TP53 Cancerous Mutations Exhibit Selection for Translation Efficiency

Yedael Y. Waldman,[1] Tamir Tuller,[1,2,3] Roded Sharan,[1] and Eytan Ruppin[1,3]

[1]Blavatnik School of Computer Science, [2]Department of Molecular Microbiology and Biotechnology, and [3]School of Medicine, Tel Aviv University, Ramat Aviv, Israel

Abstract

The tumor suppressor gene *TP53* is known to be a key regulator in cancer, and more than half of human cancers exhibit mutations in this gene. Recent evidence shows that point mutations in *TP53* not only disrupt its function but also possess gain-of-function and dominant-negative effects on wild-type copies, thus making the mutated gene an oncogene. Hence, this brings about the possibility that *TP53* mutations may be under selection for increasing the overall translation efficiency (TE) of defected *TP53* in cancerous cells. Here, we perform the first large-scale analysis of TE in human cancer mutated *TP53* variants, identifying a significant increase in TE that is correlated with the frequency of *TP53* mutations. Furthermore, mutations with a known oncogenic effect significantly increase their TE compared with the other *TP53* mutations. Further analysis shows that TE may have influence both on selecting the location of the mutation and on its outcome: codons with lower TE show stronger selection toward nonsynonymous mutations and, for each codon, frequent mutations show stronger increase in TE compared with less frequent mutations. Additionally, we find that *TP53* mutations have significantly higher TE increase in progressive versus primary tumors. Finally, an analysis of *TP53* NCI-60 cell lines points to a coadaptation between the mutations and the tRNA pool, increasing the overall *TP53* TE. Taken together, these results show that TE plays an important role in the selection of *TP53* cancerous mutations. [Cancer Res 2009;69(22):8807–13]

Introduction

The tumor suppressor gene *TP53* is known to be a key regulator in cancer, and it is estimated that mutations in the sequence of the gene occur in more than half of human cancers (1, 2). These mutations were shown to be related to prognostic features in several cancers (1, 3, 4). Being a known tumor suppressor (5), one would expect cancerous mutations would decrease the levels of *TP53* either by diminishing protein synthesis or by producing a truncated product. However, >75% of *TP53* alterations are missense point mutations that lead to the synthesis of a stable full-length protein (4, 6). Moreover, these mutated p53 proteins are often expressed in higher levels than those of the wild-type p53 (7). These intriguing results suggest that missense mutations in *TP53* may not only lead to loss-of-function but may also be related to a potential gain-of-function (GOF), contributing to cancer progression by effectively endowing the mutant *TP53* with oncogenic functions. In line with this possibility, it has been shown that some of these mutations have a dominant-negative effect (DNE) on the wild-type p53, suppressing its activity (8–11). *TP53* mutations may result in the gain of an array of different functions, including regulation of gene expression, cell growth, cell death, and resistance to chemotherapeutic agents (4, 7, 12–14). These new functional roles of mutated p53 proteins may imply that *TP53* mutations are subjected to selection forces in cancer. Indeed, it is known that, in cancerous *TP53*, the rate of nonsynonymous mutations (compared with synonymous mutations) is significantly higher than expected, testifying to a positive selection (15). In addition, the distribution of mutations shows that some mutations are more frequent than others, with several hotspots, and that mutations tend to occur in conserved sites (15).

Another possible selection force may be toward translation efficiency (TE). As each codon is read by a different set of tRNAs, codons with more abundant corresponding tRNAs are likely to be translated more quickly. Therefore, synonymous and nonsynonymous codon usage affect translation elongation rate and therefore the TE of the gene. Numerous studies have shown that the codon bias observed in different organisms plays a significant role in TE, with highly expressed and functionally important genes showing a higher level of coadaptation between the codon usage and the tRNA pool (16, 17). Analyzing >10,000 human genes, we have recently shown that codon bias plays significant role in TE in humans and other mammals as well (18). Further support to the significance of codon bias is found in heterologous protein expression experiments: in these studies, different organisms display different codon biases as well as different tRNA pools. Various small-scale studies have shown that expressing a foreign gene in different organisms (including human cells) while fitting its codons to the tRNA pool of the host organisms (without changing its protein sequence) can increase its protein levels by several orders of magnitude (Table 1 in ref. 19).

TP53 point mutations change the corresponding codon and therefore may have effect on its TE. In the current study, we aim to examine whether cancerous mutations observed in *TP53* do significantly increase its TE. Several previous studies of TE and post-transcriptional regulation in different malignancies focused on mutations altering regulatory sequences affecting mRNA stability and translation (20–22). However, they did not analyze how cancerous changes in TE are related to altered codons in the sequence of the gene. This study, focusing on *TP53*, is the first to study the effects of mutations on the TE of cancer genes on a large-scale manner.

Materials and Methods

Measuring TE. TE for each codon was measured as in ref. 23 (Supplementary Table S1). For that purpose, tRNA genomic copy numbers were downloaded from the Genomic tRNA Database (24). To measure the effect

Note: Supplementary data for this article are available at Cancer Research Online (http://cancerres.aacrjournals.org/).

Y.Y. Waldman and T. Tuller contributed equally to this work.

