Asian Journal of Andrology (2020) 22, 162–168
www.asiaandro.com; www.ajandrology.com

Open Access

ORIGINAL ARTICLE

Prostate Cancer

Integrative molecular characterization of Chinese prostate cancer specimens

Shi-Dong Lv[1,2,3], Hong-Yi Wang[3], Xin-Pei Yu[4], Qi-Liang Zhai[3], Yao-Bin Wu[1,2], Qiang Wei[3], Wen-Hua Huang[1,2]

Prostate cancer (PCa) exhibits epidemiological and molecular heterogeneity. Despite extensive studies of its phenotypic and genetic properties in Western populations, its molecular basis is not clear in Chinese patients. To determine critical molecular characteristics and explore correlations between genomic markers and clinical parameters in Chinese populations, we applied an integrative genetic/transcriptomic assay that combines targeted next-generation sequencing and quantitative real-time PCR (qRT-PCR) on samples from 46 Chinese patients with PCa. Lysine (K)-specific methyltransferase 2D (*KMT2D*), zinc finger homeobox 3 (*ZFHX3*), A-kinase anchoring protein 9 (*AKAP9*), and GLI family zinc finger 1 (*GLI1*) were frequently mutated in our cohort. Moreover, a clinicopathological analysis showed that RB transcriptional corepressor 1 (*RB1*) deletion was common in patients with a high risk of disease progression. Remarkably, four genomic events, MYC proto-oncogene (*MYC*) amplification, *RB1* deletion, APC regulator of WNT signaling pathway (*APC*) mutation or deletion, and cyclin-dependent kinase 12 (*CDK12*) mutation, were correlated with poor disease-free survival. In addition, a close link between *KMT2D* expression and the androgen receptor (AR) signaling pathway was observed both in our cohort and in The Cancer Genome Atlas Prostate Adenocarcinoma (TCGA-PRAD) data. In summary, our results demonstrate the feasibility and benefits of integrative molecular characterization of PCa samples in disease pathology research and personalized medicine.

Asian Journal of Andrology (2020) **22**, 162–168; doi: 10.4103/aja.aja_36_19; published online: 28 May 2019

Keywords: androgen receptor; molecular subtyping; next generation sequencing; prostate cancer; quantitative real-time-polymerase chain reaction

INTRODUCTION

Prostate cancer (PCa) is the second most frequent malignancy in men worldwide, accounting for more than 300 000 cancer-related deaths per year. In China, approximately 6.03% of adult males were suspected to have PCa in 2015, and the annual incidence of PCa in the Chinese population is increasing by 4.7%.[1] Actionable molecular sequencing is a plausible and promising strategy for improving treatment results by targeted therapy. However, the incidence, prognosis, and treatment outcomes differ among ethnicities due to heterogeneity at the genetic and phenotypic levels.[2,3] Genetic and transcriptional alterations have been studied extensively in Western populations; however, little is known about the genomic aberrations associated with the pathogenesis of PCa in Chinese populations.

Next-generation sequencing (NGS) has enabled extensive profiling of genome-wide alterations during PCa tumorigenesis. Recent sequencing studies have revealed various somatically acquired genetic, epigenetic, and transcriptional changes in PCa, such as alterations in phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha (*PIK3CA*), tumor protein P53 (*TP53*), speckle type BTB/POZ protein (*SPOP*), chromodomain helicase DNA-binding protein 1 (*CHD1*), and external transcribed spacer (*ETS*) transcription factor family.[4] Notably, NGS approaches not only provide numerous insights into the molecular-based pathogenesis but also identify specific biomarkers for better patient stratification and targeted therapy. The hyperactivation of oncogenic signaling pathways, such as phosphatidylinositol 3-kinase (PI3K)/phosphatase and tensin homolog (PTEN)/AKT, is considered a prognostic biomarker for PCa and is closely correlated with advanced localized or metastatic disease.[5] *SPOP*, *CHD1*, and ETS transcription factor (*ERG*) statuses can be used for segregation, as they suggest future androgen receptor (AR) loss and possibly aggressive cancer development.[6] DNA repair pathway defects indicate that patients would benefit from treatment with a poly(ADP)-ribose polymerase (PARP) inhibitor that has been approved by the US Food and Drug Administration for metastatic castration-resistant PCa monotherapy.[7] Hence, discovering and validating genomic aberrations is urgently needed for PCa prediction, prognosis, and treatment in Chinese patients.

In this study, we performed a comprehensive molecular characterization of Chinese patients with PCa and explored correlations between identified genomic markers and clinical factors. We used quantitative real-time PCR (qRT-PCR) to detect the expression of focal PCa-related genes in 46 Chinese PCa samples. Moreover, we reanalyzed previously generated gene sequencing data[8] and detected copy number variants (CNVs) by identifying clustered amplicons with significant

[1]National Key Discipline of Human Anatomy, School of Basic Medical Sciences, Southern Medical University, Guangzhou 510515, China; [2]Guangdong Engineering Research Center for Translation of Medical 3D Printing Application, Guangzhou 510515, China; [3]Department of Urology, Nanfang Hospital, Southern Medical University, Guangzhou 510515, China; [4]Department of Oncology, Affiliated Cancer Hospital and Institute of Guangzhou Medical University, Guangzhou 510095, China.
Correspondence: Dr. WH Huang (huangwenhua2009@139.com) or Dr. Q Wei (qwei@smu.edu.cn)
Received: 13 September 2018; Accepted: 21 February 2019

changes in normalized $\log_2$ read depth.[9] We provide an integrative overview of genetic/transcriptomic alterations in Chinese patients with PCa; these findings will improve our understanding of disease biology and have implications for personalized medicine.

PATIENTS AND METHODS

Patients and samples

The study included a total of 46 PCa samples collected at Nanfang Hospital (Guangzhou, China). All samples were histologically confirmed as PCa by two independent pathologists. Patients who received preoperative hormone therapy, chemotherapy, and radiotherapy were excluded. Detailed information for all samples is summarized in **Supplementary Table 1**. This study was approved by the Medical Ethics Committee of Nanfang Hospital (NFEC-2017-145), and the specimens were used with written informed consent by the Department of Pathology, Nanfang Hospital, in accordance with the Declaration of Helsinki.

DNA extraction, amplicon capture, next-generation sequencing, and analysis

Samples for targeted NGS were dissected from 4-mm unstained sections, and genomic DNA was extracted using a GeneRead DNA FFPE Tissue Kit (Qiagen, Hilden, Germany) according to the manufacturer's protocol. The extracted DNA was quantified using a Qubit fluorometer (Invitrogen, Carlsbad, CA, USA). Genomic DNA with a yield of >40 ng was used for library preparation. Amplicon enrichment was performed using the prostate cancer panel within the GeneRead DNAseq Targeted Panels V2 (Qiagen). Library preparation was performed using the GeneRead DNA Library Core Kit for Ion Torrent Proton (Qiagen) and barcoded with Ion Xpress™ Barcode Adapters (Thermo Fisher, Waltham, MA, USA). Sequencing of barcoded pools was performed using the Ion Torrent Proton, and the data were evaluated using Cloud-Based DNAseq Sequence Variant Analysis (Qiagen). Called variants were filtered by Ion Reporter (Thermo Fisher) to remove synonymous or noncoding variants and University of California, Santa Cruz (UCSC) common single-nucleotide polymorphisms (SNPs). High-confidence candidate somatic drivers were included in subsequent analyses.

For CNV determination, germline DNA (extracted from peripheral blood mononuclear cells) was collected from seven independent healthy males and sequenced to obtain the CNV baseline. CNV calling was performed using Cloud-Based DNAseq Sequence Variant Analysis. For targeted genes, CNVs were defined as clustered amplicons with significant changes in normalized $\log_2$ read depth, as previously reported.[9] In particular, CNVs were evaluated as follows. 1) Amplicons were grouped into clusters of neighbors at a genomic locus (amplicon clustering). 2) The $\log_2$ read depth at each priming site was compared between sample and control datasets (filtered clustering). *P*-values were calculated by weighted *t*-tests (total read depth was used as the weight). 3) Only CNV calls with scores of >50 indicated strong evidence (CNV identification).

RNA extraction, reverse transcription, and PCR analysis

The same tissues for NGS were used to isolate RNA using the AllPrep DNA/RNA FFPE Kit (Qiagen) according to the manufacturer's protocol. The extracted RNA was quantified using a Qubit fluorometer (Thermo Fisher). Reverse transcription (RT) of 1 μg of RNA was performed using the QuantiTect Reverse Transcription Kit (Qiagen). The RT product was mixed with nuclease-free water, primers, and the SYBR Premix Ex Taq II Kit (Takara, Shiga, Japan). Quality PCR was performed using the CFX384 Touch Real-Time PCR system (Bio-Rad, Hercules, CA, USA). Data were analyzed using the $2^{-\Delta\Delta Ct}$ method, and normalized target gene expression was visualized using R statistical software (GraphPad Software, La Jolla, CA, USA).

Statistical analyses

Kaplan–Meier analysis was performed using R statistical software. Disease-free survival (DFS) was defined as the time from diagnosis to recurrence or date of last follow-up. The log-rank test was used to calculate the *P*-values or the Kaplan–Meier analysis.

A large prostate cancer dataset from the Cancer Genome Atlas (TCGA), with both gene expression and clinical follow-up data, was used for the correlation analysis. Level 3 gene expression data were downloaded from TCGA data portal (https://tcga-data.nci.nih.gov/tcga/). RNA-seq by Expectation-Maximization (RSEM) normalized expression data were obtained.

Correlations between the recurrently altered genes and clinical parameters were evaluated using Fisher's tests; gene expression correlations were evaluated based on Pearson's correlation coefficients, as determined using R statistical software.

RESULTS

Application of integrative genetic/transcriptomic assays for Chinese clinical formalin-fixed paraffin-embedded (FFPE) PCa samples

To identify the genetic and transcriptomic spectra of genes associated with the pathogenesis of PCa in Chinese patients, we performed integrative DNA- and RNA-based assays requiring only approximately 50 ng of FFPE DNA and 500 ng of FFPE RNA. The assays consist of a multiplex PCR-based panel for gene sequencing and a qRT-PCR array for the detection of gene expression. The gene sequencing panel was used for target enrichment of the exonic regions of the 32 most commonly mutated genes in human PCa samples, and the amplification or deletion of these genes was evaluated as described previously.[9] The qRT-PCR assay included 28 genes, including robust housekeeping genes, AR transcriptional modules, cell cycle-related genes, and genes defining molecular subtypes or molecular drivers based on previous transcriptomic studies (**Supplementary Table 1**). The integrative assays were performed using samples obtained from a cohort of 46 Chinese patients with PCa after radical prostatectomy or transurethral resection of the prostate (**Supplementary Table 2**).

Mutation and copy number variant detection by targeted gene sequencing

Ion Torrent NGS after multiplex PCR-based targeted enrichment was applied to 46 FFPE PCa samples, yielding an average of 6 243 762 mapped reads and an average coverage of 3527× (1156–5419×) over targeted bases per sample. The complete coverage details are provided in **Supplementary Table 3**. An initial 5909 variants were filtered for quality, SNPs, and functional relevance. As described previously,[8] we identified 245 nonsynonymous (missense, in-frame deletion, nonsense, stop-loss, and frameshift) mutations with an allele frequency of ≥3% in all sequencing specimens (average: 5; range: 1–34) across 27 genes. All high-confidence mutations are shown in a heatmap in **Figure 1** and are detailed in **Supplementary Table 4**. The most frequently mutated genes were lysine (K)-specific methyltransferase 2D (*KMT2D*, 63.04%), zinc finger homeobox 3 (*ZFHX3*, 50.00%), A-kinase anchoring protein 9 (*AKAP9*, 32.61%), GLI family zinc finger 1 (*GLI1*, 32.61%), thrombospondin type 1 domain containing 7B (*THSD7B*, 19.57%), APC regulator of WNT signaling pathway (*APC*, 15.22%), cyclin-dependent kinase 12 (*CDK12*, 15.22%), lysine demethylase 4B (*KDM4B*, 15.22%), mediator complex subunit 12 (*MED12*, 15.22%), and zinc finger protein 595 (*ZNF595*, 15.22%).

