



# Deciphering the Increased Prevalence of TP53 Mutations in Metastatic Prostate Cancer

Wensheng Zhang<sup>1</sup>, Yan Dong<sup>2</sup>, Oliver Sartor<sup>3</sup> and Kun Zhang<sup>1,4</sup>

<sup>1</sup>Bioinformatics Core of Xavier NIH RCMI Center of Cancer Research, Xavier University of Louisiana, New Orleans, LA, USA. <sup>2</sup>Department of Structural and Cellular Biology, Tulane University School of Medicine, Tulane Cancer Center, New Orleans, LA, USA. <sup>3</sup>Department of Medicine, Tulane University School of Medicine, Tulane Cancer Center, New Orleans, LA, USA. <sup>4</sup>Department of Computer Science, Xavier University of Louisiana, New Orleans, LA, USA.

Cancer Informatics  
Volume 21: 1–10  
© The Author(s) 2022  
Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/11769351221087046



**ABSTRACT:** The prevalence of TP53 mutations in advanced prostate cancers (PCa) is 3 to 5 times of the quantity in primary PCa. By an integrative analysis of the Cancer Genome Atlas and Catalogue of Somatic Mutations in Cancer data, we revealed the supporting evidence for 2 complementary hypotheses:  $H_1$  - TP53 abnormalities promote metastasis or therapy-resistance of PCa cells, and  $H_2$ —part of TP53 mutations in PCa metastases occur after the diagnosis of original cancers. The plausibility of these hypotheses can explain the increased prevalence of TP53 mutations in PCa metastases. With  $H_1$  and  $H_2$  as the general assumptions, we developed mathematical models to decipher the change of the percentage frequency (prevalence) of TP53 mutations from primary tumors to metastases. The following results were obtained. Compared to TP53-normal patients, TP53-mutated patients had poorer biochemical relapse-free survival, higher Gleason scores, and more advanced t-stages ( $P < .01$ ). Single-nucleotide variants in metastases more frequently occurred on G bases of the coding sequence than those in primary cancers ( $P = .03$ ). The profile of TP53 hotspot mutations was significantly different between primary and metastatic PCa as demonstrated in a set of statistical tests ( $P < .05$ ). By the derived formulae, we estimated that about 40% TP53 mutation records collected from metastases occurred after the diagnosis of the original cancers. Our study provided significant insight into PCa progression. The proposed models can also be applied to decipher the prevalence of mutations on TP53 (or other driver genes) in other cancer types.

**KEYWORDS:** Prostate cancer, metastasis, TP53, somatic mutation, prevalence, modeling

**RECEIVED:** December 20, 2021. **ACCEPTED:** February 22, 2022.

**TYPE:** Original Research

**FUNDING:** The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research is supported by the NIH grant 5U54MD007595 (WZ and KZ). The funders have no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**DECLARATION OF CONFLICTING INTERESTS:** The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

**CORRESPONDING AUTHORS:** Wensheng Zhang, Bioinformatics Core of Xavier NIH RCMI Center of Cancer Research, Xavier University of Louisiana, New Orleans, LA 70125, USA. Email: wzhang@xula.edu

Kun Zhang, Bioinformatics Core of Xavier NIH RCMI Center of Cancer Research, Department of Computer Science, Xavier University of Louisiana, New Orleans, LA 70125, USA. Email: kzhang@xula.edu

## Introduction

The tumor suppressor p53 protein has a myriad of functions crucial to normal cell proliferation, apoptosis, DNA repair, and others.<sup>1,2</sup> TP53 gene, encoding p53, is the most frequently altered gene in human cancers.<sup>3</sup> Mutant-TP53 disrupts age-related accumulation patterns of somatic mutations in multiple cancer types.<sup>4</sup> However, pathogenic germline TP53 mutations are relatively populous in only a few cancer types, including inherited Li-Fraumeni syndrome, carcinomas of the breast and adrenal cortex, brain tumor, and acute leukemia.<sup>5</sup> Most somatic TP53 mutations are single-base substitutions distributed throughout exons 5 to 8.<sup>6</sup> Notably, about 20% of these mutations alter 1 of 3 codons (175 to 248, or 273) of the 393 amino acids of p53 protein.<sup>7</sup> The clinical significance of TP53 status for patient outcomes has been and continues to be a controversial topic of cancer research.<sup>8,9</sup> Many retrospective studies have associated its mutation and abnormal p53 protein expression with poor patient survival. Such an association has been demonstrated by previous studies, mostly in breast, head and neck, hematopoietic and liver cancers.<sup>10-13</sup>

Prostate cancer (PCa) is the most commonly diagnosed non-skin cancer worldwide for males. In the United States, about 30 000 men die of PCa annually.<sup>14-16</sup> Metastasis is a

primary cause of morbidity and mortality for patients with PCa or other cancers.<sup>17,18</sup> PCa progression can be predicted using transcriptomic and epigenetic signatures.<sup>19-21</sup> Androgen deprivation therapy (ADT) is a usual first-line option for men with advanced (metastatic and non-metastatic) PCa.<sup>22</sup> However, nearly all men with metastatic PCa will develop resistance to androgen deprivation therapy, a state known as metastatic castration-resistant PCa (mCRPC).<sup>23</sup> Aberrations of AR, ETS genes, TP53 and PTEN are frequent, with TP53 and AR alterations being enriched in mCRPC compared to primary PCa.<sup>24-26</sup> In particular, the percentage frequency of TP53 mutations is about 10% in primary PC samples but may be as high as 50% in advanced PCa or metastases of the disease.<sup>24,27</sup>

Cancer metastases arise in part from residual and disseminated tumor cells that originated from primary cancer. These tumor cells can survive after the initial surgery, chemotherapy, radiotherapy, and/or targeted therapy.<sup>28-30</sup> Based on such an understanding, it is logical to premise that a potential TP53 status-determined mechanism for cancer progression may contribute to the increased prevalence of TP53 mutations in metastatic PCa. That is, TP53 abnormalities could promote PCa metastasis and predispose therapeutic resistance. This



hypothesis, termed  $H_1$  hereafter, was suggested by a previous study.<sup>31</sup> As shown in the publication, biochemical recurrence (BCR), i.e. prostate-specific antigen (PSA) recurrence after prostatectomy, was more frequently observed in the patients with TP53 mutations in the primary tumor samples than in those without such mutations. A reported analysis of transcriptomic data demonstrated that abnormal p53 expression status was associated with poor overall survival, progression-free survival, and time to distant metastases for patients with locally advanced prostate cancer treated primarily by radiation therapy.<sup>32</sup>

Complementary to  $H_1$ , another hypothesis, termed  $H_2$  hereafter, for the mutation enrichment in metastatic prostate tumors is that a fraction of TP53 mutations in metastases occur after the diagnosis of original cancers. The logic underlying this novel hypothesis is that there is a substantial timespan between the initial treatment of TP53-wild-type prostate cancer and the after-therapy progression (ie, biochemical relapse and metastasis formation) such that new TP53 mutations may occur with a substantial possibility and influence the biology of the disseminated tumor cells. For example, in the patients who initially respond to abiraterone (a CYP17A1 inhibitor that reduces PSA and improves overall survival), the median time to PSA progression ranges from 5.8 to 11.1 months and a median time to radiographic progression is about 16.5 months.<sup>33-35</sup>

In this paper, via an integrative analysis of publicly available genomic data of PCa samples, we first provided supporting evidence for the 2 hypotheses. After that, we derived the mathematical models to decipher the change of the percentage frequency (prevalence) of TP53 mutations from primary cancers to metastatic ones.