Requests for reprints: Tamir Tuller, Tel Aviv University, Tel Aviv, 69978 Israel. Phone: 972-3-640-7466; Fax: 972-3-640-9357; E-mail: tamirtul@post.tau.ac.il and Eytan Ruppin, Tel Aviv University, Tel Aviv, 69978 Israel. E-mail: ruppin@post.tau.ac.il.

doi:10.1158/0008-5472.CAN-09-1653

Table 1. Tissue-specific TE comparison

Code	Tissue	TE increase (mean)	Cancerous dN/dS (mean)	*P*
C00-C14	Head and neck, not otherwise specified	1.439	4.653	
C44	Skin	1.405	1.641	0.676
C16	Stomach	1.309	3.814	0.150
C71	Brain	1.29	2.559	0.137
C20	Rectum	1.281	1.081	0.192
C15	Esophagus	1.275	6.722	0.084
C18-C20	Colorectum, not otherwise specified	1.266	2.777	0.064
C67	Bladder	1.25	1.779	0.140
C18	Colon	1.239	2.88	0.078
C61	Prostate gland	1.233	5.03	0.078
C34	Bronchus and lung	1.232	0.5	0.016
C22	Liver and intrahepatic bile ducts	1.222	0.785	0.045
C32	Larynx	1.218	3.75	0.262
C50	Breast	1.203	0.83	**0.005**
C06	Other and unspecified parts of mouth	1.181	0.569	0.024
C56	Ovary	1.152	1.336	**0.002**
C42	Hematopoietic and reticuloendothelial systems	1.124	1.375	0.033

NOTE: For each tissue, we downloaded from ref. 25 data on mutations found in primary tumors in that tissue. Wilcoxon test was used to assign *P* values for the differences between the TE of the tissue with the highest mean to that of all other tissues. Only tissues with at least 100 nonsynonymous mutations are shown in the table. Code refers to the tissue's code of *International Classification of Diseases for Oncology, Third Edition* (49). Significant results after correcting for multiple hypotheses (false discovery rate) are boldfaced.

of a point mutation on TE, we calculated the ratio between the TE of the mutated codon and that of the wild-type codon.

Mutational data. *TP53* point mutations data were downloaded from the IARC *TP53* database, Release R13 (ref. 25; Supplementary Table S2).

DNE and GOF mutations were defined based on the file "TP53MUTFunction1R13.TXT" (Supplementary Tables S3 and S4). For each functional group (DNE and GOF), we evaluated the significance of their median TE by comparing it with those of 100,000 same-sized mutation sets, randomly chosen from the set of all mutations.

Cancer progress state for each mutation was defined based on the file "TP53SomaticR13.txt." We compared the two groups of mutations by randomly selecting 100,000 sets from the larger group (primary) with the same size as the smaller group (progressive) and compared the median of the two sets.

For tissue-specific mutations, we have focused on primary tumor mutations and divided them into tissues according to the data provided in the database.

NCI-60 analysis. Nonsynonymous point mutations in NCI-60 were taken from Ikediobi and colleagues (ref. 26; Supplementary Table S5). Copy number variation of genomic locations was obtained from Barrett and colleagues.[4] Based on these data and the tRNAs loci (24), we have computed the tRNA pool of each cell line (Supplementary Table S6).

Evolutionary dN/dS. Positional nonsynonymous and synonymous substitution ratios (dN/dS) along *TP53* sequence were computed using Selecton (27) based on orthologues of 10 mammals (Fig. 1). Orthologous groups were downloaded from BioMart (28). Supplementary Table S7 and Fig. 2 show these dN/dS ratios.

Cancerous dN/dS. Similar to evolutionary rate (ER) measure, we define a measure for selection toward nonsynonymous versus synonymous mutations in cancerous *TP53*. Let *s* and *ns* be the number of synonymous and nonsynonymous point mutations, respectively, reported for a codon. Let *S* and *NS* be the number of different synonymous and nonsynonymous possible point mutations in that codon, respectively. Cancerous dN/dS of a codon is defined as $(ns/NS)/(s/S)$. Higher values of this measure imply selection toward nonsynonymous mutations, which may affect the protein structure and therefore its function. Supplementary Table S7 and Fig. 2 show cancerous dN/dS ratios.

Background model and significance test. Two background models were used to assess the significance of our results. Both models preserve the position distribution of the point mutations along the gene's sequence as reported in the database but differ in the nucleotide substitutions frequencies (e.g., from A to G). In the first model, these frequencies were estimated from *TP53* mutation data, whereas, in the second model, we used substitution frequencies based on analysis of human pseudogenes (29). For each model, we randomly sampled 18,926 point mutations (as in the original data set).

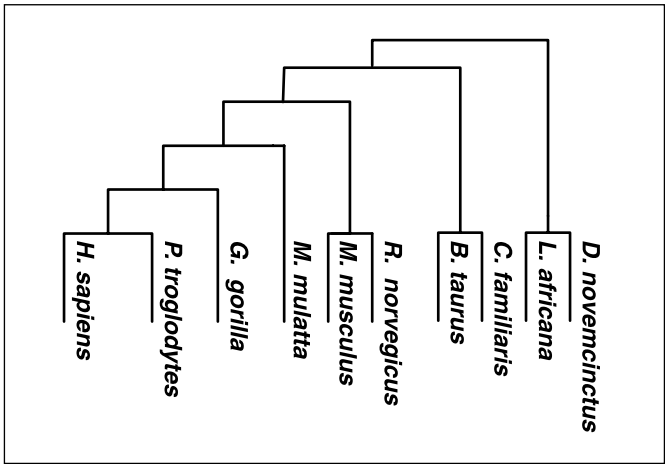

Figure 1. The phylogeny of the organisms used for estimating *TP53* ER (based on National Center for Biotechnology Information taxonomy; http://www.ncbi.nlm.nih.gov/Taxonomy/CommonTree/wwwcmt.cgi).

[4] M.T. Barrett, A. Ben-Dor, A. Tsalenko, et al. Profiling copy number aberrations in the NCI-60 cell line panel by oligonucleotide array CGH, submitted for publication.

