Klinik und Poliklinik für Strahlentherapie und Radioonkologie

Klinikum der Ludwig-Maximilians-Universität München

Vorstand: Prof. Dr. Claus Belka

Construction and Validation of Gene Signature-based Prognostic Models for Patients with Urological Cancers

Dissertation
zum Erwerb des Doktorgrades der Medizin
an der Medizinischen Fakultät der
Ludwig-Maximilians-Universität zu München

vorgelegt von
Run Shi

aus

Jiangsu, China

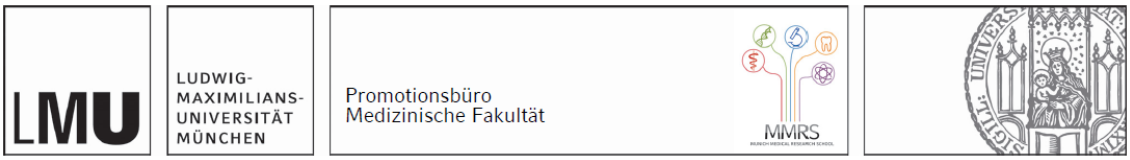
Jahr

2022

Mit Genehmigung der Medizinischen Fakultät
der Universität München

Berichterstatter:	Prof. Dr. Claus Belka
Mitberichterstatter:	Prof. Dr. Jakob Linseisen
	Prof. Dr. Michael Staehler
Mitbetreuung durch den promovierten Mitarbeiter:	PD. Dr. Minglun Li
Dekan:	Prof. Dr. med. Thomas Gudermann
Tag der mündlichen Prüfung:	20.01.2022

Affidavit



Affidavit

Shi, Run

Surname, first name

I hereby declare, that the submitted thesis entitled:

Construction and Validation of Gene Signature-based Prognostic Models for Patients with Urological Cancers

is my own work. I have only used the sources indicated and have not made unauthorised use of services of a third party. Where the work of others has been quoted or reproduced, the source is always given.

I further declare that the submitted thesis or parts thereof have not been presented as part of an examination degree to any other university.

München, 29/06/2021

place, date

Run Shi

Signature doctoral candidate

Table of contents

List of abbreviations

PCa: prostate cancer

BCR: biochemical recurrence

BCRFS: biochemical recurrence-free survival

RP: radical prostatectomy

pT: pathological T stage

PSA: prostate-specific antigen

GS: Gleason score

LASSO: least absolute shrinkage and selection operator

GEO: Gene Expression Omnibus

TCGA: The Cancer Genome Atlas

WGCNA: Weighted Gene Co-expression Network Analysis

GSEA: Gene Set Enrichment Analysis

tROC: time-dependent receiver operating characteristic

AUC: area under the curve

BCa: bladder cancer

MIBC: muscle-invasive bladder cancer

CCP: cell cycle progression

TNM: tumor-node-metastasis

LNM: lymph node metastasis

RECIST: Response Evaluation Criteria in Solid Tumors

CCPRS: cell cycle progression-related risk score

CCLE: Cancer Cell Line Encyclopedia

GDSC: Genomics of Drug Sensitivity in Cancer

ssGSEA: single-sample gene set enrichment analysis

MSigDB: Molecular Signatures Database

PCoA: principal coordinates analysis

NMF: non-negative matrix factorization

List of publications

- **Shi R**, Bao X, Weischenfeldt J, Schaefer C, Rogowski P, Schmidt-Hegemann NS, et al. A Novel Gene Signature-Based Model Predicts Biochemical Recurrence-Free Survival in Prostate Cancer Patients after Radical Prostatectomy. ***Cancers (Basel)***. 2019; 12.
- **Shi R**, Bao X, Rogowski P, Schafer C, Schmidt-Hegemann NS, Unger K, et al. Establishment and Validation of an Individualized Cell Cycle Process-Related Gene Signature to Predict Cancer-Specific Survival in Patients with Bladder Cancer. ***Cancers (Basel)***. 2020; 12.
- **Shi R**, Bao X, Sun J, Lu S, Belka C, Li M. Tumor microenvironment characterization in head and neck squamous carcinoma reveals distinct genomic alterations and clinical outcomes. ***Clin Transl Med***. 2020; 10: e187.
- **Shi R**, Bao X, Unger K, Sun J, Lu S, Manapov F, et al. Identification and validation of hypoxia-derived gene signatures to predict clinical outcomes and therapeutic responses in stage I lung adenocarcinoma patients. ***Theranostics***. 2021; 11: 5061-76.

My contribution to the publications (paper I & II)

1.1 Contribution to paper I

As the first author of paper I, Run Shi performed

i) data collection, including RNA-sequencing and clinical data from public datasets, and RNA-seq and clinical data of a cohort of 84 prostate cancer patients from a cooperative group in Charité Universitätsmedizin Berlin.
ii) formal analysis, including the bioinformatic (WGCNA, GSEA) and statistical (Cox regression analysis, LASSO regularization, meta-analysis, time-dependent ROC, decision tree) methods used in the paper.
iii) manuscript drafting.
iv) participation in revision, including response to reviewers' comments, data interpretation and language editing.

1.2 Contribution to paper II

As the first author of paper II, Run Shi performed

i) data collection, including RNA-sequencing and clinical data from public datasets.
ii) formal analysis, including the bioinformatic (ssGSEA, WGCNA) and statistical (NMF, Cox regression analysis, LASSO regularization, meta-analysis, decision tree, nomogram construction) methods used in the paper.
iii) manuscript drafting.
iv) participation in revision, including response to reviewers' comments, data interpretation and language editing.

2. Introduction

2.1 Establishment and validation of prognostic gene signatures for patients with urological malignancies including prostate cancer and bladder cancer

2.1.1 Rationale of the study

Prostate cancer (PCa) is the most common malignancy diagnosed in men worldwide [1]. More than 50% PCa patients underwent radical prostatectomy (RP) as their primary treatment [2]. After RP, about 20% patients experienced a biochemical recurrence (BCR) with an increasing prostate-specific antigen (PSA) [3]. Several clinical trials have revealed that adjuvant radiotherapy brings tremendous clinical benefits for high-risk subset including advanced stage (pT3/4) or positive surgical margin [4-6]. However, about 50% patients without adjuvant radiotherapy did not suffer BCR in a long follow-up of 5 years [5]. For these patients, adjuvant radiotherapy would be an overtreatment with potential unnecessary radiation-induced side effects. Therefore, an accurate personalized model to identify patients who have the potential of BCR after RP is an urgent issue for the optimal management of PCa.

Bladder cancer (BCa) is another common malignancy of the urological system worldwide. Among all the diagnosed cases per year, approximately 2/3 are non-muscle-invasive bladder cancer (NMIBC), while the rest 1/3 are classified into muscle-invasive bladder cancer (MIBC) [7]. In spite of improvements in BCa therapies, clinical outcomes remain unfavorable. The standard treatment for localized MIBC is radical cystectomy with bilateral pelvic lymph node dissection, but the 5-year overall survival rate is less than 50% [8]. Though TNM staging and pathological grading systems are widely applied for cancer management, prognosis remains variable among BCa patients, even in a same pathological or grading stage [9]. Therefore, establishment of a more precise prognostic model to identify high-risk subset who may benefit from systemic therapies is urgently needed.

Advancements in high-throughput techniques have provided new insight into transcriptome profiling and highlighted the utilization of molecules in disease diagnosis and prognosis [10]. Several studies have established gene signatures to

predict prognosis for patients with urological malignancies including prostate cancer and bladder cancer [11-13]. However, the clinical utility of these gene signatures remains limited, and few of them were applied to clinical practice.

In this study, we aimed to establish robust gene-expression signatures to help improve risk stratification and treatment decision making for patients with urological malignancies using a series of bioinformatic and machine learning approaches.

2.1.2 Materials and methods

A total of 903 PCa patients with comprehensive clinical records including age, Gleason score, pathological T stage, surgical margin status and follow-up BCR information from six independent cohorts were included in our study. Three cohorts come from Gene Expression Omnibus (GEO), one cohort from The Cancer Genome Atlas (TCGA), one cohort from Memorial Sloan Kettering Cancer Center (MSKCC), and a cohort from University Medical Center Hamburg-Eppendorf, Germany. GSE70769 and GSE70768 come from a same study [14], and the microarray data was produced from a same chip platform (Illumina HumanHT-12 V4.0 Array). GSE54460 was produced from Illumina HiSeq 2000, which contains 94 patients with full-scale clinical annotations [15]. In addition, 388 patients with RNA-seq data were accessed from TCGA, and microarray data of 138 patients (produced from Affymetrix Human Exon 1.0 ST Array) were obtained from MSKCC [16]. Finally, samples from 84 patients were consecutively collected at Department of Urology and the Martini Clinics at the University Medical Center Hamburg-Eppendorf from 2010 to 2016.

As regards to bladder cancer, four microarray datasets including 587 BCa patients with full-scale clinical annotations and cancer-specific survival information were downloaded from GEO. The training set: GSE13507(Illumina human-6 v2.0 expression beadchip); Three validation sets: 1) GSE31684 (Affymetrix Human Genome U133 Plus 2.0 Array), 2) GSE32894 & 3) GSE32548 (Illumina HumanHT-12 V3.0 expression beadchip). Normalized RNA-seq data, copy number data and clinical phenotypes of MIBC samples were obtained from The Cancer Genome Atlas (TCGA). Copy number data and TPM data of 22 bladder cell lines

were obtained from Cancer Cell Line Encyclopedia (CCLE) [17]. Normalized microarray data and IC50 values of different drugs were obtained from Genomics of Drug Sensitivity in Cancer (GDSC) database [18].

Probe IDs were mapping to gene symbols according to annotation files, and multiple probes towards a same gene were averaged to obtain a singular value. In addition, expression measurements of multiple samples from a same patient were averaged. All the microarray and RNA-seq data collected in this study were normalized and log2 transformed.

The weighted gene co-expression network analysis (WGCNA) algorithm [19] was used to construct a co-expression network based on transcriptome profiling data of training samples. Univariate and LASSO Cox regression analyses were combined to screen for robust candidate genes to develop prognostic signatures. Risk score formula was established for individual patients as follows: $Risk\ score = \sum_i Coef(mRNA_i) \times Expr(mRNA_i)$, in which "Coef" represents LASSO Cox coefficients, and "Expr" represents normalized gene expression value.

IBM SPSS Statistics 20 (IBM Corp., Armonk, N.Y., USA), GraphPad Prism 8.0 (GraphPad Software Inc, San Diego, CA), Stata 12 (StataCorp LLC, Texas, USA) and R software (version 3.5.2, http://www.r-project.org) were used to analyze data and plot graphs. Meta-analysis was performed to evaluate the prognostic value of the established risk score in the pooled cohort. ssGSEA scores and risk scores were scaled to Z-scores when necessary. GSEA [20] was performed to confirm the role of the established gene signatures in specific biological processes. Unsupervised hierarchical clustering method was performed to show similarity among hallmarks and pathways, and a correlation network reflecting their relationships and connectivity was generated by the Cytoscape [21]. Principal coordinates analysis (PCoA) was used to visualize the dissimilarity of two groups based on the bray-curtis distance of their expression matrix. Circos was used to visualize enrichment results of Gene Ontology analysis and overlapping genes involved in different biological processes. The webtool cBioPortal [22] was used to visualize the genomic alterations of the established gene signatures. The Kaplan-Meier method was used to draw survival curves, and the log-rank test was used to evaluate survival difference. Pearson's correlation test is used to

evaluate the correlation between two continuous variables with a normal distribution. K-means-based consensus clustering (R package 'ConsensusClusterPlus') [23] or non-negative matrix factorization (NMF) consensus clustering [24] (R package 'NMF') was performed to identify different clusters according to the gene expression matrix. Nomogram was generated to quantify the survival risk using R package 'rms'. tROC analysis was used to measure the predictive capacity of each parameter using the R package 'survivalROC', and calibration curve was plotted to measure their predictive accuracy. Recursive partitioning analysis was used to build a survival decision tree to improve risk stratification with R package 'rpart' [25]. Student's t-test or one-way ANOVA was used to analyze the differences among different groups in variables with a normal distribution. p value less than 0.05 was considered statistically significant.

2.1.3 Results and conclusions

The scoring formula of BCR risk for PCa patients is established as follows: Risk score= (-0.22345 * ALDH1A2) + (0.364318 * ASNS) + (0.67184 * FAM171B) + (-0.54351 * FREM2) + (-0.4304 * RSPO2) + (-0.17707 * SRD5A2) + (0.094559 * SSTR1) + (0.040268 * TRIM14) + (-0.77555 * VPS4A); and the scoring formula of cancer-specific survival for BCa patients is established as follows: Risk score= (-0.38760 * HIGD2A) + (-0.08920 * TRIM2) + (0.03022 * CEP72) + (0.037675 * CDKN2D) + (0.04360 * ZIC2) + (0.08159 * RCE1) + (0.08159 * GCHFR) + (0.083565 * NMU) + (0.13716 * HOXC6) + (0.15225 * ADM2) + (0.42630 * SLC6A6) + (0.49191 * WDR62). The two proposed prognostic models function well in patients with urological malignancies.

For prostate cancer patients, survival analysis revealed that significant difference of BCR (HR = 5.787, $p < 0.0001$) was observed between the low risk score and high risk score groups in the training cohort. Multivariate Cox regression analysis indicated that the risk score was an independent risk factor for BCRFS (HR = 5.084, $p < 0.0001$). The risk score was validated in five independent cohorts: validation I: HR = 4.739, $p = 0.0005$; validation II: HR = 2.684, $p = 0.0008$; validation III: HR = 4.790, $p = 0.0011$; validation IV: HR = 5.708, $p < 0.0001$; validation V: HR = 5.193, $p = 0.0004$. Multivariate Cox regression analysis was performed on the risk score and other clinicopathological features including age,

Gleason score (GS), pathological T stage (pT) and surgical margin status. Notably, the risk score was still an independent risk factor for BCRFS in all five validation cohorts: validation I: HR = 3.979, p = 0.011; validation II: HR = 2.616, p = 0.007; validation III: HR = 3.120, p = 0.037; validation IV: HR = 2.913, p = 0.020; validation V: HR = 3.241, p = 0.040. tROC analysis indicated that the risk score exhibited the strongest predictive capacity in validation I and II, while having similar predictive power with some conventional clinicopathological parameters such as Gleason score or pT in validation cohorts III, IV and V.

Meta-analysis was used to evaluate the performance of the prognostic gene signature in the pooled PCa cohort. We observed that higher risk score was significantly correlated with worse prognosis in the pooled cohort (HR = 4.84, 95% CI = 2.94 - 6.74). Risk scores were scaled to Z-score for each cohort, and we observed that Z-scores were significantly elevated in BCR patients compared to BCR-free patients ($p < 0.0001$).