## Materials and Methods

### *COSMIC data*

From the Catalogue of Somatic Mutations in Cancer (COSMIC) version-92 database,<sup>36</sup> we downloaded the table of “CosmicMutantExportCensus\_92.tsv” on August 27, 2020. It contained all the somatic genetic alterations, including single nucleotide variants (SNVs) and short inserts/deletes (indels), on 710 census cancer genes.<sup>37</sup> The information of 39,320 records of mutations on the coding sequence of TP53 gene, which did not include those annotated with “Substitution – coding silent,” was used in this study. Among them, 468 were collected from 433 primary prostate carcinomas and 312 were collected from 296 PCa metastases. The filter(s) used for a specific analysis was presented in the corresponding paragraphs of the Results section.

### *TCGA data*

The dataset generated by The Cancer Genome Atlas (TCGA) Prostate Adenocarcinoma (PRAD) project<sup>27</sup> contained 471 primary carcinoma samples with both clinical and somatic

mutation information. Among them, 46 samples each had at least one non-synonymous mutation on the TP53 gene and another 5 each had a mutation at a splice point. The tumors with  $GS \geq 7$  accounted for 91% of the sample set. In this study, the dataset was used for revealing the potential TP53 status based stratification of disease-free survival and the associations between TP53 status and cancer progression stages/Gleason scores. It was also used to estimate the percentage frequency of TP53 mutations in primary PCa. The reason was that a substantial fraction of primary cancer samples didn't have a mutation on any one of the census cancer genes, and therefore, were not collected in the relatively big COSMIC dataset.

### *Bioinformatics and statistics analysis*

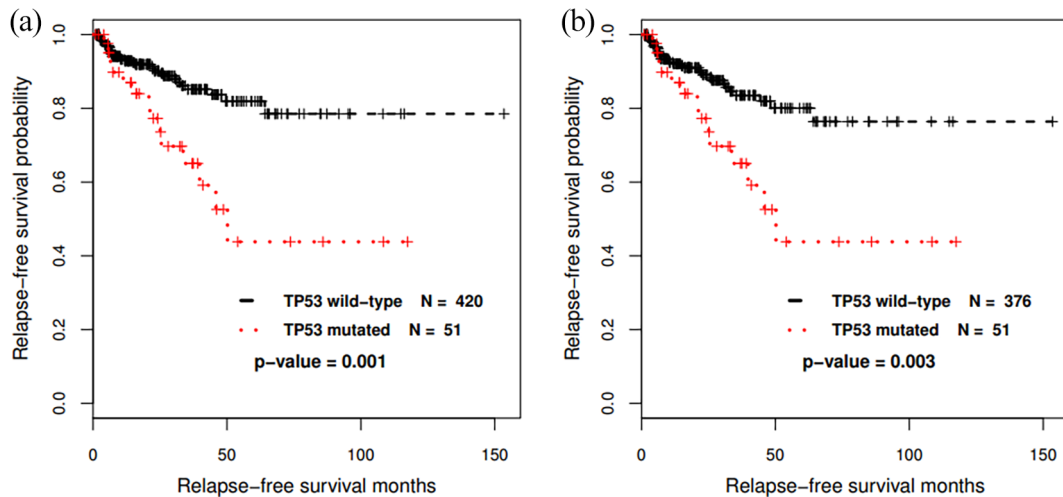
The annotation of the RefSeq gene NM001126114 (which includes 12 exons) was used as the template for mapping TP53 mutations onto individual exons. The comparison of a specific mutation feature (such as the exon or exon group where a mutation is located) between primary cancers and metastatic cancers was performed by establishing a  $2 \times k$  contingency table, where  $k$  was the category number of the feature. P-values were calculated using the Chi-squared test, Binomial test, or Proportion test, depending on the context of a specific analysis item. The Kolmogorov-Smirnov test was used to compare the distributions of ages at diagnosis between patients with TP53-mutated metastatic prostate cancers and those with TP53-wild-type cancers. The differences in survival time between the 2 patient groups were evaluated by a Cox-PH regression model, in which patient age was included as a covariate alongside TP53 status. The employed software included the relevant functions in R packages “stats” and “survival”. Two-tail p-value was used to determine the significance of a focused effect, difference or association.

### *Mathematical models*

Mathematical models were developed to decipher the change in the prevalence of TP53 mutations from primary cancers to metastases. The modeling process started from an equation that related the imbalance of TP53 mutations between primary and metastatic PCa to the disparity of progression probabilities between TP53-mutated and TP53-wild-type cancers. The underlying assumptions and the derivation of formulae were described in the Results section.

## Results

For readers' convenience, we reiterate the aforementioned hypotheses as follows:  $H_1$ : TP53 abnormalities promote metastasis or therapy-resistance of PCa cells; and  $H_2$ : A fraction of TP53 mutations in PCa metastases occur after the diagnosis of the original cancers. We also note that synonymous mutations were excluded from the following analyses.



**Figure 1.** The TP53 mutation status-based stratification of biochemical relapse-free survival. (A) All the 471 samples with completed information of Gleason score and BCR in the TCGA prostate Adenocarcinoma (PRAD) cohort was included in the analysis. (B) The sample with  $GS \leq 6$  were excluded from the analysis.  $P$ -values were calculated using the Cox-PH model, in which the patient age at the initial diagnosis was included as a covariate alongside the interested stratification variable, that is, TP53 status.

### Deriving supporting evidence from TCGA data for $H_1$ and $H_2$

**Biochemical relapse-free survival (BCRFs).** Survival analysis using the TCGA data (Figure 1) showed that TP53-mutated patients had poorer BCRFS than TP53-normal patients ( $p = 0.001$ ), even when the patients with low-grade ( $GS \leq 6$ ) PCa were excluded ( $p = 0.003$ ). This result verified the finding by Ecke et al.<sup>31</sup> and could be considered as direct evidence supporting our hypothesis  $H_1$ .

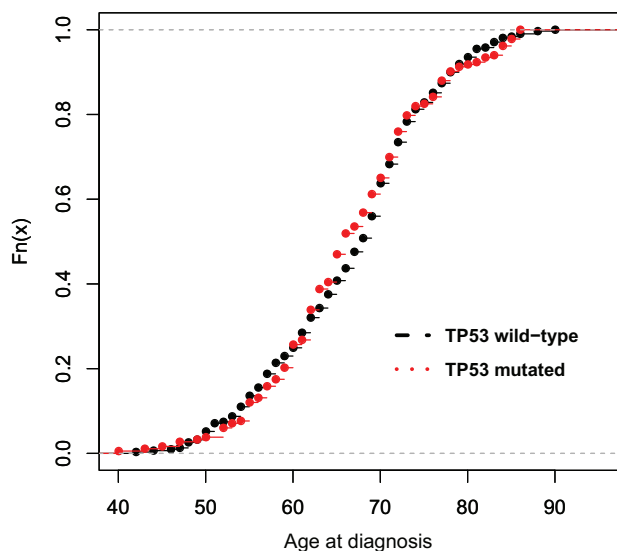
**Gleason score (GS).** The GS is the sum of the primary and secondary Gleason patterns (GPs) of a primary tumor. The GSs of the 471 TCGA samples ranged from 6 to 9+ ( $\geq 9$ ). The sizes of all the 4 GS-based groups were relatively substantial, containing 44, 238, 61, 128 samples, respectively. None of the GS-6 samples had a TP53 mutation. The mutation frequencies were 0.046 for GS-7, 0.113 for GS-8, and 0.25 for GS-9+, respectively. We performed a Chi-square test on this data, finding that the association between TP53 status and GS category was extremely significant ( $p = 4 \times 10^{-9}$ ). This association could be considered as supporting evidence for a perception equivalent to our hypothesis  $H_1$ , that is, prostate cancers with mutated TP53 are more aggressive than those of TP53-wild-types. The following are the reasons. First, mortality rarely happens among patients with GS-6(3 + 3) cancers and climbs with the increase of GS among the patients with high-grade ( $GS \geq 7$ ) PCa.<sup>38-40</sup> Second, a grade-3 GP (GP-3) cannot directly progress into a grade-4 GP (GP-4), in general.<sup>41,42</sup>