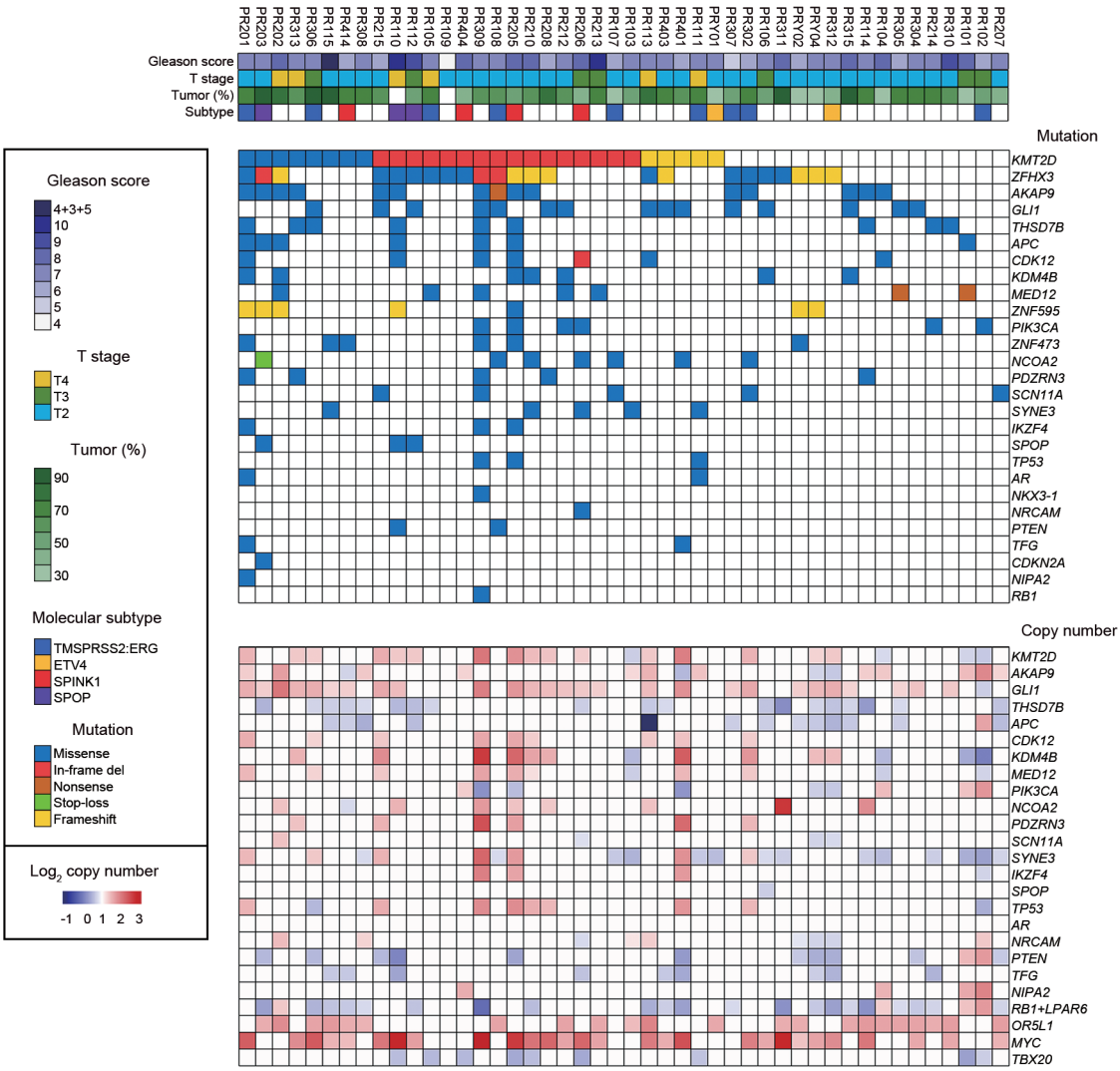


Figure 1: Integrative molecular profiling of Chinese FFPE prostate cancer (PCa) samples. A heatmap of somatic alterations for each sample is shown. Clinicopathological information is provided in the header, including Gleason score, T stage, tumor percentage, and molecular subtype based on qRT-PCR analysis. High-level somatic CNVs and mutations are indicated according to the legend. FFPE: formalin-fixed paraffin-embedded; qRT-PCR: quantitative real-time polymerase chain reaction; CNVs: copy number variants. All abbreviated names of genes are defined in **Supplementary Table 6**.

Using an approach adapted from multiplexed PCR-based targeted Ion Torrent NGS data,[9] we simultaneously assessed CNVs in 25 genes, yielding a total of 312 high-level CNVs across the 46 PCa specimens (average: 7 high-level CNVs per sample; range: 0–15). A copy number heatmap for all samples is shown in **Figure 1.**

Assessment of critical PCa transcriptomic alterations by qRT-PCR

We performed qRT-PCR using RNA isolated from the sequencing cohort of 43 FFPE PCa samples (the sequencing samples, excluding those with low RNA quality). Normalized target gene expression is shown in **Figure 2**. Combined with target sequencing, we performed basic PCa molecular subtyping for the 43 Chinese patients (**Figure 1** and **2**). *TMPRSS2:ERG* was detected in nine cancer samples, which showed approximately 70-fold greater *TMPRSS2:ERG* expression than the median of the 34 other *TMPRSS2:ERG*-negative samples. Two samples showed marked ETS variant 4 (*ETV4*) overexpression, indicating the presence of a gene fusion in this *ETS* family member. Overexpression of serine peptidase inhibitor, Kazal type 1 (*SPINK1*), was identified in four samples (the SPINK1$^+$ subtype). Moreover, by targeted gene sequencing, we observed high-confidence *SPOP* mutations in samples PR110, PR112, and PR203, each of which was *ETS*-negative by qRT-PCR. Among these samples, PR203 also showed high *SPINK1* expression, suggesting the overlap of these alterations.

AR-targeted genes were included based on previous microarray analyses of genes regulated by androgens in PCa.[10] Our results showed that the AR signature was heterogeneous in all treatment-naïve PCa specimens. AR activity might predict the response to therapies targeting the AR axis. PR315 and PRY04 had markedly lower AR activity than that of other samples, suggesting different responses to anti-androgen therapy compared with those of samples with high AR activity.

In addition, four cell cycle-related genes, nucleolar and spindle-associated protein 1 (*NUSAP1*), kinesin family member 11 (*KIF11*), cell division cycle 20 (*CDC20*), and forkhead box M1 (*FOXM1*), were evaluated as indicators of cell proliferation, a fundamental aspect of tumor biology.[11] Most of the samples showed relatively low expression of cell cycle progression genes, suggesting low aggressiveness in the

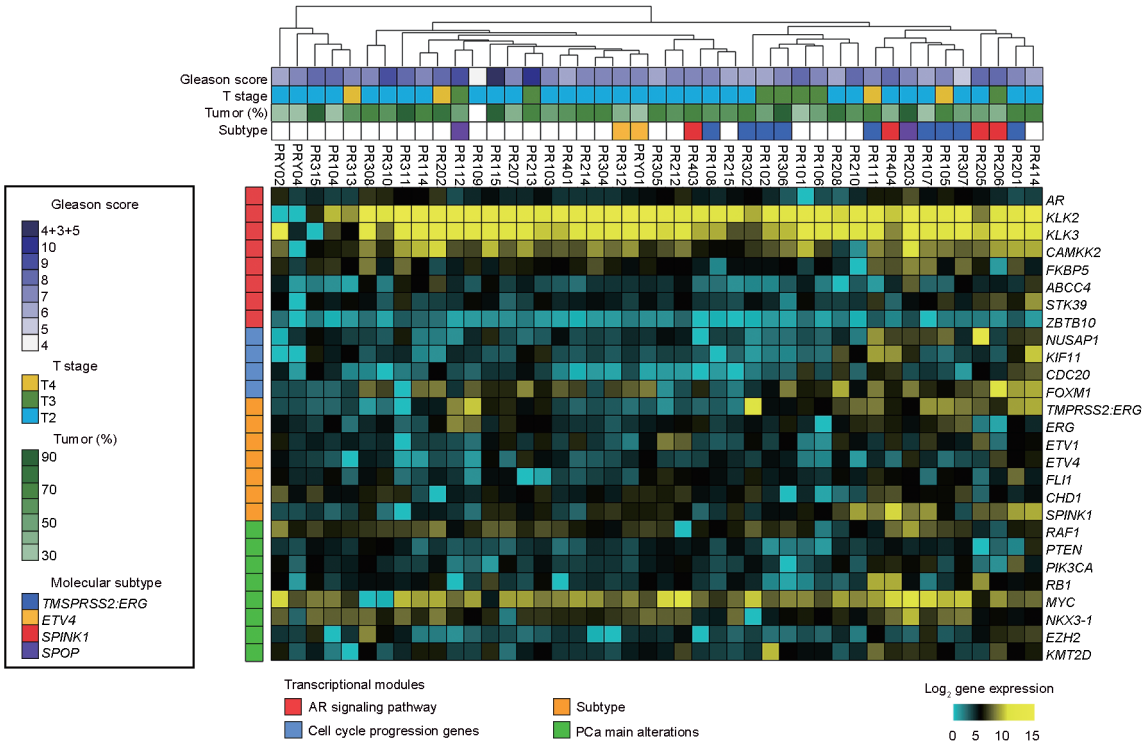


Figure 2: Gene expression signatures of PCa assessed by qRT-PCR. Unsupervised hierarchical clustering of gene expression signatures for PCa-related genes. Normalized target gene expression (log_2) for 28 robust target genes assessed by the qRT-PCR panel. qRT-PCR: quantitative real-time polymerase chain reaction; PCa: prostate cancer. All abbreviated names of genes are defined in **Supplementary Table 6**.

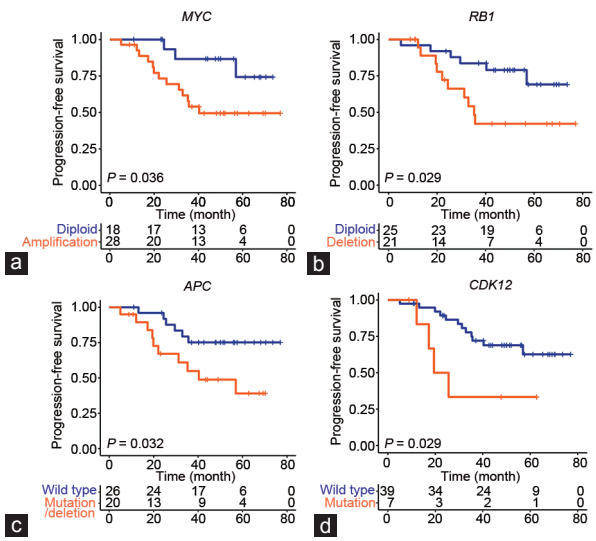


Figure 3: *MYC* amplification, *RB1* deletion, *APC* mutation/deletion, and *CDK12* mutation are linked to worse prognosis in our cohort. Survival analysis was performed in Chinese PCa patients treated according to the presence or absence of (**a**) *MYC* amplification, (**b**) *RB1* deletion, (**c**) *APC* mutation/deletion, and (**d**) *CDK12* mutation. PCa: prostate cancer; MYC: MYC proto-oncogene, bHLH transcription factor; RB1: RB transcriptional corepressor 1; APC: APC regulator of WNT signaling pathway; CDK12: Cyclin-dependent kinase 12.

early stages of PCa. Our results support the use of qRT-PCR as a complementary approach to target gene sequencing for assessing critical transcriptomic events in PCa.

Molecular features correlated with clinical characteristics and prognosis

We subdivided the genomic aberrations in Chinese PCa patients identified by integrative genetic/transcriptomic assays into nine specific pathways (**Supplementary Table 5**). These molecules are commonly altered in Western patients with PCa and have clinical implications.[12] We analyzed the correlations between these genomic aberrations and clinicopathological characteristics, including prostate-specific antigen (PSA) level, Gleason score, T stage, distant metastasis, and clinical risk stratification from the National Comprehensive Cancer Network (NCCN) and the European Association of Urology (EAU). As shown in **Supplementary Table 5**, activation of the PI3K/AKT/PTEN pathway and deletion of RB transcriptional corepressor 1 (*RB1*) were dramatically correlated with older age ($P = 0.035$ and $P = 0.041$, respectively). The *ETS* fusion was more common in patients at an advanced T stage ($P = 0.030$). Notably, an increased risk of progression was found in patients with the *RB1* deletion, both in NCCN risk stratification ($P = 0.006$) and in EAU risk classification ($P = 0.022$). In addition, recurrent mutated or copy number-altered genes were included in the correlation analysis. *GLI* mutation was notably correlated with distant metastasis ($P = 0.030$).

The prognostic impacts of the five specific pathways and recurrently altered genes were assessed. Log-rank tests were used to analyze the outcome for each individual subtype. MYC proto-oncogene (*MYC*) amplification ($P = 0.036$), *RB1* deletion ($P = 0.029$), *APC* mutation/deletion ($P = 0.032$), and *CDK12* mutation ($P = 0.029$) were associated with worse progression-free survival (PFS) in the 46 Chinese patients with PCa (**Figure 3**). Furthermore, we explored the clinical characteristics related to prognosis in Chinese populations. As expected, PFS decreased with increasing T stage ($P = 0.0024$), distant

metastasis (P = 0.0012), NCCN risk stratification (P = 0.0011), and EAU risk classification (P = 0.00039) (**Supplementary Figure 1**). These results indicated a high predictive value of these risk factors, which may contribute to clinical decision-making in Chinese patients with PCa.

KMT2D is positively correlated with the AR signaling pathway in PCa

Another advantage of the integrative genetic/transcriptomic assay is the ability to explore the correlations underlying recurrently altered genes. We performed a correlation analysis of the expression of the 27 robust target genes across all samples. We observed correlated expression of genes within the AR signaling pathway and cell cycle progression modules (**Figure 4a**). In particular, positive correlations were identified between *KMT2D* and *AR* (r = 0.31, P = 0.046), the AR downstream target calcium/calmodulin-dependent protein kinase kinase 2 (*CAMKK2*) (r = 0.45, P = 0.046), the cell cycle progression gene *CDC20* (r = 0.32, P = 0.037), and *SPINK1* (r = 0.40, P = 0.008) (**Figure 4b**). Because the AR signaling pathway is indispensable for PCa carcinogenesis, we further validated the correlation between KMT2D and AR signaling using the Cancer Genome Atlas Prostate Adenocarcinoma (TCGA-PRAD) data. Consistent with our qRT-PCR results, *KMT2D* expression was strongly correlated with *AR* mRNA levels (r = 0.70, P < 0.001; **Figure 4c** left) and AR protein levels (r = 0.32, P < 0.001; **Figure 4c** right). Patients with *AR-V7* had higher *KMT2D* expression (P = 0.036; **Figure 4d**), suggesting that KMT2D contributes to the AR splice variants. These findings highlight the close linkage between KMT2D and AR signaling and suggest that KMT2D plays a critical role in the activation of the AR axis in PCa.