For bladder cancer patients, higher ssGSEA scores of cell cycle progression predict worse cancer-specific survival (HR = 3.804, 95% CI = 1.893 - 7.643, p = 0.0004) in the training cohort. Based on this finding, WGCNA algorithm and LASSO Cox regression analysis were applied and identified a set of 12 genes that not also represent cell cycle progression, but also serve as robust prognostic genes for bladder cancer patients. Based on their relative expression values and individual LASSO Cox coefficients, the cancer-specific survival risk score was calculated for each bladder cancer patient and defined as cell cycle progression risk score (CCPRS).

CCPRS was significantly correlated with more advanced clinicopathological characteristics such as muscle-invasive (MI) status and higher grade. Survival analysis showed that BCa patients with higher CCPRS exhibited worse cancer-specific survival in each cohort: training cohort: HR = 10.20, 95% CI = 5.041 - 20.66, $p < 0.0001$; validation I: HR = 2.991, 95% CI = 1.175 - 7.614, p = 0.0008; validation II: HR = 8.468, 95% CI = 3.791 - 18.92, $p < 0.0001$; validation III: HR = 6.345, 95% CI = 2.762 - 14.58, $p < 0.0001$. Meta-analysis revealed that higher CCPRS predicted a worse cancer-specific survival in a pooled BCa cohort (overall HR = 6.93, 95% CI = 4.63 - 10.37). Furthermore, multivariate Cox regression analysis was performed to evaluate the cancer-specific survival risk in a total of 284 patients with full-scale clinical annotations including gender, grade, age,

CCPRS, lymph node metastasis (LNM) and MI status. As a result, CCPRS acts as an independent risk factor for cancer-specific survival (HR = 2.038, 95% CI = 1.291 - 3.218, p = 0.002) in BCa patients, along with MI and LNM. Among BCa patients who received systemic chemotherapy, those with higher CCPRS exhibited worse cancer-specific survival (HR = 3.415, 95% CI = 1.064 - 10.96, p = 0.0208). Among TCGA MIBC patients who received adjuvant therapies including chemo- or/and radiotherapy, those with higher CCPRS exhibited significantly worse overall survival (HR = 2.150, 95% CI = 1.082 - 4.270, p = 0.0241).

To quantify cancer-specific survival risk for individual BCa patients, a scoring nomogram was generated via the combination of CCPRS and clinicopathological parameters. The predictive capacity of the scoring nomogram is evaluated using time-dependent ROC analysis, with the time-dependent AUC value of 0.944 for 1-year cancer-specific survival and 0.932 for 3-year cancer-specific survival, respectively. In the calibration curve, the prediction of 3-year cancer-specific survival is extremely close to the actual survival probability, which indicated that the nomogram has a high accuracy of cancer-specific survival prediction.

Considering MIBC accounts for a large proportion and acts as a leading cause of death in BCa patients, we sought to improve the risk stratification for MIBC patients via construction of an integrated survival decision tree which combines CCPRS with other clinicopathological features. In detail, six variables including age, gender, LNM, grade, pT and CCPRS were submitted for recursive partitioning analysis, and finally, only CCPRS, LNM and pT remained. Three risk subgroups were defined in the survival decision tree, and we observed that patients in the high-risk subgroup exhibited the worst cancer-specific survival among all three subgroups ($p < 0.0001$). Furthermore, MIBC patients with detailed information from independent cohorts were used to test the classification capacity of the survival decision tree. As expected, significant difference in cancer-specific survival was observed among different risk subgroups identified by decision tree in MIBC patients from the GEO database (p = 0.0090).

We hope the proposed prognostic gene-expression signatures can serve as a useful tool to distinguish high-risk PCa and BCa patients who may benefit from systemic therapies, and help facilitate personalized management of patients with urological malignancies in clinical practice.

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2018; 68: 394-424.

2. Loeb S, Feng Z, Ross A, Trock BJ, Humphreys EB, Walsh PC. Can we stop prostate specific antigen testing 10 years after radical prostatectomy? J Urol. 2011; 186: 500-5.

3. Pound CR, Partin AW, Eisenberger MA, Chan DW, Pearson JD, Walsh PC. Natural history of progression after PSA elevation following radical prostatectomy. JAMA. 1999; 281: 1591-7.

4. Thompson IM, Tangen CM, Paradelo J, Lucia MS, Miller G, Troyer D, et al. Adjuvant Radiotherapy for Pathological T3N0M0 Prostate Cancer Significantly Reduces Risk of Metastases and Improves Survival: Long-Term Followup of a Randomized Clinical Trial. The Journal of urology. 2009; 181: 956-62.

5. Wiegel T, Bottke D, Steiner U, Siegmann A, Golz R, Storkel S, et al. Phase III postoperative adjuvant radiotherapy after radical prostatectomy compared with radical prostatectomy alone in pT3 prostate cancer with postoperative undetectable prostate-specific antigen: ARO 96-02/AUO AP 09/95. J Clin Oncol. 2009; 27: 2924-30.

6. Bolla M, van Poppel H, Collette L, van Cangh P, Vekemans K, Da Pozzo L, et al. Postoperative radiotherapy after radical prostatectomy: a randomised controlled trial (EORTC trial 22911). Lancet. 2005; 366: 572-8.

7. Kamat AM, Hahn NM, Efstathiou JA, Lerner SP, Malmstrom PU, Choi W, et al. Bladder cancer. Lancet. 2016; 388: 2796-810.

8. Mak RH, Hunt D, Shipley WU, Efstathiou JA, Tester WJ, Hagan MP, et al. Long-term outcomes in patients with muscle-invasive bladder cancer after selective bladder-preserving combined-modality therapy: a pooled analysis of Radiation Therapy Oncology Group protocols 8802, 8903, 9506, 9706, 9906, and 0233. J Clin Oncol. 2014; 32: 3801-9.

9. Fonteyne V, Ost P, Bellmunt J, Droz JP, Mongiat-Artus P, Inman B, et al. Curative Treatment for Muscle Invasive Bladder Cancer in Elderly Patients: A Systematic Review. Eur Urol. 2018; 73: 40-50.

10. Kogenaru S, Qing Y, Guo Y, Wang N. RNA-seq and microarray complement each other in transcriptome profiling. BMC Genomics. 2012; 13: 629.

11. Lee JS, Leem SH, Lee SY, Kim SC, Park ES, Kim SB, et al. Expression signature of E2F1 and its associated genes predict superficial to invasive progression of bladder tumors. J Clin Oncol. 2010; 28: 2660-7.

12. van der Heijden AG, Mengual L, Lozano JJ, Ingelmo-Torres M, Ribal MJ, Fernandez PL, et al. A five-gene expression signature to predict progression in T1G3 bladder cancer. Eur J Cancer. 2016; 64: 127-36.

13. Le Goux C, Vacher S, Pignot G, Sibony M, Barry Delongchamps N, Terris B, et al. mRNA Expression levels of genes involved in antitumor immunity: Identification of a 3-gene signature associated with prognosis of muscle-invasive bladder cancer. Oncoimmunology. 2017; 6: e1358330.

14. Ross-Adams H, Lamb AD, Dunning MJ, Halim S, Lindberg J, Massie CM, et al. Integration of copy number and transcriptomics provides risk stratification in prostate cancer: A discovery and validation cohort study. EBioMedicine. 2015; 2: 1133-44.

15. Long Q, Xu J, Osunkoya AO, Sannigrahi S, Johnson BA, Zhou W, et al. Global transcriptome analysis of formalin-fixed prostate cancer specimens identifies biomarkers of disease recurrence. Cancer Res. 2014; 74: 3228-37.

16. Taylor BS, Schultz N, Hieronymus H, Gopalan A, Xiao Y, Carver BS, et al. Integrative genomic profiling of human prostate cancer. Cancer Cell. 2010; 18: 11-22.

17. Ghandi M, Huang FW, Jane-Valbuena J, Kryukov GV, Lo CC, McDonald ER, 3rd, et al. Next-generation characterization of the Cancer Cell Line Encyclopedia. Nature. 2019; 569: 503-8.

18. Yang W, Soares J, Greninger P, Edelman EJ, Lightfoot H, Forbes S, et al. Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. Nucleic Acids Res. 2013; 41: D955-61.

19. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008; 9: 559.

20. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005; 102: 15545-50.

21. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003; 13: 2498-504.

22. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov. 2012; 2: 401-4.

23. Wilkerson MD, Hayes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. Bioinformatics. 2010; 26: 1572-3.

24. Lee DD, Seung HS. Learning the parts of objects by non-negative matrix factorization. Nature. 1999; 401: 788-91.

25. Strobl C, Malley J, Tutz G. An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests. Psychol Methods. 2009; 14: 323-48.

3. Summary

The current tumor-node-metastasis (TNM) classification system is insufficient for precise treatment decision-making or accurate survival prediction for patients with urological malignancies, such as prostate cancer and bladder cancer. Therefore, novel reliable biomarkers are urgently needed to identify the high-risk subset who can benefit from adjuvant therapy after tumor resection.

In recent years, advancements in high-throughput techniques such as microarray and RNA-Seq have provided new insight into transcriptome profiling, highlighting the role of molecule markers in cancer diagnosis and prognosis. However, the exact biological function of each gene in such a gene signature was often unclear, nor the interaction of them. Till now, the clinical utility of these signatures remains limited, and few of them were applied to clinical practice.

In our studies, we combined a series of bioinformatic and statistical analyses to develop gene signatures to predict prognosis in patients with prostate cancer and bladder cancer. In combination of gene signatures and clinicopathological features, survival decision tree or nomogram was established to optimize risk stratification and survival prediction for these patients. The prognostic and predictive capacities of these gene signature-based models were also validated and compared with traditional clinicopathological features in different patient cohorts from public databases.

Some limitations in our study should be acknowledged. First, this is a retrospective study, so the prognostic robustness and clinical utilization of the gene signatures need further validation in prospectively designed clinical trials. Second, further experimental studies are needed to reveal the regulatory role of the gene signatures in urological malignancies.

4. Zusammenfassung

Die derzeitige Klassifizierung der Tumor-Knoten-Metastasierung (TNM) reicht nicht aus, um Patienten mit urologischen Malignitäten wie Prostatakrebs und Blasenkrebs präzise über die Behandlung zu entscheiden oder das Überleben genau vorherzusagen. Daher werden dringend neuartige zuverlässige Biomarker benötigt, um die Hochrisiko-Untergruppe zu identifizieren, die nach einer Tumorresektion von einer adjuvanten Therapie profitieren kann.

In den letzten Jahren haben Fortschritte bei Hochdurchsatztechniken wie Microarray und RNA-Sequenzierung (RNA-seq) neue Einblicke in die Transkriptomprofilierung geliefert und die Rolle von Molekülmarkern bei der Krebsdiagnose und -prognose hervorgehoben. Die genaue biologische Funktion jedes Gens in einer solchen Gensignatur war jedoch oft nicht klar, noch die Wechselwirkung zwischen ihnen. Bis jetzt ist der klinische Nutzen dieser Signaturen begrenzt, und nur wenige davon wurden in der klinischen Praxis angewendet.

In unseren Studien haben wir eine Reihe von bioinformatischen und statistischen Analysen kombiniert, um Gensignaturen zu entwickeln, um die Prognose bei Patienten mit Prostatakrebs und Blasenkrebs vorherzusagen. In Kombination von Gensignaturen und klinisch-pathologischen Merkmalen wurde ein Überlebensentscheidungsbaum oder ein Nomogramm erstellt, um die Risikostratifizierung und Überlebensvorhersage für diese Patienten zu optimieren. Die prognostischen und prädiktiven Fähigkeiten dieser auf Gensignaturen basierenden Modelle wurden ebenfalls validiert und mit traditionellen klinisch-pathologischen Merkmalen in verschiedenen Patientenkohorten aus öffentlichen Datenbanken verglichen.

Einige Einschränkungen in unserer Studie sollten anerkannt werden. Erstens handelt es sich um eine retrospektive Studie, sodass die prognostische Robustheit und der klinische Nutzen der Gensignaturen in prospektiv konzipierten klinischen Studien weiter validiert werden müssen. Zweitens sind weitere experimentelle Studien erforderlich, um die regulatorische Rolle der Gensignaturen bei urologischen Malignitäten aufzudecken.

5. Paper I

cancers

MDPI

Article

A Novel Gene Signature-Based Model Predicts Biochemical Recurrence-Free Survival in Prostate Cancer Patients after Radical Prostatectomy

Run Shi [1,†], Xuanwen Bao [2,3,†], Joachim Weischenfeldt [4,5], Christian Schaefer [1], Paul Rogowski [1], Nina-Sophie Schmidt-Hegemann [1], Kristian Unger [1,6,7], Kirsten Lauber [1,7], Xuanbin Wang [8], Alexander Buchner [9], Christian Stief [9], Thorsten Schlomm [10], Claus Belka [1,7] and Minglun Li [1,*]

1 Department of Radiation Oncology, University Hospital, LMU Munich, D-81377 Munich, Germany; Run.Shi@med.uni-muenchen.de (R.S.); Christian.Schaefer@med.uni-muenchen.de (C.S.); Paul.Rogowski@med.uni-muenchen.de (P.R.); Nina-Sophie.Hegemann@med.uni-muenchen.de (N.-S.S.-H.); unger@helmholtz-muenchen.de (K.U.); Kirsten.Lauber@med.uni-muenchen.de (K.L.); Claus.Belka@med.uni-muenchen.de (C.B.)

2 Institute of Radiation Biology, Helmholtz Center Munich, German Research Center for Environmental Health, D-85764 Neuherberg, Germany; xuanwen.bao@tum.de

3 Technical University of Munich, D-80333 Munich, Germany

4 Biotech Research & Innovation Centre (BRIC) and Finsen Laboratory, University of Copenhagen, Rigshospitalet, DK-2200 Copenhagen, Denmark; joachim.weischenfeldt@bric.ku.dk

5 Charité Universitätsmedizin Berlin, Charité platz 1, D-10117 Berlin, Germany

6 Research Unit Radiation Cytogenetics, Helmholtz Center Munich, German Research Center for Environmental Health GmbH, D-85764 Neuherberg, Germany

7 Clinical Cooperation Group 'Personalized Radiotherapy in Head and Neck Cancer', Helmholtz Zentrum München, German Research Center for Environmental Health GmbH, D-85764 Neuherberg, Germany

8 Laboratory of Chinese Herbal Pharmacology, Oncology Center, Renmin Hospital, Hubei University of Medicine, Shiyan 442000, China; wangxb@hbmu.edu.cn

9 Department of Urology, University Hospital, LMU Munich, D-81377 Munich, Germany; Alexander.Buchner@med.uni-muenchen.de (A.B.); Christian.Stief@med.uni-muenchen.de (C.S.)

