at these very low counts should perhaps be excluded in conclusions about structural and functional trends.

Additional Mechanisms Contributing to the Different Frequencies of TP53 Mutational Alleles

Environmental mutagens and p53 mutations. A number of environmental mutagens have been identified that react with a specific set of DNA sequences, which are found in the TP53 gene. As such they could contribute to an enhanced frequency of specific p53 mutant alleles. It is common that these mutagens further show tissue specificity for the cancers they promote and that enhances the evidence for a role by these mutagens. However, the high frequencies of most of the hotspot mutations in the TP53 gene in human cancers appear to be independent of the tissue specificity observed in cancers, indicating a less important role than the mutations initiated by environmental mutagens.

An example of an environmental mutagen acting preferentially at R249S has a codon (a GC-rich region) that reacts preferentially with aflatoxin, an environmental mutagen produced by the fungus *Aspergillus*, and the R249S mutation is commonly associated with liver cancers (which are prevalent in COSMIC).¹¹ To VIPUR, R249S appears wild-type like in its structure and only modestly unfolded (0.2–0.3), yet occurs in high numbers of cancers (in the top 10), likely owing to common exposures to this mutagen (mostly in China and Africa). Importantly the R249 codon (AGG) does not contain a methylated CpG dinucleotide ruling out this source of an increased mutation frequency. Similarly benzo(a)pyrene diol epoxide, a mutagen in cigarette smoke, reacts with codon 157¹² and is a common mutation in lung cancers producing an unfolded protein according to its VIPUR score (0.6–0.7). The V157F mutant allele occurs with a 0.97% frequency in human cancers. Aristolochic acid is present in plants that are chewed and eaten by many humans. It is a mutagen that preferentially reacts with A:T base pairs producing T:A transversions and is commonly associated with urinary tract cancers at a frequency of about 0.89 percent (two different alleles) at the R280 codon, which has been identified as a sensitive site for mutation.¹³ These few examples impact modestly upon the top 30% of the hotspot mutations observed in human cancers (Table 1).

There are, however, examples of mutagens that do contribute to the formation of hotspot mutations. Solar UV light can give rise

to mutations in skin cancers (non-melanoma) at hotspot mutations R248W and R282W,¹ and tobacco smoke (benzo(a)pyrene diol epoxide) has been shown to cause some of the mutations at hotspot mutations G245V, G245C and R249M.¹ Just how much the frequencies of these mutations is due to an environmental mutagen or the presence of a methylated CpG in the TP53 gene remains to be sorted out with environmental mutagens tending to be tissue-specific, whereas methylated CpG residues appear in all or most human tissues.

Methylation of CpG residues within the TP53 gene. There are 30 CpG-associated codons in the exons encoding the DNA-binding domain of the p53 protein, all of which contain a 5'-methylated cytosine, based upon the Epigenome Roadmap, which has identified CpG positions with a methylated C residue among a large number of human tissues and cell lines. Interestingly, exon 1 and the adjacent part of intron 1 (~1.8 kilobases of DNA sequence) of the TP53 gene also have CpG dinucleotides but *none* of these have methylated cytosine residues. Table 1 shows that seven of these methylated CpGs in the DNA-binding domain are found among the 10 most common hotspot mutations and three others are found in the next group of hotspot mutations (0.35–0.95%). The 10-fold increase (or greater when tissue inflammation is involved¹) in C to T transitions that occurs when a cytosine in a CpG dinucleotide is methylated could contribute to the higher frequency of mutations at these codons and help explain the prevalence of methylated CpGs in 7 of the 10 most common p53 mutant alleles. This also helps to explain why so many of the p53 hotspot mutations alter arginine residues (top seven mutant alleles). Four of the six codons encoding arginine have CpG dinucleotides in the first two positions of their respective codons. These codons, CG (U,C,A or G in the third position), are the codons mutated in the p53 gene at the seven hotspots that encode for an arginine (Table 1). The observation that these CpG sites are methylated across diverse tissue types is consistent with the high frequency of these CpG codon mutations across diverse cancer types. This impressive correlation between the presence of methylated cytosine residues and many of the TP53 hotspot mutations also demonstrates that additional variables must account for why the other 23 methylated CpG dinucleotides (of the 30 in the gene) reside at positions in the p53 gene that do not have mutations selected at very high frequencies in human cancers.