**Progression stage.** The T-stage information of 382 TCGA cancer samples was publicly available. The numbers of T1, T2, T3, and T4 samples were 167, 162, 51, 2, respectively. We firstly combined the T3 and T4 samples into a single group (ie, T3&4), and then calculated the t-stage specific percentage

frequencies of TP53-mutated samples. With a linear pattern, the quantities increased from 0.054 for T1, 0.13 for T2, to 0.189 for T3&4. The Chi-square test showed that the association between TP53 status and t-stage was significant ( $p = 0.008$ ). This result was compatible with the hypothesis  $H_1$  since the relative enrichment of TP53 mutations in late-stage cancer cases means that the variants promote cancer progression. It also could be considered as indirect evidence for  $H_2$  because t-stage is a feature that reflects the progression level of primary cancers, determined by the spreading, extension, and invasion.<sup>43</sup> The rationale of the last statement can be further scrutinized in the following manner. The aforementioned statistics suggest that, for a TP53-mutated patient (patient-X) whose PCa was diagnosed at the T3 stage, the mutation likely occurred between T1 and T3 stages with a probability over 70% ( $\frac{0.189 - 0.054}{0.189} \approx 0.7$ ). If patient-X had been early diagnosed with PCa at the T1 stage rather than the T3 stage, it would be logical to state that the mutation was acquired after the “initial diagnosis.”

### Deriving supporting evidence from COSMIC data for $H_1$ and $H_2$

**Ages of patients with metastatic cancers.** We compared the distribution of patient ages at the diagnosis of TP53-mutated metastatic prostate cancers (Group-A) and the corresponding age distribution for TP53-wild-type cancers (Group-B). We conceived that a piece of strong (but not necessary) supporting evidence for the hypothesis  $H_1$  could be that Group-A patients were younger than Group-B patients on average. To perform the comparison, we extracted the information of 763 metastatic PCa samples from the COSMIC dataset to establish these 2 groups, that was Group-A ( $N_1 = 295$ ) and Group-B ( $N_2 = 468$ ). A sample was selected once it met the following 2 criteria. First, its



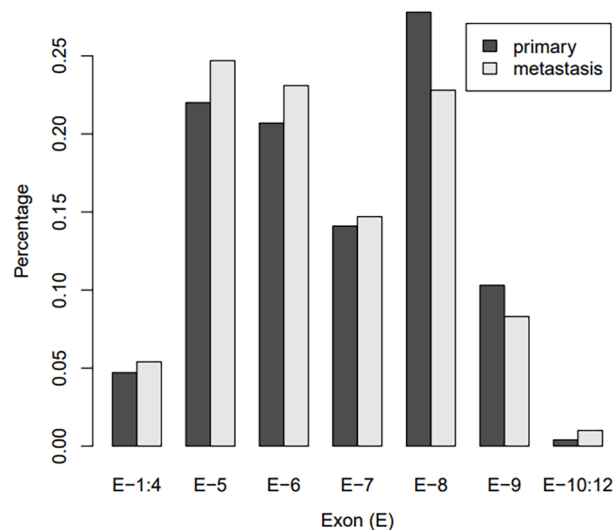
**Figure 2.** The distributions of ages, at dates of diagnosis or tumor sampling, for patients diagnosed with TP53 mutated metastatic PCa and patients with TP53 wild-type metastatic PCa in the COSMIC data. The  $F_n(x)$  on the y-axis represents the empirical accumulation probability.

molecular and clinical information was documented by a previous study archived in the PubMed database; second, the TP53 status (ie, mutated or wild-type) of the sample was known. In particular, of the 11 samples from the publication indexed with the PubMed ID “PMID24135135,”<sup>44</sup> only one was included due to the repeated sampling from a 42-year-old participant. Advanced statistical analysis was performed on the 183 Group-A samples and 289 Group-B samples with the age information. As shown in Figure 2, there was a moderate difference in the cumulative distribution of patient ages between these 2 groups. In terms of median age, Group-A was 2-year younger than Group-B. However, the Kolmogorov-Smirnov test showed that the difference was not significant ( $p > 0.05$ ).

*Mutations exclusively observed in metastatic cancers.* In the COSMIC dataset, an indexed mutation was uniquely determined by the physical position and the involved DNA base alteration (or indel) such as  $G > C$ . It was common that, for the same mutation, multiple mutation records were collected from different tumor samples. In particular, 272 (and 172) mutations were shared by 468 (and 312) TP53 mutation records from primary (and metastatic) PCa samples. Eighty-four mutations were in both lists of primary PCa and metastatic PCa. Eighty-eight mutations exclusively existed in metastatic PCa, accounting for 36.2% of mutation records of this cancer category. *This result could be considered as supporting evidence for the hypothesis  $H_2$ .*

#### *Suggestive evidence for $H_2$ derived from COSMIC data*

In this subsection, we show some differences in the profiles of TP53 mutations between primary and metastatic PCa. These



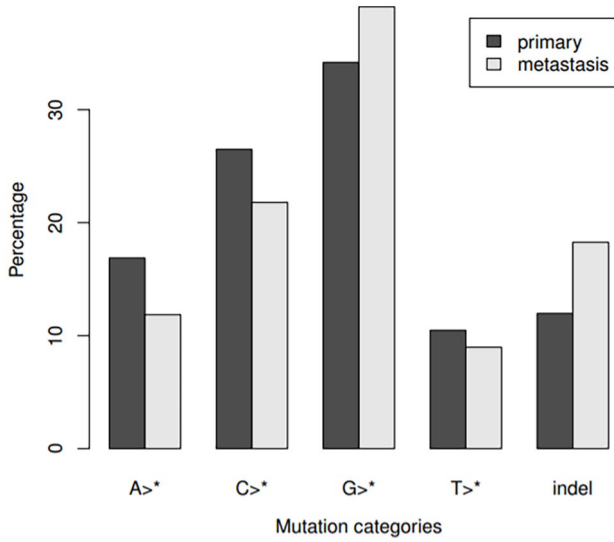
**Figure 3.** The distributions of TP53 mutation records over exons (and exon clusters) for primary and metastatic PCa samples in the COSMIC data.

results somewhat suggest the plausibility of our hypothesis  $H_2$  (see the Discussion section).

*Physical position.* We depicted the distribution pattern of mutation records over the 12 exons of the TP53 gene, among which the exons 1 to 4 encode the transcriptional activation domain of p53 protein, the exons 5 to 8 encode the sequence-specific DNA-binding domain and the exons 9 to 11 encode the tetramerization domain. Because mutation events in the 4 exons at the upstream end and the 3 exons at the downstream end were relatively rare (in particular, no mutation record was in exon 12 that is 10754 bases away from exon 11), we combined them into 2 exon clusters, that is, E-1:4 and E-10:12. As shown in Figure 3, the recorded mutations in primary PCa most frequently (28%) occurred on exon 8 (E-8) and the percentage frequency decreased to 23% in metastatic PCa. However, the difference was not significant ( $p = 0.14$ ). This result was obtained from the Chi-squared test in which the mutation records of each cancer category were partitioned into 2 groups, that is, E-8 (exon 8) and E-(-8) (other exons except for exon 8).