10 Martini-Clinic Prostate Cancer Center at the University Medical Center Hamburg-Eppendorf, D-20246 Hamburg, Germany; thorsten.schlomm@charite.de

* Correspondence: Minglun.li@med.uni-muenchen.de; Tel.: +49-(089)-4400-73770; Fax: +49-(089)-4400-76770

† These authors contributed equally to this paper.

Received: 19 November 2019; Accepted: 14 December 2019; Published: 18 December 2019

check for updates

Abstract: Currently, decision-making regarding biochemical recurrence (BCR) following prostatectomy relies solely on clinical parameters. We therefore attempted to develop an integrated prediction model based on a molecular signature and clinicopathological features, in order to forecast the risk for BCR and guide clinical decision-making for postoperative therapy. Using high-throughput screening and least absolute shrinkage and selection operator (LASSO) in the training set, a novel gene signature for biochemical recurrence-free survival (BCRFS) was established. Validation of the prognostic value was performed in five other independent datasets, including our patient cohort. Multivariate Cox regression analysis was performed to evaluate the importance of risk for BCR. Time-dependent receiver operating characteristic (tROC) was used to evaluate the predictive power. In combination with relevant clinicopathological features, a decision tree was built to improve the risk stratification. The gene signature exhibited a strong capacity for identifying high-risk BCR patients, and multivariate Cox regression analysis demonstrated that the gene signature consistently acted as a risk factor for BCR. The decision tree was successfully able to identify the high-risk subgroup. Overall, the gene signature established in the present study is a powerful predictor and risk factor for BCR after radical prostatectomy.

Keywords: prostate cancer; radical prostatectomy; gene signature; risk stratification; biochemical recurrence-free survival

1. Introduction

Prostate cancer (PCa) is the second most commonly diagnosed cancer in men worldwide [1]. Over half of PCa patients will undergo radical prostatectomy as their primary treatment choice [2]. After radical prostatectomy, approximately 20% of patients experience a biochemical recurrence (BCR) with a rising prostate-specific antigen (PSA) level [3]. Several randomized phase III trials have shown that adjuvant radiotherapy is beneficial for patients with high-risk factors such as pathological T3/4 (pT3/4) or R1 resection status [4–6]. However, about 50% of these patients did not suffer a biochemical recurrence without adjuvant radiotherapy, even after a long follow-up of 5 years [7]. For these patients, adjuvant radiotherapy would be an overtreatment, with some risk of unnecessary radiation-induced side effects. Hence, a more precise method to identify patients suffering BCR after radical prostatectomy is a critical issue for the optimal management of PCa.

Nowadays, advancements in high-throughput techniques such as microarray and RNA-sequencing (RNA-seq) have provided new insights into transcriptome profiling, which facilitate the utilization of molecules as diagnostic and prognostic biomarkers [8,9]. Some studies have established gene signatures to help distinguish aggressive PCa tumors or improve survival prediction in PCa patients [10–12]. However, most of these signatures exhibit a prognostic value without having a direct impact on treatment decision-making.

In this study, we established a gene expression-based signature to improve the prediction of BCR after radical prostatectomy, using a univariate and least absolute shrinkage and selection operator (LASSO) Cox model. Then, the prognostic value of the gene signature was further validated in five independent datasets across multiple platforms and our patient cohort. As regards clinical application, the gene signature was combined with clinicopathological features to build a decision tree to improve risk stratification for BCR. In addition, bioinformatic analyses were performed to reveal the biological processes and potential pathways underlying the gene signature.

2. Methods

2.1. Dataset Preparation and Sample Collection

In total, 903 PCa samples with full-scale clinical annotations (age, Gleason score, pathological T stage, surgical margin status, and follow-up BCR information) from six independent cohorts were included in our study. Three cohorts were from Gene Expression Omnibus (GEO), one cohort from The Cancer Genome Atlas (TCGA), one cohort from Memorial Sloan Kettering Cancer Center (MSKCC, Manhattan, NY, US), and our patient cohort was collected from University Medical Center Hamburg-Eppendorf, Germany (Hamburg, Germany). GSE70769 and GSE70768 were involved in the same research [13], and the microarray data were produced with the same chip platform (Illumina HumanHT-12 V4.0 Array). The RNA-seq data of GSE54460 were produced with Illumina HiSeq 2000, for 94 patients with full-scale clinical records [14]. Additionally, RNA-seq data of 388 patients were accessed from TCGA, and microarray data of 138 patients (produced with Affymetrix Human Exon 1.0 ST Array) were obtained from MSKCC [15]. Probe IDs were mapped to gene symbols according to the corresponding annotation file, and expression measurements of all probes linking to the same gene were averaged to obtain a single value. Finally, samples from 84 patients were consecutively collected at the Department of Urology and the Martini Clinics at the University Medical Center Hamburg-Eppendorf (UKE) from 2010 to 2016. Informed consent and an ethical vote in Ethics Commission University Hamburg (ethic codes WF-049/09 and PV3652) were obtained according to the current International Cancer Genome Consortium (ICGC) guidelines (see http://www.icgc.org). Written informed consent

was obtained from each patient as described in our previous study [16]. GSE70769 was used as a training set, while the other five cohorts were used to validate. All microarray and RNA-seq data in our study were normalized and log2 transformed, and expression measurements of multiple samples taken from the same patient were averaged to a single value.

2.2. Candidate Selection and Signature Establishment

The weighted gene co-expression network analysis (WGCNA) R package [17] was used to construct a scale-free co-expression network based on the microarray data of the training cohort. The weighted network adjacency was defined by the formula $a_{i,j} = s_{i,j}^{\beta}$, $s_{i,j} = \left|cor\left(x_i, x_j\right)\right|$. ($x_i$, x_j: each pair of genes; cor: Pearson's correlation; β: soft-power threshold). The topological overlap matrix (TOM) was constructed based on the adjacency, and the corresponding dissimilarity (1-TOM) was used as the distance measure, with deepSplit of 2 and minModuleSize of 30, to assign whole-genome genes into different modules via hierarchical clustering analysis. Unassigned genes were categorized into the gray module. Gene significance quantifies the association of individual genes with biochemical recurrence-free survival (BCRFS) status, and module membership represents the correlation between the module eigengene and the gene expression profile. Among non-gray modules, the modules which had the highest absolute correlations with BCRFS status were selected as candidate modules for further selection. Genes from these modules were submitted for high-throughput univariate Cox regression analysis to screen for prognostic candidates. Subsequently, the LASSO Cox regression model was used to further screen for the most robust prognostic markers [18]. Finally, a risk score (RS) formula was established with individual normalized gene expression values weighted by their LASSO Cox coefficients as follows: $\sum_i Coefficient(mRNA_i) \times Expression(mRNA_i)$.

2.3. Bioinformatic Analyses

WGCNA was used to construct a scale-free co-expression network and to identify the most significant modules, with a risk score based on TCGA RNA-seq data. Hub genes with gene significance >0.3 in the black module were extracted and submitted for Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis, and a Circos diagram was used to visualize outputs [19]. Moreover, gene set enrichment analysis (GSEA) [20] was performed to analyze the potential signaling pathways underlying the gene signature, using gene set "hallmark.all.v6.1.symbols.gmt", based on TCGA RNA-seq data.

2.4. Statistical Analyses

IBM SPSS Statistics version 20 (IBM Corp., Armonk, NY, USA), GraphPad Prism 8.0 (GraphPad Software Inc., San Diego, CA, USA), Stata 12 (StataCorp LLC, College Station, TX, USA) and R software (version 3.5.2, http://www.r-project.org) were used to analyze data and plot graphs. The Kaplan–Meier method was applied to draw survival curves, and the log-rank test was used to evaluate survival difference. The Cox proportional-hazards regression model was used to evaluate the significance of each parameter for biochemical recurrence-free survival (BCRFS). Time-dependent receiver operating characteristic (tROC) analysis was performed to measure the predictive power, using the "survivalROC" package [21], and areas under the curve (AUC) of each variable at different time nodes were compared. Meta-analysis ($I^2 < 30\%$, fixed-effect model) was performed to evaluate the prognostic value in the pooled cohort. The Z-score method was used to normalize risk scores in each cohort. Recursive partitioning analysis (RPA) was performed to construct decision trees using the "rpart" package [22]. Student's t-test or one-way analysis of variance was used to analyze differences between groups in variables with a normal distribution.

3. Results

3.1. Establishment of a Prognostic Gene Signature for BCRFS

First, WGCNA was performed with microarray data and BCRFS status on the training cohort. Sample clustering showed that no outlier was detected (Figure S1). With a power of $\beta = 2$ set as the optimal soft threshold to construct a scale-free network, a total of 31 non-grey modules were identified (Figure 1A). Among these non-grey modules, two modules (darkorange and tan) with the highest absolute correlation values with BCRFS were picked out (Figure 1B). Then, 455 genes from these two modules were submitted for high-throughput univariate Cox regression analysis. With a threshold of $p < 0.01$, 73 promising candidate genes (32 protective and 41 risk markers) were identified (Figure 1C). Next, the LASSO Cox regression model was used to identify robust markers among the 73 candidates. Cross-validation was applied to prevent overfitting, and the optimal λ value of 0.1614 with $\log(\lambda) = -1.8239$ was selected (Figure 1D). Nine genes (*ALDH1A2*, *ASNS*, *SSTR1*, *FAM171B*, *FREM2*, *RSPO2*, *SRD5A2*, *TRIM14*, and *VPS4A*) remained with their individual nonzero LASSO coefficients (Figure 1E). The distribution of LASSO coefficients of the gene signature is demonstrated in Figure 1F. Finally, the risk score (RS) of the gene signature was established as follows:

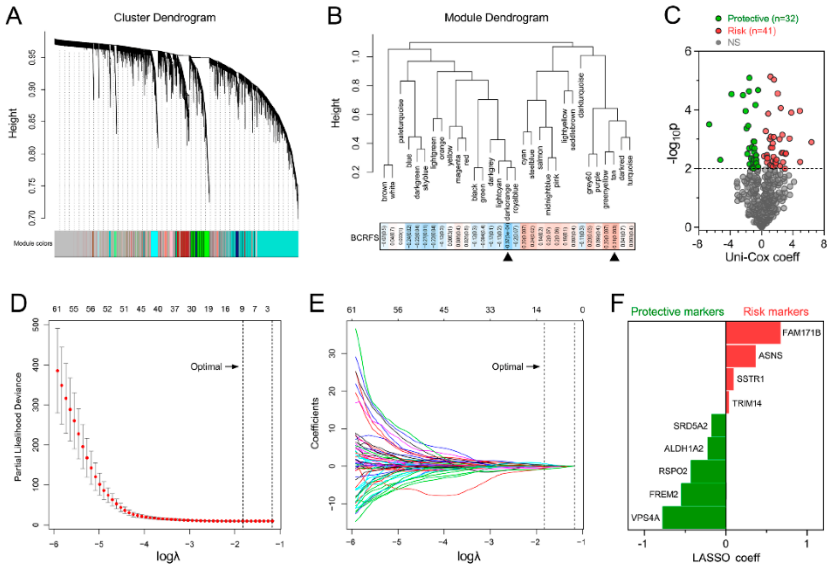


Figure 1. Selection of robust biomarkers to establish a prognostic gene signature. (**A**) Weighted gene co-expression network analysis (WGCNA) was performed to construct a scale-free network, and whole-genome genes from the training cohort were assigned to different modules. (**B**) Two modules (darkorange and tan) were mostly correlated with biochemical recurrence (BCR), and 455 candidates were extracted for further study. (**C**) Univariate Cox regression analysis was performed to screen for significant candidates. (**D**) Cross-validation was applied to prevent overfitting, and an optimal λ value of 0.1614 with $\log(\lambda) = -1.8239$ was selected. (**E**) Nine genes finally remained with their nonzero LASSO coefficients. (**F**) Distribution of least absolute shrinkage and selection operator (LASSO) coefficients of the gene signature.

Risk score = (−0.22345 * expression level of *ALDH1A2*) + (0.364318 * expression level of *ASNS*) + (0.67184 * expression level of *FAM171B*) + (−0.54351 * expression level of *FREM2*) + (−0.4304 * expression level of *RSPO2*) + (−0.17707 * expression level of *SRD5A2*) + (0.094559 * expression level of *SSTR1*) + (0.040268 * expression level of *TRIM14*) + (−0.77555 * expression level of *VPS4A*).

The expression levels of each gene were log2 normalized. Additionally, the expression profiles of the gene signature were mainly dysregulated across 497 tumors and 52 adjacent normal tissues from the TCGA data (Figure S2).

3.2. Gene Signature Serves as a Risk Factor and Promising Predictor for BCRFS

We ranked the risk scores of all patients in the training cohort, and the risk scores of BCR patients were significantly elevated compared with those of BCR-free (BCR-F) ones. Kaplan–Meier survival analysis demonstrated that the two groups exhibited significantly different outcomes (Hazard Ratio (HR) = 5.787, $p < 0.0001$). Multivariate Cox regression analysis showed that the risk score was an independent risk factor for BCRFS (HR = 5.084, $p < 0.0001$). tROC analysis indicated that the risk score also functioned as a powerful predictor for BCR, with an average AUC(t) of 0.836 at 36 months follow-up (Figure 2A).