There are at least two possible explanations for this observation that involve mutation rate and evolutionary fitness selection. Arginine has four codons encoded by CG* (where * is any of the four nucleotides). The methylated CpG leads to both a high mutation rate for C to T as well as G to A because the antisense of G is a methylated C. This makes arginine unique in that in four of its codons there are high mutation rates in both its first and second codon positions leading to altered amino acids. Amino acids encoded by *CG can have high mutation rates in the second and third position of the codon if the CpG is methylated but the third 'wobble' base usually does not alter the amino acid encoded. Similar reasoning applies to codons **C where the G is now the first base of the next codon and G** where the C is the third base of the previous codon (these codons would only have one base with high mutation

rate and should have an overall lower codon mutation rate than CG* and *CG). One exceptional codon is CGC in the context of two tandem CpGs (only two such codons are found in the DNA-binding domain). The second possible explanation is that not all amino acids in the DNA-binding domain of p53 confer the same degree of function and evolutionary fitness. An examination of amino-acid conservation at each position in the p53 DNA-binding domain (residues 102–292) shows that the seven p53 hotspot mutations all contain a conserved arginine (R) at that codon (R175H, R248Q, R273H, R248W, R273C, R282W, R249S) over large evolutionary time scales (see methods), whereas the remainder of the methylated CpG codons in the p53 gene show some variation in amino acids encoded at their positions in the TP53 gene among many diverse organisms. This suggests that the seven p53 hotspot mutations with methylated CpGs in their codons have a higher mutation rate and a greater impact upon the loss of function phenotype of the p53 protein than similar changes among the remaining 23 methylated CpG residues in the TP53 gene. In addition, nucleotide sequence contexts surrounding a CpG dinucleotide in the p53 gene and chromatin structural differences could also have a role giving rise to the differences observed in mutation frequency.

p53 Gain-of-Function Mutations

The notion that a p53 protein with a missense mutation could in fact contribute additional new functions to a cancer cell was first tested by adding a c-DNA with a TP53 missense mutation to a cell that had its p53 genes deleted (a null-p53 cell) and testing these cells for a variety of new properties (phenotypes) consistently found in independent isolates of those cells.¹⁴ The new properties conferred by the p53 missense protein were more rapid cell division in cell culture, loss of contact inhibition, growth of the cells in agar or suspension media, a higher tumorigenic potential in nude mice and larger tumors in nude mice.¹⁴ In addition, there was an impressive correlation between the presence of TP53 missense mutations in a great majority of newly immortalized cell lines (that previously had wild-type p53) produced in cell culture.¹⁵ Interestingly, these are all the phenotypes employed to recognize the transformed cell *in vitro* and *in vivo*. Over time a wide variety of pTP53 missense mutant alleles were shown to contribute to gain-of-function phenotypes, which included cellular invasion and migration, epithelial-mesenchymal transitions, spheroid disorganization, cytoskeletal alterations, wound healing, altered metabolism, drug resistance and altered transcriptional patterns.¹⁶ There have been a number of studies that have uncovered various mechanisms of action to help explain the gain-of-function phenotype. A common theme is that the mutant missense p53 protein forms a complex with a cellular transcription factor (p63, p73, vitamin D nuclear receptor, Ets) where the p53 protein contributes its active transactivation domain and the cellular transcription factor contributes its DNA-binding domain, creating a new hybrid transcription factor that alters the patterns of transcription in a cell and adds new phenotypes.^{17–20}

We do not know if all mutant p53 proteins can accomplish a gain-of-function phenotype, or if the phenotypes of different mutant p53 proteins are very different and so some are selected for more than others in human cancers, or if this depends upon the cell type or

environment in which the cancer cell finds itself (drug treatments, immunological activity, the extracellular matrix, etc).

An examination of the literature does appear to indicate that many of the p53 missense mutant proteins do confer some common phenotypes upon a variety of cells (cellular invasion and migration, cytoskeletal alterations, wound healing, metastasis and enhanced tumor size). At present the possibility that common p53 mutant hotspot mutations contribute a set of gain of new functions to a cancer cell that is preferentially selected for by the cancers than p53 mutations that occur at low frequencies, remains a reasonable hypothesis.