DISCUSSION

We used a rapid, robust, and high-throughput approach for the characterization of gene mutations, copy number variants, and crucial transcriptomic events in Chinese patients with PCa by combined multiple-PCR-based deep NGS and qRT-PCR. This integrative assay required only 50 ng of DNA and 500 ng of RNA, which can be co-isolated from most routine clinical samples, making it practical for academic research settings as well as routine clinical settings. In this study, we added to the current knowledge of the genetic heterogeneity of prostate tumors among ethnic populations and identified correlations between clinical and prognostic factors and recurrently altered genes, providing a basis for therapeutic decisions. Moreover, we observed a close relationship between *KMT2D* and AR signaling in a gene co-expression analysis, highlighting the important role of *KMT2D* in the activation of the AR axis in PCa.

Our findings expand our understanding of the ethnicity-specific molecular landscape of PCa. As a heterogeneous disease, previous studies have reported notable disparities in PCa epidemiology among ethnic groups.[13] Emerging evidence suggests that these differences might be the result of distinct molecular characteristics.[14,15] Ren

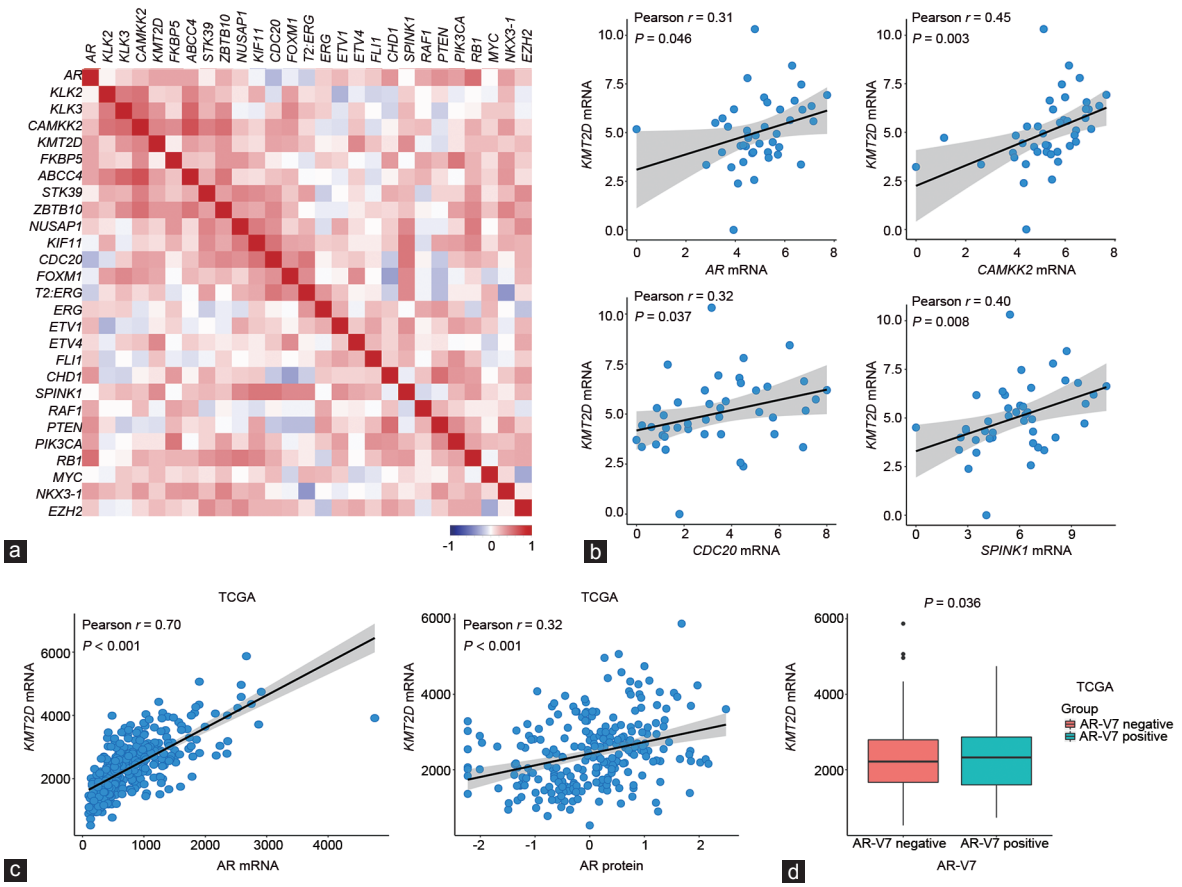


Figure 4: *KMT2D* is positively correlated with the AR signaling pathway in PCa. (**a**) Heatmap of correlations (r) between normalized gene expression for the 27 robust target genes across all samples shown in **Figure 2**. (**b**) *KMT2D* mRNA levels were positively correlated with *AR, CAMKK2, CDC20*, and *SPINK1* expression in PCa tissues in our cohort. (**c**) *KMT2D* mRNA levels were positively correlated with both gene and protein expression of AR in TCGA-PRAD. (**d**) mRNA expression of *KMT2D* in AR-V7-positive and AR-V7-negative cases from TCGA-PRAD. PCa: prostate cancer; TCGA-PRAD: The Cancer Genome Atlas Prostate Adenocarcinoma. All abbreviated names of genes are defined in **Supplementary Table 6**.

et al.[16] identified salient genetic characteristics of PCa in Chinese patients. Similarly, the molecular profiles of Chinese patients with PCa identified in our cohort differed from that of Western patients. The most frequently mutated genes were *KMT2D, ZFHX3, AKAP9*, and *GLI1* in our cohort; recurrently mutated genes in Western populations are *SPOP,* titin (*TTN*), *TP53,* mucin 16 (*MUC16*), *MED12*, and Forkhead Box A1 (*FOXA1*).[17] These results might be due to either *de novo* ethnicity-specific mutations in Chinese patients or high-depth sequencing effects. Asian and Western populations have distinct genetic polymorphisms,[18,19] familial aggregation,[20] and diet.[21] Basic science and epidemiological studies have demonstrated that these factors have a role in the clinical heterogeneity of PCa.[22] Thus, the differences in these factors might influence the tumor genome and contribute to the ethnic genetic heterogeneity of PCa. Moreover, unlike other large-scale sequencing studies in Western patients, in which the sequencing depth is about 100×, in this study, we focused on 32 PCa-correlated genes and the average sequencing depth was approximately 3527×. The higher sequencing depth can remarkably increase the sensitivity of single-nucleotide variant detection;[23] hence, the high coverage might account for the high mutation frequency of some genes in our cohort.

In addition to mutation, using an algorithm developed by Grasso *et al*.,[9] we detected CNVs in our amplicon-based sequencing data. We observed copy number gains of *KMT2D, AKAP9, MYC,* and *GLI* and copy number losses of *PTEN*, *RB1*, and *APC* in Chinese patients. *PTEN* inactivation and copy number deletions account for the majority of PTEN loss-of-function cases.[5] We found a low frequency of these genomic changes, *i.e.*, 23.9% (11/46) in our cohort, compared with approximately 40% in Western samples.[15] *MYC* amplification and *RB1* deletion are two major CNVs detected in multiple genomic analyses.[12] We also found *MYC* gains (60.9%, 28/46) and *RB1* deletions (45.7%, 21/46) in our samples. However, the frequencies were higher than those in Western populations, in which *MYC* amplification and *RB1* were deleted in 7.4% (37/498) and 16.3% (81/498) cases from TCGA-PRAD, respectively. For critical transcriptomic events, we identified a lower frequency of the *ETS* fusion (23.9%, 11/46) in our patients than in Caucasian patients (with a prevalence of approximately 50%).[24] The results of other large-scale studies further support the low frequency of the *ETS* fusion in Asian patients,[14,15,25] indicating regional differences in molecular subtypes.

Furthermore, we identified clinical associations for the recurrently altered genes. Copy number deletion of *RB1* was remarkably correlated with a high risk of disease progression. *ETS* fusion and *GLI1* mutations were more common in patients with a higher T stage and distant metastasis, respectively, than in other patients. Importantly, poor prognosis was associated with various alterations, including *MYC* amplification, *RB1* deletion, *APC* mutation or deletion, and *CDK12* mutation. MYC and RB1 are involved in the regulation of the cell cycle; amplification of *MYC* prompts progression from G1 to S phase,[17,26] and *MYC* amplification status is predictive of biochemical recurrence after prostatectomy.[27] However, its mRNA or protein levels show weak prognostic ability.[28] RB1 is a critical negative regulator of the G1–S checkpoint and is responsible for repressing E2F family transcription factors.[29] According to a recent study, *RB1* deletion occurs early in PCa, prior to subclonal diversification.[30] Consistent with our results, a large PCa tissue microarray study indicated that the deletion of *RB1* is linked to an adverse phenotype and poor prognosis in this disease.[31] APC is a multifunctional protein and is as a negative regulator of the Wnt pathway.[32] The inactivation of APC by mutations or hypermethylation is associated with tumor aggressiveness.[33,34] In Chinese populations, Geng *et al*.[35] showed that *APC* mutations are associated with both PCa progression and all-cause mortality. CKD12 is a cyclin-dependent kinase; however, Wu *et al*.[36] found that inactivation of CDK12 could define a distinct subtype of PCa, leading to increased neoantigen burden and T cell infiltration. Moreover, the loss of CDK12 is associated with genomic instability and predicts a worse poor prognosis in breast cancer.[37] These findings further indicated that these genomic events are associated with clinical parameters and might serve as biomarkers in Chinese patients. However, given the small sample size, an independent cohort of patients is needed to confirm these results in the future.

Our integrative genetic/transcriptomic assays might also have utility for personalized medicine. The five main pathways involved in our analysis have clinical implications, and drugs targeting these alterations have been developed.[12] For instance, the PI3K/AKT/PTEN pathway was upregulated in 34.78% of samples in our cohort. Therapeutically, to target this pathway, pan-PI3K and dual PI3K–mTOR inhibitors have been developed, and the AKT-targeting drugs BKM120 and MK-2206 show clinical efficacy in PCa.[6] Cancer immunotherapies have become a revolutionized approach to the treatment of cancers, including PCa.[38] The inactivation of CDK12 delineates a distinct immunogenic class of PCa, and a clinical study showed that patients with metastatic PCa who harbor a biallelic *CDK12* loss have a higher likelihood of response to immunotherapy.[36] In our cohort, 15.22% (7/46) of patients had *CDK12* mutations. Among them, PR110, PR113, PR206, and PR309 were in an advanced state and relapsed after radical prostatectomy. These patients might benefit from immune checkpoint inhibitors after progression to the final stage of this disease. In addition, we evaluated AR signaling pathway activity. Consistent with previous results, most of our treatment-naïve cases showed high AR activity, indicating a better response to anti-androgen therapy.[10] Compared with other cases, PR305 and PRY04 showed decreased AR activity, suggesting a heterogeneous response. Therefore, careful attention is needed when administering hormonal therapy. Above all, our findings suggest that molecular profiling using a targeted panel could provide valuable information, *i.e.*, personalized mutation profiles for actionable targets, thereby contributing to clinical treatment decisions.

Notably, we detected a close link between *KMT2D* and the AR signaling pathway in our cohort using the integrative assays. These results were further confirmed using an additional dataset, TCGA-PRAD, showing that KMT2D expression was positively correlated with AR expression at the mRNA and protein level. Moreover, AR-V7-positive patients had high KMT2D expression levels. KMT2D is a histone lysine methyltransferase involved in the monomethylation of H3K4 residues (histone H3 lysine 4-monomethylation, H3K4me1). We previously concluded that, in PCa, KMT2D epigenetically upregulates LIF receptor alpha (*LIFR*) and Kruppel-like factor 4 (*KLF4*) expression, activating PI3K/AKT and epithelial–mesenchymal transition (EMT) oncogenic pathways and promoting PCa outgrowth and metastasis.[8] Recent studies have reported that KMT2D is a transcriptional coactivator that recognizes target genes via DNA-bound transcription factors.[39] In breast cancer, KMT2D is required for FOXA1, PBX homeobox 1 (PBX1), and estrogen receptor (ER) recruitment and activation and contributes to resistance to PI3Kα inhibitors.[40] Thus, the positive correlation between KMT2D and the AR signaling pathway raises the possibility that KMT2D serves as a transcriptional coactivator and controls AR activation through posttranslational modification of epigenetic regulators. Further research is required to comprehensively clarify the role of KMT2D in PCa.

CONCLUSION

We applied an integrative approach to identify the critical genomic and transcriptomic events in 46 routine clinical Chinese PCa samples.

This approach is helpful in characterizing retrospective material and analyzing correlations between clinical parameters and molecular changes. More importantly, focused assessments of genomic and transcriptomic features are important for disease pathology research as well as for advancements in personalized medicine.

AUTHOR CONTRIBUTIONS

SDL carried out the study design, data acquisition, statistical analysis, and interpretation and drafted and revised the manuscript. HYW carried out the data acquisition and statistical analysis. XPY participated in the study design and coordination and helped draft the manuscript. QLZ helped data acquisition. YBW provided the important suggestions for manuscript revision. QW and WHH supervised the study and provided the fund support. All authors read and approved the final manuscript.

COMPETING INTERESTS

All authors declared no competing interests.

ACKNOWLEDGMENTS

This work was mainly supported by the National Key R&D Program of China (No. 2017YFC1103403), the National Natural Science Foundation of China (No. 21773199, No. 61427807 and No. 81872092), the Sanming Project of Medicine in Shenzhen (No. SZSM201612019), and the PhD Start-up Fund of Natural Science Foundation of Guangdong Province of China (No. 2016A030310287).

Supplementary Information is linked to the online version of the paper on the *Asian Journal of Andrology* website.

