Comprehensive analysis of tumour mutational burden and its clinical significance in prostate cancer

Lijuan Wang, Shucheng Pan, Binbin Zhu, Zhenliang Yu and Wei Wang*

Abstract

Background: The tumorigenesis of prostate cancer involves genetic mutations. Tumour mutational burden (TMB) is an emerging biomarker for predicting the efficacy of immunotherapy.

Results: Single-nucleotide polymorphisms were the most common variant type, and C>T transversion was the most commonly presented type of single-nucleotide variant. The high-TMB group had lower overall survival (OS) than the low-TMB group. TMB was associated with age, T stage and N stage. Functional enrichment analysis of differentially expressed genes (DEGs) showed that they are involved in pathways related to the terms spindle, chromosomal region, nuclear division, chromosome segregation, cell cycle, oocyte meiosis and other terms associated with DNA mutation and cell proliferation. Six hub genes, PLK1, KIF2C, MELK, EXO1, CEP55 and CDK1, were identified. All the genes were associated with disease-free survival, and CEP55 and CDK1 were associated with OS.

Conclusions: The present study provides a comprehensive analysis of the significance of TMB and DEGs and infiltrating immune cells related to TMB, which provides helpful information for exploring the significance of TMB in prostate cancer.

Keywords: Prostate cancer, Tumour mutational burden, Gene, Clinical
to be one of the candidates for predicting the efficacy of immunotherapy.

Next-generation sequencing (NGS) profiling of patients has enabled advancements, and The Cancer Genome Atlas (TCGA) offers convenient access to relevant information. In this study, the gene expression and mutation profiles of prostate cancer samples were extracted from TCGA, and the data were used to investigate the clinical significance of TMB and its related differentially expressed genes (DEGs) and immune cell infiltration signatures.

Methods
Data collection and processing
Transcriptome data in the HTSeq‐FPKM format from 524 samples, including 499 prostate and 25 adjacent normal tissue samples, were acquired from the TCGA databank (https://portal.gdc.cancer.gov/). The data were in masked somatic mutation files, and these data were analysed and visualized using the R software package maftools.

Estimation of TMB and its associations with clinical factors
TMB was defined as the sum of mutations in coding regions per megabase and was calculated as the total number of mutations/the length of exons. The length of exons was estimated to be 38 Mb in previous research [8]. First, Perl scripts were written to obtain the estimated TMB data, which were amalgamated with inter-related survival profiles for each patient. Then, we set the median TMB value as the cutoff, according to which samples were classified into a group with high TMB and a group with low TMB. We utilized the survival R package to analyse the overall survival (OS) differences between the two groups. The differences in TMB between groups categorized based on clinical features (age, T stage and N stage) were analysed using the limma R package.

Identification of DEGs and DEG pathway analysis
First, we performed DEG analysis with \( \log_2 \) fold change > 1 and false discovery rate (FDR) < 0.05 as cut-offs. We employed the limma package in the R package to analyse the differences and used heatmap in the R package to generate a heatmap. Subsequently, the gene symbols were transferred into the ID of Entrez using the org.Hs.eg.db package of the R package. Moreover, we employed the ggplot2, enrichplot, and clusterProfiler of R packages to perform Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses of the DEGs. We used GSEA software (https://www.gsea-msigdb.org/gsea/index.jsp) to perform GSEA. We used c2.cp.kegg.v7.0.symbols.gmt as the databank of gene sets. In addition, the TMB level was set as the phenotype labels. The pathways were considered to be statistically enriched according to a cutoff of FDR < 0.25.

Protein–protein interaction (PPI) network and classification of core genes
We constructed a PPI network of DEGs using STRING (https://string-db.org/). Evidence-based interactions were generated with a minimum required interaction score > 0.4 [9]. Then, the PPI network was visualized by utilizing Cytoscape 3.8.0 software [10]. Using the CytoHubba plugin, we obtained five protein groups, including the top 30 proteins, with five analytic algorithms, including MCC, MNC, Degree, EPC and EcCentricity, as previously reported [11]. In addition, we acquired the overlapping proteins of the five groups and identified pivotal proteins with more interactions than others as potential hub proteins. These intersections were visualized by a Venn diagram, which was generated online (http://bioinformatics.psb.ugent.be/webtools/Venn/). GEPIA (http://gepia.cancer-pku.cn/) was used to investigate the prognostic significance of the hub proteins.