Conclusions

Somatic TP53 mutations in human cancers were first identified by Bert Vogelstein and his colleagues.²¹ It soon became clear that these mutations were common in many cancers and once systematic studies were carried out it was found that p53 mutations in both alleles were the most common mutational signature of all human cancers.² Among human cancers with p53 mutations, the great majority are missense mutations with identical p53 mutant alleles on all copies of chromosome 17. The first mutation appears to be spontaneous and the second identical mutation occurs through loss of the chromosome with the wild-type p53 allele and duplication of the chromosome with the mutant allele or by gene conversion of the p53 locus and selection of the mutant alleles. In this way most cancers have a mutant p53 allele (no wild-type copies) and two or more alleles in a cell are identical. Because the p53 protein is a tetramer *in vivo*, cells with a wild-type p53 protein (Li-Fraumeni patients or early forms of cancer) and a mutant p53 protein produce tetrameric proteins with various mixtures of mutant and wild-type proteins that can be poisoned for p53 transcriptional activity at high levels of mutant p53 (dominant negative mutant p53 activity). This was the reason that a mutant TP53 c-DNA acted as a dominant negative transforming gene.

In this review, the reasons why a selected group of eight TP53 missense mutations (hotspot mutations) are up to 60-fold more common than the other 182 missense mutant alleles (Table 1) was considered. Four hypotheses were explored; (1) the structure of the p53 protein produced from the top eight mutant alleles was least like the wild-type protein, (2) environmental mutagens acting at specific DNA sequences in the TP53 gene result in selection of the eight most common alleles, (3) there are 30 methylated CpG residues in exons of the DNA-binding domain of the TP53 gene and these residues have a 10-fold higher mutation rate than unmethylated CpG residues and (4) some TP53 mutant alleles produce proteins that add a 'gain-of-function phenotype' to cancerous cells that could be selected for in an allele-specific fashion. Evidence that each of the first three hypotheses contribute to the selection of the hotspot mutant TP53 alleles is presented and the fourth idea clearly requires additional experimental data to test this concept.

The protein structures produced by the hotspot alleles can be clearly divided into two groups; those mutant proteins that are more wild-type like in structure occur when a amino acid is changed that makes a contact with the DNA base in the p53 DNA-binding domain (Figure 1). Other hotspot mutations produce proteins that are more denatured and less like wild-type p53 proteins. Both hotspot mutants have in

common that they fail to bind to DNA specifically at the p53 DNA-regulatory sequence. Clearly some environmental mutagens can react with bases in the TP53 gene that produce hotspot mutations. However, in most cases this will produce the hotspot mutation in a selected tissue, so this does not explain why hotspot mutations are found in many human cancers from many tissue types. Seven of the top eight hotspot mutations derive from mutations of a methylated C residue in CpG dinucleotides. Six of codons encode for an arginine in the wild-type protein that is conserved throughout evolutionary time scales. Other methylated CpG dinucleotides in the p53 gene that are not found in hotspot alleles encode arginines that are not as well conserved in an evolutionary comparison of amino-acid changes in the p53 protein. This suggests that an enhanced mutation rate plus selection for conservation of a key amino acid drives hotspot mutations in the TP53 gene. It remains possible that several differences in the gain-of-function phenotypes of TP53 alleles are selected for by cancers, giving rise to hotspot mutations. This idea needs to be tested.

Materials and Methods

Obtaining distributions of VIPUR Scores for TP53 variants and hotspot residues. The VIPUR pipeline was run for 3686 mutants of the TP53 DNA-binding domain, making single amino-acid substitutions. We consider the distribution of VIPUR scores for 1147 variants that may arise from a single-nucleotide change to the TP53 DNA sequence. The PDB 1TUP (chain B) was used as a structural model for TP53, covering 194 residues (positions 96–289).

TP53 hotspot positions are derived from the COSMIC database, considering the positions of the 10 most common mutations in the COSMIC database. These positions include R175, V157F, Y220C, G245, R248, R249, R273 and R282.