REFERENCES

1 Chen W, Zheng R, Baade PD, Zhang S, Zeng H, *et al*. Cancer statistics in China, 2015. *CA Cancer J Clin* 2016; 66: 115–32.
2 Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, *et al*. Global cancer statistics, 2012. *CA Cancer J Clin* 2015; 65: 87–108.
3 Kimura T. East meets West: ethnic differences in prostate cancer epidemiology between East Asians and Caucasians. *Chin J Cancer* 2012; 31: 421–9.
4 Cancer Genome Atlas Research Network. The molecular taxonomy of primary prostate cancer. *Cell* 2015; 163: 1011–25.
5 Barbieri CE, Bangma CH, Bjartell A, Catto JW, Culig Z, *et al*. The mutational landscape of prostate cancer. *Eur Urol* 2013; 64: 567–76.
6 Yadav SS, Li J, Lavery HJ, Yadav KK, Tewari AK. Next-generation sequencing technology in prostate cancer diagnosis, prognosis, and personalized treatment. *Urol Oncol* 2015; 33: 267.e1–13.
7 Mateo J, Boysen G, Barbieri CE, Bryant HE, Castro E, *et al*. DNA repair in prostate cancer: biology and clinical implications. *Eur Urol* 2017; 71: 417–25.
8 Lv S, Ji L, Chen B, Liu S, Lei C, *et al*. Histone methyltransferase KMT2D sustains prostate carcinogenesis and metastasis via epigenetically activating LIFR and KLF4. *Oncogene* 2017; 37: 1354–68.
9 Grasso C, Butler T, Rhodes K, Quist M, Neff TL, *et al*. Assessing copy number alterations in targeted, amplicon-based next-generation sequencing data. *J Mol Diagn* 2015; 17: 53–63.
10 Mendiratta P, Mostaghel E, Guinney J, Tewari AK, Porrello A, *et al*. Genomic strategy for targeting therapy in castration-resistant prostate cancer. *J Clin Oncol* 2009; 27: 2022–9.
11 Cuzick J, Swanson GP, Fisher G, Brothman AR, Berney DM, *et al*. Prognostic value of an RNA expression signature derived from cell cycle proliferation genes in patients with prostate cancer: a retrospective study. *Lancet Oncol* 2011; 12: 245–55.
12 Spratt DE, Zumsteg ZS, Feng FY, Tomlins SA. Translational and clinical implications of the genetic landscape of prostate cancer. *Nat Rev Clin Oncol* 2016; 13: 597–610.
13 Zhau HE, Li Q, Chung LW. Interracial differences in prostate cancer progression among patients from the United States, China and Japan. *Asian J Androl* 2013; 15: 705–7.
14 Ren S, Wei GH, Liu D, Wang L, Hou Y, *et al*. Whole-genome and transcriptome sequencing of prostate cancer identify new genetic alterations driving disease progression. *Eur Urol* 2017. Doi: 10.1016/j.eururo.2017.08.027. [Epub ahead of print].
15 Mao X, Yu Y, Boyd LK, Ren G, Lin D, *et al*. Distinct genomic alterations in prostate cancers in Chinese and Western populations suggest alternative pathways of prostate carcinogenesis. *Cancer Res* 2010; 70: 5207–12.
16 Ren S, Wei GH, Liu D, Wang L, Hou Y, *et al*. Whole-genome and transcriptome sequencing of prostate cancer identify new genetic alterations driving disease progression. *Eur Urol* 2018; 48: 86.
17 Fraser M, Sabelnykova VY, Yamaguchi TN, Heisler LE, Livingstone J, *et al*. Genomic hallmarks of localized, non-indolent prostate cancer. *Nature* 2017; 541: 359–64.
18 Edwards A, Hammond HA, Jin L, Caskey CT, Chakraborty R. Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups. *Genomics* 1992; 12: 241–53.
19 Irvine RA, Yu MC, Ross RK, Coetzee GA. The CAG and GGC microsatellites of the androgen receptor gene are in linkage disequilibrium in men with prostate cancer. *Cancer Res* 1995; 55: 1937–40.
20 Fukabori Y, Ohtake N, Nakata S, Suzuki K. [Characteristics of hereditary and familial prostate cancer in Japan 2011]. *Nihon Rinsho* 2011; 69 Suppl 5: 197–202. [Article in Japanese].
21 Norie K, Motoki I, Manami I, Shizuka S, Shoichiro T. Plasma isoflavones and subsequent risk of prostate cancer in a nested case-control study: the Japan Public Health Center. *J Clin Oncol* 2008; 26: 5923–9.
22 Kazuto I. Prostate cancer in Asian men. *Nat Rev Urol* 2014; 11: 197–212.
23 Ming-Tseh L, Mosier SL, Michele T, Beierl KF, Marija D, *et al*. Clinical validation of *KRAS, BRAF,* and *EGFR* mutation detection using next-generation sequencing. *Am J Clin Pathol* 2014; 141: 856–66.
24 Magigalluzzi C, Tsusuki T, Elson P, Simmerman K, Lafargue C, *et al*. *TMPRSS2-ERG* gene fusion prevalence and class are significantly different in prostate cancer of Caucasian, African-American and Japanese patients. *Prostate* 2011; 71: 489–97.
25 Miyagi Y, Sasaki T, Fujinami K, Sano J, Senga Y, *et al*. *ETS* family-associated gene fusions in Japanese prostate cancer: analysis of 194 radical prostatectomy samples. *Mod Pathol* 2010; 23: 1492.
26 Boutros PC, Fraser M, Harding NJ, De BR, Trudel D, *et al*. Spatial genomic heterogeneity within localized, multifocal prostate cancer. *Nat Genet* 2015; 47: 736–45.
27 Fromont G, Godet J, Peyret A, Irani J, Celhay O, *et al*. 8q24 amplification is associated with Myc expression and prostate cancer progression and is an independent predictor of recurrence after radical prostatectomy. *Hum Pathol* 2013; 44: 1617–23.
28 Pettersson A, Gerke T, Penney KL, Lis RT, Stack EC, *et al*. MYC overexpression at the protein and mRNA level and cancer outcomes among men treated with radical prostatectomy for prostate cancer. *Cancer Epidemiol Biomarkers Prev* 2018; 27: 201–7.
29 Knudsen ES, Knudsen KE. Tailoring to RB: tumour suppressor status and therapeutic response. *Nat Rev Cancer* 2008; 8: 714–24.
30 Smg E, Liu LY, Rubanova Y, Bhandari V, Holgersen EM, *et al*. The evolutionary landscape of localized prostate cancers drives clinical aggression. *Cell* 2018; 173:1003–13.e15.
31 Kluth M, Scherzai S, Büschek F, Fraune C, Möller K, *et al*. 13q deletion is linked to an adverse phenotype and poor prognosis in prostate cancer. *Genes Chromosomes Cancer* 2018; 57: 504–12.
32 Näthke IS. The adenomatous polyposis coli protein: the Achilles heel of the gut epithelium. *Annu Rev Cell Dev Biol* 2004; 20: 337–66.
33 Cho NY, Kim JH, Moon KC, Kang GH. Genomic hypomethylation and CpG island hypermethylation in prostatic intraepithelial neoplasm. *Virchows Arch* 2009; 454: 17–23.
34 Costa VL, Henrique R, Jerónimo C. Epigenetic markers for molecular detection of prostate cancer. *Dis Markers* 2007; 23: 31–41.
35 Geng JH, Lin VC, Yu CC, Huang CY, Yin HL, *et al*. Inherited variants in Wnt pathway genes influence outcomes of prostate cancer patients receiving androgen deprivation therapy. *Int J Mol Sci* 2016; 17: 1970.
36 Wu YM, Cieślik M, Lonigro RJ, Vats P, Reimers MA, *et al*. Inactivation of CDK12 delineates a distinct immunogenic class of advanced prostate cancer. *Cell* 2018; 173: 1770–82.e14.
37 Guffanti F, Fruscio R, Rulli E, Damia G. The impact of DNA damage response gene polymorphisms on therapeutic outcomes in late stage ovarian cancer. *Sci Rep* 2016; 6: 38142.
38 Sharma P, Allison JP. Immune checkpoint targeting in cancer therapy: toward combination strategies with curative potential. *Cell* 2015; 161: 205–14.
39 Kim DH, Kim J, Kwon JS, Sandhu J, Tontonoz P, *et al*. Critical roles of the histone methyltransferase MLL4/KMT2D in murine hepatic steatosis directed by ABL1 and PPARγ2. *Cell Rep* 2016; 17: 1671–82.
40 Toska E, Osmanbeyoglu HU, Castel P, Chan C, Hendrickson RC, *et al*. PI3K pathway regulates ER-dependent transcription in breast cancer through the epigenetic regulator KMT2D. *Science* 2017; 355: 1324–30.

Supplementary Table 1: Primer sequences used in our quantitative real-time polymerase chain reaction analysis

Gene	*Type*	*Forward primer*	*Reverse primer*
AR	AR pathway	CCAGATCAGGGTTGAAGAGAAA	ATATGCTGGGTGACAAGAAGAG
KLK2	AR pathway	GATAAGGCCGTGAGCAGAAA	GACCTGAACAAACCTCCTGTAA
KLK3	AR pathway	GCTGGGAACTGCTATCTGTTAT	AGCAGGGAGAGAGTGAGATAG
CAMKK2	AR pathway	CCCTAGACTCCACACAATAACC	AAGCCCTGGTTTCCTCATAC
FKBP5	AR pathway	GTCTCCCACGTGTGTATTATGG	AATGGGCACCCTGTAGTTATTT
ABCC4	AR pathway	GGAACTCCCACACTAAGGAATC	CTCTCCAGAGCACCATCTTTC
STK39	AR pathway	CAGTTGAGTGTCAGCTGATGTA	GGAGGAAATGGGCAGAAAGA
ZBTB10	AR pathway	GACATGGATCTCGACGTTATGG	CATCCCGAGGGAATTCTGTATC
NUSAP1	Cell cycle	CCTCTTGTGATGAGACTGAGATAC	CAGTCTTGCACCTTCTCCTT
KIF11	Cell cycle	CCGTTCTGGAGCTGTTGATAA	TGTTCTTTCTACAAGGGCAGTAA
CDC20	Cell cycle	TTTGGCCAGTGGTGGTAATG	CCTTGATGCTGGGTGAATGT
FOXM1	Cell cycle	GGGAACAACAAAGGCAATGG	GCAAAGATCAGGGAAGGTAAGA
TMPRSS2: ERG	Subtype	CTGGAGCGCGGCAGGAA	GTAGGCACACTCAAACAACGACTGG
ERG	Subtype	GAAGACTGGACTCAGGACATTT	GCTGTGTCCTTTCTCCTAACA
ETV1	Subtype	GGTGGGTGGGAGTAATCTAAAC	CCCATCCCAAGCCTAAGTAAA
ETV4	Subtype	GAAGAAAGGAACCTGGGATGAG	CCTAGGAAAGGGCAGAAGAAAG
FLI1	Subtype	CATCTCCTACATGCCTTCCTAC	TGATGCGGCTCCAAAGAA
CHD1	Subtype	TGGCAAAGGGAGAGAATATGG	CAACCTTGGCAGGAAGAGAT
SPINK1	PCa main alterations	GCTTCTGAAGAGACGTGGTAAG	CCAAGGCACTGAGAAGAAAGA
RAF1	PCa main alterations	CTCGTGGACAGAGAGATTCAAG	CTCCGTGCCATTTACCCTTAT
PTEN	PCa main alterations	CTGCCAGCTAAAGGTGAAGATA	ATCACCACACACAGGTAACG
PIK3CA	PCa main alterations	CTGGTTCAGCAGTGTGGTAA	CCTGCGTGGGAATAGCTAAA
RB1	PCa main alterations	CTGTCTGAGCACCCAGAATTAG	GTCCAAATGCCTGTCTCTCAT
MYC	PCa main alterations	CATACATCCTGTCCGTCCAAG	GAGTTCCGTAGCTGTTCAAGT
NKX3-1	PCa main alterations	CGGAGACCCAAGTGAAGATATG	CAAAGAGGAGTGCTTCTCCAA
EZH2	PCa main alterations	CTCGGTGTCAAACACCAATAAAG	AGTGCCAATGAGGACTCTAAAC
KMT2D	PCa main alterations	ACCATGTGAAGAACAGGAAGAG	TCACCCTGGCTCAGATTAGA
GAPDH	House keeping	CTCCTCACAGTTGCCATGTA	GTTGAGCACAGGGTACTTTATTG

AR: androgen receptor; PTEN: phosphatase and tensin homolog

Supplementary Table 2: 46 Chinese prostate cancer samples assessed by target gene sequence and quantitative real-time polymerase chain reaction

Sample ID	*Target NGS*	*RT-qPCR*	*Age*	*Site*	*Percentage tumor*	*Gleason score*	*TNM stage*	*tPSA*	*Disease free status*	*NCCN risk stratification*	*EAU risk classification*
PR101	Yes	Yes	68	RP	30	8	T3aN0Mx	158.69	Disease free	High	Locally advanced
PR102	Yes	Yes	73	RP	50	6	T3aN0Mx	>50	Recurred	High	Locally advanced
PR103	Yes	Yes	66	RP	50	7	T2aN0M0	16.9	Disease free	Favorable intermediate	Intermediate risk
PR104	Yes	Yes	61	RP	30	8	T2aN0M0	7.027	Disease free	High	High risk
PR105	Yes	Yes	49	RP	70	7	T4N0Mx	>100	Disease free	Very high	Locally advanced
PR106	Yes	Yes	68	RP	50	7	T3aN0Mx	>100	Recurred	High	Locally advanced
PR107	Yes	Yes	71	RP	30	6	T2aN0M0	9.3	Disease free	Low	Low risk
PR108	Yes	Yes	67	RP	60	7	T2aN0M0	9.85	Disease free	Favorable intermediate	Intermediate risk
PR109	Yes	Yes	67	RP	70	4	T2aN0Mx	3.143	Disease free	Low	Low risk
PR110	Yes	No*	67	RP	80	10	T4N0Mx	21.523	Recurred	Very high	Locally advanced
PR111	Yes	Yes	68	RP	70	7	T4N0Mx	9.101	Recurred	Very high	Locally advanced
PR112	Yes	Yes	71	RP	50	9	T3bNxM1	17.95	Recurred	Metastatic	Metastatic
PR113	Yes	No*	66	RP	80	7	T4N0Mx	9.66	Recurred	Very high	Locally advanced
PR114	Yes	Yes	67	RP	70	7	T2bNxM0	43.57	Disease free	High	High risk
PR115	Yes	Yes	63	RP	90	4+3+5	T2cN0M0	20.37	Disease free	High	High risk
PR201	Yes	Yes	73	RP	70	7	T2bN0M0	18.58	Disease free	Unfavorable intermediate	Intermediate risk
PR202	Yes	Yes	54	RP	80	8	T4N0Mx	114.93	Recurred	Very high	Locally advanced
PR203	Yes	Yes	68	RP	90	7	T2cN0M0	8.76	Disease free	Unfavorable intermediate	High risk
PR205	Yes	Yes	65	RP	50	8	T2cN0M0	13.86	Disease free	High	High risk
PR206	Yes	Yes	73	RP	40	8	T3bN0M0	35.16	Recurred	Very high	Locally advanced
PR207	Yes	Yes	62	RP	40	7	T2cN0M0	6.41	Recurred	Unfavorable intermediate	High risk

Contd...