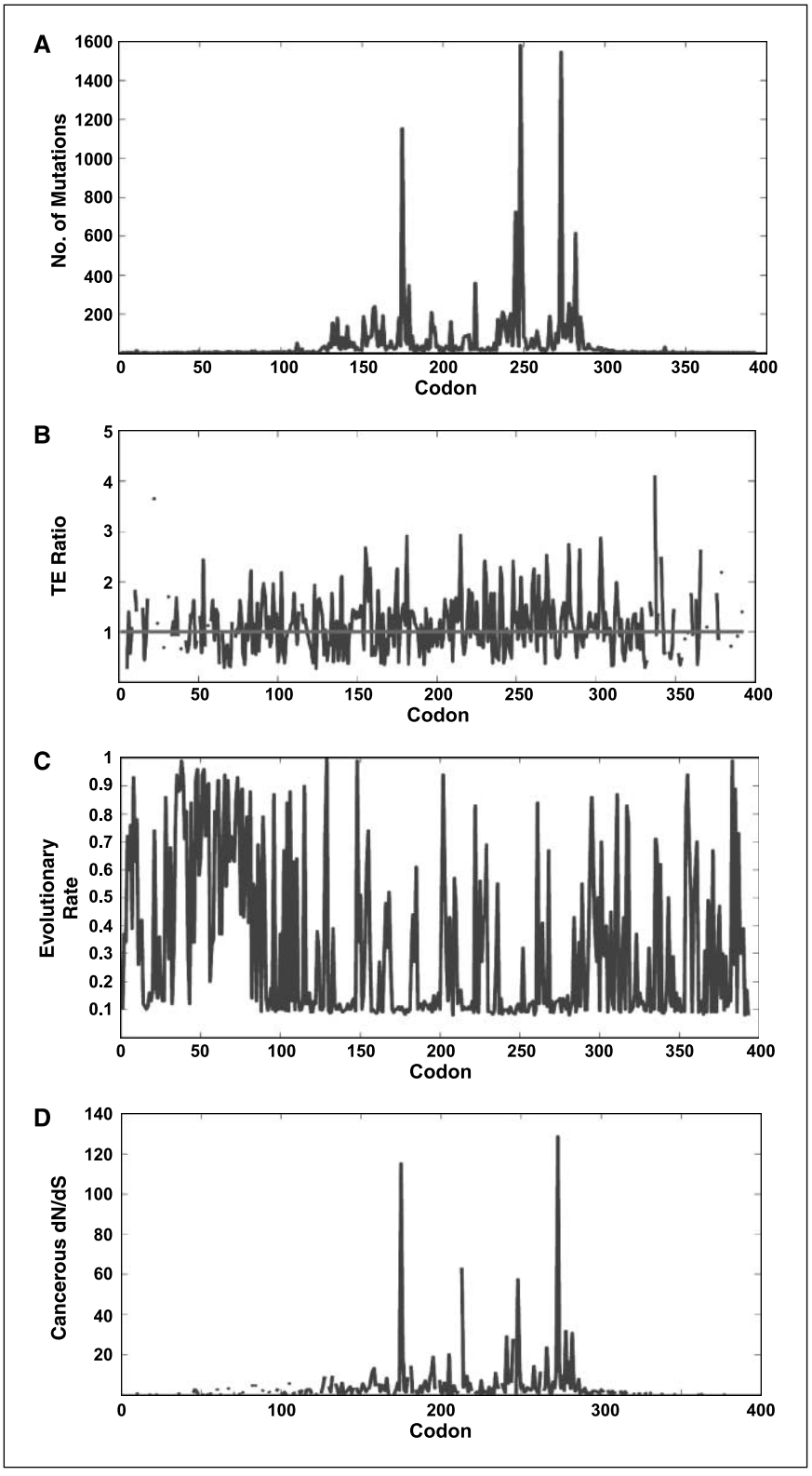

Figure 2. Various features examined across the *TP53* coding sequence. For each codon (*X axis*), we calculated the following measures (*Y axis*). *A,* number of nonsynonymous mutations; nonsynonymous mutations are found in 330 of the 393 codons of *TP53*. Most mutations are located in the DNA-binding domain (codons 102-292). *B,* TE ratio (the ratio of TE of the mutated codon and the wild-type codon). When several different nonsynonymous mutations were reported for the same codon, we took the mean TE ratio, weighted by their frequency. The horizontal line, indicating TE ratio of 1, is used as a reference. *C,* ER values calculated using Selecton (27) based on 10 orthologues. *D,* cancerous dN/dS ratio. This ratio measures positive selection toward nonsynonymous mutations versus synonymous mutations (see Materials and Methods).

Next, we compared the TE of the original data set to those of the background models. An optimal examination should compare, for each codon and mutation, mutations that result in the same amino acid substitution (e.g., from aspartic acid to valine), thus preserving the functionality of the mutated codon. However, the nature of the genetic code is such that, for most amino acid substitutions, there is not more than one possible point mutation leading to it. Therefore, such a comparison would not be applicable for most mutations. Instead, we compared between mutations resulting in similar (but not necessarily identical) amino acids. Two amino acids were defined as similar if their distance was below a certain threshold; to