*Nucleotide acid substitutions and indels.* With reference to the coding sequence, we partitioned TP53 mutation records into 5 categories, that is,  $A > *$ ,  $T > *$ ,  $C > *$ ,  $G > *$ , and indel. The last one stood for short inserts and deletes. The other 4 were defined by the DNA bases (in the coding sequences) at which single nucleotide substitutions arose. As shown by Figure 4 and according to the results of Chi-squared tests, the mutation categories were not independent of cancer categories ( $p = 0.02$ ). In particular, the mutations of metastatic PCa were relatively enriched with  $G > *$  substitutions ( $p = 0.03$ ) and indels ( $p = 0.02$ ) compared to those of primary PCa.





**Figure 4.** The distributions of TP53 mutation records over 5 alteration categories, defined by single nucleotide substitutions and indels, for primary and metastatic PCa samples in the COSMIC data. The asterisk \* represents any member of single nucleotides except for the wild-type one.

**Hotspot mutations.** From the COSMIC dataset, we selected a set ( $N = 18$ ) of TP53 hotspot mutations, each of which contributed over 1% of mutation records to at least one of 3 sample categories, that is, primary PCa, metastatic PCa or panCancer (containing all cancer types, alongside PCa). The information and statistical analysis results of those mutations were summarized in Table 1. The top 4 genetic substitutions in panCancer and metastatic PCa (but not in primary PCa) were ENST00000269305.8:c.524G>A (p.R175H), c.743G>A (p.R248Q), c.818G>A (p.R273H), and c.817C>T (p.R273C), consistent with the statistics in literature.<sup>6</sup> We further inferred the significance of the inter-group difference in the frequencies of individual mutations. For a comparison between primary (or metastatic) PCa and panCancer, we performed the Chi-squared goodness of fit test, in which the former was considered as the “sample set” and the latter was treated as the “population” to be fit. For a comparison between primary PCa and metastatic PCa, a proportion test was used, in which the null hypothesis was that the proportions of the focused mutation in the 2 PCa categories were equal. The results indicated that, compared to primary PCa, the hotspot mutation profile of metastatic PCa was more similar to that of panCancer. Three (or eight) mutations showed significantly different frequencies ( $P < .05$ ) between metastatic (or primary) PCa and panCancer. Here, the genetic substitution ENST00000269305.8:c.743G>A was worth special attention. It was the top one mutation in metastatic PCa with the percentage frequency being over 8.0%, nearly 2 times of the quantity in panCancer. Because the involved mutation records were collected from multiple studies, the observed high percentage frequency should be free from a severe sampling bias and might indicate a unique point of the mutation spectrum for metastatic PCa.

### Modeling the prevalence of TP53 mutations in metastatic prostate tumors

Based on the hypotheses  $H_1$  and  $H_2$  and several assumptions about the relationship between the metastasis-promoting effect of TP53 mutations and their timespans, we propose 4 mathematical models to decipher the change of the percentage frequency (prevalence) of somatic TP53 mutations in PCa progression. The symbols and terms used in our model equations and the related description are defined as follows.

$g^+$ : TP53 mutated.

$g^-$ : TP53 wild-type.

$f_p(g^+)$ : Percentage frequency of  $g^+$  primary cancers.

$f_p(g^-)$ : Percentage frequency of  $g^-$  primary cancers.

$f_m(g^+)$ : Percentage frequency of  $g^+$  metastatic cancers.

$f_m(g^-)$ : Percentage frequency of  $g^-$  metastatic cancers.

$S_1$ : Probability that  $g^+$  primary cancers metastasize after the original diagnosis.

$S_2$ : Probability that  $g^-$  primary cancers metastasize after the original diagnosis if the cancerous cells and their descendants don't acquire TP53 mutation(s) since then.

$S_3$ : Probability that  $g^-$  primary cancers metastasize after the original diagnosis regardless whether the cancerous cells and their descendants acquire or don't acquire TP53 mutation(s) since then.

$m$ : Probability that  $g^-$  primary cancers acquire TP53 mutations after the original diagnosis.

$m^*$ : Proportion of  $g^+$  metastatic cancers that acquire their TP53 mutations after the original diagnosis among all  $g^+$  metastatic cancers.

$N_p$ : Speculated total number of primary cancers.

$N_m$ : Speculated total number of metastatic cancers.

A-O-D: After the Original Diagnosis.

$$R_p = \frac{f_p(g^+)}{f_p(g^-)}$$

$$R_m = \frac{f_m(g^+)}{f_m(g^-)}$$

$$R_s = \frac{S_1}{S_2}$$

**Model-1.** This model is based on the assumption that the probability of the  $g^+$  primary tumor cells' metastasis is independent of the time when the TP53 mutation(s) occurs. In other words, it is speculated that TP53 mutations occurring in

**Table 1.** TP53 Hotspot mutations in panCancer, primary PCa and metastatic PCa.\*

MUTATION DESCRIPTION			PERCENTAGE <sup>†</sup>			P-VALUE		
CDS SUBSTITUTION	AMINO ACID SUBSTITUTION	GENOME POSITION	PANCANCER (PN)	PRIMARY (PR)	METASTASIS (ME)	PN VERSUS PR	PN VERSUS ME	PR VERSUS ME
c.524G>A	p.R175H	17:7675088	4.86	2.14	4.17	0.004	0.692	0.101
c.743G>A	p.R248Q	17:7674220	3.26	3.42	8.01	0.794	<0.001	0.005
c.818G>A	p.R273H	17:7673802	3.06	1.71	4.17	0.105	0.247	0.038
c.817C>T	p.R273C	17:7673803	2.93	4.91	3.85	0.018	0.312	0.480
c.742C>T	p.R248W	17:7674221	2.54	1.28	1.92	0.103	0.716	0.476
c.844C>T	p.R282W	17:7673776	2.3	2.35	2.24	0.877	1	0.922
c.637C>T	p.R213*	17:7674894	1.73	1.07	1.6	0.372	1	0.516
c.733G>A	p.G245S	17:7674230	1.62	1.71	1.28	0.853	0.823	0.635
c.659A>G	p.Y220C	17:7674872	1.45	1.5	3.53	0.846	0.007	0.064
c.536A>G	p.H179R	17:7675076	0.71	0.21	1.28	0.274	0.291	0.067
c.734G>A	p.G245D	17:7674229	0.55	1.07	0.32	0.118	1	0.242
c.473G>A	p.R158H	17:7675139	0.38	1.5	0.64	0.002	0.332	0.274
c.641A>G	p.H214R	17:7674890	0.36	1.07	0	0.028	0.634	0.067
c.451C>T	p.P151S	17:7675161	0.32	1.07	0.32	0.018	1	0.242
c.487T>C	p.Y163H	17:7675125	0.09	1.07	0	<0.001	1	0.067
c.313G>T	p.G105C	17:7676056	0.08	0.21	1.28	0.312	<0.001	0.067
c.639A>G	p.R213=	17:7674892	0.07	2.56	0.32	<0.001	0.196	0.016
c.108G>A	p.P36=	17:7676261	0.05	2.14	0	<0.001	1	0.009
Total number of mutation records			39320	468	312	—	—	—

\*Each "hotspot" mutation contributes over 1% of TP53 mutation records for at least one of 3 sample categories, that is, primary PCa, metastatic PCa or panCancer. The selected mutations are sorted according to their contribution percentages to the records of the panCancer category.

<sup>†</sup>The quantity is the percentage of the records of the corresponding mutation among the total (mutation) records.