Supplementary Table 2: Contd...

Sample ID	*Target NGS*	*RT-qPCR*	*Age*	*Site*	*Percentage tumor*	*Gleason score*	*TNM stage*	*tPSA*	*Disease free status*	*NCCN risk stratification*	*EAU risk classification*
PR208	Yes	Yes	49	RP	80	6	T2cN0M0	123.12	Recurred	High	High risk
PR210	Yes	Yes	73	RP	60	8	T2bN0M0	22.9	Recurred	High	High risk
PR212	Yes	Yes	60	RP	60	7	T2cN0M0	43.03	Disease free	High	High risk
PR213	Yes	Yes	69	RP	70	10	T3aN0M0	3.79	Disease free	Very high	Locally advanced
PR214	Yes	Yes	57	RP	70	7	T2bN0M0	12	Disease free	Favorable intermediate	Intermediate risk
PR215	Yes	Yes	65	RP	60	6	T2cN0M0	8.11	Disease free	Favorable intermediate	High risk
PR302	Yes	Yes	60	RP	70	6	T2aN0M0	10.88	Disease free	Favorable intermediate	Intermediate risk
PR304	Yes	Yes	75	RP	70	7	T2bN0M0	4.65	Disease free	Unfavorable intermediate	Intermediate risk
PR305	Yes	Yes	80	RP	70	6	T2bN0M0	7.01	Disease free	Favorable intermediate	Intermediate risk
PR306	Yes	Yes	66	RP	90	7	T3aN0M0	98.16	Disease free	High	Locally advanced
PR307	Yes	Yes	67	RP	50	5	T2bN0M0	5.81	Disease free	Favorable intermediate	Intermediate risk
PR308	Yes	Yes	56	RP	70	7	T2aN0M0	13.7	Disease free	Favorable intermediate	Intermediate risk
PR309	Yes	No*	59	RP	60	7	T2N1M1	7.69	Recurred	Metastatic	Metastatic
PR310	Yes	Yes	63	RP	60	9	T2cN0M0	9.22	Recurred	High	High risk
PR311	Yes	Yes	63	RP	95	8	T2cN1M0	67.61	Disease free	Regional	Locally advanced
PR312	Yes	Yes	52	RP	40	7	T2cN0M0	19.22	Recurred	Unfavorable intermediate	High risk
PR313	Yes	Yes	77	TURP	65	7	T4N1Mx	157.86	Disease free	Regional	Locally advanced
PR315	Yes	Yes	77	TURP	95	8	T2cNxM1	>500	Disease free	Metastatic	Metastatic
PR401	Yes	Yes	67	RP	60	6	T2aN0M0	9.27	Disease free	Low	Low risk
PR403	Yes	Yes	49	RP	70	6	T2bN0M0	9.21	Disease free	High	High risk
PR404	Yes	Yes	44	RP	70	7	T2bN0M0	24.4	Disease free	High	High risk
PR414	Yes	No	72	RP	40	8	T2cN0M0	91.48	Recurred	Favorable intermediate	Intermediate risk
PRY01	Yes	Yes	59	RP	30	7	T2N0M0	14.082	Disease free	Favorable intermediate	Intermediate risk
PRY02	Yes	Yes	63	RP	30	6	T2bN0M0	12.18	Disease free	Favorable intermediate	Intermediate risk
PRY04	Yes	Yes	72	RP	30	7	T2cN0M0	20.2	Recurred	High	High risk

*The samples didn't perform gene expression analysis by qRT-PCR due to low RNA quality. Specimen type: RRP: radical retropubic prostatectomy; TURP: transurethral resection of the prostate. qRT-PCR: quantitative real-time polymerase chain reaction; NGS: next-generation sequencing; TNM: Tumor-node-metastasis; tPSA: total PSA; RP: radical prostatectomy; NCCN: National Comprehensive Cancer Network; EAU: European Association of Urology

Supplementary Table 3: Sequencing statistics for Chinese prostate cancer samples assessed on target gene sequence

Sample ID	*Total reads*	*Read depth*	*Percentage of bases covered at ≥100×*	*SNPs/MNPs*	*Insertions/deletions*
PR101	5068352	2830	98	98	6
PR102	5947960	3373	98	104	4
PR103	5286208	3106	98	105	6
PR104	6744522	3889	98	95	8
PR105	6929606	4076	99	109	6
PR106	5894989	3467	98	105	6
PR107	5911356	3381	98	112	9
PR108	5918910	3321	98	117	11
PR109	6543896	3709	98	104	5
PR110	5653814	3064	96	111	7
PR111	6314382	3550	98	114	10
PR112	4459067	2496	98	111	9
PR113	6677205	3636	98	150	11
PR114	5470762	3094	98	100	7
PR115	9498650	5419	99	126	8
PR201	3145871	1584	92	176	9
PR202	2801865	1572	95	129	9
PR203	2633500	1488	98	114	7
PR205	9233792	4882	96	153	11
PR206	7980130	4619	98	121	10
PR207	7660193	4497	99	141	11
PR208	8228356	4574	98	112	9
PR210	7929129	4429	98	121	11
PR212	7932509	4595	98	123	11

Contd...

Supplementary Table 3: Contd..

Sample ID	*Total reads*	*Read depth*	*Percentage of bases covered at ≥100×*	*SNPs/MNPs*	*Insertions/deletions*
PR213	7707601	4476	98	121	14
PR214	6895340	4048	98	117	7
PR215	8254890	4630	97	107	8
PR302	5687529	3129	96	113	7
PR304	7384117	4227	98	134	8
PR305	5473565	3106	98	127	9
PR306	6107461	3432	97	105	7
PR307	6090712	3486	98	139	5
PR308	6220821	3516	98	102	5
PR309	7586732	4021	95	206	8
PR310	4196760	2343	97	116	5
PR311	6496237	3710	98	119	9
PR312	6120013	3559	98	126	6
PR313	6981974	3865	97	137	8
PR315	5432214	3133	98	102	5
PR401	5855582	3217	95	121	7
PR403	5291636	2727	98	111	10
PR404	9418627	5355	98	130	10
PR414	7055318	4060	99	106	6
PRY01	8761095	5121	99	146	13
PRY02	1989740	1156	98	104	10
PRY04	2340079	1305	97	94	7

SNPs: single-nucleotide polymorphisms; MNPs: multiple-nucleotide polymorphisms

Supplementary Table 4: High confidence somatic nonsynonymous mutations identified by target gene sequence

Sample	*Location*	*Gene*	*Variant classification*	*Reference allele*	*Variant allele*	*AA change*	*Filtered coverage*	*Variant allele frequency*	*RefSeq*	*dbSNP ID*	*COSMIC ID*
PR101	chrX:70360666	MED12	Stop_Gained	C	T	p.Q2076*	52	0.077	NM_005120.2		
PR101	chr5:112162851	APC	Non_Synonymous_Coding	G	A	p.M485I	5924	0.469	NM_001127510.2		
PR102	chr3:178922301	PIK3CA	Non_Synonymous_Coding	G	A	p.R357Q	3977	0.137	NM_006218.2		COSM276752
PR103	chr12:49446101	KMT2D	Codon_Change_Plus_Coding	TTCCTCAGGTGGTGGGGACAGGGGTGAC	T	p.ESPLSPPPEE446E	2846	0.223	NM_003482.3		
PR103	chr12:49432245	KMT2D	Non_Synonymous_Coding	T	C	p.N2965S	6929	0.489	NM_003482.3		
PR103	chr14:95916416	SYNE3	Non_Synonymous_Coding	C	T	p.R434H	2238	0.492	NM_152592.3	rs183604014	
PR104	chr7:91731945	AKAP9	Non_Synonymous_Coding	G	A	p.R3712Q	4052	0.456	NM_005751.4	rs186148498	
PR104	chr17:37673806	CDK12	Non_Synonymous_Coding	C	T	p.S987F	8795	0.479	NM_016507.2	rs201661022	
PR105	chr12:49446101	KMT2D	Codon_Change_Plus_Coding	TTCCTCAGGTGGTGGGGACAGGGGTGAC	T	p.ESPLSPPPEE446E	5408	0.224	NM_003482.3		
PR105	chr16:72828867	ZFHX3	Non_Synonymous_Coding	T	C	p.M2572V	4624	0.51	NM_006885.3	rs147672861	
PR105	chrX:70360579	MED12	Non_Synonymous_Coding	A	G	p.I2047V	548	0.969	NM_005120.2		
PR106	chr12:57865558	GLI1	Non_Synonymous_Coding	G	T	p.G1012V	312	0.417	NM_005269.2	rs2229300	
PR106	chr16:72984669	ZFHX3	Non_Synonymous_Coding	G	A	p.S972L	942	0.499	NM_006885.3	rs77124117	
PR106	chr19:5131419	KDM4B	Non_Synonymous_Coding	G	A	p.A550T	1370	0.503	NM_015015.2		
PR107	chr12:49421002	KMT2D	Non_Synonymous_Coding	G	A	p.P4916L	110	0.073	NM_003482.3		
PR107	chr15:23019819	NIPA2	Non_Synonymous_Coding	C	G	p.W59C	34	0.118	NM_030922.6		
PR107	chr12:49446101	KMT2D	Codon_Change_Plus_Coding	TTCCTCAGGTGGTGGGGACAGGGGTGAC	T	p.ESPLSPPPEE446E	4141	0.205	NM_003482.3		
PR107	chr3:38888862	SCN11A	Non_Synonymous_Coding	A	G	p.C1567R	4641	0.467	NM_014139.2	rs201595463	
PR108	chr7:91715706	AKAP9	Stop_Gained	T	A	p.C3063*	420	0.043	NM_005751.4		
PR108	chr10:89653841	PTEN	Non_Synonymous_Coding	A	G	p.R47G	91	0.077	NM_000314.4		COSM28887
PR108	chr12:49446101	KMT2D	Codon_Change_Plus_Coding	TTCCTCAGGTGGTGGGGACAGGGGTGAC	T	p.ESPLSPPPEE446E	4216	0.228	NM_003482.3		
PR108	chr7:91731945	AKAP9	Non_Synonymous_Coding	G	A	p.R3712Q	2572	0.478	NM_005751.4	rs186148498	
PR108	chr12:57865558	GLI1	Non_Synonymous_Coding	G	T	p.G1012V	365	0.499	NM_005269.2	rs2229300	
PR108	chr7:91646406	AKAP9	Non_Synonymous_Coding	G	A	p.R1276Q	2288	0.505	NM_005751.4	rs146797353	
PR108	chr8:71037059	NCOA2	Non_Synonymous_Coding	T	A	p.T1320S	1423	0.515	NM_006540.2		
PR108	chr16:72831357	ZFHX3	Coding_Deletion	CTTGTTG	C	p.QQ1740-	2233	0.563	NM_006885.3		COSM1731926
PR109	chr12:49446101	KMT2D	Codon_Change_Plus_Coding	TTCCTCAGGTGGTGGGGACAGGGGTGAC	T	p.ESPLSPPPEE446E	5095	0.242	NM_003482.3		
PR109	chr16:72829385	ZFHX3	Non_Synonymous_Coding	T	C	p.N2399S	5581	0.467	NM_006885.3		
PR110	chr12:49421003	KMT2D	Non_Synonymous_Coding	G	A	p.P4916S	148	0.061	NM_003482.3		
PR110	chr7:91682269	AKAP9	Non_Synonymous_Coding	T	A	p.S1866R	113	0.071	NM_005751.4		
PR110	chr16:72828408	ZFHX3	Non_Synonymous_Coding	C	A	p.A2725S	5104	0.073	NM_006885.3		
PR110	chr17:37619359	CDK12	Non_Synonymous_Coding	T	A	p.S345R	32	0.094	NM_016507.2		
PR110	chr17:47696426	SPOP	Non_Synonymous_Coding	A	C	p.F133V	3470	0.223	NM_001007230.1		COSM219965
PR110	chr12:49446101	KMT2D	Codon_Change_Plus_Coding	TTCCTCAGGTGGTGGGGACAGGGGTGAC	T	p.ESPLSPPPEE446E	3688	0.368	NM_003482.3		
PR110	chr7:91712512	AKAP9	Non_Synonymous_Coding	A	G	p.Q2730R	4012	0.44	NM_005751.4	rs80191629	
PR110	chr5:112174665	APC	Non_Synonymous_Coding	T	C	p.V1125A	2958	0.494	NM_001127510.2	rs377278397	
PR110	chr10:89711909	PTEN	Non_Synonymous_Coding	A	G	p.Y176C	1004	0.536	NM_000314.4		
PR110	chr2:138000044	THSD7B	Non_Synonymous_Coding	G	A	p.R692Q	3741	0.985	NM_001080427.1	rs76693568	COSM148945
PR111	chr17:7578427	TP53	Non_Synonymous_Coding	T	G	p.H168P	4091	0.058	NM_000546.5		COSM44808
PR111	chr12:49446101	KMT2D	Codon_Change_Plus_Coding	TTCCTCAGGTGGTGGGGACAGGGGTGAC	T	p.ESPLSPPPEE446E	3900	0.255	NM_003482.3		
PR111	chr14:95916404	SYNE3	Non_Synonymous_Coding	G	A	p.A438V	3236	0.5	NM_152592.3	rs376830751	COSM198717
PR111	chr12:49444996	KMT2D	Frame_shift	A	AG	p.-823?	93	0.677	NM_003482.3		

Contd...