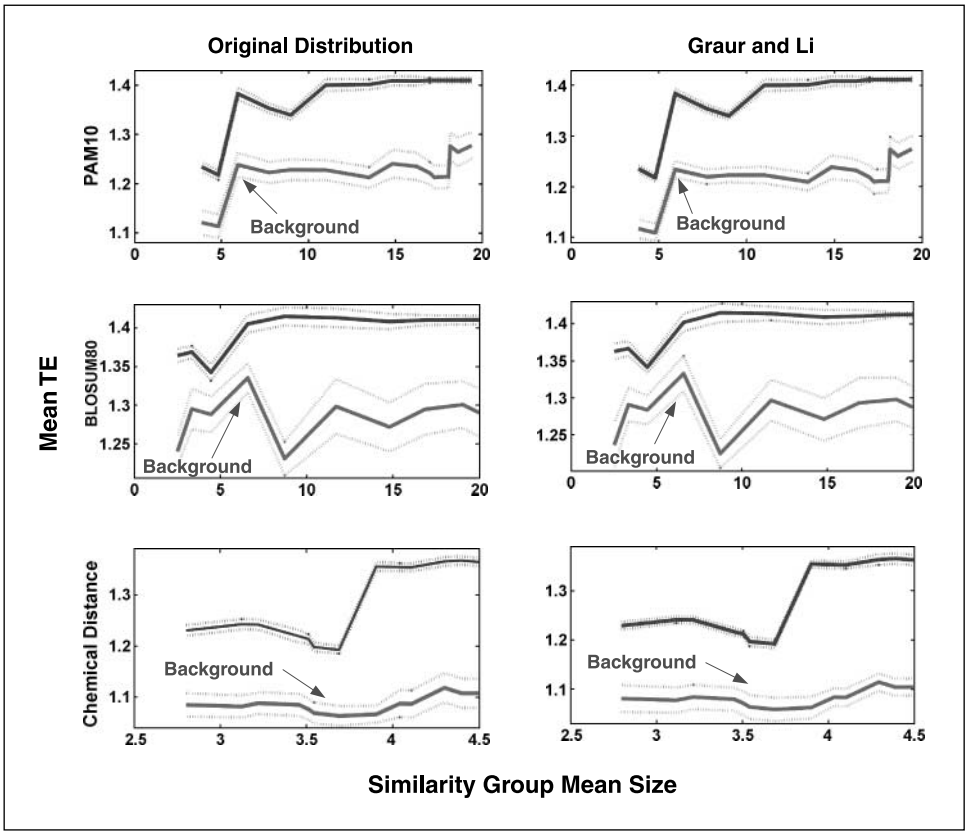

Figure 3. Comparison between observed *TP53* mutations and background models: assessing the significance of TE increase in *TP53* mutations, we used two background models based on different distributions of mutations [original distribution and Graur and Li (29); see Materials and Methods]. Each mutation was compared to a similar mutation in the background model. Similarity between amino acids was defined according to amino acid distance matrix and a threshold. For the two background models, we used three amino acid distance matrices (PAM10, BLOSUM80, and chemical distance), resulting in the six plots depicted in the figure. Each threshold in the amino acid similarity matrices defines groups of amino acid substitution that can be compared with each other (similarity groups). Tighter thresholds define smaller similarity groups. These graphs plot the mean TE (*Y axis*) as a function of the mean size of the similarity groups (*X axis*; defined by the threshold) for both the observed data and the background model. The dashed lines represent 2 SDs from the mean TE. In all cases, the TE ratio is significantly higher than the background models ($P < 10^{-4}$, Wilcoxon test). For additional details, see Materials and Methods.

this end, we used several distance matrices for amino acids from the literature: PAM10 (30), BLOSUM80 (ref. 31; both downloaded from the National Center for Biotechnology Information Blast FTP site),[5] and Grantham chemical distance (ref. 32; downloaded from the AAindex database; ref. 33). Lowering the similarity threshold results in less comparable random mutations for each mutation and therefore decreases the number of reported mutations for which we can assess significance. We used thresholds that allowed a comparison for at least half of the reported mutations in the database. For each comparable mutation, we calculated its TE ratio (the ratio of the TE obtained when considering the cancerous mutation versus the wild-type) and the mean TE ratio for similar mutations in the same codon in the background model (the ratio obtained when mutating from the wild-type codon to the functionally similar one in the background model). Concatenating these values (weighted by their frequencies), we obtained two vectors of TE ratios (one for the real data and one for the background model). Using a Wilcoxon test, we calculated *P* values for the significance of the difference between the two distributions. We repeated the analysis 20 times (each time a different set of random mutations was generated), thus obtaining data on the variance of the background models. Figure 3 also shows some small variance for the IARC *TP53* data set itself as in different background models; slightly different subsets of mutations were compared.

TE selection. For each codon, we tested whether the mutation with the highest TE is also the one that is most frequent. For mutation *a*, let *TE(a)* and *C(a)* be the TE and the number of reported occurrences, respectively, of *a*. For any codon *i*, let M_i be the mutation of codon *i* with the highest TE. For any other mutation m_i in codon *i*, we calculated two ratios: $TE(M_i)/TE(m_i)$ and $C(M_i)/C(m_i)$. We calculated the overall Spearman correlation between these two ratios for all codons (and, similarly, for all codons with >100 mutations).

[5] ftp://ftp.ncbi.nih.gov/blast/matrices

Results

Mutational effects on *TP53* TE are selected for. We have downloaded *TP53* mutation data from the IARC *TP53* database (25). The database contains data for 1,710 different reported somatic mutations. As each mutation may be reported more than once in the database, there are 18,926 point mutations in total. The vast majority of these mutations are nonsynonymous, composing 1,342 (78.5%) different mutations and a total of 17,851 (94.3%) mutations.

Focusing on *TP53* nonsynonymous somatic mutations that are more likely to be under positive selection pressures, we asked how these mutations affect *TP53* TE (see Supplementary Note 1 for analysis of synonymous somatic mutations). Thus, we adopted the approach that selection for TE can be obtained not only by optimizing the codon usage of the same amino acid but also by considering substitutions to functionally similar amino acids that are translated more efficiently (34, 35). Figure 2*A* shows the distribution of nonsynonymous mutations along the *TP53* sequence.