CIBERSORT
CIBERSORT is an algorithm for characterizing the immune cell composition of certain tissues according to their gene expression profiles [12]. We used CIBERSORT in the R package to analyse 22 types of immune cells. The comparison of the levels of immune cells between the group with low TMB and the group with high TMB was conducted using the Wilcoxon rank-sum test, and the results were presented in a violin diagram generated with the vioplot package of the R package.

Statistical analysis
We used the Kaplan–Meier method to generate the survival curve. The comparisons of OS between different groups categorized according to TMB, age, T stage and N stage was performed using the log-rank test. We used the Shapiro–Wilk test to determine whether the groups categorized according to TMB, age, T stage and N stage had normal distributions. Through the Shapiro–Wilk test, we found that all the groups were not normally distributed, and thus, nonparametric tests were required. The Wilcoxon rank-sum test was used to compare groups classified by age and N stage. The Kruskal–Wallis test was used to compare groups classified by T stage. We used R software (version 3.6.1) to perform the Kaplan–Meier analysis, log-rank test, Wilcoxon rank-sum test and Kruskal–Wallis test. We used SPSS (version 25.0) to perform the Shapiro–Wilk test. The criterion for statistical significance was a P value < 0.05.
Results

Mutations in prostate cancer
We obtained mutation data from TCGA, which were analysed and visualized with the maftools package. Missense mutations were the most common type of variant (Fig. 1a). The frequency of single-nucleotide polymorphisms (SNPs) was greater than that of other variant types (Fig. 1b). C>T transversions represented the largest proportion of single-nucleotide variants (SNVs) in prostate cancer (Fig. 1c). TTN, TP53, SPOP, KMT2D, SYNE1, MUC16, FOXA1, KMT2C, SPTA1 and ATM were the top ten genes with high mutation frequencies.

Associations of TMB with prognostic and clinical factors
Based on the median TMB, we stratified the patients with prostate cancer into a group with high TMB and a group with low TMB. The prognostic analysis revealed that the group with low TMB had increased OS compared with the group with high TMB (Fig. 2a, \(P=0.026\)). There was a relationship between age, T stage, N stage and TMB level (Fig. 2b–d). The median age of the patients was 61 years. It was demonstrated that patients older than 61 years had a higher TMB than those who were 61 years or younger (Fig. 2b, \(P<0.001\)). In addition, the higher the T stage was, the higher the TMB (Fig. 2c, \(P<0.001\)). Similar results were seen for N stage, with the N1 stage group having a higher TMB than the group with other N stages (Fig. 2d, \(P<0.001\)).

Comparison of DEGs and functional enrichment analysis
A total of 257 DEGs were identified, and the top 20 DEGs were displayed in the heatmap (Fig. 3a). Functional enrichment analysis of the DEGs was conducted. GO enrichment analysis, including three major categories, was utilized (Fig. 3b). In the biological process (BP) category, the terms nuclear division, organelle fission and chromosome segregation were enriched. The cellular component (CC) category terms included spindle, chromosomal region, kinetochore, midbody and microtubule. The molecular function (MF) category terms involved receptor-ligand activity, growth factor activity and hormone activity. Moreover, we identified pathways related to the terms microRNA in cancer, cell cycle, oocyte meiosis and ECM-receptor interaction in the KEGG enrichment analysis (Fig. 3c). There were 37 pathways enriched in the GSEA, and the top 10 pathways were displayed in a diagram (Fig. 3d). These pathways were related to DNA-level cell proliferation, including mechanisms such as pyrimidine metabolism, DNA replication, DNA degradation, and aminoacyl tRNA biosynthesis, and the findings were in accordance with the above results.

PPI network of DEGs and selected hub genes
A PPI network of 189 nodes and 1789 edges was generated by STRING, and online tool for analysing proteins, and the results were visualized by Cytoscape (Fig. 4a). Five algorithms, MCC, MNC, Degree, EPC and EcCentricity, were utilized, and the overlapping proteins in the results generated from each algorithm were identified with Venn diagrams (Fig. 4b). PLK1, KIF2C, MELK, EXO1, CEP55 and CDK1 were identified as hub genes through this method. High expression of PLK1 and KIF2C was related to poor overall survival, with a \(P\) value < 0.05. High expression of PLK1, KIF2C, MELK, EXO1, CEP55 and CDK1 was related to poor disease-free survival, with a \(P\) value < 0.05.

Comparison of differential immune cell signatures
A violin diagram was generated to visualize the differences in the proportions of 22 infiltrating immune cells between the group with low TMB and the group with high TMB (Fig. 5). The group with high TMB had higher levels of CD8 T cells and activated CD4 memory cells than the low TMB group (\(P<0.05\)). However, there were no other statistically significant findings or trends for the other infiltrating immune cells.