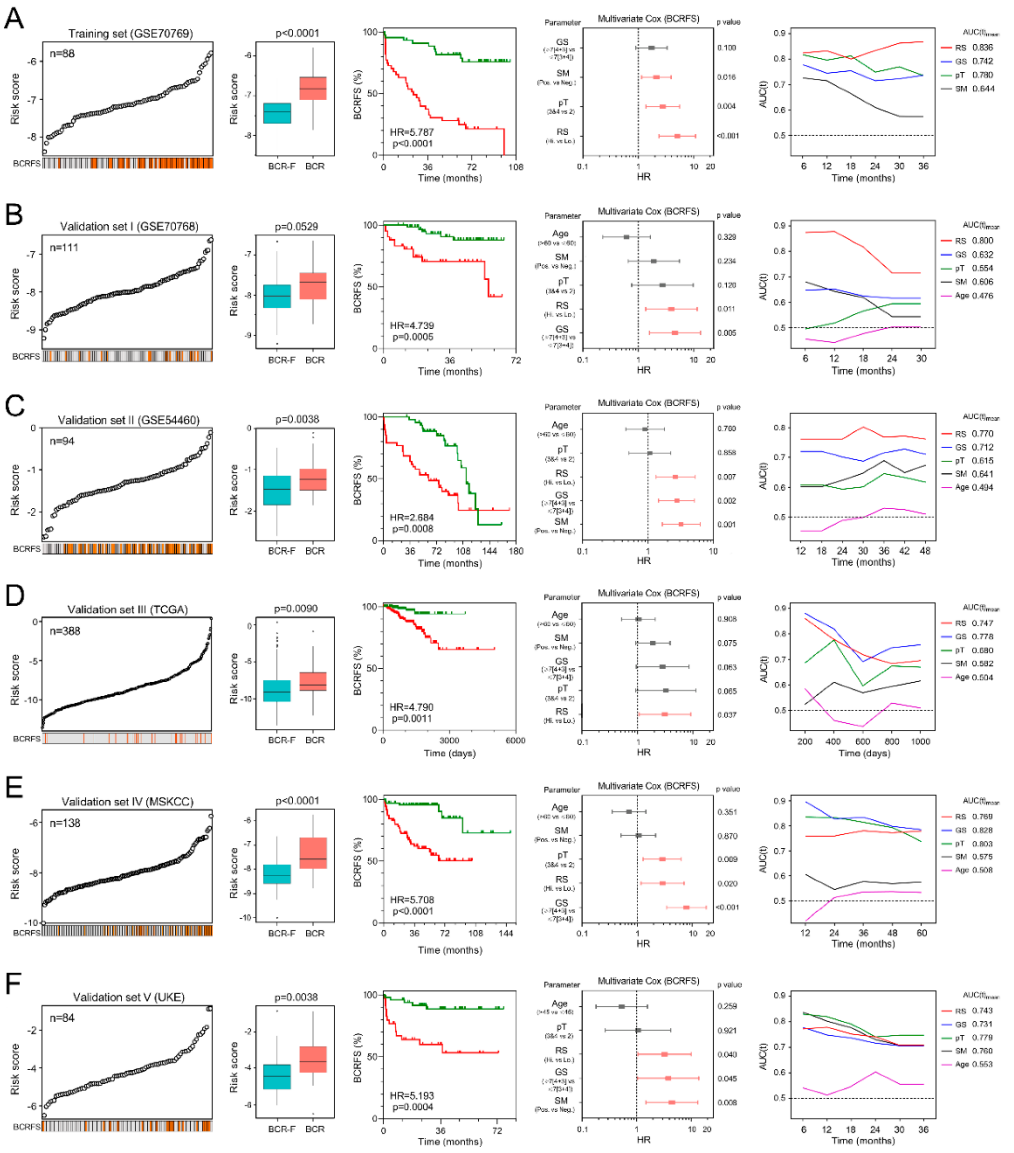


Figure 2. Gene signature serves as a risk factor and promising predictor for biochemical recurrence-free survival (BCRFS) in each cohort. (**A**–**F**) In each cohort, the risk score was significantly elevated in BCR patients compared with BCR-free (BCR-F) ones. Kaplan–Meier analysis showed patients with higher scores exhibited a worse prognosis. The multivariate Cox regression model indicated that the risk score was an independent risk factor for BCRFS in each cohort. Time-dependent receiver operating characteristic (ROC) analysis showed the risk score was a powerful and stable predictor for BCR in each cohort.

To confirm the prognostic robustness of the gene signature, it was further validated in five other independent cohorts (Figure 2B–F). Consistently, patients with higher risk scores exhibited significantly worse BCRFS than patients with lower risk scores in Kaplan–Meier analysis in all five cohorts (validation I: HR = 4.739, $p = 0.0005$; validation II: HR = 2.684, $p = 0.0008$; validation III: HR = 4.790, $p = 0.0011$; validation IV: HR = 5.708, $p < 0.0001$; validation V: HR = 5.193, $p = 0.0004$). Furthermore, multivariate Cox regression analysis was performed on the risk score and clinicopathological features including age, Gleason score (GS), pathological T stage (pT) and surgical margin (SM), to evaluate the significance of each for BCR risk. Notably, the risk score was always an independent risk factor for BCRFS in all five validation series (validation I: HR = 3.979, $p = 0.011$; validation II: HR = 2.616, $p = 0.007$; validation III: HR = 3.120, $p = 0.037$; validation IV: HR = 2.913, $p = 0.020$; validation V: HR = 3.241, $p = 0.040$). tROC analysis demonstrated that the risk score exhibited the most powerful prediction in validations I and II, while having similar predictive power to some clinicopathological parameters such as Gleason score or pT in validations III, IV, and V. Interestingly, among all the clinical variables, age was neither a risk factor nor a promising predictor for BCR in all five validation cohorts.

Next, meta-analysis was used to analyze the prognostic value of the gene signature in the pooled cohort. Our result indicated that a higher risk score was correlated with a significantly worse prognosis in the pooled cohort (HR = 4.84, 95% CI = 2.94–6.74; Figure 3A). Additionally, we normalized risk scores to Z-scores for each cohort and found that Z-scores were significantly elevated in BCR patients compared to BCR-free (BCR-F) patients. Further, Z-score was more sensitive for the prediction of an early biochemical recurrence, as demonstrated in Figure 3B ($p < 0.0001$).

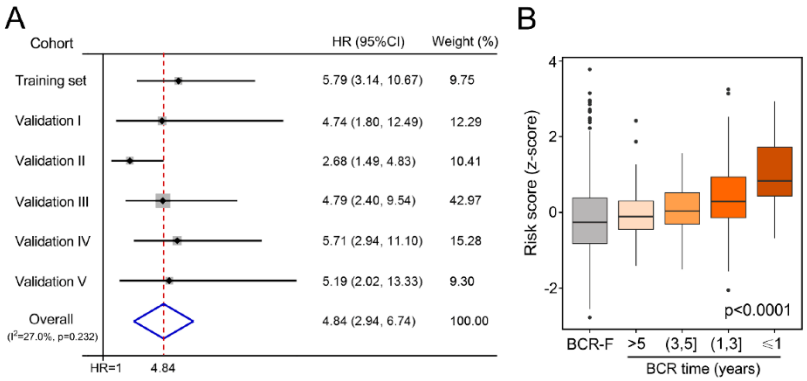


Figure 3. Gene signature-derived risk score could identify high-risk patients in the pooled cohort. (**A**) Meta-analysis indicated that patients with higher risk scores exhibited worse prognosis compared to those with lower ones (HR = 4.84, 95% CI = 2.94–6.74) in the pooled cohort. Additionally, risk scores were normalized to Z-scores in each cohort, and we observed that (**B**) Z-scores of risk scores were significantly elevated in BCR patients compared with BCR-free (BCR-F) patients, especially in shorter-term BCR patients.

3.3. Combination with Clinical Variables to Improve Risk Stratification

Recursive partitioning analysis (RPA) was performed to construct a decision tree to further improve risk stratification for BCR. Based on the pooled cohort, four parameters, namely, GS, pT, SM, and RS, were used as inputs for decision tree construction. Clusters 1–4 (C1–4) with different labels were identified as the outputs of the decision tree. C1 was considered as the low-risk subgroup, C2–3 as intermediate, and C4 as high-risk. The Sankey diagram shows the outcomes of different risk subgroups (Figure 4A). Risk score acted as the dominant factor in the decision tree. Moreover, the low-risk subgroup was labeled with low risk score and negative SM, while the high-risk subgroup was labeled with high risk score and positive SM, further suggesting our signature-derived risk score is the most important factor for risk stratification. Among decision tree-defined subgroups, the high-risk group exhibited the highest BCR rate (Figure 4B) and the worst prognosis (Figure 4C).

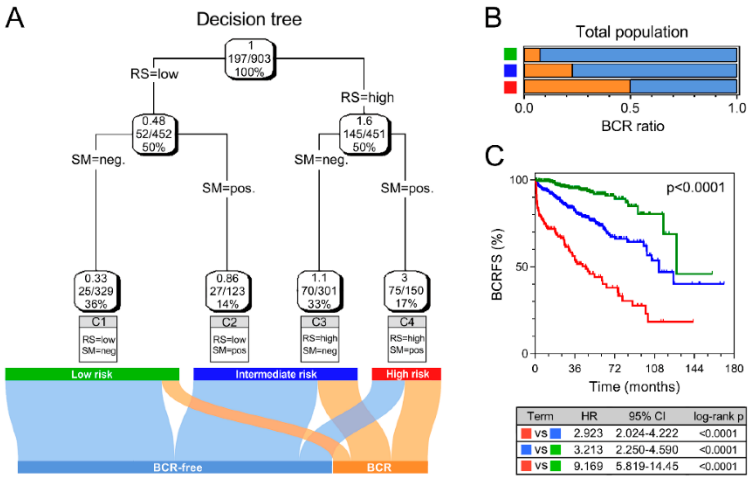


Figure 4. Combination of risk score and clinicopathological features to improve risk stratification and survival prediction. (**A**) A decision tree was generated to optimize risk stratification in the pooled cohort, and risk score served as the dominant component. (**B**,**C**) The high-risk subgroup exhibited the highest BCR rate and worst prognosis.

3.4. Bioinformatic Analyses to Explore Biological Processes Underlying the Gene Signature

First, sample clustering was performed to exclude outliers (Figure S3). Then, the remaining TCGA samples with RNA-seq data and corresponding risk scores were submitted for WGCNA to construct a scale-free co-expression network. Whole-genome cluster dendrogram trees were generated, and a total of 15 non-grey modules were identified (Figure 5A). A heatmap, as shown in Figure 5B, showed the correlations between the risk score and different modules, and the black module presented the highest correlation with the risk score ($r = 0.52$, $p = 8 \times 10^{-28}$). With a threshold of gene significance >0.3, hub genes extracted from the black module were submitted for KEGG enrichment analysis. The Circos diagram showed that hub genes were mainly enriched in terms of "Cell cycle", "Oocyte meiosis", and "Oocyte maturation" (Figure 5C). In addition, GSEA was performed to explore potential pathways using low- and high-risk score samples. As shown in Figure 5D, with the "hallmark" gene set, GSEA showed that the significant predicted signaling pathways are: "E2F targets", "G2M checkpoint", "MYC targets", and "Mitotic spindle".

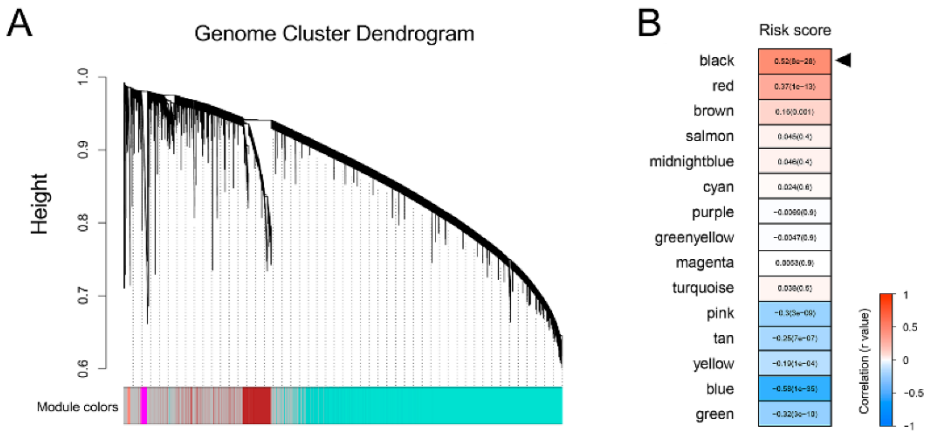


Figure 5. *Cont.*

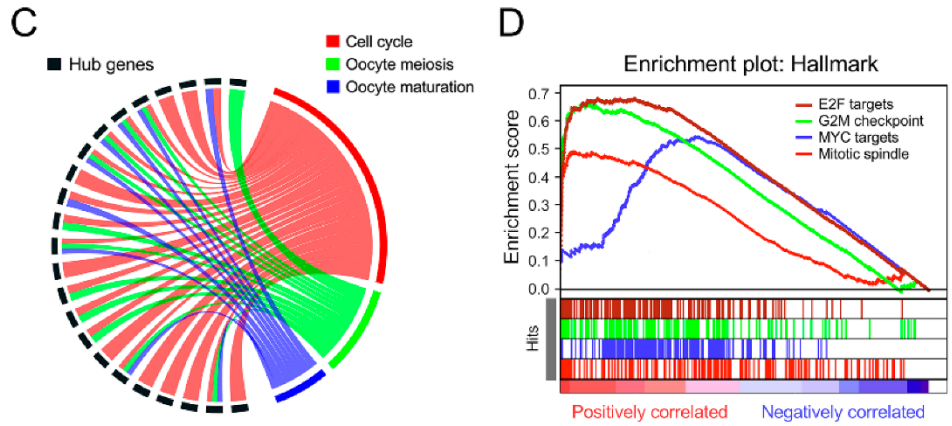


Figure 5. Bioinformatic analyses indicated that the gene signature was correlated with cell cycle-related processes in prostate cancer (PCa). (**A**) WGCNA was performed, and 15 non-grey modules were identified. (**B**) The black module presented the highest correlation with the risk score. (**C**) The Circos diagram showed that hub genes were mainly enriched in cell cycle-related processes. (**D**) GSEA showed that significant predicted signaling pathways were labeled with cell cycle-related hallmarks.

4. Discussion

In recent years, high-throughput transcriptome profiling techniques have been widely applied to identify promising biomarkers for disease diagnosis and prognosis [8,9]. Though some gene signatures have been established to predict the prognosis of PCa patients, few of them have direct relevance to treatment decision-making. In the postoperative setting, the use of adjuvant vs. salvage radiotherapy is until now an unsolved issue permanently under debate. Although three randomized phase III trials demonstrated the benefit of adjuvant radiotherapy in patients with high-risk factors, such as pT3a/b or R1 resection status [4–6], about half of these patients will not suffer BCR, for whom adjuvant radiotherapy would be an overtreatment, with an unnecessary risk of radiation-induced side effects. On the other side, salvage radiotherapy is associated with worse prognosis, particularly in patients with high-risk factors, or in patients with high PSA values at initiation of salvage radiotherapy [23,24]. There is a growing body of evidence that the effectivity of salvage radiotherapy is inversely correlated with increases in the salvage treatment PSA [23,25]. Another issue related to the salvage approach is the unclear definition of BCR which should trigger initiation of salvage radiotherapy. Its PSA-value threshold varied from 0.05 to 0.5 in different clinical trials and guidelines [26]. Taken together, optimally, postoperative radiotherapy should be performed in patients who suffer, albeit with possible low PSA values, or in those who are developing BCR with an unmeasurable PSA value at the initiation of radiotherapy. Until now, the prediction of BCR has been based upon clinical parameters, all of them displaying a low predictive accuracy. Thus, any novel biomarkers for a more accurate prediction of BCR would be of high clinical value.