To measure TE, we used the tRNA adaptation index (24). In tRNA adaptation index calculation, each codon is assigned a normalized value, which is based on the genomic copy numbers of the corresponding tRNAs available for its translation and their affinity. Thus, this value reflects the efficiency of translation for each codon. Next, the tRNA adaptation index measure is calculated for each gene based on its codon composition. The tRNA adaptation index measure was shown to be highly correlated with protein abundance in *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe* (35) and was previously used to analyze TE both in yeast species (17) and in humans (18). Following these studies, we have used the value assigned to each codon in the tRNA adaptation index calculation as a measure for its TE. Notably, we find the

TE measure to correspond to known *in vivo* measurements of translation rate of specific codons made in *Escherichia coli*. For example, the TE ratio between GAA and GAG codons was 3.125 compared with a ratio of 3.34 reported experimentally (36).

We measured the ratio between the TE of the mutated codon and the wild-type codon for each of the 17,851 mutations reported in the database. We found that the mean ratio is 1.412, indicating an overall increase in TE for these mutations (see Fig. 2*B* for TE ratios along the *TP53* sequence). Further, there is a low but significant correlation between the increase in TE of the mutation and the number of its occurrences in the database ($R = 0.0706$; $P = 0.0097$). Similar signal was observed when measuring the correlation between the mean TE ratio of a codon and the number of mutations reported for that codon ($R = 0.113$; $P = 0.039$). In addition, when we divide the mutations according to their TE change, we find a significant difference in frequency. For mutations with decreased TE (TE ratio below 1), the mean frequency was 9.65 counts per mutation, whereas, for mutations with increased TE (TE ratio above 1), the mean frequency was almost twice as high: 17.85 ($P = 4.71 \times 10^{-5}$, Wilcoxon test). To evaluate the statistical significance of the TE ratio results reported above, we compared our results to background model (Materials and Methods). As can be seen in Fig. 3, the increase in TE was highly significant compared with the random models. Interestingly, the results reported above were not observed in germ-line mutations, where selection forces are less direct and hence expected to be lower (Supplementary Note 2).

Additionally, we found that mutations with higher TE are more frequent: there is a positive correlation between the mutation with the highest TE for each codon and its frequency (relative to the frequency of the other mutations in that codon; $R = 0.0714$; $P = 0.0231$; Materials and Methods). When focusing on hotspots, codons with >100 reported mutations (which comprise >70% of the reported mutations), where selection forces are presumably stronger, the correlation increases ($R = 0.2283$; $P = 4.31 \times 10^{-4}$). To further study the potential role of TE in selecting *TP53* mutations, we next examined whether positions with relatively lower TE are mutated more frequently to increase their TE. Examining all mutations, we did not find a significant correlation between the TE of the wild-type codon and the number of mutations reported in it. Yet, when focusing on hotspots, we found a significant negative correlation ($R = -0.4325$; $P = 0.0042$). Furthermore, we found a significant negative correlation between the TE of the codon and the ratio of nonsynonymous to synonymous mutations (cancerous dN/dS ratio; Fig. 2*D*; Materials and Methods; $R = -0.169$; $P = 0.013$). The correlation was much more significant for hotspots ($R = -0.5$; $P = 0.0019$). These results show that sites with relatively low TE were under stronger selection for nonsynonymous mutations.

Next, we calculated the TE ratios of mutations that are known to have either a DNE or a GOF effect on p53 and therefore are presumably selected for. Functionality data were available from the IARC *TP53* database for 49 DNE mutations and 92 GOF mutations (Supplementary Tables S3 and S4). The mean TE ratio was significantly higher for both sets of mutations than for non-DNE/non-GOF mutations (empirical P values = 0.0424 and 0.00002 for DNE and GOF mutations, respectively). Although both results were significant, GOF mutations were more significantly separated from the other mutations.

TE and cancer progress. A plausible assumption is that as cancer progresses, the mutations it acquires and retains are more beneficial to the survival of the cancer cells. Therefore, if TE is under selection, *TP53* mutations in progressive cancer should have a higher TE. To test this hypothesis, we collected data on 11,422 (1,419 unique) mutations in primary tumors and 755 (329 unique) mutations in recurrent, secondary, and metastasis tumors (Materials and Methods). We found that although in both cases there is an increase in TE (compared to random), the increase in progressive tumors is significantly higher than primary cancers (mean, 1.503 versus 1.417 in progressive versus primary tumors, respectively; $P = 3 \times 10^{-5}$). The significant difference between the TE increase in the different stages of tumor progress strongly testifies to the selective advantage that TE increase has in emerging *TP53* mutations.

***TP53* exhibits different levels of selection for TE in different tumors.** Previous studies have shown that tumors from different tissues exhibit different *TP53* mutation patterns (4, 37, 38). To study whether we can find tissue-specific differences in the TE changes, we have compiled a list of mutations from primary tumors from different tissues. Although *TP53* exhibited an increase in TE in all tissues examined, its TE ratios varied between 1.12 and 1.44 in different tissues, with significant differences between several tissues (Table 1; Supplementary Table S8). These results suggest that the origin of the tumor might have an effect on the strength of TE selection. In addition, we found a marked correlation between each tissue's mean cancerous dN/dS (Materials and Methods) and its mean TE ($R = 0.4975$; $P = 0.0442$). However, we should note that, for several of the tissues, the dN/dS ratio was only available for a relatively small number of codons (Supplementary Table S8).