VIPUR scores versus COSMIC counts. The COSMIC database contains 884 TP53 mutations with available count data. We removed 45 of these mutations that were not within the DNA-binding domain or contained amino-acid transitions, which required two or more nucleotide substitutions to the wild-type p53 sequence. For each mutation in the COSMIC database, we plotted the VIPUR score versus COSMIC count. The COSMIC counts vary from 1 to ~2000 and are scaled logarithmically (base 10) to demonstrate correlation with VIPUR scores.

Obtaining DNA methylation of CpG-associated codons in TP53.

For the analysis of DNA methylation, the Whole Genome Bisulfite Sequencing database was used, as provided by NIH Roadmap Epigenomics Mapping Consortium at <http://egg2.wustl.edu/roadmap/data/byDataType/dnamethylation/README>. This database provides fractional methylation calls at the CpG sites across the genome of 37 cell lines/tissues: H1 Cell Line, H1 BMP4 Derived Mesoderm Cultured Cells, H1 BMP4 Derived Trophoblast Cultured Cells, H1 Derived Mesenchymal Stem Cells, H1 Derived Neuronal Progenitor Cultured Cells, H9 Cell Line, hESC Derived CD184+ Endoderm Cultured Cells, hESC Derived CD56+ Ectoderm Cultured Cells, hESC Derived CD56+ Mesoderm Cultured Cells, HUES64 Cell Line, IMR90 Cell Line, iPS DF 6.9 Cell Line, iPS DF 19.11 Cell Line, 4star, Mobilized CD34 Primary Cells Female, Neurosphere Cultured Cells Cortex Derived, Neurosphere Cultured Cells Ganglionic Eminence Derived, Penis Foreskin Keratinocyte Primary Cells skin03, Aorta, Adult Liver, Brain Germinal Matrix, Brain Hippocampus Middle, Esophagus, Fetal Intestine Large, Fetal Intestine Small, Gastric, Left Ventricle, Lung, Ovary, Pancreas, Psoas Muscle, Right Atrium, Right Ventricle, Sigmoid Colon, Small Intestine, Thymus, Spleen. The CpG sites located inside the p53 gene were extracted with the information provided by IARC at <http://p53.iarc.fr/p53Sequences.aspx>.

Codon mutation rates versus evolutionary conservation. The mutation rates of each base along p53 amino acids were computed using the data set, 'somaticMutationDataIARC TP53 Database, R17.txt', provided by IARC at <http://p53.iarc.fr/p53Sequences.aspx>. There were totally 29 711 mutation records. In the DNA-binding domain of p53, 30 amino acids are associated with CpG sites and they are divided into five groups according to the locations of CpG sites: CG*, *CG, **C, G** and CGC. For example, only amino acid 107 (Tyr: TAC) is classified as **C

group; and amino acids 125, 152, 170 and 222 is grouped as *CG. The mutation rate of a codon is the sum of all mutations in that codon.

The degree of evolutionary conservation of amino acids in p53 was estimated using ConSurf.²² The data set for the calculation of conservation scores is comprised of 33 p53 amino-acid sequences, including human, mouse, rat, sheep, pig, rabbit, cow, cat, channel catfish, blind subterranean mole rat, chicken, Chinese hamster, common tree shrew, Congo puffer, crab eating macaque, dog, European flounder, golden hamster, green monkey, guinea pig, Japanese macaque, Mongolian jird, rainbow trout, rhesus macaque, southern platyfish, swordtail, western clawed frog, woodchuck, zebrafish, zebu, African clawed frog, barbel and beluga whale, which are provided at <http://p53.bii.a-star.edu.sg/analysis/aaspecConsV/index.php>.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgements. This work was partly supported by NIH grant Award 2P01CA087497-16A1.