Discussion
Immunotherapies have shown preliminary results in prostate cancer. TMB is considered an emerging biomarker for response evaluation. Therefore, it is meaningful to investigate the relationship between TMB and prostate cancer. This study provides an overview of mutations, the clinical significance of TMB, and DEGs and infiltrating immune cells related to TMB in prostate cancer.

Among the top 10 mutated genes, TP53, SPOP, FOXA1 and ATM showed pivotal functions in the initiation and development of malignant prostate cancer. As a tumour suppressor, TP53 has a high mutation frequency among various kinds of tumours, and the mutant form is equipped with antiproliferative functions and is related to the metastasis and progression of prostate cancer [13, 14]. Mutations in SPOP are considered the most common recurrent point mutations in prostate cancer [15]. SPOP is crucial for the preservation of nuclear genome stability and is essential for the degradation of multiple proteins [16, 17]. FOXA1 is necessary for androgen receptor-mediated activation of prostate genes [18]. Furthermore, the expression of FOXA1 is related to tumorigenesis and the progression of prostate cancer [19]. ATM is considered one of the
Fig. 1 Overview of mutation profile in prostate cancer patients.

(a) Variant classification, (b) variant type, (c) SNV class, (d) variants per sample, (e) variant classification summary, (f) top 10 mutated gene, (g) waterfall diagram of top 30 mutated genes.

Altered in 290 (59.92%) of 484 samples.
DNA damage repair genes, and its activation can be seen in the earlier stages of prostate tumorigenesis [20].

The clinical significance of TMB in prostate cancer was assessed. Similar to a previous prostate and renal cancer study, our study found that the group of patients with prostate cancer with high TMB had lower OS than the group of patients with prostate cancer with low TMB[1][21]. Furthermore, we observed higher TMB levels in patients older than 61 years than in those who were 61 years or younger and in the higher T stage and N1 stage groups. According to reports from the TCGA database, the prognostic role of TMB is unclear. In a study of bladder cancer, the high TMB group exhibited increased OS compared with the low TMB group [22]. Two reasons probably account for the results of our research.

One reason is that not all patients with a high TMB have an increased treatment response, as not every generated neoantigen has immunogenicity [23]. That is, high TMB does not always initiate an antitumour response. Another reason is that the high TMB group in this study was older and had a more advanced stage than the low TMB group.

GO and KEGG analyses revealed that the DEGs between the group with high TMB and the group with low TMB were related to the terms spindle, chromosomal region, kinetochore, nuclear division, chromosome segregation, cell cycle, oocyte meiosis, receptor-ligand activity, growth factor activity and ECM-receptor interaction, all of which are associated with DNA mutation and cell proliferation. Furthermore, the results of GSEA also supported this observation, showing enrichment of
Fig. 3 DEGs and its corresponding enrichment analysis. 

a Heatmap of the top 20 DEGs, 
b GO enrichment of the DEGs, 
c KEGG enrichment of the DEGs, 
d top 10 pathways of the GSEA enrichment.
pathways related to pyrimidine metabolism, DNA replication, DNA degradation and aminoacyl tRNA biosynthesis. None of the analyses showed associations with pathways related to immune mediation or response. This phenomenon is likely because the immunogenicity of prostate cancer cases with low TMB is poorer than that of lung cancer and melanoma [24]. Because of this poor immunogenicity, the numbers of neoantigens generated by prostate cancer patients may be less than those generated by high-TMB cancer patients. The subsequent immune response in patients with low TMB is probably also weaker than that of high-TMB cancer patients.

Through CytoHubba analysis of the PPI network, we identified six hub genes: PLK1, KIF2C, MELK, EXO1, CEP55 and CDK1. All the genes were correlated with DFS, and CEP55 and CDK1 were associated with OS. PLK1 and MELK have been suggested to be potential targets in prostate cancer. As PLK1 plays a critical role in the proliferation of cells, centrosome abnormalities, mediation of the cell cycle and apoptosis, it is considered a potential treatment target in prostate cancer [25]. It has been reported that targeting PLK1 can enhance the response to androgen signalling inhibitors or olaparib in CRPC [26, 27]. MELK is upregulated in prostate cancer and related to aggressiveness. Furthermore, in vitro silencing of MELK can weaken the proliferation of prostate cancer cells, and in vivo tests also proved that an inhibitor of MELK could repress the growth of prostate cancer.