***TP53* point mutations and chromosomal aberrations modifying the tRNA pool.** Many cancers exhibit chromosomal aberrations (39), changing the copy number of many genes. Similar to other genes, the copy number of various tRNA genes can be altered as well, changing the actual levels of these genes (40). Overall, these variations do not change dramatically the tRNA pool, thus allowing us to use wild-type genomic copy numbers for our calculations (Supplementary Fig. S1). Nevertheless, we aimed to study whether these rather minor-scale changes coadapt with the *TP53* mutations to increase the overall TE. Using copy number variation data on 60 cancerous cell lines (NCI-60), we calculated the tRNA pool for each of these cell lines (Supplementary Table S7). We then measured the TE ratio for the 25 nonsynonymous point mutations present in these cell lines (ref. 26; Supplementary Table S6) using both the original wild-type (before the chromosomal abnormalities have occurred) and the mutated tRNA pool. We found that the mutated tRNA pool leads to a significantly higher TE (mean, 1.825 versus 1.640; $P = 0.03$, Wilcoxon signed rank test). This result (yet on relatively small scale) suggests that there may be highly intriguing process of coevolution between tRNA copy number aberrations and point mutations in *TP53*.

ER and *TP53* TE. Glazko and colleagues (15) have shown that *TP53* substitution hotspots in cancer tend to be evolutionary conserved and that there is negative correlation between ER and mutation frequency in cancer cells. This implies that selection forces in cancer may be very different than evolutionary selection forces acting on healthy tissues. To study the relation between evolutionary conservation and *TP53* TE, we calculated the ER values for all codons in *TP53* based on orthologues from 10 species (ref. 27; Materials and Methods). Figure 2*C* shows ER values along the *TP53* sequence. In this analysis, ER was calculated based on the ratio between evolutionary synonymous and nonsynonymous substitutions across the phylogenetic tree of the species.

First, adding to the results of Glazko and colleagues (15), we found significant negative correlation between ER and cancerous dN/dS ($R = -0.183$; $P = 5.86 \times 10^{-3}$). This correlation is significant

even after controlling for the TE of the codon and the number of nonsynonymous mutations ($R = -0.233$; $P = 4.60 \times 10^{-4}$ and $R = -0.141$; $P = 0.0353$, respectively). Second, we found that conserved sites tend to be translated more efficiently: there is a negative correlation between the ER of the codon and its TE ($R = -0.185$; $P = 2.33 \times 10^{-4}$). This result is similar to previous results showing that conserved sites and genes tend to be translated more efficiently in other organisms (41, 42) and that conserved human genes are translated more efficiently (18). As mentioned above, there is also a negative correlation between the TE of the codon and its cancerous dN/dS ($R = -0.169$; $P = 0.013$). This correlation is even more significant after controlling for ER and for the number of nonsynonymous mutations ($R = -0.222$; $P = 8.66 \times 10^{-4}$ and $R = -0.3$; $P = 5.41 \times 10^{-6}$ for ER and number of nonsynonymous mutations, respectively). Focusing on hotspots, we found a significant correlation between cancerous dN/dS and the increase in TE ($R = 0.368$; $P = 0.02$). We did not find such a correlation between ER and the increase in TE ($R = 0.02$; $P = 0.71$), further implying that the above correlation is not indirect.

Discussion

Numerous studies in the last two decades have shown that *TP53* has a major role in cancer. As more and more data have been gathered, it was shown that its mutations not only disrupt its wild-type activity but also may have additional effects of DNE and GOF, effectively making it an oncogene. This study is the first to analyze the effects of *TP53* mutations on its TE on a large scale. We focus on nonsynonymous mutations, comprising the vast majority of mutations reported but control for their functional effects when studying TE by looking only on mutations resulting in functionally similar amino acids and by a careful choice of random background models. Our results show that, overall, *TP53* mutations significantly increase its TE. We find that there is a significant correlation between the frequency of a mutation and its effect on TE and that mutations with known oncogenic function (DNE and GOF) significantly increase TE compared with other mutations. These results further support the hypothesis of TE selection in *TP53* mutations.

We find that TE increases with cancer progression, further supporting the role of TE selection in these mutations. This gives rise to the possibility that TE analysis of existing *TP53* mutations can perhaps help predicting the outcome of the cancer, similar to other studies using mutational data (1, 3, 4). In addition, our analysis also shows that there are significant differences in TE between cancers from different tissues. A highly interesting future work would be to further analyze the differences between the tissues and see whether they are correlated with other differences between these tumors (e.g., the aggressiveness of tumors in these tissues). Focusing on NCI-60 cell lines, we find an intriguing coadaptation between the *TP53* mutations and chromosomal aberrations that change the tRNA pool. Taken together, these results indicate that TE has a role in shaping mutation patterns of *TP53* in tumors.

A basic assumption in this study is that the changes in TE we analyze lead to significant effects on the protein activity levels and thus are under selection. Although it has been reported previously that a single synonymous codon change can affect protein folding via TE (43), can a single codon choice have a significant effect on the protein abundance of a 393-codon protein? The answer is likely to be positive: mutations with relatively small effects on fitness are not "neutral" and may be under selection forces when occurring within relatively large effective population size (44). Previous studies estimated that different codons can have as much as 6-fold difference in their translation rate and that a change of a single codon can change protein levels via TE regulation in ~1% (45). Another analysis showed that even a small decrease in fitness (0.01 and 0.001) due to TE can be under selection in effective population size of 10^5 to 10^7 organisms (46). Given that the effective population size of cancerous cells within a tumor is relatively large (~10^9 in a cubic centimeter; ref. 47), the changes analyzed here (as well as even smaller changes) may indeed be under selection. This conclusion is in line with the numerous other pertaining results presented in this study.