Supplementary Table 4: Contd...

Sample	*Location*	*Gene*	*Variant classification*	*Reference allele*	*Variant allele*	*AA change*	*Filtered coverage*	*Variant allele frequency*	*RefSeq*	*dbSNP ID*	*COSMIC ID*
PR111	chrX:66937337	AR	Non_Synonymous_Coding	G	A	p.V731M	2059	0.997	NM_000044.3	rs137852571	COSM1468985
PR112	chr17:47696425	SPOP	Non_Synonymous_Coding	A	C	p.F133C	2480	0.189	NM_001007230.1		COSM242645
PR112	chr12:57865558	GLI1	Non_Synonymous_Coding	G	T	p.G1012V	137	0.263	NM_005269.2	rs2229300	
PR112	chr12:49446101	KMT2D	Codon_Change_Plus_Coding	TTCCTCAGGTGGTGGGGACAGGGGTGAC	T	p.ESPLSPPPEE446E	3359	0.286	NM_003482.3		
PR112	chr16:72832153	ZFHX3	Non_Synonymous_Coding	C	T	p.M1476I	1266	0.509	NM_006885.3	rs200133878	
PR113	chr17:37687207	CDK12	Non_Synonymous_Coding	G	C	p.V1371L	272	0.04	NM_016507.2		
PR113	chr12:57864588	GLI1	Non_Synonymous_Coding	C	T	p.P689S	178	0.067	NM_005269.2		
PR113	chr16:72828408	ZFHX3	Non_Synonymous_Coding	C	A	p.A2725S	5110	0.068	NM_006885.3		
PR113	chr12:49446101	KMT2D	Codon_Change_Plus_Coding	TTCCTCAGGTGGTGGGGACAGGGGTGAC	T	p.ESPLSPPPEE446E	10 048	0.241	NM_003482.3		
PR113	chr12:49445021	KMT2D	Frame_shift	C	CG, CGG	p.-815?	173	0.422	NM_003482.3		
PR114	chr2:137988662	THSD7B	Non_Synonymous_Coding	G	T	p.R560M	124	0.04	NM_001080427.1		
PR114	chr3:73433708	PDZRN3	Non_Synonymous_Coding	G	A	p.P670L	261	0.05	NM_015009.1		
PR114	chr7:91739463	AKAP9	Non_Synonymous_Coding	T	C	p.M3905T	4221	0.589	NM_005751.4	rs77447750	
PR115	chr14:95932477	SYNE3	Non_Synonymous_Coding	T	C	p.I140V	7018	0.496	NM_152592.3		
PR115	chr12:49424475	KMT2D	Non_Synonymous_Coding	C	T	p.C4583Y	3132	0.522	NM_003482.3		
PR201	chr19:5144388	KDM4B	Non_Synonymous_Coding	T	C	p.S956P	9471	0.031	NM_015015.2		
PR201	chr17:37619118	CDK12	Non_Synonymous_Coding	G	T	p.S265I	75	0.04	NM_016507.2		
PR201	chr3:100467047	TFG	Non_Synonymous_Coding	G	T	p.G292V	197	0.041	NM_006070.5		
PR201	chr2:138400097	THSD7B	Non_Synonymous_Coding	A	C	p.K1251Q	95	0.042	NM_001080427.1		
PR201	chr12:56428856	IKZF4	Non_Synonymous_Coding	A	G	p.K500R	88	0.045	NM_022465.3		
PR201	chr7:91714912	AKAP9	Non_Synonymous_Coding	C	T	p.P2979L	2084	0.047	NM_005751.4		
PR201	chr16:72821079	ZFHX3	Non_Synonymous_Coding	T	G	p.D3699A	1342	0.047	NM_006885.3		
PR201	chr5:112173687	APC	Non_Synonymous_Coding	A	C	p.Y799S	105	0.048	NM_001127510.2		
PR201	chr2:138163363	THSD7B	Non_Synonymous_Coding	G	T	p.R863L	348	0.052	NM_001080427.1		
PR201	chr17:37687207	CDK12	Non_Synonymous_Coding	G	T	p.V1371L	111	0.054	NM_016507.2		
PR201	chr19:50549280	ZNF473	Non_Synonymous_Coding	G	A	p.G527D	3327	0.057	NM_001006656.1		
PR201	chr3:100467073	TFG	Non_Synonymous_Coding	G	C	p.A301P	198	0.061	NM_006070.5		
PR201	chr5:112111327	APC	Non_Synonymous_Coding	T	A	p.S142T	63	0.063	NM_001127510.2		
PR201	chr12:49421002	KMT2D	Non_Synonymous_Coding	G	A	p.P4916L	126	0.063	NM_003482.3		
PR201	chr12:49425616	KMT2D	Non_Synonymous_Coding	G	C	p.A4291G	157	0.064	NM_003482.3		
PR201	chr19:5071040	KDM4B	Non_Synonymous_Coding	C	T	p.H216Y	240	0.071	NM_015015.2		
PR201	chrX:66766249	AR	Non_Synonymous_Coding	G	A	p.G421R	84	0.071	NM_000044.3		
PR201	chr4:87217	ZNF595	Non_Synonymous_Coding	C	A	p.T607N	626	0.073	NM_182524.2		
PR201	chr12:56428449	IKZF4	Non_Synonymous_Coding	C	G	p.S364R	112	0.08	NM_022465.3		
PR201	chr3:73437120	PDZRN3	Non_Synonymous_Coding	T	C	p.Q506R	285	0.084	NM_015009.1		
PR201	chr16:72828408	ZFHX3	Non_Synonymous_Coding	C	A	p.A2725S	2338	0.147	NM_006885.3		
PR201	chr4:86822	ZNF595	Coding_Deletion	CGAGAAACCCTACAAATGTGAAGAATGT GGCAAAGCTTTCATATGGTCCGCAAGC CTGAATGAACATAAGAATATTCATACTGGA	C	p.EKPYKCEECGKAF IWSASLNEHKNIH TG476-	212	0.302	NM_182524.2		
PR201	chr17:37673806	CDK12	Non_Synonymous_Coding	C	T	p.S987F	10 151	0.506	NM_016507.2	rs201661022	
PR201	chr4:87313	ZNF595	Frame_shift	G	GA	p.-639?	184	0.783	NM_182524.2	rs60154095	
PR201	chr7:91714911	AKAP9	Non_Synonymous_Coding	C	T	p.P2979S	2076	0.947	NM_005751.4	rs1063242	
PR201	chrX:66765516	AR	Non_Synonymous_Coding	C	A	p.S176R	1550	0.995	NM_000044.3		

Contd...

Supplementary Table 4: Contd...

Sample	*Location*	*Gene*	*Variant classification*	*Reference allele*	*Variant allele*	*AA change*	*Filtered coverage*	*Variant allele frequency*	*RefSeq*	*dbSNP ID*	*COSMIC ID*
PR202	chr19:5150397	KDM4B	Non_Synonymous_Coding	C	T	p.T1017M	161	0.043	NM_015015.2		
PR202	chr12:49444986	KMT2D	Non_Synonymous_Coding	T	C	p.Q827R	27	0.074	NM_003482.3		
PR202	chr5:112175043	APC	Non_Synonymous_Coding	T	C	p.V1251A	107	0.075	NM_001127510.2		
PR202	chrX:70339696	MED12	Non_Synonymous_Coding	G	T	p.G122V	53	0.075	NM_005120.2		
PR202	chr16:72821887	ZFHX3	Frame_shift	C	CG	p.-3429?	1432	0.214	NM_006885.3		
PR202	chr7:91646406	AKAP9	Non_Synonymous_Coding	G	A	p.R1276Q	1608	0.631	NM_005751.4	rs146797353	
PR202	chr4:87313	ZNF595	Frame_shift	G	GA	p.-639?	142	0.965	NM_182524.2	rs60154095	
PR203	chr16:72991626	ZFHX3	Non_Synonymous_Coding	C	T	p.E807K	216	0.042	NM_006885.3		
PR203	chr15:23006222	NIPA2	Stop_Lost	T	A	p.*361L	836	0.045	NM_030922.6		
PR203	chr12:49445310	KMT2D	Non_Synonymous_Coding	G	A	p.P719L	59	0.085	NM_003482.3	rs185660524	
PR203	chr5:112173644	APC	Non_Synonymous_Coding	C	G	p.H785D	1781	0.127	NM_001127510.2		
PR203	chr17:47696425	SPOP	Non_Synonymous_Coding	A	G	p.F133S	1070	0.133	NM_001007230.1		COSM247573
PR203	chr9:21974694	CDKN2A	Non_Synonymous_Coding	C	G	p.G45R	2296	0.501	NM_001195132.1		
PR203	chr16:72831357	ZFHX3	Coding_Deletion	CTTGTTG	C	p.QQ1740-	1886	0.512	NM_006885.3		COSM1731926
PR203	chr7:91690697	AKAP9	Non_Synonymous_Coding	G	A	p.A1909T	2874	0.558	NM_005751.4	rs200844952	
PR203	chr4:87313	ZNF595	Frame_shift	G	GA	p.-639?	110	0.936	NM_182524.2	rs60154095	
PR203	chr7:91714911	AKAP9	Non_Synonymous_Coding	C	T	p.P2979S	1553	0.958	NM_005751.4	rs1063242	
PR205	chr19:50549280	ZNF473	Non_Synonymous_Coding	G	A	p.G527D	14 424	0.041	NM_001006656.1		
PR205	chr12:49427950	KMT2D	Non_Synonymous_Coding	C	T	p.R3547H	6749	0.045	NM_003482.3		COSM240685
PR205	chr17:37687204	CDK12	Non_Synonymous_Coding	C	A	p.L1370M	437	0.046	NM_016507.2		
PR205	chr2:138400100	THSD7B	Non_Synonymous_Coding	G	C	p.E1252Q	312	0.048	NM_001080427.1		
PR205	chr5:112176407	APC	Non_Synonymous_Coding	T	A	p.S1706T	548	0.049	NM_001127510.2		
PR205	chr12:49424422	KMT2D	Non_Synonymous_Coding	G	T	p.P4601T	201	0.05	NM_003482.3		
PR205	chr3:178942538	PIK3CA	Non_Synonymous_Coding	A	G	p.N782S	791	0.053	NM_006218.2		
PR205	chr12:56428447	IKZF4	Non_Synonymous_Coding	A	T	p.S364C	431	0.053	NM_022465.3		
PR205	chr12:56428445	IKZF4	Non_Synonymous_Coding	A	T	p.H363L	404	0.054	NM_022465.3		
PR205	chr19:5119672	KDM4B	Non_Synonymous_Coding	G	A	p.R375Q	403	0.062	NM_015015.2		
PR205	chr16:72991377	ZFHX3	Non_Synonymous_Coding	C	T	p.G890R	496	0.067	NM_006885.3		
PR205	chr12:49444990	KMT2D	Non_Synonymous_Coding	G	T	p.P826T	292	0.068	NM_003482.3		
PR205	chr2:138400097	THSD7B	Non_Synonymous_Coding	A	C	p.K1251Q	318	0.097	NM_001080427.1		
PR205	chr16:72828408	ZFHX3	Non_Synonymous_Coding	C	A	p.A2725S	8996	0.105	NM_006885.3		
PR205	chr16:72828405	ZFHX3	Frame_shift	GA	G	p.-2725	8808	0.107	NM_006885.3		
PR205	chr12:49445334	KMT2D	Non_Synonymous_Coding	G	A	p.P711L	44	0.136	NM_003482.3		
PR205	chr17:7577147	TP53	Non_Synonymous_Coding	A	T	p.L264Q	98	0.184	NM_000546.5		
PR205	chr7:91682269	AKAP9	Non_Synonymous_Coding	T	A	p.S1866R	83	0.193	NM_005751.4		
PR205	chr17:37627784	CDK12	Non_Synonymous_Coding	C	G	p.P567A	92	0.196	NM_016507.2		
PR205	chr12:49446101	KMT2D	Codon_Change_Plus_Coding	TTCCTCAGGTGGTGGGGACAGGGGTGAC	T	p.ESPLSPPPEE446E	7007	0.463	NM_003482.3		
PR206	chr8:71078935	NCOA2	Non_Synonymous_Coding	T	C	p.N199S	4974	0.035	NM_006540.2		
PR206	chr3:178928073	PIK3CA	Non_Synonymous_Coding	G	A	p.G451R	4034	0.037	NM_006218.2		
PR206	chr14:95932477	SYNE3	Non_Synonymous_Coding	T	C	p.I140V	8892	0.041	NM_152592.3		
PR206	chr3:178916833	PIK3CA	Non_Synonymous_Coding	A	G	p.T74A	7255	0.106	NM_006218.2		
PR206	chr17:37649079	CDK12	Coding_Deletion	CATT	C	p.I729-	13 486	0.22	NM_016507.2		

Contd...

Supplementary Table 4: Contd...