(post-treatment) residual primary tumor cells are equally efficient in driving metastasis as those occurring before the treatment. Accordingly, we establish the following proportion equation.

$$\frac{f_p(g^+)S_1N_p + f_p(g^-)S_1mN_p}{f_p(g^-)S_2(1-m)N_p} = \frac{f_m(g^+)N_m}{f_m(g^-)N_m} \quad (1)$$

In (1),  $N_p$  and  $N_m$  are included to improve the logic and clarity but can be dropped (as done in the following text). After some mathematical transformations, we obtain the following formula for calculating  $m$ .

$$m = \frac{R_m - R_p R_s}{R_m + R_s} \quad (2)$$

Then, the formula to calculate  $m^*$  is derived as follows.

$$m^* = \frac{f_p(g^-)S_1m}{f_p(g^+)S_1 + f_p(g^-)S_1m} = \frac{R_m - R_p R_s}{R_m(1 + R_p)} \quad (3)$$

*Model-2.* This model is based on one general and 3 specific assumptions. The general assumption is that the probability of the  $g^+$  primary tumor cells' metastasis depends on the time when the TP53 mutation(s) occurs. The specific assumptions include: (i) The timespan ( $t$ ) between the diagnosis of primary cancer and the occurrence of the A-O-D TP53 mutation(s) follows the uniform distribution  $t \sim U(0, T)$  with the density function  $u(t) = \frac{1}{T}$ , where  $T$  is the speculated maximum follow-up time after the diagnosis of primary cancer; (ii) For a

$g^-$  primary cancer, A-O-D TP53 mutation(s) increases its metastasis probability but the increment quantity descends as the timespan increases; and (iii) The probability increment, denoted by  $h(t)$ , and timespan have a linear relationship, that is,  $h(t) = (S_1 - S_2)(1 - \frac{t}{T})$ . Let  $S_2^*$  denote the mathematical expectation of metastasis probability of  $g^-$  primary cancers with A-O-D TP53 mutations, then, it can be evaluated by

$$S_2^* = S_2 + \int_0^T (u(t)h(t))dt = \frac{1}{2}(S_1 + S_2)$$

Using  $S_2^*$  to replace  $S_1$  in the second term of the numerator on the left hand of the equation (1), we had the following equation.

$$\frac{f_p(g^+)S_1 + f_p(g^-)(S_1 + S_2)m/2}{f_p(g^-)S_2(1-m)} = \frac{f_m(g^+)}{f_m(g^-)} \quad (4)$$

From the equation (4), we derive the formulae for calculating  $m$  and  $m^*$ :

$$m = \frac{2(R_m - R_p R_s)}{2R_m + R_s + 1} \quad (5)$$

and

$$m^* = \frac{(R_s + 1)(R_m - R_p R_s)}{R_m(2R_p R_s + R_s + 1)} \quad (6)$$

*Model-3.* This model had the same general assumption and the specific assumptions (i) and (ii) as the Model-2. However, the relationship between the metastasis probability increment and mutation timespan is modeled by a cosine function, that is,  $h(t) = (S_1 - S_2)\cos(t)$ . The timespan is rescaled such that the maximum  $T$  is equal to  $\pi/2$ . Accordingly, we had the following formulae.

$$S_2^* = S_2 + \int_0^{\pi/2} (u(t)h(t))dt = S_2 + \frac{2}{\pi}(S_1 - S_2)$$

$$\frac{f_p(g^+)S_1 + f_p(g^-)[S_2 + 2(S_1 - S_2)/\pi]m}{f_p(g^-)S_2(1-m)} = \frac{f_m(g^+)}{f_m(g^-)} \quad (7)$$

$$m = \frac{\pi(R_m - R_p R_s)}{\pi R_m + 2R_s + \pi - 2} \quad (8)$$

$$m^* = \frac{(2R_s + \pi - 2)(R_m - R_p R_s)}{R_m(\pi R_p R_s + 2R_s + \pi - 2)} \quad (9)$$

*Model-4.* This model had the same general assumption and the specific assumptions (i) and (ii) as the Model-2. However, the relationship between the probability increment and mutation timespan is modeled by an exponential function, that is,  $h(t) = (S_1 - S_2)\frac{e^{-t} - e^{-1}}{1 - e^{-1}}$ . The timespan is rescaled such that the

maximum  $T$  is equal to 1. Accordingly, we had the following simplified formulae, in which  $\frac{1 - 2e^{-1}}{1 - e^{-1}}$  is denoted by  $\alpha$ .

$$S_2^* = S_2 + \int_0^1 (u(t)h(t))dt = S_2 + \alpha(S_1 - S_2)$$

$$\frac{f_p(g^+)S_1 + f_p(g^-)[S_2 + \alpha(S_1 - S_2)]m}{f_p(g^-)S_2(1-m)} = \frac{f_m(g^+)}{f_m(g^-)} \quad (10)$$

$$m = \frac{R_m - R_p R_s}{R_m + \alpha R_s + 1 - \alpha} \quad (11)$$

$$m^* = \frac{[1 + \alpha(R_s - 1)](R_m - R_p R_s)}{R_s R_m (\alpha + R_p) + R_m(1 - \alpha)} \quad (12)$$

Here, 2 things are worth noting. First, the equations (11) and (12) can be considered as the general formulae for calculating  $m$  and  $m^*$ , applicable to all 4 models. That is, they are equivalent to the equations (2) and (3), the equations (5) and (6), or the equations (8) and (9) when  $\alpha$  is 1,  $1/2$  or  $2/\pi$ , respectively. Second, while different  $h(t)$  functions are defined in Model-2, -3 to -4, they have a common property, that is, the function value is 1 when  $t=0$  and the value is 0 when  $t$  is equal to the upper limit.

*Inferring  $S_2$ .* The assumedly known  $S_2$  in our models cannot be directly retrieved from the available datasets. As such, we designed an iterative post-hoc contribution decomposition procedure to obtain an estimate ( $\widetilde{S}_2$ ) of  $S_2$  for model implementation. Assume that  $m$  (i.e. the probability that  $g^-$  primary cancers acquire TP53 mutations after the original diagnosis) is known, then, based on the equations of Model-4, we had,

$$(1-m)\widetilde{S}_2 + m[\widetilde{S}_2 + \alpha(S_1 - \widetilde{S}_2)] = S_3 \quad (13)$$

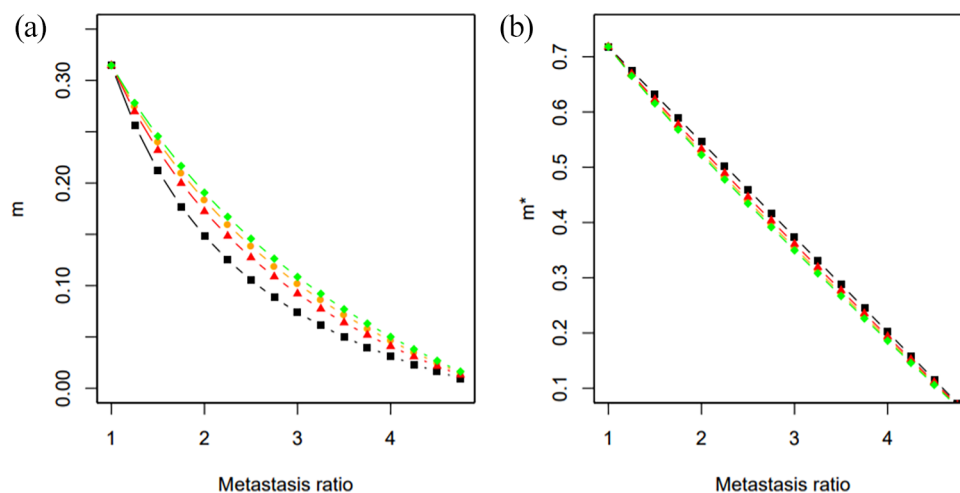
After some mathematical transformations, we had the following formula for  $\widetilde{S}_2$ .

$$\widetilde{S}_2 = \frac{(S_3 - \alpha m S_1)}{1 - \alpha m} \quad (14)$$

In this setting, the iteration procedure took the following steps.