Different kinds of immune cells play a pivotal role in the tumour microenvironment and are also determinants of immunotherapy efficacy. Therefore, our study explored differences in the levels of immune cells between the group with high TMB and the group with low TMB. The analysis indicated that the group with high TMB had a significantly higher proportion of CD8

![Fig. 4](See figure on previous page.)

**Fig. 4** PPI network and hub genes with their correlation to survival. a PPI network of DEGs generated by Cytoscape with hub genes showing red color, b Venn dots of intersections from five methodology involved MCC, MNC, Degree, EPC and EcCentricity. c–j Logrank analysis of hubgenes. The relation of PLK1 (c) and KIF2C (d) with overall survival in prostate cancer patients. The relation of PLK (e), KIF2C (f), MELK (g), EXO1 (h), CEP55 (i) and CDK1 (j) with disease free survival in prostate cancer patients.

![Fig. 5](See figure on previous page.)

**Fig. 5** Comparisons of 22 immune cell infiltrations between low-TMB groups and high-TMB groups.
T cells and activated CD4 memory cells than the group with low TMB. These findings were similar to those in previous reports of bladder cancer [22]. Generally, a high TMB can produce more neoantigens, eliciting a subsequent immune response. CD8 T cells are one of the determinants of antigen-specific responses. Therefore, the high TMB group with higher levels of CD8 T cells is likely to experience superior immunotherapy efficacy. However, a recent study demonstrated that metastatic CRPC with low TMB but a high density of CD8 T cells could also benefit from immune checkpoint inhibitors [28]. Our findings suggesting that TMB has an interaction with immune infiltration are only preliminary, and further validation in clinical cohorts and investigations to explore the underlying mechanisms of the correlations are needed.

The study has some limitations that should be considered. Firstly, all the data were retrospectively collected which had bias. Secondly, the results of our study were preliminary exploration and should be further tested through in vitro or in vivo experiments. Thirdly, our study lacked sub-group analysis including CRPC patients and non-CRPC patients.

**Conclusion**

In conclusion, we performed a comprehensive and systematic analysis of TMB in prostate cancer and analysed its clinical significance. Furthermore, we also identified enriched pathways of DEGs, hub genes with prognostic roles and infiltrating immune cells related to TMB. Our results elucidate the association between TMB and infiltrating immune cells in prostate malignancy and will likely be useful for future investigations of TMB in prostate cancer.

**Supplementary Information**

The online version contains supplementary material available at https://doi.org/10.1186/s12894-021-00795-7.

**Acknowledgments**

Not applicable.

**Authors’ contributions**

LJW performed data analysis work and wrote the manuscript. SCR, BBZ and ZLY edited the manuscript. WW designed the study. All authors read and approved the final manuscript.

**Funding**

The project is supported by the Natural Science Foundation of Zhejiang Province, China (LGJ19H050007) and the Medical Health Science and Technology Project of Zhejiang Provincial health Commission (2018ZD006).

**Availability of data and materials**

The datasets used and/or analyzed during the current study are available from the TCGA databank (https://portal.gdc.cancer.gov/). Transcriptome data, clinical data, and mutation data could be downloaded according to the supplementary file's detailed setting items (Additional file 1: Table S1, Table S2 and Table S3).

**Ethics approval and consent to participate**

Not applicable.

**Consent for publication**

Not applicable.

**Competing interests**

The authors declare that there are no competing interests associated with the manuscript.