In the present study, we established a nine-gene expression-based signature for BCRFS prediction in PCa patients after prostatectomy and validated it in five other independent datasets, including our own patient cohort. With the transcriptome profiling data and BCRFS status in the training cohort, WGCNA was performed to identify gene modules mostly correlated with BCR, and subsequently, univariate Cox analysis and a LASSO algorithm were applied to overcome overfitting and thus to screen for the most robust biomarkers. Then, the risk score of each patient was calculated with individual normalized expression level and LASSO coefficient according to the established formula. Overall, Kaplan–Meier analysis indicated that patients with higher risk scores exhibited worse BCRFS in each cohort. Moreover, the risk score always serves as an independent risk factor for BCRFS among all the clinicopathological variables in the multivariate Cox regression model. In addition, time-dependent ROC was performed to evaluate the predictive power at different time nodes during follow-up. We

observed that the risk score was the most powerful predictor in the GSE70769, GSE70768, and GSE54460 datasets, and was an important predictor beside two clinicopathological features (Gleason score and pT stage) in TCGA, MSKCC, and our cohort. Notably, the risk score was the only significant predictor in all six cohorts, with an extremely stable AUC(t) value of at least 0.75 for each cohort. Furthermore, the prognostic value of the risk score was also validated in the pooled cohort with Z-score normalization.

A decision tree was generated to further optimize risk stratification by combining the risk score with traditional clinicopathological factors. In the pooled cohort-derived decision tree, the risk score functioned as the dominant factor for risk stratification. When stratified by the decision tree, BCRFS varied dramatically in different risk subgroups.

Some biomarkers involved in our gene signature have been investigated in cancer, even in prostate cancer. For instance, *ASNS*, one risk biomarker in our study, was reported to function as a therapeutic target in castration-resistant prostate cancer [27]. *SSTR1* was widely related to the progression of various cancers [28–30], and also functions as a prognostic marker in prostate cancer [31,32]. *TRIM14* has been reported to promote invasion in glioblastoma [33] and colorectal cancer [34]. *SRD5A2*, one protective biomarker in our study, inhibits the invasion of prostate cancer cells via regulating the ERK/MAPK pathway [35], and polymorphism in *SRD5A2* contributes to resistance to androgen-deprivation therapy [36]. Regarding *ALDH1A2*, another protective biomarker in our study, it has been reported that the promoter region was significantly hypermethylated in prostate cancer, and overexpression of *ALDH1A2* resulted in decreased colony growth, suggesting that *ALDH1A2* serves as a tumor suppressor in prostate cancer [37]. In addition, *VPS4A* repressed growth and invasion in hepatocellular carcinoma, acting as a tumor suppressor [38]. In a word, the biological roles and clinical significance of the nine genes still need further investigation in PCa.

As our gene signature showed considerable power in risk stratification, the potential biological process and signaling pathways need to be investigated. Using WGCNA co-expression network construction and KEGG enrichment analysis, we observed that the gene signature-related hub genes were mainly enriched in terms of cell cycle. In addition, GSEA indicated that the predicted results that correlated with high risk score were shown as "E2F targets", "G2M checkpoint", "MYC targets", and "Mitotic spindle", which are also mainly involved in cell cycle-related processes. We suppose that the gene signature-derived cell cycle alteration might contribute to cancer progression and poor prognosis in PCa patients.

Some limitations of our study should be acknowledged. First, this is a retrospective study, so the robustness of the predictive value of the gene signature should be further validated in large prospective clinical trials. Second, experimental studies are required to further elucidate the biological functions underlying the gene signature in PCa.

5. Conclusions

In conclusion, we established a gene-expression signature to predict BCRFS in PCa after radical prostatectomy. Integrated with clinicopathological features, a decision tree was generated to further improve the risk stratification for BCR after radical prostatectomy. Our model could be a useful tool for personalized management of PCa patients.

Supplementary Materials: The following are available online at http://www.mdpi.com/2072-6694/12/1/1/s1, Figure S1: Sample clustering showed no outlier was detected. Figure S2: Expression profiles of the gene signature in primary tumor tissues and adjacent normal tissues from TCGA. Figure S3: Sample clustering was performed to exclude outliers.

Author Contributions: Conceptualization, C.B. and M.L.; Formal analysis, R.S. and X.B.; Validation, C.S. (Christian Schaefer) and P.R.; Resources, T.S. and J.W.; Writing—original draft preparation, R.S. and M.L.; Writing—review and editing, A.B. and C.S. (Christian Stief); Investigation, N.-S.S.-H., K.U., K.L., and X.W.; Supervision, J.W., T.S., C.B., and M.L. All authors have read and agreed to the published version of the manuscript.

Funding: We greatly thank the China Scholarship Council (CSC) for supporting the research and work of Run Shi (No. 201708320347) and Xuanwen Bao (No. 201608210186).

Conflicts of Interest: The authors declare no competing interests.

References

1. Bray, F.; Ferlay, J.; Soerjomataram, I.; Siegel, R.L.; Torre, L.A.; Jemal, A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **2018**, *68*, 394–424. [CrossRef] [PubMed]
2. Loeb, S.; Feng, Z.; Ross, A.; Trock, B.J.; Humphreys, E.B.; Walsh, P.C. Can we stop prostate specific antigen testing 10 years after radical prostatectomy? *J. Urol.* **2011**, *186*, 500–505. [CrossRef] [PubMed]
3. Pound, C.R.; Partin, A.W.; Eisenberger, M.A.; Chan, D.W.; Pearson, J.D.; Walsh, P.C. Natural history of progression after PSA elevation following radical prostatectomy. *JAMA* **1999**, *281*, 1591–1597. [CrossRef] [PubMed]
4. Thompson, I.M.; Tangen, C.M.; Paradelo, J.; Lucia, M.S.; Miller, G.; Troyer, D.; Messing, E.; Forman, J.; Chin, J.; Swanson, G.; et al. Adjuvant radiotherapy for pathological T3N0M0 prostate cancer significantly reduces risk of metastases and improves survival: Long-term follow up of a randomized clinical trial. *J. Urol.* **2009**, *181*, 956–962. [CrossRef]
5. Wiegel, T.; Bartkowiak, D.; Bottke, D.; Bronner, C.; Steiner, U.; Siegmann, A.; Golz, R.; Störkel, S.; Willich, N.; Semjonow, A.; et al. Adjuvant radiotherapy versus wait-and-see after radical prostatectomy: 10-year follow-up of the ARO 96-02/AUO AP 09/95 trial. *Eur. Urol.* **2014**, *66*, 243–250. [CrossRef]
6. Bolla, M.; van Poppel, H.; Tombal, B.; Vekemans, K.; Da Pozzo, L.; de Reijke, T.M.; Verbaeys, A.; van Velthoven, R.; Colombel, M.; van de Beek, C.; et al. Postoperative radiotherapy after radical prostatectomy for high-risk prostate cancer: Long-term results of a randomised controlled trial (EORTC trial 22911). *Lancet* **2012**, *380*, 2018–2027. [CrossRef]
7. Wiegel, T.; Bottke, D.; Steiner, U.; Siegmann, A.; Golz, R.; Storkel, S.; Willich, N.; Semjonow, A.; Souchon, R.; Althaus, P.; et al. Phase III postoperative adjuvant radiotherapy after radical prostatectomy compared with radical prostatectomy alone in pT3 prostate cancer with postoperative undetectable prostate-specific antigen: ARO 96-02/AUO AP 09/95. *J. Clin. Oncol.* **2009**, *27*, 2924–2930. [CrossRef]
8. Byron, S.A.; Van Keuren-Jensen, K.R.; Engelthaler, D.M.; Carpten, J.D.; Craig, D.W. Translating RNA sequencing into clinical diagnostics: Opportunities and challenges. *Nat. Rev. Genet.* **2016**, *17*, 257–271. [CrossRef]
9. Simon, R.; Radmacher, M.D.; Dobbin, K.; McShane, L.M. Pitfalls in the use of DNA microarray data for diagnostic and prognostic classification. *J. Natl. Cancer Inst.* **2003**, *95*, 14–18. [CrossRef]
10. Peng, Z.; Skoog, L.; Hellborg, H.; Jonstam, G.; Wingmo, I.L.; Hjalm-Eriksson, M.; Harmenberg, U.; Cedermark, G.C.; Andersson, K.; Ährlund-Richter, L.; et al. An expression signature at diagnosis to estimate prostate cancer patients' overall survival. *Prostate Cancer Prostatic Dis.* **2014**, *17*, 81–90. [CrossRef]
11. Jin, R.; Yi, Y.; Yull, F.E.; Blackwell, T.S.; Clark, P.E.; Koyama, T.; Smith, J.A.; Matusik, R.J. NF-kappaB gene signature predicts prostate cancer progression. *Cancer Res.* **2014**, *74*, 2763–2772. [CrossRef] [PubMed]
12. Pellegrini, K.L.; Sanda, M.G.; Patil, D.; Long, Q.; Santiago-Jimenez, M.; Takhar, M.; Erho, N.; Yousefi, K.; Davicioni, E.; Klein, E.A.; et al. Evaluation of a 24-gene signature for prognosis of metastatic events and prostate cancer-specific mortality. *BJU Int.* **2017**, *119*, 961–967. [CrossRef] [PubMed]
13. Ross-Adams, H.; Lamb, A.D.; Dunning, M.J.; Halim, S.; Lindberg, J.; Massie, C.M.; Egevad, L.A.; Russell, R.; Ramos-Montoya, A.; Vowler, S.L.; et al. Integration of copy number and transcriptomics provides risk stratification in prostate cancer: A discovery and validation cohort study. *EBioMedicine* **2015**, *2*, 1133–1144. [CrossRef]
14. Long, Q.; Xu, J.; Osunkoya, A.O.; Sannigrahi, S.; Johnson, B.A.; Zhou, W.; Gillespie, T.; Park, J.Y.; Nam, R.K.; Sugar, L.; et al. Global transcriptome analysis of formalin-fixed prostate cancer specimens identifies biomarkers of disease recurrence. *Cancer Res.* **2014**, *74*, 3228–3237. [CrossRef] [PubMed]
15. Taylor, B.S.; Schultz, N.; Hieronymus, H.; Gopalan, A.; Xiao, Y.; Carver, B.S.; Arora, V.K.; Kaushik, P.; Cerami, E.; Reva, B.; et al. Integrative genomic profiling of human prostate cancer. *Cancer Cell* **2010**, *18*, 11–22. [CrossRef]
16. Gerhauser, C.; Favero, F.; Risch, T.; Simon, R.; Feuerbach, L.; Assenov, Y.; Heckmann, D.; Sidiropoulos, N.; Waszak, S.M.; Hübschmann, D.; et al. Molecular Evolution of Early-Onset Prostate Cancer Identifies Molecular Risk Markers and Clinical Trajectories. *Cancer Cell* **2018**, *34*, 996–1011.e8. [CrossRef]
17. Langfelder, P.; Horvath, S. WGCNA: An R package for weighted correlation network analysis. *BMC Bioinform.* **2008**, *9*, 559. [CrossRef]

18. Wu, T.T.; Chen, Y.F.; Hastie, T.; Sobel, E.; Lange, K. Genome-wide association analysis by lasso penalized logistic regression. *Bioinformatics* **2009**, *25*, 714–721. [CrossRef]
19. Krzywinski, M.; Schein, J.; Birol, I.; Connors, J.; Gascoyne, R.; Horsman, D.; Jones, S.J.; Marra, M.A. Circos: An information aesthetic for comparative genomics. *Genome Res.* **2009**, *19*, 1639–1645. [CrossRef]
20. Subramanian, A.; Tamayo, P.; Mootha, V.K.; Mukherjee, S.; Ebert, B.L.; Gillette, M.A.; Paulovich, A.; Pomeroy, S.L.; Pomeroy, S.L.; Golub, T.R.; et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 15545–15550. [CrossRef]
21. Heagerty, P.J.; Lumley, T.; Pepe, M.S. Time-dependent ROC curves for censored survival data and a diagnostic marker. *Biometrics* **2000**, *56*, 337–344. [CrossRef] [PubMed]
22. Strobl, C.; Malley, J.; Tutz, G. An introduction to recursive partitioning: Rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychol. Methods* **2009**, *14*, 323–348. [CrossRef] [PubMed]
23. Tendulkar, R.D.; Agrawal, S.; Gao, T.; Efstathiou, J.A.; Pisansky, T.M.; Michalski, J.M.; Koontz, B.F.; Hamstra, D.A.; Feng, F.Y.; Liauw, S.L.; et al. Contemporary Update of a Multi-Institutional Predictive Nomogram for Salvage Radiotherapy After Radical Prostatectomy. *J. Clin. Oncol.* **2016**, *34*, 3648–3654. [CrossRef] [PubMed]
24. Gandaglia, G.; Briganti, A.; Clarke, N.; Karnes, R.J.; Graefen, M.; Ost, P.; Zietman, A.L.; Roach, M., III. Adjuvant and Salvage Radiotherapy after Radical Prostatectomy in Prostate Cancer Patients. *Eur. Urol.* **2017**, *72*, 689–709. [CrossRef] [PubMed]
25. King, C.R. Adjuvant radiotherapy after prostatectomy: Does waiting for a detectable prostate-specific antigen level make sense? *Int. J. Radiat. Oncol. Biol. Phys.* **2011**, *80*, 1–3. [CrossRef]
26. Drost, F.H.; Osses, D.; Nieboer, D.; Bangma, C.H.; Steyerberg, E.W.; Roobol, M.J.; Schoots, I.G. Prostate Magnetic Resonance Imaging, with or Without Magnetic Resonance Imaging-targeted Biopsy, and Systematic Biopsy for Detecting Prostate Cancer: A Cochrane Systematic Review and Meta-analysis. *Eur. Urol.* **2019**. [CrossRef]
27. Sircar, K.; Huang, H.; Hu, L.; Cogdell, D.; Dhillon, J.; Tzelepi, V.; Efstathiou, E.; Koumakpayi, I.H.; Saad, F.; Luo, D.; et al. Integrative molecular profiling reveals asparagine synthetase is a target in castration-resistant prostate cancer. *Am. J. Pathol.* **2012**, *180*, 895–903. [CrossRef]
28. Zhao, J.; Liang, Q.; Cheung, K.F.; Kang, W.; Dong, Y.; Lung, R.W.; Tong, J.H.-M.; To, K.-F.; Sung, J.J.Y.; Yu, J. Somatostatin receptor 1, a novel EBV-associated CpG hypermethylated gene, contributes to the pathogenesis of EBV-associated gastric cancer. *Br. J. Cancer* **2013**, *108*, 2557–2564. [CrossRef]
29. Vesterinen, T.; Leijon, H.; Mustonen, H.; Remes, S.; Knuuttila, A.; Salmenkivi, K.; Vainio, P.; Arola, J.; Haglund, C. Somatostatin Receptor Expression Is Associated With Metastasis and Patient Outcome in Pulmonary Carcinoid Tumors. *J. Clin. Endocrinol. Metab.* **2019**, *104*, 2083–2093. [CrossRef]
30. Misawa, K.; Misawa, Y.; Kondo, H.; Mochizuki, D.; Imai, A.; Fukushima, H.; Uehara, T.; Kanazawa, T.; Mineta, H. Aberrant methylation inactivates somatostatin and somatostatin receptor type 1 in head and neck squamous cell carcinoma. *PLoS ONE* **2015**, *10*, e0118588. [CrossRef]
31. Kosari, F.; Munz, J.M.; Savci-Heijink, C.D.; Spiro, C.; Klee, E.W.; Kube, D.M.; Tillmans, L.; Slezak, J.; Karnes, J.; Cheville, J.C.; et al. Identification of prognostic biomarkers for prostate cancer. *Clin. Cancer Res.* **2008**, *14*, 1734–1743. [CrossRef] [PubMed]
32. Pedraza-Arevalo, S.; Hormaechea-Agulla, D.; Gomez-Gomez, E.; Requena, M.J.; Selth, L.A.; Gahete, M.D.; Castaño, J.P.; Luque, R.M. Somatostatin receptor subtype 1 as a potential diagnostic marker and therapeutic target in prostate cancer. *Prostate* **2017**, *77*, 1499–1511. [CrossRef] [PubMed]
33. Feng, S.; Cai, X.; Li, Y.; Jian, X.; Zhang, L.; Li, B. Tripartite motif-containing 14 (TRIM14) promotes epithelial-mesenchymal transition via ZEB2 in glioblastoma cells. *J. Exp. Clin. Cancer Res.* **2019**, *38*, 57. [CrossRef] [PubMed]
34. Jin, Z.; Li, H.; Hong, X.; Ying, G.; Lu, X.; Zhuang, L.; Wu, S. TRIM14 promotes colorectal cancer cell migration and invasion through the SPHK1/STAT3 pathway. *Cancer Cell Int.* **2018**, *18*, 202. [CrossRef]
35. Aggarwal, S.; Singh, M.; Kumar, A.; Mukhopadhyay, T. SRD5A2 gene expression inhibits cell migration and invasion in prostate cancer cell line via F-actin reorganization. *Mol. Cell Biochem.* **2015**, *408*, 15–23. [CrossRef] [PubMed]