- (1) Initialize  $\widetilde{S}_2$  with a prior value (such as 0.15).
- (2) Replace  $S_2$  with  $\widetilde{S}_2$  to calculate  $R_s$  by  $S_1/S_2$  and calculate  $m$  using the equation (11).
- (3) Calculate  $\widetilde{S}_2$  using the equation (14).
- (4) Repeat (2) and (3) until convergence for  $m$  and  $\widetilde{S}_2$ .

*Model comparison.* In all 4 models, the required inputs for calculating  $m$  and  $m^*$  are the values of  $R_p$ ,  $R_m$  and  $R_s$ . Based on the TCGA dataset,  $g^+$  cancers accounted for 11% ( $\frac{51}{471}$ ) primary PCa samples. Based on the filtered COSMIC dataset (See "Ages of patients with metastatic cancers" subsection),  $g^+$  cancers accounted for 39% ( $\frac{296}{763}$ ) of metastatic PCa samples.



**Figure 5.** The relationships between TP53 mutation-caused fold change of metastasis probability and 2 metrics (ie,  $m$  and  $m^*$ ) for TP53 mutations arising after diagnosis of the original cancers. Metastasis ratio ( $R_s$ ), on the x-axis, represents the ratio of the probability that  $\mathcal{G}^+$  (TP53-mutated) primary cancers metastasize after the original diagnosis to the corresponding probability for  $\mathcal{G}^-$  (TP53 wild-type) primary cancers. The  $m$ , on the y-axis of (A) represents the probability that  $\mathcal{G}^-$  primary cancers acquire TP53 mutations after the original diagnosis. The  $m^*$ , on the y-axis of (B) represents the proportion of  $\mathcal{G}^+$  metastatic cancers that acquire their TP53 mutations after the original diagnosis among all  $\mathcal{G}^+$  metastatic cancers. The results of the Model-1, -2, -3 and -4 are presented with black, orange, red, and green curve (or lines), respectively.

Accordingly, we had an estimate of  $0.12 \left( \frac{0.11}{1-0.11} \right)$  for  $R_p$  and  $0.64 \left( \frac{0.39}{1-0.39} \right)$  for  $R_m$ . In this context, we depicted the relationships of  $R_s$  versus  $m$  and  $R_s$  versus  $m^*$ . As shown in Figure 5, for  $m$  versus  $R_s$ , the curve of Model-1 is consistently below those of the other models. This indicates that the value of  $m$  might be underestimated if the time of the TP53 mutation occurrence were not taken into account. The relationship between  $R_s$  and  $m^*$  is linear in Model-1 and the regression line almost overlaps with the curves of the other models, implying that the estimate of  $m^*$  is less sensitive to the related model assumptions.

**Model application.** The implementation procedure of the proposed models includes 4 steps: estimate  $R_p$  and  $R_m$ ; estimate  $S_1$  and  $S_3$ ; infer  $S_2$  and calculate  $R_s$ ; and calculate  $m$  and  $m^*$ . Except for the second step where a survival analysis may be required, the other calculation can be achieved using the explicit formulae. As mentioned above, we got an estimate of 0.12 for  $R_p$  and 0.64 for  $R_m$  from the TCGA and COSMIC data, respectively. While the data suitable for exactly estimating  $S_1$  and  $S_3$  has not yet been available, we used the TCGA dataset to derive substitutes for the 2 metrics. In particular, biochemical recurrence (BCR) was used as the proxy measure of metastasizing. This manner is largely appropriate because BCR is the first sign for PCa relapse and the subsequent metastases,<sup>45</sup> and the cases of cancer progression with undetectable or low prostate-specific antigen levels have been rarely observed.<sup>46,47</sup> As shown in Figure 1, the BCR probability, that is, 1 minus the disease-free survival probability of primary PCa patients, approached a plateau after 5 years from the initial diagnosis. At that time point, the BCR probability was 0.225 for the patients with TP53 wild-type cancers

and 0.56 for those with TP53-mutated cancers. Hereby, we obtained an estimate of 0.56 for  $S_1$  and 0.225 for  $S_3$ . Introducing these values, along with the estimates for  $R_p$  and  $R_m$ , into our formulae resulted in a range of 0.081 to 0.129 for  $m$  and the same value of 0.397 for  $m^*$ . This result indicated that 8.1% to 12.9% of wild-type TP53 primary cancers acquired TP53 mutations after the original diagnosis, and 39.7% of TP53 mutation records collected from metastases occurred after the diagnosis of original cancers.

## Discussion

The plausibility of the complementary hypotheses  $H_1$  and  $H_2$  was the first issue addressed in this study. For  $H_1$ , the significant supporting evidence revealed by our analysis included the associations between TP53 status and a few clinical characteristics (or outcome), that is, Gleason score, progression stage ( $t$ -stage) and disease-free survival time. The supporting evidence for  $H_2$  included the association between TP53-status and  $t$ -stage, and the substantial existence of the mutations solely observed in metastatic PCa samples. In addition, we found that, at the diagnosis dates, patients with TP53-mutated metastases were 2 years younger than those with TP53-wild-type metastases in terms of median ages. While the statistical significance level of such a difference was modest (one tail  $P= .07$ ), we expect that this could prove to be direct supporting evidence for  $H_1$ , as more data is accumulated. This perception is based on the following reasons. First, the limited sample sizes in the current analysis might impact the statistical power, especially in the context that cancer patients had a quite wide age range. Second, the earlier onset of TP53-mutated metastases implies that abnormal p53 protein can facilitate tumor metastasis, which is consistent with a recent study about the effect of mutant p53 on ovarian cancer progression in mice.<sup>48</sup>



Regarding TP53 mutation features, we found that the single nucleotide variants in PCa metastases more frequently occurred on the G bases of the coding sequence of the gene compared to those in primary cancers, and the percentage frequency profile of hotspot mutations was different between the 2 PCa categories. We deemed these results as “suggestive” evidence for  $H_2$ . The reason was that, only in the case that individual TP53 mutation was equally efficient in promoting cancer progression, the observed changes in the mutation profile from primary PCa to metastatic PCa could be convincingly attributed to the mutation events that occurred after the diagnosis of original cancers. However, the “equal efficiency” assumption might be questionable. We have this concern because previous studies showed that mutations within the exon 4 of TP53 were particularly associated with poor prognosis in breast cancer patients, and mutations in exons 1 to 4 were more lethal than those in exons 5 to 9 for the patients with lung adenocarcinomas.<sup>9,49</sup> In particular, the poor prognosis associated with exon 4 mutations was probably related to the importance of this region in cell apoptosis.<sup>50</sup> At present, due to the lack of necessary data, it is still challenging to conduct a similar survival analysis in PCa to clarify this issue. In other words, much larger cohort data (compared to the TCGA one) would be needed to evaluate the relative effects of individual mutations and mutation clusters on cancer-free survival.

A novel finding in this study was that, compared to primary PCa, the profile of the TP53 hotspot mutations in metastatic PCa was more similar to that in panCancer. This observation, together with the well-known understanding that the cancer types with high TP53 mutation rates (such as bladder cancer and colorectal cancer) are generally more lethal than primary PCa,<sup>51</sup> suggests that the occurrence of TP53 mutations in tumor cells represents a crucial driving force in the process from less aggressive PCa to TP53 mutation-enriched metastatic PCa. In particular, because PCa coincidence rate was as high as 70% among the patients with bladder cancer,<sup>52</sup> it could be interesting to investigate the potential association between the coincidence and TP53-status in these 2 cancer types.