Nevertheless, some limitations of the current analysis should be pointed out. First, other selection forces obviously play an important role in shaping the mutation distribution found in *TP53* in tumor tissues. No doubt, a major influence should be the effect of the mutation on the structure of the protein and therefore on the function of the mutated p53. Yet, TE selection may prefer one mutation over functionally similar one. In addition, although significant, the overall effect of a single codon on the entire protein level may be fairly limited; therefore, the selection force can be relatively small. This may explain why some of the correlations found are quite weak. Nevertheless, these correlations are statistically significant, implying that TE is a nonnegligible selection force in the evolution of cancerous *TP53*. In that respect, some TE changes may have effect also on the protein folding (43). Bias in the mutation database due to methodologic and other biases (6) can also affect our results. Nevertheless, we obtained similar significant results also when using another database (Catalogue of Somatic Mutations in Cancer database; ref. 48). Analyzing the 360 nonsynonymous mutations found in this database, we find that they increase the TE (TE ratio of 1.41) and that there is a significant correlation between the increase in TE and the frequency of the mutation ($R = 0.187$; $P = 0.0228$).

As mentioned, in this work, we focused on *TP53* because, in addition to its central importance in cancer and its intriguing pattern of mutations, there are extensive data on *TP53* mutations observed in human cancers. Currently, data about mutations in other central cancer genes are insufficient for an extensive analysis as that done here for *TP53*. However, a preliminary analysis done on a few other genes has failed to identify a selection pressure that is akin to what we find regarding *TP53*, suggesting that the pattern of cancerous mutations in *TP53* may perhaps be quite unique possibly due to the DNE/GOF effect that takes place with its regard (details in Supplementary Note 3 and Supplementary Table S9). When more data will be available, it would be interesting to conduct a large-scale TE analysis of mutations in other central cancer-related genes as well. The basic methodology and tools are hereby laid out.

Taken together, we believe that the various significant results reported here encourage an experimental analysis of TE in *TP53* mutations to further validate the main findings of this paper.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgments

Received 5/8/09; revised 9/3/09; accepted 9/10/09; published OnlineFirst 11/3/09.

Grant support: Edmond J. Safra Bioinformatics Program at Tel Aviv University (Y.Y. Waldman), Converging Technologies grant from the Israel Science Foundation (R. Sharan and E. Ruppin), and partial support from a Koshland Scholar Award from the Weizmann Institute of Science (T. Tuller).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

We thank the anonymous referees for valuable comments.

References

1. Sousi T, Beroud C. Assessing TP53 status in human tumours to evaluate clinical outcome. Nat Rev Cancer 2001;1:233–9.

2. Vousden KH, Lu X. Live or let die: the cell's response to p53. Nat Rev Cancer 2002;2:594–604.

3. Russo A, Bazan V, Iacopetta B, Karr D, Sousi T, Gebbia N. The TP53 Colorectal Cancer International Collaborative study on the prognostic and predictive significance of p53 mutation: influence of tumor site, type of mutation, and adjuvant treatment. J Clin Oncol 2005;23:7518–28.

4. Petitjean A, Achatz MI, Borresen-Dale AL, Hainaut P, Olivier M. TP53 mutations in human cancers: functional selection and impact on cancer prognosis and outcomes. Oncogene 2007;26:2157–65.

5. Levine AJ, Finlay CA, Hinds PW. p53 is a tumor suppressor gene. Cell 2004;116:S67–70.

6. Sousi T, Asselain B, Hamroun D, et al. Meta-analysis of the p53 mutation database for mutant p53 biological activity reveals a methodologic bias in mutation detection. Clin Cancer Res 2006;12:62–9.

7. Vousden KH, Prives C. p53 and prognosis: new insights and further complexity. Cell 2005;120:7–10.

8. Oren M. p53: the ultimate tumor suppressor gene? FASEB J 1992;6:169–76.

9. de Vries A, Flores ER, Miranda B, et al. Targeted point mutations of p53 lead to dominant-negative inhibition of wild-type p53 function. Proc Natl Acad Sci U S A 2002;99:2948–53.

10. Wijnhoven SWP, Speksnijder EN, Liu X, et al. Dominant-negative but not gain-of-function effects of a p53.R270H mutation in mouse epithelium tissue after DNA damage. Cancer Res 2007;67:4648–56.

11. Lee MK, Sabapathy K. The R246S hot-spot p53 mutant exerts dominant-negative effects in embryonic stem cells *in vitro* and *in vivo*. J Cell Sci 2008;121:1899–906.

12. Sigal A, Rotter V. Oncogenic mutations of the p53 tumor suppressor: the demons of the guardian of the genome. Cancer Res 2000;60:6788–93.

13. Cadwell C, Zambetti GP. The effects of wild-type p53 tumor suppressor activity and mutant p53 gain-of-function on cell growth. Gene 2001;277:15–30.

14. Olivier M, Petitjean A, Marcel V, et al. Recent advances in p53 research: an interdisciplinary perspective. Cancer Gene Ther 2009;16:1–12.

15. Glazko GV, Koonin EV, Rogozin IB. Mutation hotspots in the p53 gene in tumors of different origin: correlation with evolutionary conservation and signs of positive selection. Biochim Biophys Acta 2004;1679:95–106.

16. Duret L. tRNA gene number and codon usage in the *C. elegans* genome are co-adapted for optimal translation of highly expressed genes. Trends Genet 2000;16:287–9.

17. Man O, Pilpel Y. Differential translation efficiency of orthologous genes is involved in phenotypic divergence of yeast species. Nat Genet 2007;39:415–21.