- Hainaut P, Pfeifer G. Somatic TP53 Mutations in the Era of Genome Sequencing. In: Lozano G, Levine AJ (eds). *The p53 Protein, From Cell Regulation to Cancer*. 1st edn, Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, USA, 2016 pp 279–300.
- Bouaoun L, Sonkin D, Ardin M, Hollstein M, Brynes G, Zaudail J *et al*. TP53 variations in human cancers: new lessons from the IARC database and genomics. *Hum Mutat* 2016; **37**: 865–876.
- Lee J-K, Wang J, Sa JK, Ladewig E, Lee H-O, Lee I-H *et al*. Spatiotemporal genomic architecture informs precision oncology in glioblastoma. *Nat Genet* 2017; **49**: 594–599.
- Joerger AC, Fersht AR. The tumor suppressor p53: from structures to drug discovery. In: Levine AJ, Lane DP (eds). *The p53 Family*. 1st edn, Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, USA, 2016 pp 19–38.
- Campomenosi P, Monti P, Aprile A, Abbondandolo A, Frebourg T, Gold B *et al*. P53 mutants can often transactivate promoters containing a p21 but not Bax or PIG3 responsive elements. *OncoGene* 2001; **20**: 3573–3579.
- Leaver-Fay A, Tyka M, Lewis SM, Lange OF, Thompson J, Jacak R *et al*. ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. *Methods Enzymol* 2011; **487**: 548–574.
- Moult J, Fidelis K, Kryshchukovych A, Schwede T, Tramontano A. Critical assessment of methods of protein structure predictions (CASP) -round x. *Proteins* 2014; **82**: 1–6.
- Baugh E, Simmons-Edler R, Muller CL, Alford RF, Volfovsky N, Lash AE *et al*. Robust classification of protein variation using structural modelling and large-scale data integration. *Nucleic Acids Res* 2016; **44**: 2501–2513.
- Yu X, Vazquez A, Levine A, Carpizo DR. Allele-specific p53 mutant reactivation. *Cancer Cell* 2012; **21**: 614–625.
- Blanden AR, Yu X, Loh SN, Levine AJ, Carpizo DR. Reactivating mutant p53 using small molecules as zinc metallochaperones: awakening a sleeping giant in cancer. *Drug Discov Today* 2015; **20**: 1391–1397.
- Bressac B, Kew M, Wands J, Ozturk M. Selective G to T mutations of p53 gene in hepatocellular carcinoma from southern Africa. *Nature* 1991; **350**: 429–432.
- Denissenko M, Pao A, Tang M, Pfeifer G. Preferential formations of Benzo[a]pyrene adducts at lung cancer mutational hot spots in p53. *Science* 1996; **274**: 430–432.
- Hollstein M, Moiyra M, Grollman AP, Oliviera M. 2013. Analysis of TP53 mutation spectra reveals the fingerprint of the potent environmental carcinogen, aristolochic acid. *Mutat Res* 2013; **753**: 41–49.
- Dittmer D, Pati S, Zambetti G, Chu S, Teresky AK, Moore M *et al*. P53 gain of function mutations. *Nat Genet* 1993; **4**: 42–46.
- Harvey D, Levine AJ. p53 alteration is a common event in the spontaneous immortalization of primary Balb/c murine embryo fibroblasts. *Genes Dev* 1991; **5**: 2375–2385.
- Muller P, Vousden KH. P53 mutations in cancer. *Nat Cell Biol* 2012; **15**: 2–8.
- Lozano G, Levine A (eds). *The p53 protein from cell regulation to cancer*. In: *A subject collection from Cold Spring Harbor Perspectives in Medicine*. CSHL Press: NY, 2016.
- Hanel W, Marchenko N, Xu S, Weng W, Moll U. Two hot spot mutant p53 models display differential gain of function in tumorigenesis. *Cell Death Differ* 2013; **20**: 898–909.
- Alexandrova E, Mirza S, Xu S, Schulz-Heddergott R, Marchenko N, Moll U. p53 loss of heterozygosity is a necessary prerequisite for mutant p53 stabilization and gain of function in vivo. *Cell Death Dis* 2017; **8**: e2661.
- Alexandrova E, Yallowitz A, Li D, Xu S, Schulz R, Proia D *et al*. Improving Survival by exploiting tumor dependence on stabilized mutant p53 for treatment 2015 *Nature* **523**: 352–356.
- Baker SJ, Fearon ER, Nigro JM, Hamilton SR, Preisinger AC, Jessup JM *et al*. Chromosome 17 deletions and p53 gene mutations in colorectal carcinomas. *Science* 1989; **244**: 217–221.
- Ashkenazy H, Abadi S, Martz E, Chay O, Mayrose I, Pupko T *et al*. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res* 2016; **44**: W344–W350.