In this paper, we propose a set of mathematical models to decipher the prevalence change of somatic TP53 mutations in PCa progression. Using these models, we estimated that 39.7% of TP53 mutation records collected from metastases arose after the diagnosis of original cancers. According to the results from analyzing the COSMIC data, 36.2% of TP53 mutation records of metastatic PCa were consisted of the “unique mutations” present in the metastatic PCa samples but not in the primary cancers. These quantities indicate that the increment of the prevalence of TP53 mutations in metastatic PCa could be mostly attributed to the hits of those unique mutations. We also estimated that the probability that TP53 wild-type primary cancers acquire TP53 mutations (during the follow-up periods) after the original diagnosis ranged from 8% to 13%. The quantity is comparable to the mutation prevalence observed in primary cancer. Previous studies showed that there

was a growing period of ~10 years between the genesis of initial tumorous cells and a tumor that can be detected by transvaginal ultrasound,<sup>53</sup> close to the timespan from a primary PCa to its distant metastases.<sup>54</sup> These observations and findings suggest that TP53 mutation (and mutation accumulation) rate over time is largely consistent in the growing period and progression period of advanced prostate cancer.

Besides the aforementioned insights into PCa progression, our results uncover a potential pitfall in the study of tumor evolution. Phylogenetic trees were often used to infer the temporal order of multiple driver mutations of individual cancer drivers.<sup>55-60</sup> When this approach is applied to static tumor sample data, it typically leads to such a conclusion (or a similar one) that the genetic alterations on the most frequently mutated driver gene(s) (for a specific cancer type) occur before those on the other drivers. However, the plausibility of our hypothesis  $H_2$  indicates that, from a predominant driver gene (such as TP53 for advanced PCa), mutations may substantially arise in both the early and later time of cancer development.

Our mathematical models can also be applied to decipher the prevalence of the somatic mutations on TP53 (or other main driver genes) in other cancer types. The most subjective assumption of these models is the function  $h(t)$  that describes the relationship between the increment of metastasizing probability caused by a (TP53) mutation and its timespan. However, as indicated by the empirical results, the estimated proportion of  $g^+$  metastatic cancers that acquire the TP53 mutations after the original diagnosis among all  $g^+$  metastatic cancers is not sensitive to the  $h(t)$  options.



## Acknowledgements

The analyses presented here are based on the data generated by the TCGA Research Network and the data collected by Catalogue of Somatic Mutations in Cancer (COSMIC). The authors downloaded the TCGA and COSMIC datasets from <https://portal.gdc.cancer.gov/legacy-archive/search/f> and <https://cancer.sanger.ac.uk/cosmic>, respectively. The authors are grateful to the 2 reviewers for their constructive comments which significantly improved this paper.

## Author Contributions

Study conceiving: WZ, KZ. Method design: WZ, KZ, YD, OS. Experiments performing: WZ. Data analysis: WZ, KZ. Writing: WZ, KZ, YD, OS. All authors read and approved the final manuscript.

## ORCID iDs

Wensheng Zhang  <https://orcid.org/0000-0002-5043-8794>  
Kun Zhang  <https://orcid.org/0000-0002-1915-788X>

## REFERENCES

1. Harris CC, Hollstein M. Clinical implications of the p53 tumor-suppressor gene. *New Engl J Med*. 1993;329:1318-1327.
2. Lane DP. The regulation of p53 function: Steiner Award Lecture. *Int J Cancer*. 1994;57:623-627.