Sample	*Location*	*Gene*	*Variant classification*	*Reference allele*	*Variant allele*	*AA change*	*Filtered coverage*	*Variant allele frequency*	*RefSeq*	*dbSNP ID*	*COSMIC ID*
PR206	chr12:49446101	KMT2D	Codon_Change_Plus_Coding	TTCCTCAGGTGGTGGGGACAGGGGTGAC	T	p.ESPLSPPPEE446E	5541	0.243	NM_003482.3		
PR206	chr7:107823354	NRCAM	Non_Synonymous_Coding	C	T	p.G772D	5502	0.486	NM_001037132.2		
PR207	chr3:38888625	SCN11A	Non_Synonymous_Coding	G	T	p.L1646I	6475	0.127	NM_014139.2		
PR208	chr12:49445341	KMT2D	Non_Synonymous_Coding	G	A	p.P709S	100	0.04	NM_003482.3		
PR208	chr12:57865549	GLI1	Non_Synonymous_Coding	C	T	p.P1009L	81	0.074	NM_005269.2		
PR208	chr16:72828408	ZFHX3	Non_Synonymous_Coding	C	A	p.A2725S	8350	0.082	NM_006885.3		
PR208	chr12:49446101	KMT2D	Codon_Change_Plus_Coding	TTCCTCAGGTGGTGGGGACAGGGGTGAC	T	p.ESPLSPPPEE446E	7658	0.303	NM_003482.3		
PR208	chr3:73433697	PDZRN3	Non_Synonymous_Coding	C	G	p.G674R	390	0.454	NM_015009.1		
PR210	chr16:72828408	ZFHX3	Non_Synonymous_Coding	C	A	p.A2725S	9425	0.06	NM_006885.3		
PR210	chr19:5144410	KDM4B	Non_Synonymous_Coding	C	G	p.P963R	105	0.105	NM_015015.2		
PR210	chr16:72821887	ZFHX3	Frame_shift	C	CG	p.-3429?	3674	0.201	NM_006885.3		
PR210	chr12:49446101	KMT2D	Codon_Change_Plus_Coding	TTCCTCAGGTGGTGGGGACAGGGGTGAC	T	p.ESPLSPPPEE446E	7636	0.323	NM_003482.3		
PR210	chr14:95942049	SYNE3	Non_Synonymous_Coding	C	T	p.R37H	7754	0.401	NM_152592.3		
PR210	chr14:95932477	SYNE3	Non_Synonymous_Coding	T	C	p.I140V	10 415	0.499	NM_152592.3		
PR210	chr7:91646406	AKAP9	Non_Synonymous_Coding	G	A	p.R1276Q	2452	0.529	NM_005751.4	rs146797353	
PR210	chr8:71078935	NCOA2	Non_Synonymous_Coding	T	C	p.N199S	4847	0.56	NM_006540.2		
PR210	chr16:72991697	ZFHX3	Codon_Change_Plus_Coding	GCCGCCGCCGCCGCAGCCA	G	p.VAAAAAA777A	390	0.985	NM_006885.3		
PR212	chr3:178917586	PIK3CA	Non_Synonymous_Coding	G	C	p.R154T	30	0.1	NM_006218.2		
PR212	chrX:70348499	MED12	Non_Synonymous_Coding	A	T	p.I1136F	799	0.149	NM_005120.2		
PR212	chr19:5144410	KDM4B	Non_Synonymous_Coding	C	A	p.P963H	83	0.205	NM_015015.2		
PR212	chr12:49446101	KMT2D	Codon_Change_Plus_Coding	TTCCTCAGGTGGTGGGGACAGGGGTGAC	T	p.ESPLSPPPEE446E	6707	0.211	NM_003482.3		
PR212	chr12:57865558	GLI1	Non_Synonymous_Coding	G	T	p.G1012V	168	0.357	NM_005269.2	rs2229300	
PR212	chr12:49420078	KMT2D	Non_Synonymous_Coding	C	T	p.R5224H	3045	0.5	NM_003482.3	rs3782356	
PR213	chr12:49446101	KMT2D	Codon_Change_Plus_Coding	TTCCTCAGGTGGTGGGGACAGGGGTGAC	T	p.ESPLSPPPEE446E	5577	0.203	NM_003482.3		
PR213	chrX:70349256	MED12	Non_Synonymous_Coding	T	G	p.V1223G	3524	0.442	NM_005120.2		
PR213	chr12:49420078	KMT2D	Non_Synonymous_Coding	C	T	p.R5224H	4335	0.493	NM_003482.3	rs3782356	
PR214	chr3:178917586	PIK3CA	Non_Synonymous_Coding	G	T	p.R154M	86	0.058	NM_006218.2		
PR214	chr2:138000044	THSD7B	Non_Synonymous_Coding	G	A	p.R692Q	6899	0.492	NM_001080427.1	rs76693568	COSM148945
PR215	chr16:72828408	ZFHX3	Non_Synonymous_Coding	C	A	p.A2725S	12 278	0.059	NM_006885.3		
PR215	chr7:91699374	AKAP9	Non_Synonymous_Coding	A	C	p.K2121Q	55	0.091	NM_005751.4		
PR215	chr12:57865558	GLI1	Non_Synonymous_Coding	G	T	p.G1012V	45	0.178	NM_005269.2	rs2229300	
PR215	chr12:49446101	KMT2D	Codon_Change_Plus_Coding	TTCCTCAGGTGGTGGGGACAGGGGTGAC	T	p.ESPLSPPPEE446E	8010	0.342	NM_003482.3		
PR215	chr3:38962696	SCN11A	Non_Synonymous_Coding	T	G	p.N255H	1634	0.523	NM_014139.2	rs371647806	
PR302	chr8:23540243	NKX3-1	Non_Synonymous_Coding	G	A	p.P54S	96	0.052	NM_006167.3		
PR302	chr16:72828408	ZFHX3	Non_Synonymous_Coding	C	A	p.A2725S	7963	0.061	NM_006885.3		
PR302	chr7:91682269	AKAP9	Non_Synonymous_Coding	T	A	p.S1866R	60	0.117	NM_005751.4		
PR302	chr8:71036150	NCOA2	Non_Synonymous_Coding	G	A	p.T1421M	5550	0.451	NM_006540.2		
PR302	chr3:38941459	SCN11A	Non_Synonymous_Coding	G	A	p.L650F	7862	0.483	NM_014139.2		COSM136332
PR304	chr12:57865558	GLI1	Non_Synonymous_Coding	G	T	p.G1012V	115	0.461	NM_005269.2	rs2229300	
PR305	chrX:70360666	MED12	Stop_Gained	C	T	p.Q2076*	50	0.06	NM_005120.2		
PR305	chr12:57865558	GLI1	Non_Synonymous_Coding	G	T	p.G1012V	309	0.469	NM_005269.2	rs2229300	
PR306	chr12:49444990	KMT2D	Non_Synonymous_Coding	G	C	p.P826A	106	0.057	NM_003482.3		

Contd...

Supplementary Table 4: Contd...

Sample	*Location*	*Gene*	*Variant classification*	*Reference allele*	*Variant allele*	*AA change*	*Filtered coverage*	*Variant allele frequency*	*RefSeq*	*dbSNP ID*	*COSMIC ID*
PR306	chr2:137988661	THSD7B	Non_Synonymous_Coding	A	C, G	p.R560G	32	0.094	NM_001080427.1		
PR306	chr12:57865558	GLI1	Non_Synonymous_Coding	G	T	p.G1012V	117	0.496	NM_005269.2	rs2229300	
PR307	chr12:57865558	GLI1	Non_Synonymous_Coding	G	T	p.G1012V	377	0.44	NM_005269.2	rs2229300	
PR307	chr7:91668025	AKAP9	Non_Synonymous_Coding	G	C	p.C1544S	2296	0.472	NM_005751.4		
PR307	chr7:91709065	AKAP9	Non_Synonymous_Coding	A	G	p.I2540V	1978	0.502	NM_005751.4		
PR307	chr16:72832320	ZFHX3	Non_Synonymous_Coding	G	C	p.Q1421E	7856	0.516	NM_006885.3		
PR307	chr16:72829373	ZFHX3	Non_Synonymous_Coding	G	A	p.S2403F	4584	0.517	NM_006885.3		
PR308	chr12:49425853	KMT2D	Non_Synonymous_Coding	C	T	p.R4212Q	3047	0.493	NM_003482.3		
PR309	chr2:138417295	THSD7B	Non_Synonymous_Coding	A	C	p.N1450T	2559	0.032	NM_001080427.1		
PR309	chr16:72821079	ZFHX3	Non_Synonymous_Coding	T	G	p.D3699A	3002	0.033	NM_006885.3		
PR309	chr19:50549280	ZNF473	Non_Synonymous_Coding	G	A	p.G527D	10 005	0.034	NM_001006656.1		
PR309	chr12:49431013	KMT2D	Non_Synonymous_Coding	C	G	p.V3376L	594	0.042	NM_003482.3		
PR309	chr3:178916786	PIK3CA	Non_Synonymous_Coding	T	C	p.L58P	1431	0.043	NM_006218.2		
PR309	chr3:38904766	SCN11A	Non_Synonymous_Coding	T	C	p.T1326A	2150	0.045	NM_014139.2		
PR309	chr5:112173291	APC	Non_Synonymous_Coding	A	G	p.Q667R	736	0.045	NM_001127510.2		
PR309	chr3:73450149	PDZRN3	Non_Synonymous_Coding	G	A	p.A393V	1574	0.047	NM_015009.1		COSM1425294
PR309	chr12:49436006	KMT2D	Non_Synonymous_Coding	T	C	p.E1992G	64	0.047	NM_003482.3		
PR309	chr2:138163363	THSD7B	Non_Synonymous_Coding	G	A	p.R863Q	915	0.048	NM_001080427.1		
PR309	chr16:72991476	ZFHX3	Non_Synonymous_Coding	C	T	p.A857T	1570	0.049	NM_006885.3	rs146607776	COSM1379726
PR309	chr5:112176066	APC	Non_Synonymous_Coding	A	G	p.K1592R	1204	0.052	NM_001127510.2		
PR309	chr13:48934201	RB1	Non_Synonymous_Coding	T	C	p.M219T	538	0.052	NM_000321.2		
PR309	chr13:48934204	RB1	Non_Synonymous_Coding	T	C	p.L220P	534	0.052	NM_000321.2		
PR309	chr17:37687207	CDK12	Non_Synonymous_Coding	G	C	p.V1371L	323	0.053	NM_016507.2		
PR309	chrX:70346837	MED12	Non_Synonymous_Coding	G	A	p.V902I	668	0.054	NM_005120.2		COSM1124669
PR309	chr2:138425344	THSD7B	Non_Synonymous_Coding	A	G	p.M1522V	1208	0.058	NM_001080427.1		
PR309	chr16:72828408	ZFHX3	Non_Synonymous_Coding	C	A	p.A2725S	9128	0.058	NM_006885.3		
PR309	chr5:112174845	APC	Non_Synonymous_Coding	C	T	p.T1185I	461	0.059	NM_001127510.2		
PR309	chr5:112179113	APC	Non_Synonymous_Coding	G	A	p.A2608T	494	0.059	NM_001127510.2		
PR309	chr5:112177895	APC	Non_Synonymous_Coding	A	G	p.K2202E	275	0.062	NM_001127510.2		
PR309	chr7:91631131	AKAP9	Non_Synonymous_Coding	G	A	p.E634K	699	0.062	NM_005751.4	rs368006874	
PR309	chr3:178943760	PIK3CA	Non_Synonymous_Coding	A	T	p.Q809H	302	0.063	NM_006218.2		
PR309	chr12:56428449	IKZF4	Non_Synonymous_Coding	C	G	p.S364R	560	0.066	NM_022465.3		
PR309	chrX:70347260	MED12	Non_Synonymous_Coding	T	C	p.L975P	2086	0.07	NM_005120.2		
PR309	chr7:91609645	AKAP9	Non_Synonymous_Coding	G	C	p.E117Q	113	0.08	NM_005751.4		
PR309	chr8:23540200	NKX3-1	Non_Synonymous_Coding	C	T	p.R68H	191	0.089	NM_006167.3		
PR309	chr16:72828173	ZFHX3	Non_Synonymous_Coding	G	A	p.S2803F	357	0.098	NM_006885.3		
PR309	chr17:7577147	TP53	Non_Synonymous_Coding	A	T	p.L264Q	61	0.098	NM_000546.5		
PR309	chr12:57865549	GLI1	Non_Synonymous_Coding	C	T	p.P1009L	29	0.172	NM_005269.2		
PR309	chr7:91682269	AKAP9	Non_Synonymous_Coding	T	G	p.S1866R	47	0.255	NM_005751.4		
PR309	chr12:49446101	KMT2D	Codon_Change_Plus_Coding	TTCCTCAGGTGGTGGGGACAGGGGTGAC	T	p.ESPLSPPPEE446E	8861	0.35	NM_003482.3		
PR309	chr16:72831357	ZFHX3	Coding_Deletion	CTTGTTG	C	p.QQ1740-	2881	0.431	NM_006885.3		COSM1731926
PR309	chr16:72832153	ZFHX3	Non_Synonymous_Coding	C	T	p.M1476I	1414	0.592	NM_006885.3	rs200133878	

Contd...

Supplementary Table 4: Contd...