36. Shiota, M.; Fujimoto, N.; Yokomizo, A.; Takeuchi, A.; Itsumi, M.; Inokuchi, J.; Tatsugami, K.; Uchiumi, T.; Naito, S. SRD5A gene polymorphism in Japanese men predicts prognosis of metastatic prostate cancer with androgen-deprivation therapy. *Eur. J. Cancer* **2015**, *51*, 1962–1969. [CrossRef]
37. Kim, H.; Lapointe, J.; Kaygusuz, G.; Ong, D.E.; Li, C.; van de Rijn, M.; Brooks, J.D.; Pollack, J.R. The retinoic acid synthesis gene ALDH1a2 is a candidate tumor suppressor in prostate cancer. *Cancer Res.* **2005**, *65*, 8118–8124. [CrossRef]
38. Wei, J.X.; Lv, L.H.; Wan, Y.L.; Cao, Y.; Li, G.L.; Lin, H.M.; Zhou, R.; Cao, J.; He, H.; Zhou, G.; et al. Vps4A functions as a tumor suppressor by regulating the secretion and uptake of exosomal microRNAs in human hepatoma cells. *Hepatology* **2015**, *61*, 1284–1294. [CrossRef]

6. Paper II

cancers

MDPI

Article

Establishment and Validation of an Individualized Cell Cycle Process-Related Gene Signature to Predict Cancer-Specific Survival in Patients with Bladder Cancer

Run Shi [1,†], Xuanwen Bao [2,†], Paul Rogowski [1], Christian Schäfer [1], Nina-Sophie Schmidt-Hegemann [1], Kristian Unger [1,3], Shun Lu [4], Jing Sun [1], Alexander Buchner [5], Christian Stief [5], Claus Belka [1] and Minglun Li [1,*]

1 Department of Radiation Oncology, University Hospital, LMU Munich, D-81377 Munich, Germany; Run.Shi@med.uni-muenchen.de (R.S.); paul.Rogowski@med.uni-muenchen.de (P.R.); Christian.Schaefer@med.uni-muenchen.de (C.S.); Nina-Sophie.Hegemann@med.uni-muenchen.de (N.-S.S.-H.); unger@helmholtz-muenchen.de (K.U.); Jing.Sun@med.uni-muenchen.de (J.S.); Claus.Belka@med.uni-muenchen.de (C.B.)

2 Technical University of Munich, D-80333 Munich, Germany; xuanwen.bao@tum.de

3 Research Unit Radiation Cytogenetics, Helmholtz Center Munich, German Research Center for Environmental Health GmbH, D-85764 Neuherberg, Germany

4 Department of Radiotherapy, Sichuan Cancer Hospital & Institute, Sichuan Cancer Center, School of Medicine, University of Electronic Science and Technology of China, Chengdu 610041, China; lushun1982@live.cn

5 Department of Urology, University Hospital, LMU Munich, D-81377 Munich, Germany; Alexander.Buchner@med.uni-muenchen.de (A.B.); Christian.Stief@med.uni-muenchen.de (C.S.)

* Correspondence: minglun.li@med.uni-muenchen.de

† These authors contributed equally to this paper.

Received: 7 April 2020; Accepted: 24 April 2020; Published: 2 May 2020

Abstract: More accurate models are essential to identify high-risk bladder cancer (BCa) patients who will benefit from adjuvant therapies and thus helpful to facilitate personalized management of BCa. Among various cancer-related hallmarks and pathways, cell cycle process (CCP) was identified as a dominant risk factor for cancer-specific survival (CSS) in BCa. Using a series of bioinformatic and statistical approaches, a CCP-related gene signature was established, and the prognostic value was validated in other independent BCa cohorts. In addition, the risk score derived from the gene signature serves as a promising marker for therapeutic resistance. In combination with clinicopathological features, a nomogram was constructed to provide more accurate prediction for CSS, and a decision tree was built to identify high-risk subgroup of muscle invasive BCa patients. Overall, the gene signature could be a useful tool to predict CSS and help to identify high-risk subgroup of BCa patients, which may benefit from intensified adjuvant therapy.

Keywords: bladder cancer; cell cycle process; gene signature; cancer-specific survival; therapeutic resistance

1. Introduction

Bladder cancer (BCa) is a common malignancy in the urological system worldwide, with an estimated 430,000 newly diagnosed cases per year. Among these cases, about two-thirds are non-muscle-invasive bladder cancer (NMIBC), while the rest are classified as muscle-invasive bladder cancer (MIBC) [1]. Despite improved understanding of BCa biology and advances in treatments,

outcomes of BCa patients remain suboptimal. For example, the standard treatment for localized MIBC is radical cystectomy with bilateral pelvic lymph node dissection, which provides a 5-year overall survival rate less than 50% [2]. Although Tumor-Node-Metastasis (TNM) staging and pathological grading systems are widely used for cancer management and survival prediction, clinical outcomes remain variable in BCa patients, even with similar characteristics [3]. Hence, establishment of a more precise model is essential for individual patients to identify high-risk subgroup who may benefit from systemic adjuvant therapies.

In recent years, advancements in high-throughput techniques such as microarray and RNA-sequencing (RNA-seq) have provided new insight into transcriptome profiling, highlighting the role of molecule markers in cancer diagnosis and prognosis [4,5]. Several groups have developed gene signatures to predict progression or survival in BCa patients [6–8]. However, the exact biological function of each gene in such a gene signature was often not clear, nor the interaction of them. Till now, the clinical utility of these signatures remains limited, and few of them were applied to clinical practice.

In this study, we mined public databases and developed a cell cycle process-related risk score (CCPRS) to predict cancer-specific survival (CSS) in BCa patients, which was further validated in other independent cohorts from Gene Expression Omnibus (GEO) and The Cancer Genome Atlas (TCGA). In combination with clinicopathological features, we aimed at constructing an integrated model to improve predictive power and risk stratification for CSS of individual BCa patients.

2. Materials and Methods

2.1. Dataset Preparation and Data Processing

Four microarray datasets including 587 BCa patients with clinical annotations and CSS information were downloaded from Gene Expression Omnibus (GEO). GSE13507 was produced by Illumina human-6 v2.0 expression beadchip and was used as the training set in our study. GSE31684 (Affymetrix Human Genome U133 Plus 2.0 Array) and GSE32894 and GSE32548 (Illumina HumanHT-12 V3.0 expression beadchip) were used as independent validation cohorts. Probe IDs were mapped to gene symbols according to the corresponding annotation file, and expression measurements of all probes linking to a same gene were averaged to obtain a single value. RSEM-normalized RNA-seq data, copy number data, and clinical phenotypes of MIBC samples were obtained from The Cancer Genome Atlas (TCGA). Copy number data and TPM data of 22 bladder cell lines were obtained from Cancer Cell Line Encyclopedia (CCLE) [9]. RMA normalized microarray expression data and IC50 values of different drugs for 1018 cell lines were accessed from Genomics of Drug Sensitivity in Cancer (GDSC) [10]. All microarrays and RNA-seq data included in this study were normalized and log2 transformed.

2.2. Candidate Selection and Signature Establishment

In brief, the levels of cancer-related hallmarks and pathways in each sample from the training set were quantified by a single-sample gene set enrichment analysis (ssGSEA) [11] algorithm based on the transcriptome profiling data and corresponding gene sets retrieved from Molecular Signatures Database (MSigDB) [12]. Cox proportional-hazards regression model was used to evaluate the risk of each hallmark and pathway for CSS in BCa patients using R package 'survival'. Weighted correlation network analysis (WGCNA) was used to construct a scale-free co-expression network using R package 'wgcna' [13] and identify the module which is mostly correlated with CCP based on transcriptome profiling data and CCP ssGSEA score. Gene significance (GS) represents the association of individual genes with CCP ssGSEA score, and module membership (MM) depicts the correlation between module eigengenes and gene expression profiles. With a threshold of p value of GS < 0.0001 and p value of univariate Cox regression < 0.0001, 64 candidates from CCP module remained. Subsequently, a least absolute shrinkage and selection operator (LASSO) Cox regression model was used to further screen out the most robust prognostic genes [14]. A detailed screening diagram is shown in Figure S1. A cell

cycle process-related risk score (CCPRS) was calculated by including normalized gene expression value weighed by LASSO Cox coefficient as follows:

$$CCPRS = \sum_{i} Coefficient(mRNA_i) \times Expression(mRNA_i)$$

2.3. Additional Bioinformatic and Statistical Analyses

IBM SPSS Statistics 20 (IBM Corp., Armonk, NY, USA), GraphPad Prism 8.0 (GraphPad Software Inc, San Diego, CA, USA), Stata 12 ((StataCorp LLC, College Station, TX, USA), and R software (version 3.5.2, http://www.r-project.org) were used to analyze data and plot graphs. Z-score method was used to normalize ssGSEA scores and CCPRS when necessary. GSEA [15] was performed to confirm the positive regulation role in CCP. Unsupervised hierarchical clustering was performed to show distance between hallmarks and pathways, and a network depicting their relationships and connectivity was generated by the Cytoscape software [16]. Principal coordinates analysis (PCoA) was used to visualize dissimilarity of two groups based on Bray–Curtis distance matrix. Circos plot was used to visualize enrichment results of Gene Ontology analysis and overlapping genes involved in different biological processes. The webtool cBioPortal for Cancer Genomics [17] was used to visualize the genomic alterations of the established gene signature in BCa samples from TCGA. The Kaplan–Meier method was used to draw survival curves, and the log-rank test was performed to evaluate survival difference. Cox proportional-hazards regression model was used to evaluate the significance of each parameter for the risk of CSS. Meta-analysis was performed to evaluate the prognostic value of CCPRS in the pooled cohort. Pearson correlation coefficient is used to evaluate the relationship between two continuous variables with a normal distribution. K-means-based consensus clustering using R package 'ConsensusClusterPlus' [18] or non-negative matrix factorization (NMF) consensus clustering [19] using R package 'nmf' was performed to obtain subgroups based on a gene expression matrix. Nomogram and calibration analysis were generated using 'rms' package. Time-dependent receiver operating characteristic (tROC) analysis was performed to measure the predictive power of the nomogram using 'survivalROC' package, and a calibration curve was plotted to visualize the predictive accuracy of the nomogram. Recursive partitioning analysis was performed to construct a decision tree of risk stratification for CSS with R package 'rpart' [20]. Student's *t*-test or one-way analysis of variance (ANOVA) was used to analyze differences between groups in variables with a normal distribution. A p value less than 0.05 was considered statistically significant.

3. Results

3.1. Schematic Diagram of the Study Design

First, cell cycle process (CCP) was identified as the dominant risk factor for CSS in BCa patients among various cancer-related hallmarks and pathways (Figure 1A). Then, WGCNA was performed to identify a CCP module. A series of screening methods including LASSO algorithm were used to screen out most promising candidates and to develop a robust CCP-related gene signature for CSS prediction (Figure 1B). Subsequently, a CCPRS formula was established to quantify risk assessment for BCa patients, and the prognostic value was evaluated in the training and independent validation cohorts. Meta-analysis was performed to evaluate CCPRS in the pooled cohort (Figure 1C). In addition, response to anti-cancer therapies was evaluated to investigate whether the gene signature is a valuable marker for therapeutic resistance. Regarding clinical application, based on the combination of CCPRS and traditional prognostic variables, a nomogram was generated to quantify the risk assessment for individuals and a decision tree was constructed to improve risk stratification for CSS in BCa patients (Figure 1D).

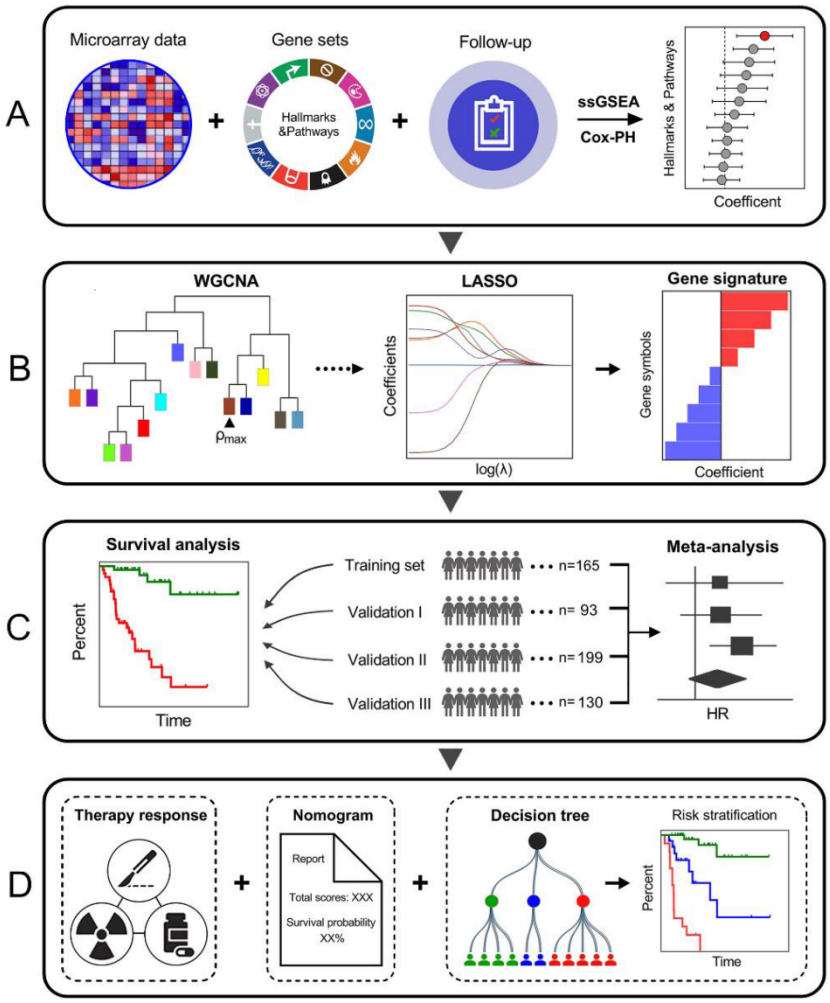


Figure 1. Schematic diagram of the study design. (**A**) Among various cancer-related hallmarks and pathways, cell cycle process (CCP) was identified as the dominant risk factor for cancer-specific survival (CSS) in bladder cancer (BCa) patients. (**B**) Combined methods were used to establish a robust CCP-related gene signature for CSS. (**C**) The prognostic value of the gene signature was further validated in different cohorts. (**D**) Clinical application.