3. Vogelstein B, Sur S, Prives C. p53: The most frequently altered gene in human cancers. *Nat Educ*. 2010;3:6.
4. Zhang W, Flemington EK, Zhang K. Mutant TP53 disrupts age-related accumulation patterns of somatic mutations in multiple cancer types. *Cancer Genet*. 2016;209:376-380.
5. Nichols KE, Malkin D, Garber JE, Fraumeni JF Jr., Li FP. Germ-line p53 mutations predispose to a wide spectrum of early-onset cancers. *Cancer Epidemiol Biomarkers Prev*. 2001;10:83-87.
6. Olivier M, Hollstein M, Hainaut P. TP53 mutations in human cancers: origins, consequences, and clinical use. *Cold Spring Harb Perspect Biol*. 2010;2:a001008.
7. Bunz F. *Principles of Cancer Genetics*. Springer; 2008.
8. Levine AJ, Lane D. *The p53 Family: A Subject Collection From Cold Spring Harbor Perspectives in Biology*. Cold Spring Harbor Laboratory Press; 2010.
9. Zhang W, Edwards A, Flemington EK, Zhang K. Significant prognostic features and patterns of somatic TP53 mutations in human cancers. *Cancer Inform*. 2017;16:1176935117691267.
10. Shahbandi A, Nguyen HD, Jackson JG. TP53 mutations and outcomes in breast cancer: reading beyond the headlines. *Trends Cancer*. 2020;6:98-110.
11. Zhang W, Edwards A, Fang Z, Flemington EK, Zhang K. Integrative Genomics and Transcriptomics Analysis reveals potential mechanisms for favorable prognosis of patients with HPV-Positive head and neck carcinomas. *Sci Rep*. 2016;6:24927.
12. Kharfan-Dabaja MA, Komrokji RS, Zhang Q, et al. TP53 and IDH2 somatic mutations are associated with inferior overall survival after allogeneic hematopoietic cell transplantation for myelodysplastic syndrome. *Clin Lymphoma Myeloma Leuk*. 2017;17:753-758.
13. Zhan P, Ji YN, Yu LK. TP53 mutation is associated with a poor outcome for patients with hepatocellular carcinoma: evidence from a meta-analysis. *Hepatobiliary Surg Nutr*. 2013;2:260-265.
14. Rawla P. Epidemiology of prostate cancer. *World J Oncol*. 2019;10:63-89.
15. Siegel R, Ma J, Zou Z, Jemal A. Cancer statistics, 2014. *CA Cancer J Clin*. 2014;64:9-29.
16. PDQ® Adult Treatment Editorial Board. Prostate Cancer Treatment (PDQ®) –Health Professional Version. 2018; [https://www.cancer.gov/types/prostate/hp/prostate-treatment-pdq#cit/section\\_1.21](https://www.cancer.gov/types/prostate/hp/prostate-treatment-pdq#cit/section_1.21). (accessed 30 May 2020).
17. Seyfried TN, Huysentruyt LC. On the origin of cancer metastasis. *Crit Rev Oncog*. 2013;18:43-73.
18. Bittner N, Merrick GS, Galbreath RW, et al. Primary causes of death after permanent prostate brachytherapy. *Int J Radiat Oncol Biol Phys*. 2008;72:433-440.
19. Zhao S, Geybels MS, Leonardson A, et al. Epigenome-wide tumor DNA methylation profiling identifies novel prognostic biomarkers of metastatic-lethal progression in men diagnosed with clinically localized prostate cancer. *Clin Cancer Res*. 2017;23:311-319.
20. Alkhatieb A, Rezaeian I, Singireddy S, Cavallo-Medved D, Porter LA, Rueda L. Transcriptomics signature from next-generation sequencing data reveals new transcriptomic biomarkers related to prostate cancer. *Cancer Inform*. 2019;18:1176935119835522.
21. Zhang W, Flemington EK, Deng HW, Zhang K. Epigenetically silenced candidate tumor suppressor genes in prostate cancer: Identified by modeling methylation stratification and applied to progression prediction. *Cancer Epidemiol Biomarkers Prev*. 2019;28:198-207.
22. Crawford ED, Heidenreich A, Lawrentschuk N, et al. Androgen-targeted therapy in men with prostate cancer: evolving practice and future considerations. *Prostate Cancer Prostatic Dis*. 2019;22:24-38.
23. Karantanos T, Corn PG, Thompson TC. Prostate cancer progression after androgen deprivation therapy: mechanisms of castrate resistance and novel therapeutic approaches. *Oncogene*. 2013;32:5501-5511.
24. Robinson D, Van Allen EM, Wu YM, et al. Integrative Clinical Genomics of Advanced Prostate Cancer. *Cell*. 2015;162:454.
25. Dessel LF, van Riet J, Smits M, et al. The genomic landscape of metastatic castration-resistant prostate cancers reveals multiple distinct genotypes with potential clinical impact. *Nat Commun*. 2019;10:5251.
26. Nyquist MD, Corella A, Coleman I, et al. Combined TP53 and RB1 loss promotes prostate cancer resistance to a spectrum of therapeutics and confers vulnerability to replication stress. *Cell Rep*. 2020;31:107669.
27. Cancer Genome Atlas Research Network. The molecular taxonomy of Primary Prostate Cancer. *Cell*. 2015;163:1011-1025.
28. Hall C, Krishnamurthy S, Lodhi A, et al. Disseminated tumor cells predict survival after neoadjuvant therapy in primary breast cancer. *Cancer*. 2012;118:342-348.
29. Franci C, Zhou J, Jiang Z, et al. Biomarkers of residual disease, disseminated tumor cells, and metastases in the MMTV-PyMT breast cancer model. *PLoS One*. 2013;8:e58183.
30. van der Toom EE, Verdone JE, Pienta KJ. Disseminated tumor cells and dormancy in prostate cancer metastasis. *Curr Opin Biotechnol*. 2016;40:9-15.
31. Ecke TH, Schlechte HH, Schiemenz K, et al. TP53 gene mutations in prostate cancer progression. *Anticancer Res*. 2010;30:1579-1586.
32. Grignon DJ, Caplan R, Sarkar FH, et al. p53 status and prognosis of locally advanced prostatic adenocarcinoma: a study based on RTOG 8610. *J Natl Cancer Inst*. 1997;89:158-165.
33. de Bono JS, Logothetis CJ, Molina A, et al. Abiraterone and increased survival in metastatic prostate cancer. *N Engl J Med*. 2011;364:1995-2005.
34. de Bono JS, Oudard S, Ozguroglu M, et al. Prednisone plus cabazitaxel or mitoxantrone for metastatic castration-resistant prostate cancer progressing after docetaxel treatment: a randomised open-label trial. *Lancet*. 2010;376:1147-1154.
35. Ryan CJ, Smith MR, de Bono JS, et al. Abiraterone in metastatic prostate cancer without previous chemotherapy. *New Engl J Med*. 2013;368:138-148.
36. Tate JG, Bamford S, Jubb HC, et al. COSMIC: the Catalogue of somatic mutations in cancer. *Nucleic Acids Res*. 2019;47:D941-D947.
37. Sondka Z, Bamford S, Cole CG, Ward SA, Dunham I, Forbes SA. The COSMIC Cancer Gene Census: describing genetic dysfunction across all human cancers. *Nat Rev Cancer*. 2018;18:696-705.
38. Zhang W, Dong Y, Sartor O, Flemington EK, Zhang K. SEER and gene expression data analysis deciphers racial disparity patterns in prostate cancer mortality and the public health implication. *Sci Rep*. 2020;10:6820.
39. Stamey TA, McNeal JE, Yemoto CM, Sigal BM, Johnstone IM. Biological determinants of cancer progression in men with prostate cancer. *JAMA*. 1999;281:1395-1400.
40. Cheng L, Davidson DD, Lin H, Koch MO. Percentage of Gleason pattern 4 and 5 predicts survival after radical prostatectomy. *Cancer*. 2007;110:1967-1972.
41. Sowalsky AG, Kissick HT, Gerrin SJ, et al. Gleason score 7 prostate cancers emerge through branched evolution of clonal Gleason Pattern 3 and 4. *Clin Cancer Res*. 2017;23:3823-3833.
42. Lavery HJ, Droller MJ. Do Gleason patterns 3 and 4 prostate cancer represent separate disease states? *J Urol*. 2012;188:1667-1675.
43. Amin MB; American Joint Committee on Cancer and American Cancer Society. *AJCC Cancer Staging Manual*. eighth ed. American Joint Committee on Cancer, Springer; 2017.
44. Haffner MC, Mosbrugger T, Esopi DM, et al. Tracking the clonal origin of lethal prostate cancer. *J Clin Invest*. 2013;123:4918-4922.
45. Paller CJ, Antonarakis ES. Management of biochemically recurrent prostate cancer after local therapy: evolving standards of care and new directions. *Clin Adv Hematol Oncol*. 2013;11:14-23.
46. Pound CR, Partin AW, Eisenberger MA, Chan DW, Pearson JD, Walsh PC. Natural history of progression after PSA elevation following radical prostatectomy. *JAMA*. 1999;281:1591-1597.
47. Leibovici D, Spiess PE, Agarwal PK, et al. Prostate cancer progression in the presence of undetectable or low serum prostate-specific antigen level. *Cancer*. 2007;109:198-204.
48. Ren YA, Mullany LK, Liu Z, Herron AJ, Wong KK, Richards JS. Mutant p53 promotes epithelial ovarian cancer by regulating tumor differentiation, metastasis, and responsiveness to steroid hormones. *Cancer Res*. 2016;76:2206-2218.
49. Powell B, Soong R, Iacopetta B, Seshadri R, Smith DR. Prognostic significance of mutations to different structural and functional regions of the p53 gene in breast cancer. *Clin Cancer Res*. 2000;6:443-451.
50. Venot C, Maratrat M, Dureuil C, Conseiller E, Bracco L, Debussche L. The requirement for the p53 proline-rich functional domain for mediation of apoptosis is correlated with specific PIG3 gene transactivation and with transcriptional repression. *EMBO J*. 1998;17:4668-4679.
51. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. *CA Cancer J Clin*. 2019;69:7-34.
52. Kinoshita Y, Singh A, Rovito Pm Jr, Wang CY, Haas GP. Double primary cancers of the prostate and bladder: a literature review. *Clin Prostate Cancer*. 2004;3:83-86.
53. Hori SS, Gambhir SS. Mathematical model identifies blood biomarker-based early cancer detection strategies and limitations. *Sci Transl Med*. 2011;3:109ra116.
54. Popiolek M, Rider JR, Andr n O, et al. Natural history of early, localized prostate cancer: a final report from three decades of follow-up. *Eur Urol*. 2013;63:428-435.
55. Beerenwinkel N, Schwarz RF, Gerstung M, Markowitz F. Cancer evolution: mathematical models and computational inference. *Syst Biol*. 2015;64:e1-25.
56. Shah SP, Roth A, Goya R, et al. The clonal and mutational evolution spectrum of primary triple-negative breast cancers. *Nature*. 2012;486:395-399.
57. Van Loo P, Wedge DC, Nik-Zainal S, Stratton MR, Futreal PA, Campbell PJ. 5 proffered paper: the life history of 21 breast cancers. *Cell*. 2012;48:S2-1007.
58. Roth A, Khattraj J, Yap D, et al. PyClone: statistical inference of clonal population structure in cancer. *Nat Methods*. 2014;11:396-398.
59. Sprouffske K, Pepper JW, Maley CC. Accurate reconstruction of the temporal order of mutations in neoplastic progression. *Cancer Prev Res*. 2011;4:1135-1144.
60. H glund M, Gisselsson D, Mandahl N, et al. Multivariate analyses of genomic imbalances in solid tumors reveal distinct and converging pathways of karyotypic evolution. *Genes Chromosomes Cancer*. 2001;31:156-171.