Received: 13 October 2020 Accepted: 10 February 2021

Published online: 25 February 2021

**References**

1. Center MM, Jemal A, Lorett-Tieulent J, et al. International variation in prostate cancer incidence and mortality rates. Eur Urol. 2011;61(6):1079–92.  
2. Zhu Y, Ye D. Chinese Expert Consensus on the Diagnosis and Treatment of Castration-Resistant Prostate Cancer (2019 Update). Cancer Manag Res. 2020;12:2127–40.  
3. Kwon ED, Drake CG, Scher HI, et al. Ipilimumab versus placebo after radiotherapy in patients with metastatic castration-resistant prostate cancer that had progressed after docetaxel chemotherapy (CA184-045): a multicentre, randomised, double-blind, phase 3 trial. Lancet Oncol. 2014;15(7):709–12.  
4. Beer TM, Kwon ED, Drake CG, et al. Randomized, double-blind, phase III trial of ipilimumab versus placebo in asymptomatic or minimally symptomatic patients with metastatic chemotherapy-naive castration-resistant prostate cancer. J Clin Oncol. 2017;35(1):40–7.  
5. Riaz N, Morris L, Havel JJ, Makarov V, Desrichard A, Chan TA. The role of neoantigens in response to immune checkpoint blockade. Int Immunol. 2016;28(8):411–9.  
6. Snyder A, Makarov V, Merghoub T, et al. Genetic basis for clinical response to CTLA-4 blockade in melanoma. N Engl J Med. 2014;371(23):2189–99.  
7. Rizvi NA, Hellmann MD, Snyder A, et al. Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. Science. 2015;348(6230):124–8.  
8. Chalmers ZR, Connelly CF, Fabrizio D, et al. Analysis of 100,000 human cancer genomes reveals the landscape of tumor mutational burden. Genome Med. 2017;9(1):34.  
9. Szklarczyk D, Morris JH, Cook H, et al. The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. Nucleic Acids Res. 2017;45(D1):D362–8.  
10. Shannon P, Markiel A, Ozier O, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13(11):2498–504.  
11. Liu Z, Meng J, Li X, et al. Identification of Hub genes and key pathways associated with two subtypes of diffuse large B-cell lymphoma based on gene expression profiling via integrated bioinformatics. Biomed Res Int. 2018;2018:3574534.  
12. Newman AM, Liu CL, Green MR, et al. Robust enumeration of cell subsets from tissue expression profiles. Nat Methods. 2015;12(5):453–7.
13. Kandoth C, McLellan MD, Vandin F, et al. Mutational landscape and significance across 12 major cancer types. Nature. 2013;502(7471):333–9.
14. 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(15):2157–65.
15. Barbieri CE, Baca SC, Lawrence MS, et al. Exome sequencing identifies recurrent SPOP, FOXA1 and MED12 mutations in prostate cancer. Nat Genet. 2012;44(6):685–9.
16. Geng C, He B, Xu L, et al. Prostate cancer-associated mutations in speckle-type POZ protein (SPOP) regulate steroid receptor coactivator 3 protein turnover. Proc Natl Acad Sci USA. 2013;110(17):6997–7002.
17. Boysen G, Barbieri CE, Prandi D, et al. SPOP mutation leads to genomic instability in prostate cancer. Elife. 2015;4:e09207.
18. Gao N, Zhang J, Rao MA, et al. The role of hepatocyte nuclear factor-3 alpha (Forkhead Box A1) and androgen receptor in transcriptional regulation of prostatic genes. Mol Endocrinol. 2003;17(8):1484–507.
19. Alvarez-Cubero MJ, Martinez-Gonzalez LJ, Robles-Fernandez I, et al. Somatic mutations in prostate cancer: closer to personalized medicine. Mol Diagn Ther. 2017;21(2):167–78.
20. Fan C, Quan R, Feng X, et al. ATM activation is accompanied with earlier stages of prostate tumorigenesis. Biochim Biophys Acta. 2006;1763(10):1090–7.
21. Zhang C, Li Z, Qi F, Hu X, Luo J. Exploration of the relationships between tumor mutation burden with immune infiltrates in clear cell renal cell carcinoma. Ann Transl Med. 2019;7(22):648.
22. Zhang C, Shen L, Qi F, Wang J, Luo J. Multi-omics analysis of tumor mutation burden combined with immune infiltrates in bladder urothelial carcinoma. J Cell Physiol. 2020;235(4):3849–63.
23. Chabanon RM, Pedrero M, Lefebvre C, Marabelle A, Soria JC, Postel-Vinay S. Mutational landscape and sensitivity to immune checkpoint blockers. Clin Cancer Res. 2016;22(17):4309–21.
24. De Velasco MA, Uemura H. Prostate cancer immunotherapy: where are we and where are we going? Curr Opin Urol. 2018;28(1):15–24.
25. Ahmad N. Polo-like kinase (Plk) 1: a novel target for the treatment of prostate cancer. FASEB J. 2004;18(1):5–7.
26. Zhang Z, Hou X, Shao C, et al. Plk1 inhibition enhances the efficacy of androgen signaling blockade in castration-resistant prostate cancer. Cancer Res. 2014;74(22):6635–47.
27. Li J, Wang R, Kong Y, et al. Targeting Plk1 to enhance efficacy of olaparib in castration-resistant prostate cancer. Mol Cancer Ther. 2017;16(3):469–79.
28. Subudhi SK, Vence L, Zhao H, et al. Neocitigen responses, immune correlates, and favorable outcomes after ipilimumab treatment of patients with prostate cancer. Sci Transl Med. 2020;12(537):eaaz3377.

Publisher’s Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.