18. Waldman YY, Tuller T, Shlomi T, Karni S, Sharan R, Ruppin E. Gene translation efficiency in healthy and cancerous human tissues. In: Kellis M, Califano A, Stolovitzky G, editors. RECOMB Satellite on Regulatory Genomics 2008: Proceedings of the 5th Annual RECOMB Satellite on Regulatory Genomics; 2008, Oct 29-Nov 2; Cambridge, Massachusetts. Cambridge: MIT/Broad Institute/CSAIL; 2008, p. 84.

19. Gustafsson C, Govindarajan S, Minshull J. Codon bias and heterologous protein expression. Trends Biotechnol 2004;22:346–53.

20. Chabanon H, Persson L, Wallace HM, Ferrara M, Brachet P. Increased translation efficiency and antizyme-dependent stabilization of ornithine decarboxylase in amino acid-supplemented human colon adenocarcinoma cells, Caco-2. Biochem J 2000;348:401–8.

21. Jiang Y, Muschel RJ. Regulation of matrix metalloproteinase-9 (MMP-9) by translational efficiency in murine prostate carcinoma cells. Cancer Res 2002;62:1910–4.

22. Audic Y, Hartley RS. Post-transcriptional regulation in cancer. Biol Cell 2004;96:479–98.

23. dos Reis M, Savva R, Wernisch L. Solving the riddle of codon usage preferences: a test for translational selection. Nucleic Acids Res 2004;32:5036–44.

24. Chan PP, Lowe TM. GtRNAdb: a database of transfer RNA genes detected in genomic sequence. Nucleic Acids Res 2009;37:D93–7.

25. Olivier M, Eeles R, Hollstein M, Khan MA, Harris CC, Hainaut P. The IARC TP53 database: new online mutation analysis and recommendations to users. Hum Mutat 2002;19:607–14.

26. Ikediobi ON, Davies H, Bignell G, et al. Mutation analysis of 24 known cancer genes in the NCI-60 cell line set. Mol Cancer Ther 2006;5:2606–12.

27. Stern A, Doron-Faigenboim A, Erez E, Martz E, Bacharach E, Pupko T. Selecton 2007: advanced models for detecting positive and purifying selection using a Bayesian inference approach. Nucleic Acids Res 2007;35:W506–11.

28. Durinck S, Moreau Y, Kasprzyk A, et al. BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. Bioinformatics 2005;21:3439–40.

29. Graur D, Li W-H. Fundamentals of molecular evolution. 2nd ed. Sunderland: Sinauer Associates; 2000.

30. Dayhoff MO, Schwartz RM, Orcutt BC. A model of evolutionary change in proteins. Matrices for detecting distant relationships. In: Dayhoff MO, editor. Atlas of protein sequence and structure. Washington: National Biomedical Research Foundation; 1978, p. 345–58.

31. Henikoff S, Henikoff JG. Amino acid substitution matrices from protein blocks. Proc Natl Acad Sci U S A 1992;89:10915–9.

32. Grantham R. Amino acid difference formula to help explain protein evolution. Science 1974;185:862–4.

33. Kawashima S, Ogata H, Kanehisa M. AAindex: amino acid index database. Nucleic Acids Res 1999;27:368–9.

34. Akashi H. Translational selection and yeast proteome evolution. Genetics 2003;164:1291–303.

35. Tuller T, Kupiec M, Ruppin E. Determinants of protein abundance and translation efficiency in *S. cerevisiae*. PLoS Comput Biol 2007;3:e248.

36. Sørensen MA, Pedersen S. Absolute *in vivo* translation rates of individual codons in *Escherichia coli*. The two glutamic acid codons GAA and GAG are translated with a threefold difference in rate. J Mol Biol 1991;222:265–80.

37. Forrester K, Lupold SE, Ott VL, et al. Effects of p53 mutants on wild-type p53-mediated transactivation are cell type dependent. Oncogene 1995;10:2103–11.

38. Soussi T, Wiman KG. Shaping genetic alterations in human cancer: the p53 mutation paradigm. Cancer Cell 2007;12:302–12.

39. Mertens F, Johansson B, Höglund M, Mitelman F. Chromosomal imbalance maps of malignant solid tumors: a cytogenetic survey of 3185 neoplasms. Cancer Res 1997;57:2765–80.

40. Tsafrir D, Bacolod M, Selvanayagam Z, et al. Relationship of gene expression and chromosomal abnormalities in colorectal cancer. Cancer Res 2006;66:2129–37.

41. Akashi H. Synonymous codon usage in *Drosophila melanogaster*: natural selection and translational accuracy. Genetics 1994;136:927–35.

42. Stoletzky N, Eyre-Walker A. Synonymous codon usage in *Escherichia coli*: selection for translational accuracy. Mol Biol Evol 2007;24:374–81.

43. Kimchi-Sarfaty C, Oh JM, Kim IW, et al. A "silent" polymorphism in the MDR1 gene changes substrate specificity. Science 2007;318:1382–3.

44. Ohta T, Gillespie JH. Development of neutral and nearly neutral theories. Theor Popul Biol 1996;49:128–42.

45. Bulmer M. The selection-mutation-drift theory of synonymous codon usage. Genetics 1991;129:897–907.

46. Li W-H. Models of nearly neutral mutations with particular implications for nonrandom usage of synonymous codons. J Mol Evol 1987;24:337–45.

47. Friberg S, Mattson S. On the growth rates of human malignant tumors: implications for medical decision making. J Surg Oncol 1997;64:284–97.

48. Bamford S, Dawson E, Forbes S, et al. The COSMIC (Catalogue of Somatic Mutations in Cancer) database and website. Br J Cancer 2004;91:355–8.

49. In: Fritz A, Percy C, Jackm A, Shanmugaratnam K, Sobin LH, Parkin MD, editors. International classification of diseases for oncology (ICD-O). 3rd ed. Geneva: WHO; 2000.