Sample	*Location*	*Gene*	*Variant classification*	*Reference allele*	*Variant allele*	*AA change*	*Filtered coverage*	*Variant allele frequency*	*RefSeq*	*dbSNP ID*	*COSMIC ID*
PR310	chr2:137988662	THSD7B	Non_Synonymous_Coding	G	C	p.R560T	49	0.102	NM_001080427.1		
PR311	chr16:72993731	ZFHX3	Non_Synonymous_Coding	G	A	p.P105L	1658	0.154	NM_006885.3	rs199521581	COSM130307
PR312	chr16:72821887	ZFHX3	Frame_shift	C	CG	p.-3429?	3536	0.154	NM_006885.3		
PR313	chr2:137988662	THSD7B	Non_Synonymous_Coding	G	A	p.R560K	39	0.077	NM_001080427.1		
PR313	chr7:91712775	AKAP9	Non_Synonymous_Coding	A	C	p.T2818P	62	0.113	NM_005751.4		
PR313	chr12:49432474	KMT2D	Non_Synonymous_Coding	C	T	p.G2889R	31	0.129	NM_003482.3		
PR313	chr3:73657760	PDZRN3	Non_Synonymous_Coding	G	A	p.R267W	1517	0.477	NM_015009.1	rs373980887	
PR315	chr19:5144410	KDM4B	Non_Synonymous_Coding	C	T	p.P963L	21	0.238	NM_015015.2		
PR315	chr7:91739463	AKAP9	Non_Synonymous_Coding	T	C	p.M3905T	2857	0.476	NM_005751.4	rs77447750	
PR315	chr19:5134026	KDM4B	Non_Synonymous_Coding	C	T	p.T680M	3383	0.514	NM_015015.2	rs149752315	
PR315	chr12:57865558	GLI1	Non_Synonymous_Coding	G	T	p.G1012V	498	0.552	NM_005269.2	rs2229300	
PR401	chr12:49444992	KMT2D	Non_Synonymous_Coding	G	C	p.S825C	155	0.052	NM_003482.3		
PR401	chr12:57865558	GLI1	Non_Synonymous_Coding	G	T	p.G1012V	63	0.349	NM_005269.2	rs2229300	
PR401	chr8:71071746	NCOA2	Non_Synonymous_Coding	A	G	p.L373P	24	0.375	NM_006540.2		
PR401	chr12:49444996	KMT2D	Frame_shift	A	AG	p.-823?	127	0.701	NM_003482.3		
PR403	chr16:72821887	ZFHX3	Frame_shift	C	CG	p.-3429?	3159	0.172	NM_006885.3		
PR403	chr12:49445021	KMT2D	Frame_shift	C	CG, CGG	p.-815?	85	0.376	NM_003482.3		
PR403	chr12:57865558	GLI1	Non_Synonymous_Coding	G	T	p.G1012V	517	0.472	NM_005269.2	rs2229300	
PR404	chr16:72991476	ZFHX3	Non_Synonymous_Coding	C	T	p.A857T	2706	0.04	NM_006885.3	rs146607776	COSM1379726
PR404	chr16:72991494	ZFHX3	Non_Synonymous_Coding	C	T	p.G851S	2731	0.044	NM_006885.3		
PR404	chr16:72991473	ZFHX3	Non_Synonymous_Coding	C	T	p.E858K	2757	0.051	NM_006885.3		
PR404	chr12:49436073	KMT2D	Non_Synonymous_Coding	C	T	p.D1970N	138	0.094	NM_003482.3		
PR404	chr16:72991700	ZFHX3	Non_Synonymous_Coding	G	A	p.A782V	54	0.148	NM_006885.3	rs62639304	
PR404	chr16:72991706	ZFHX3	Non_Synonymous_Coding	G	A	p.A780V	53	0.151	NM_006885.3		
PR404	chr12:49446101	KMT2D	Codon_Change_Plus_Coding	TTCCTCAGGTGGTGGGGACAGGGGTGAC	T	p.ESPLSPPPEE446E	6052	0.282	NM_003482.3		
PR414	chr12:49416542	KMT2D	Non_Synonymous_Coding	C	T	p.R5390Q	5172	0.485	NM_003482.3		
PR414	chr19:50549441	ZNF473	Non_Synonymous_Coding	G	A	p.E581K	10 668	0.487	NM_001006656.1	rs150885598	
PRY01	chr12:49445021	KMT2D	Frame_shift	C	CG, CGG	p.-815?	191	0.424	NM_003482.3		
PRY02	chr19:50549280	ZNF473	Non_Synonymous_Coding	G	A	p.G527D	2445	0.034	NM_001006656.1		
PRY02	chr16:72821887	ZFHX3	Frame_shift	C	CG	p.-3429?	1603	0.224	NM_006885.3		
PRY02	chr4:87313	ZNF595	Frame_shift	G	GA	p.-639?	62	0.968	NM_182524.2	rs60154095	
PRY04	chr16:72827613	ZFHX3	Non_Synonymous_Coding	C	T	p.V2990I	509	0.063	NM_006885.3		
PRY04	chr16:72821887	ZFHX3	Frame_shift	C	CG	p.-3429?	1684	0.249	NM_006885.3		
PRY04	chr16:72821887	ZFHX3	Frame_shift	C	CG	p.-3429?	1684	0.249	NM_006885.3		
PRY04	chr4:87313	ZNF595	Frame_shift	G	GA	p.-639?	66	0.879	NM_182524.2	rs60154095	

PTEN: phosphatase and tensin homolog; AR: androgen receptor

Supplementary Table 5: Clinical features and driver alterations in Chinese prostate cancer patients

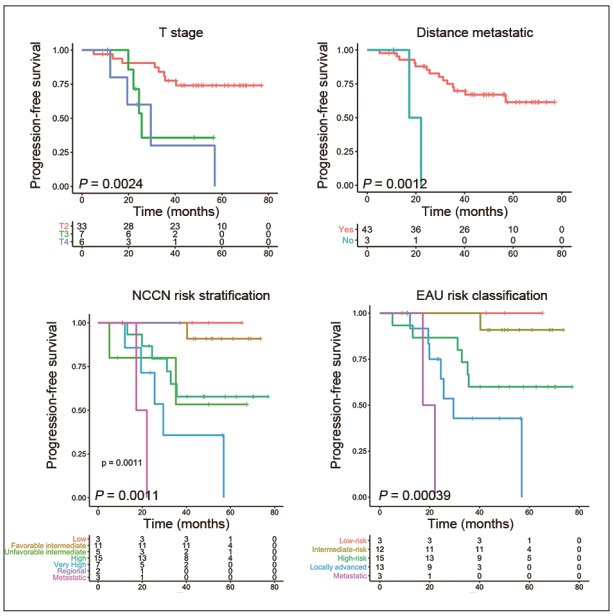
Classification	*All (n)*	*PTEN loss function mutation and deletions PIK3CA amplification*			*NCOA2 amplification SPOP mutation*			*AR activity*			*MYC amplification*			*RB1 deletion*			*Cell cycle-related gene expression*			*ETS fusion status*			*APC mutation/ deletion*			*GLI1 mutation*		
		Yes (n)	*No (n)*	*P*	*Yes (n)*	*No (n)*	*P*	*High (n)*	*Low (n)*	*P*	*Yes (n)*	*No (n)*	*P*	*Yes (n)*	*No (n)*	*P*	*High (n)*	*Low (n)*	*P*	*Yes (n)*	*No (n)*	*P*	*Yes (n)*	*No (n)*	*P*	*Yes (n)*	*No (n)*	*P*
Age (year)																												
≤60	12	1	11	0.035*	4	8	0.416	7	4	0.310	7	5	1.000	2	10	0.041*	6	5	0.736	2	10	0.701	5	7	1.000	4	8	1.000
>60	34	15	19		6	28		14	18		21	13		19	15		15	17		9	25		15	19		11	23	
Total PSA levels (ng ml^{-1})																												
<10	17	7	10	0.787	6	11	0.110	8	7	0.538	9	8	0.687	6	11	0.146	8	7	0.925	5	12	0.824	7	10	0.477	8	9	0.155
10–20	10	3	7		0	10		6	4		6	4		3	7		5	5		2	8		6	4		1	9	
>20	19	6	13		4	15		7	11		13	6		12	7		8	10		4	15		7	12		6	13	
Gleason score																												
<7	11	4	7	1.000	4	7	0.427	6	5	0.335	6	5	0.860	4	7	0.555	5	6	0.632	4	7	0.505	4	7	0.865	5	6	0.207
7	21	7	14		4	17		11	8		13	8		9	12		8	11		5	16		9	12		8	13	
>7	14	5	9		2	12		4	9		9	5		8	6		8	5		2	12		7	7		2	12	
T stage																												
pT2	33	12	21	0.703	9	24	0.335	16	16	0.349	20	13	0.714	12	21	0.066	15	17	0.687	6	27	0.030*	14	19	1.000	11	22	0.699
pT3	7	3	4		0	7		2	5		5	2		6	1		3	4		1	6		3	4		3	4	
pT4	6	1	5		1	5		3	1		3	3		3	3		3	1		4	2		3	3		1	5	
Distance metastasis																												
Yes	3	0	3	0.541	1	2	0.529	1	1	1.000	2	1	1.000	2	1	0.162	2	0	0.233	0	3	1.000	3	0	0.075	3	0	0.030*
No	43	16	27		9	34		20	21		26	17		10	33		19	22		11	32		17	26		12	31	
NCCN risk stratification																												
Low	3	1	2	0.173	0	3	0.607	3	0	0.314	1	2	0.378	2	1	0.006*	2	1	0.314	1	2	0.352	0	3	0.099	1	2	0.263
Favorable intermediate	11	3	8		3	8		6	5		4	7		0	11		4	7		3	8		5	6		4	7	
Unfavorable intermediate	5	4	1		2	3		3	2		4	1		3	2		3	2		1	4		4	1		1	4	
High	15	7	8		2	13		4	11		11	4		10	5		6	9		2	13		5	10		5	10	
Very high	7	1	6		1	6		3	2		4	3		3	4		4	1		4	3		3	4		1	6	
Regional	2	0	2		1	1		1	1		2	0		1	1		0	2		0	2		0	2		0	2	
Metastatic	3	0	3		1	2		1	1		2	1		2	1		2	0		0	3		3	0		3	0	
EAU risk classificiton																												
Low risk	3	1	2	0.376	0	3	0.862	3	0	0.078	1	2	0.242	2	1	0.022*	2	1	0.512	1	2	0.183	0	3	0.210	1	2	0.165
Intermediate risk	12	3	9		3	9		8	4		5	7		1	11		4	8		4	8		6	6		4	8	
High risk	15	8	7		4	11		4	11		12	3		8	7		8	7		1	14		6	9		4	11	
Locally advanced	13	4	9		2	11		5	6		8	5		8	5		5	6		5	8		5	8		3	10	
Metastatic	3	0	3		1	2		1	1		2	1		2	1		2	0		0	3		3	0		3	0	

*$P < 0.05$. EAU: European Association of Urology; AR: androgen receptor; ETS: external transcribed spacer; PSA: prostate-specific antigen; NCCN: National Comprehensive Cancer Network; APC: APC regulator of WNT signaling pathway; GLI1: GLI family zinc finger 1

Supplementary Table 6: The official full name of genes listed in the figures

Gene abbreviation	*Official full name*
KMT2D	Lysine methyltransferase 2D
ZFHX3	Zinc finger homeobox 3
AKAP9	A-kinase anchoring protein 9
GLI1	GLI family zinc finger 1
THSD7B	Thrombospondin type 1 domain containing 7B
APC	APC regulator of WNT signaling pathway
CDK12	Cyclin-dependent kinase 12
KDM4B	Lysine demethylase 4B
MED12	Mediator complex subunit 12
ZNF595	Zinc finger protein 595
PIK3CA	Phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha
ZNF473	Hypoxia inducible factor 1 subunit alpha
NCOA2	Nuclear receptor coactivator 2
PDZRN3	PDZ domain containing ring finger 3
SCN11A	Sodium voltage-gated channel alpha subunit 11
SYNE3	Spectrin repeat containing nuclear envelope family member 3
IKZF4	IKAROS family zinc finger 4
SPOP	Speckle type BTB/POZ protein
TP53	Tumor protein p53
AR	Androgen receptor
NKX3-1	NK3 homeobox 1
NRCAM	Neuronal cell adhesion molecule
PTEN	Phosphatase and tensin homolog
TFG	TRK-fused gene
CDKN2A	Cyclin-dependent kinase inhibitor 2A
NIPA2	NIPA magnesium transporter 2
RB1	RB transcriptional corepressor 1
OR5L1	Olfactory receptor family 5 subfamily L member 1
MYC	MYC proto-oncogene, bHLH transcription factor
TBX20	T-box 20
KLK2	Kallikrein-related peptidase 2
KLK3	Kallikrein-related peptidase 3
CAMKK2	Calcium/calmodulin-dependent protein kinase kinase 2
FKBP5	FKBP prolyl isomerase 5
ABCC4	ATP-binding cassette subfamily C member 4
STK39	Serine/threonine kinase 39
ZBTB10	Zinc finger and BTB domain containing 10
NUSAP1	Nucleolar and spindle-associated protein 1
KIF11	KIF11
CDC20	Cell division cycle 20
FOXM1	Forkhead box M1
TMPRSS2:ERG	transmembrane serine protease 2: ETS transcription factor ERG
ERG	ETS transcription factor ERG
ETV1	ETS variant 1
ETV4	ETS variant 4
FLI1	Fli-1 proto-oncogene
CHD1	Chromodomain helicase DNA-binding protein 1
SPINK1	Serine peptidase inhibitor, Kazal type 1
RAF1	Raf-1 proto-oncogene
EZH2	Enhancer of zeste 2 polycomb repressive complex 2 subunit

PTEN: phosphatase and tensin homolog; AR: androgen receptor; ETS: external transcribed spacer



Supplementary Figure 1: T stage, distance metastatic, NCCN risk stratification and EAU risk classification are linked to worse prognosis in our cohort. Survival analysis was performed on Chinese PCa patients treated according to the presence or absence of T stage (**a**), or distance metastatic (**b**), or NCCN risk stratification (**c**), and EAU risk classification (**d**).