3.2. Cell Cycle Process was Identified as the Primary Risk Factor for CSS

The levels of each cancer-related hallmark and pathway were quantified by ssGSEA. A hierarchical clustering dendrogram was generated to show the distance between different hallmarks and pathways, and WNT/β-catenin signaling appeared to be the most distinct (Figure 2A). Moreover, a network depicting their relationships is shown in Figure 2B. The network was constructed based on the soft threshold of connectivity derived from Pearson correlation between any two nodes (coefficient < 0.5 was ignored). In the network, bigger size represented a more significant role, and solid lines represented higher correlation. Subsequently, Cox proportional-hazards regression analysis was performed based on ssGSEA scores of cancer-related hallmarks and pathways and CSS information in the training set, and bubble heatmap indicated that CCP serves as the most powerful risk factor for CSS (Figure 2C). Kaplan–Meier curve demonstrated that patients with higher CCP ssGSEA scores exhibited worse CSS compared to those with lower scores when the median value was the cut-off value (HR = 3.804, 95% CI = 1.893–7.643, $p = 0.0004$; Figure 2D).

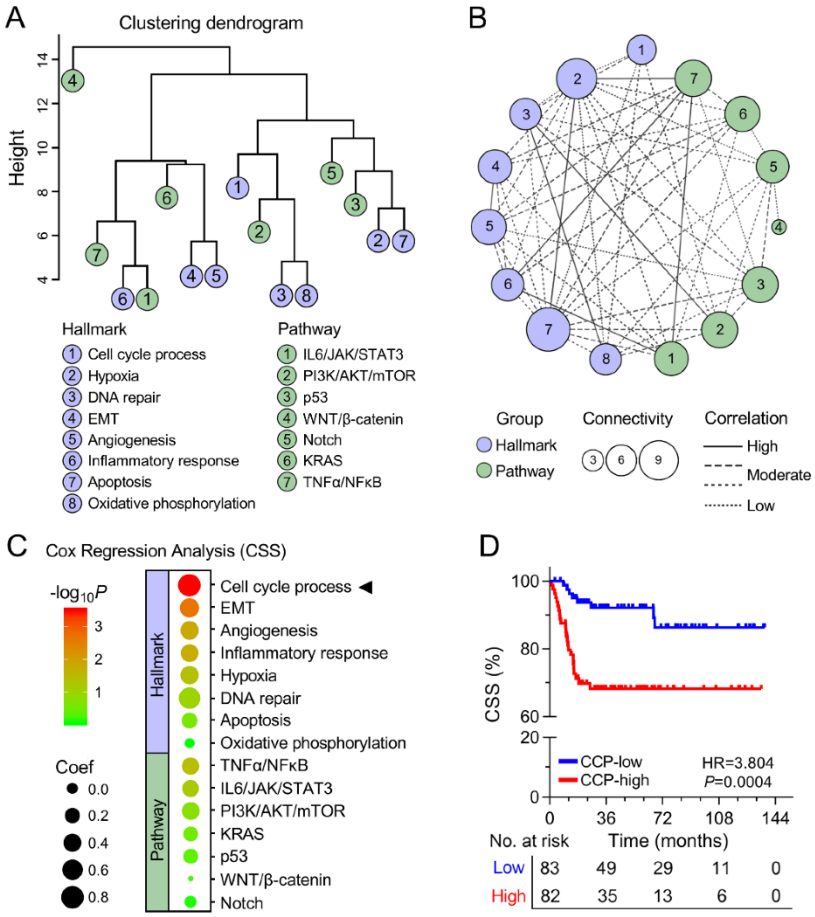


Figure 2. CCP was identified as the primary risk factor for CSS. (**A**) Hierarchical clustering dendrogram was generated to show the distance between different hallmarks and pathways. (**B**) A network was constructed to depict relationships between different hallmarks and pathways. **(C)** Bubble heatmap of Cox results indicated that CCP serves as the most powerful risk factor for CSS. (**D**) Kaplan–Meier curve demonstrated that patients with higher CCP single-sample gene set enrichment analysis (ssGSEA) scores exhibited worse CSS compared to those with lower scores.

3.3. Identification of a Cell Cycle Process-Related Gene Module

Firstly, WGCNA was performed with transcriptome profiling data and CCP ssGSEA Z-scores in the training set (Figure 3A). Sample clustering showed that one outlier was detected and excluded (Figure S2). A total of 25 non-grey modules were generated with a power of $\beta = 9$ as the optimal soft threshold to ensure a scale-free co-expression network (Figure S3). Among these modules, the brown module depicting the highest correlation with CCP ssGSEA scores was considered as "CCP module" ($r = 0.87$, $p = 7 \times 10^{-52}$; Figure 3B). The scatter diagram showed a highly positive correlation between GS and MM in the brown module ($r = 0.87$, $p < 1 \times 10^{-200}$; Figure S4), indicating this module is highly correlated with CCP. Based on the Bray–Curtis distance matrix derived from the expression pattern of the brown module genes of BCa samples, PCoA was performed to visualize the dissimilarity. We observed that samples of low CCP and high CCP were clearly separated into two discrete groups (Figure 3C). Then, all the 1126 genes involved in the brown module were submitted to Gene Ontology for enrichment analysis. Circos plot demonstrated that five most significant processes were labelled with cell cycle-related features such as cell division, DNA replication, G1/S transition, mitotic nuclear division, and sister chromatid cohesion, with a high proportion of overlapping genes involved (Figure 3D).

To further confirm whether the brown module could represent CCP, a K-means-based consensus clustering was performed to classify the training group into different subgroups according to the expression patterns of the 1126 genes involved in the brown module. Cumulative distribution function (CDF) plot showed the cumulative distribution functions of the consensus matrix for each k (from 2 to 8, indicated by colors), which is a quantification of how entries of the consensus matrix are distributed within the range from 0 to 1 (Figure 3E). The difference between two CDF curves (k and k − 1) is illustrated by measuring the area under the CDF curves from k = 2 to 8 in the delta area plot (Figure 3F). Then, we chose k = 2 and 3 as optimal parameters to divide the training cohort into different subgroups (Figure 3G,J). The violin plot showed that CCP ssGSEA scores were significantly elevated in cluster 2 (C2) compared to cluster 1 (C1) ($p < 0.0001$; Figure 3H), and patients in C2 exhibited worse CSS ($p = 0.0002$; Figure 3I). Similar CCP ssGSEA distribution ($p < 0.0001$; Figure 3K) and CSS difference ($p = 0.0011$; Figure 3L) were observed in three different subgroups derived from consensus clustering when k = 3. These results demonstrated that genes involved in the brown module could represent the feature of CCP.

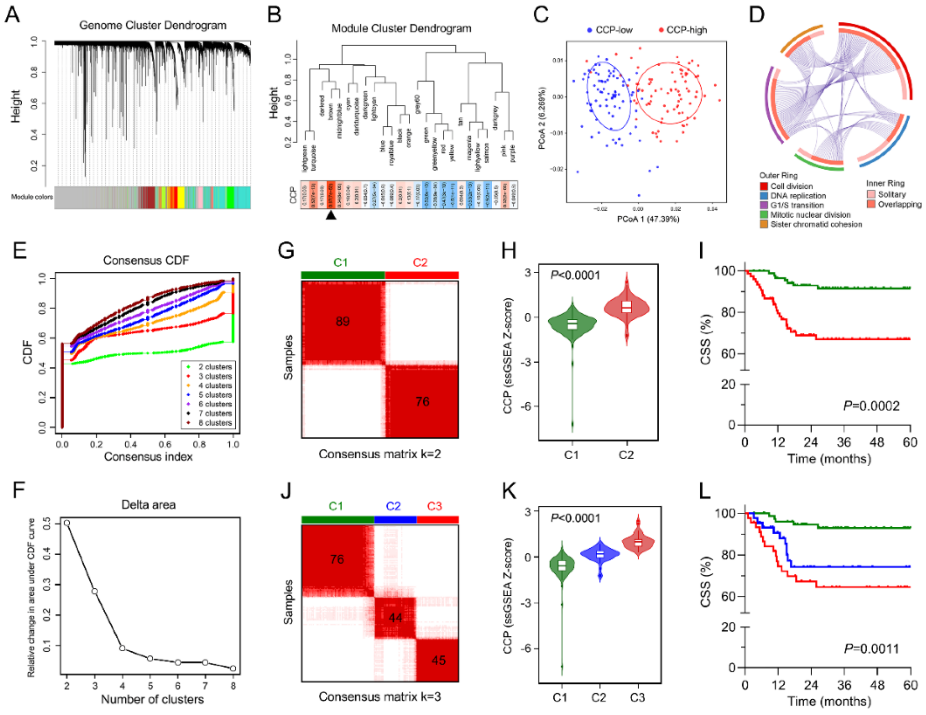


Figure 3. Identification of a CCP-related gene module. (**A**) Weighted correlation network analysis (WGCNA) was performed with transcriptome profiling data and CCP ssGSEA Z-scores in the training set. (**B**) The brown module depicting the highest correlation (r = 0.87, p = 7e–52) with CCP ssGSEA scores was considered as "CCP module". (**C**) Principal coordinates analysis (PCoA) was performed to visualize the dissimilarity of CCP-low and CCP-high samples based on the Bray–Curtis distance matrix. (**D**) Circos plot demonstrated that all of five most significant processes were labelled with cell cycle-related features. (**E**) Based on expression pattern of genes involved in the brown module, K-means-based consensus clustering was performed to classify the training group into different subgroups. Cumulative distribution function (CDF) plot showed the cumulative distribution functions of the consensus matrix for each k (from 2 to 8, indicated by colors). (**F**) The difference between two CDF curves (k and k-1) is illustrated by measuring the area under the CDF curves from k = 2 to 8 in the delta area plot. (**G**,**J**) k = 2 and 3 were chosen as optimal parameters to divide the training cohort into different subgroups. (**H**,**K**) Violin plots showed that CCP ssGSEA scores were differentially distributed in identified subgroups. (**I**,**L**) Patients exhibited different CSS in identified subgroups.

3.4. Establishment of a CCP-Related Gene Signature for CSS

With a threshold of p value for GS less than 0.0001, hub genes from the brown module were submitted to univariate Cox proportional-hazards model. The volcano plot showed that with a threshold of p value for Cox regression analysis less than 0.0001, 64 promising candidates were filtered out (Figure 4A). Subsequently, the LASSO algorithm was used to identify the most robust prognostic genes for CSS. Cross-validation was applied to overcome over-fitting effect (Figure 4B), and the optimal λ value of 0.0585 was selected (Figure 4C). An ensemble of 12 genes remained with their individual LASSO coefficients, and the distribution of LASSO coefficients of the gene signature is shown in Figure 4D. NMF consensus clustering was used to divide the training cohort into two subgroups based on the expression matrix of the established gene signature when k = 2 (Figure 4E). According to the expression pattern of the two subgroups, cluster 1 was labeled with CCP-high and cluster 2 as CCP-low. GSEA analysis confirmed the positive regulation role of the gene signature in cell cycle process in BCa samples by comparing cluster 1 and cluster 2 (Nominal $p < 0.0001$; Figure 4F). Considering cell cycle process is tightly correlated with cancer cell proliferation, we investigated the correlation between CCPRS and MKI67 expression. Pearson correlation test indicated that CCPRS was significantly positively correlated with MKI67 in 165 BCa samples (r = 0.8033, $p < 0.0001$; Figure 4G). In addition, CCPRS was significantly elevated in MIBC samples, especially in more advanced stages ($p < 0.0001$; Figure 4H). These results suggested that the gene signature enhanced proliferative ability via positive regulation of CCP in BCa.

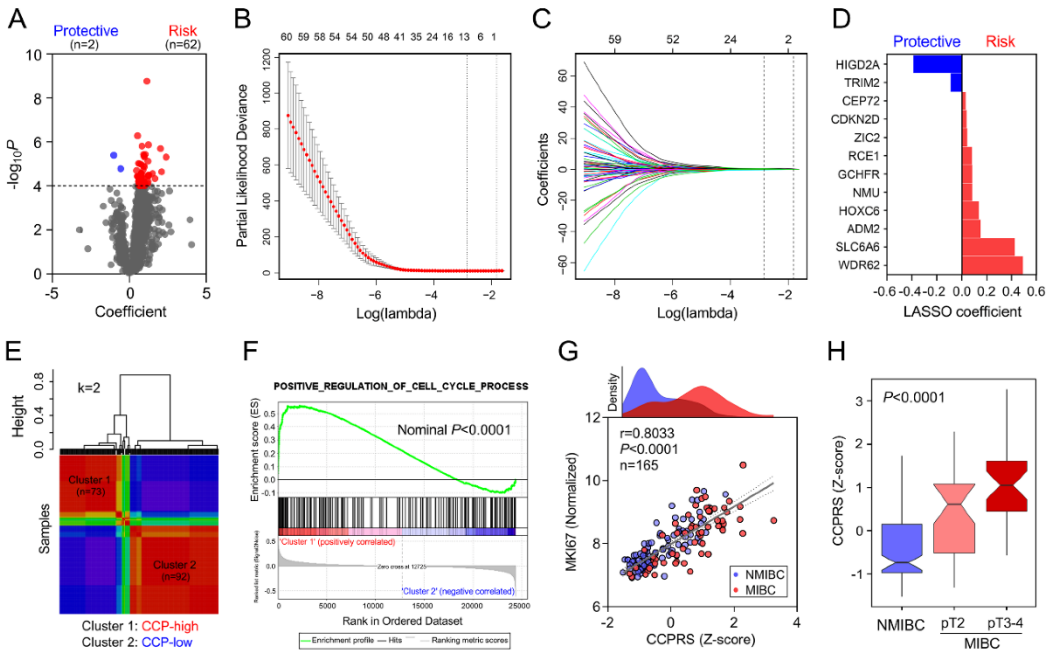


Figure 4. Establishment of a CCP-related gene signature for CSS. (**A**) Volcano plot showed that 64 promising candidates were filtered out. (**B**) Least absolute shrinkage and selection operator (LASSO) algorithm was used to identify the most robust prognostic genes for CSS. Cross-validation was applied to overcome over-fitting effect. (**C**) An optimal λ value of 0.0585 was selected. (**D**) Distribution of LASSO coefficients of the established gene signature. (**E**) Non-negative matrix factorization (NMF) consensus clustering was used to divide the training cohort into two subgroups based on the expression matrix of the established gene signature when k = 2. (**F**) GSEA analysis confirmed the positive regulation role of the gene signature in cell cycle process in BCa samples. (**G**) Cell cycle process-related risk score (CCPRS) was significantly positively correlated with MKI67 in 165 BCa samples (r = 0.8033, $p < 0.0001$). (**H**) CCPRS was significantly elevated in muscle-invasive bladder cancer (MIBC) samples, especially in more advanced stages ($p < 0.0001$).

3.5. Copy Number Alteration is Closely Correlated with Dysregulated Expression of the Gene Signature

We investigated the expression profiles of the gene signature in 19 paired BCa and adjacent normal tissues from TCGA. As shown in Figure 5A, paired *t*-test indicated most genes [11–13] involved in the gene signature were dysregulated in BCa samples compared to normal. The webtool cBioPortal for Cancer Genomics was used to generate Oncoprint to visualize the genomic alterations including mutation and copy number alteration of the gene signature in 413 TCGA BCa samples. Mutation was rarely detected but copy number alterations of the gene signature frequently occurred in BCa tissues (Figure 5B). We integrated copy number and RNA-seq data from TCGA, and an overview of relationships between copy number and mRNA expression is plotted in Figure 5C. Bubble size represents Pearson correlation coefficient, and color presents the significance. Two representative genes RCE1 and CEP72 with r > 0.5 are shown in detail: mRNA expression values of both genes were progressively and significantly elevated in copy number amplification groups ($p < 0.0001$ for both genes). In addition, copy numbers of RCE1 and CEP72 were frequently amplified in most BCa cell lines, and highly positive correlation between copy number and mRNA expression of RCE1 (r = 0.7196, $p = 0.0002$) and CEP72 (r = 0.6728, $p = 0.0006$) was observed in 22 BCa cell lines (Figure 5D).

3.6. Higher CCPRS Predicts Worse CSS in BCa

CCPRS for each sample was calculated and normalized to Z-scores in each cohort. Heatmaps of the associations between CCPRS Z-scores and different clinicopathological features in each cohort were plotted (Figure 6A,C,E,G). In general, an overview of these heatmaps indicated that CCPRS correlates with more advanced clinicopathological features such as muscle-invasive (MI) status and higher grade, while no significant correlation was observed between CCPRS and non-risk factors such as gender. Kaplan–Meier analysis demonstrated that BCa patients with higher CCPRS exhibited worse CSS in each cohort (Training cohort: HR = 10.20, 95% CI = 5.04–20.66, $p < 0.0001$; Validation I: HR = 2.991, 95% CI = 1.175–7.614, $p = 0.0008$; Validation II: HR = 8.468, 95% CI = 3.791–18.92, $p < 0.0001$; Validation III: HR = 6.345, 95% CI = 2.762–14.58, $p < 0.0001$; Figure 6B,D,F,H, respectively). Meta-analysis using fixed-effect model ($I^2 = 34.0\%$, $p = 0.209$) of the four cohorts showed that higher CCPRS was correlated with a significant worse CSS (pooled HR = 6.93, 95% CI = 4.63–10.37; Figure 6I). Kaplan–Meier analysis also showed that BCa patients with positive CCPRS Z-scores exhibited significant worse CSS compared to negative Z-scores (HR = 3.083, 95% CI = 2.144–4.433, $p < 0.0001$; Figure 6J) in the pooled cohort of 587 patients. In addition, CCPRS Z-scores were significantly elevated in those patients who deceased during follow-up, with progressively increasing Z-scores as survival time decreased ($p < 0.0001$; Figure 6K). Furthermore, as shown in Figure 6L, multivariate Cox regression analysis was performed on a total of 284 patients with full-scale information including gender, grade, age, CCPRS, lymph node metastasis (LNM), and MI status. Results indicated that CCPRS was an independent risk factor for CSS (HR = 2.038, 95% CI = 1.291–3.218, $p = 0.002$) along with MI and LNM.

3.7. The Gene Signature Serves as A Promising Marker for Therapeutic Resistance

NMF consensus clustering was performed to divide 1018 cell lines into two clusters based on the gene signature expression matrix when k = 2, and cluster 1 was identified as CCP-high cluster according to the distribution of expression pattern (Figure 7A). We observed that IC50 values of different routine chemotherapeutic drugs (including cisplatin, vinblastine, 5-fluorouracil, and gemcitabine) were significantly elevated in cluster 1 compared to cluster 2. In addition, cell lines in cluster 1 exhibited a significant increased resistance to cell cycle-targeting drugs including palbociclib (CDK4, 6) and AZD7762 (CHEK1, 2) (Figure 7B). Among BCa patients who received systemic chemotherapy in the training cohort, those with higher CCPRS exhibited worse CSS (HR = 3.415, 95% CI = 1.064–10.96, $p = 0.0208$; Figure 7C). Among TCGA MIBC patients who received adjuvant therapies including chemo- or/and radiotherapy, those with higher CCPRS exhibited worse overall survival (HR = 2.150, 95% CI = 1.082–4.270, $p = 0.0241$; Figure 7D). Then we investigated the association between CCPRS and Response

Evaluation Criteria in Solid Tumors (RECIST) among TCGA patients. After initial treatment, CCPRS was progressively and significantly elevated in groups with worse outcomes ($p = 0.0050$; CR: complete remission, PR: partial remission, SD: stable disease, PD: progressive disease; Figure 7E).

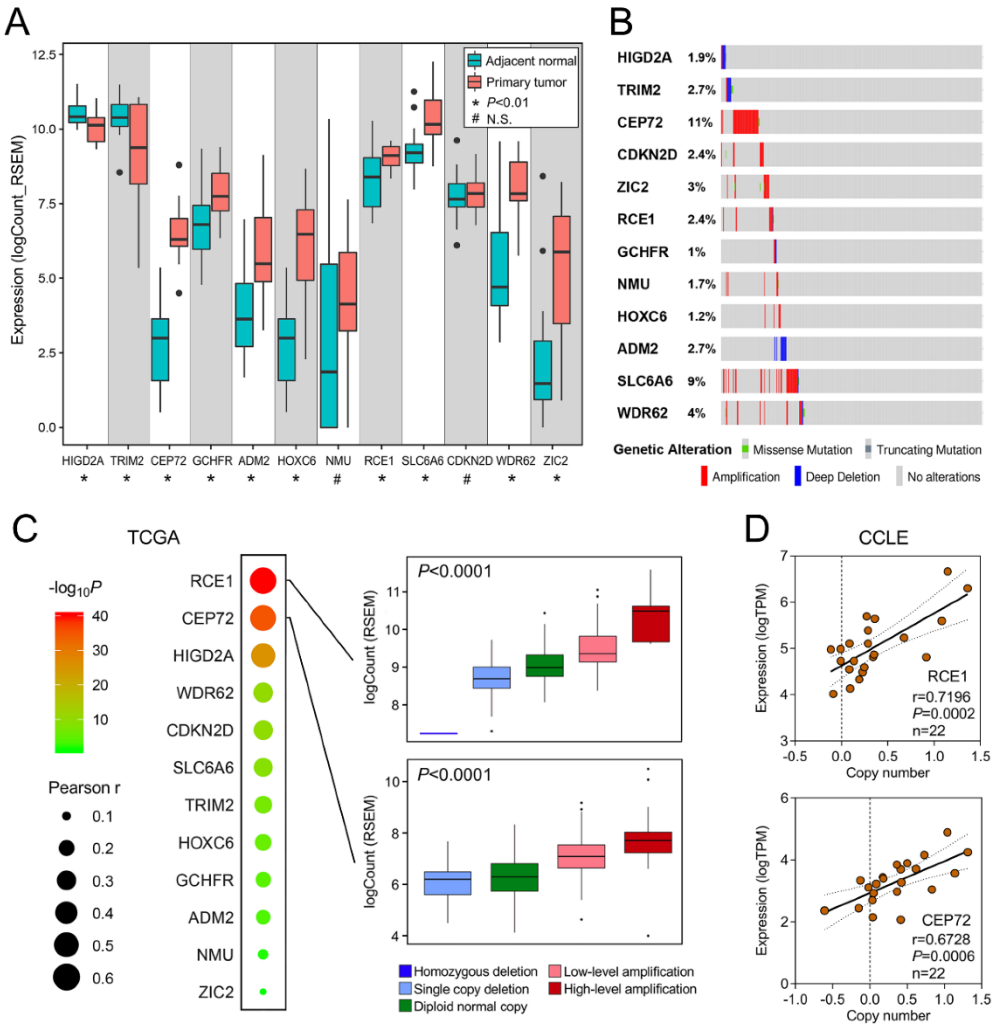


Figure 5. Copy number alteration is closely correlated with dysregulated expression of the gene signature. (**A**) Most genes (10/12) involved in the gene signature were dysregulated in 19 paired BCa and adjacent normal tissues from The Cancer Genome Atlas (TCGA). (**B**) An Oncoprint was plotted to visualize the genomic alterations including mutation and copy number alteration of the gene signature in 413 TCGA BCa samples. (**C**) Overview of relationships between copy number and mRNA expression is plotted. Bubble size represents Pearson correlation coefficient, and color presents the significance. Two representative genes RCE1 and CEP72 with r > 0.5 are shown in detail. (**D**) Copy numbers of RCE1 and CEP72 were frequently amplified in most BCa cell lines, and highly positive correlation between copy number and mRNA expression of RCE1 (r = 0.7196, $p = 0.0002$) and CEP72 (r = 0.6728, $p = 0.0006$) was observed in 22 BCa cell lines. * $p < 0.01$; N.S., no significance.

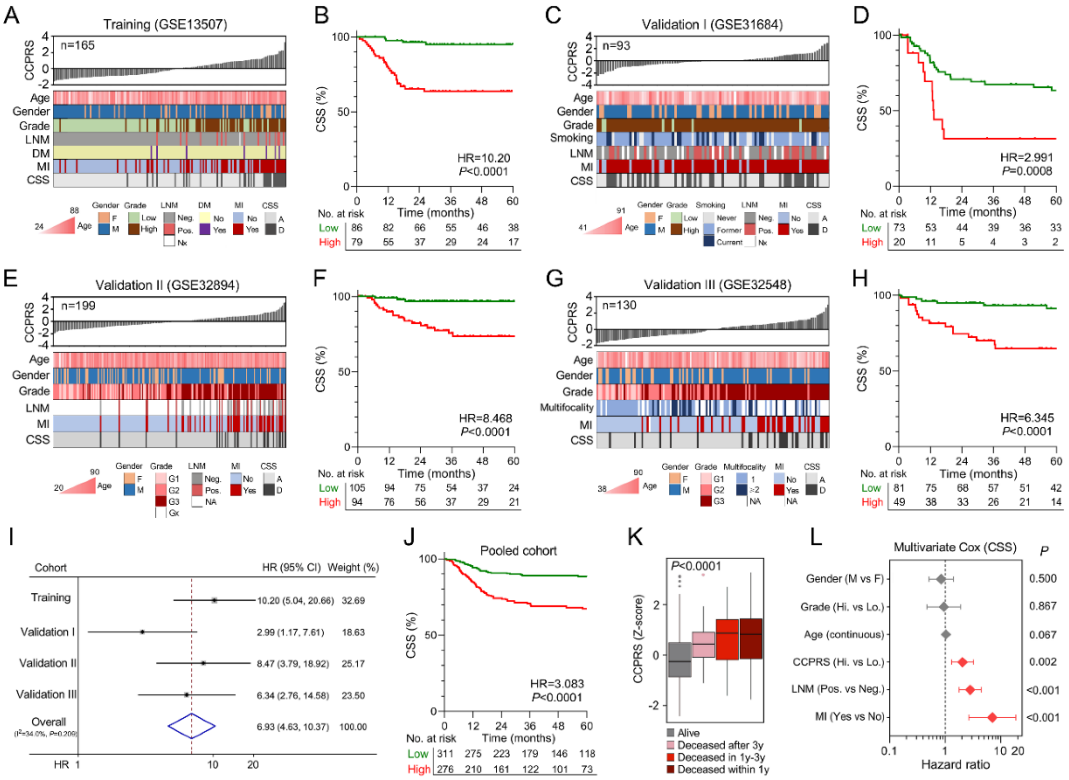


Figure 6. CCPRS correlates with more advanced clinicopathological features and predicts worse CSS. (**A**,**C**,**E**,**G**) Overview of heatmaps depicted associations between CCPRS and different clinicopathological features in each cohort. (**B**,**D**,**F**,**H**) Patients with higher CCPRS exhibited worse CSS in each cohort. (**I**) Meta-analysis on four cohorts included. (**J**) Kaplan–Meier analysis showed that BCa patients with higher CCPRS Z-scores exhibited significantly worse CSS compared to lower scores in the pooled cohort of 587 patients. (**K**) CCPRS Z-scores were significantly elevated in those patients who died during follow-up, with progressively increasing Z-scores as survival time decreased ($p < 0.0001$). (**L**) Multivariate Cox regression analysis on a total of 284 patients with full-scale information including CCPRS, gender, grade, age, lymph node metastasis (LNM) and muscle-invasive (MI) status, and results indicated CCPRS was an independent risk factor for CSS along with MI and LNM.

3.8. Combination of CCPRS and Clinical Variables Improves Risk Assessment and Stratification

To quantify risk assessment and predict CSS probability for individual BCa patients, a nomogram was constructed in combination of CCPRS and clinicopathological features (Figure 8A). The predictive power of nomogram was evaluated using time-dependent ROC analysis, with AUC of 0.944 for 1-year CSS and 0.932 for 3-year CSS, respectively (Figure 8B). In the calibration analysis, the prediction of 3-year CSS closely matched the observed situation, suggesting the nomogram has a high accuracy of CSS prediction (Figure 8C).

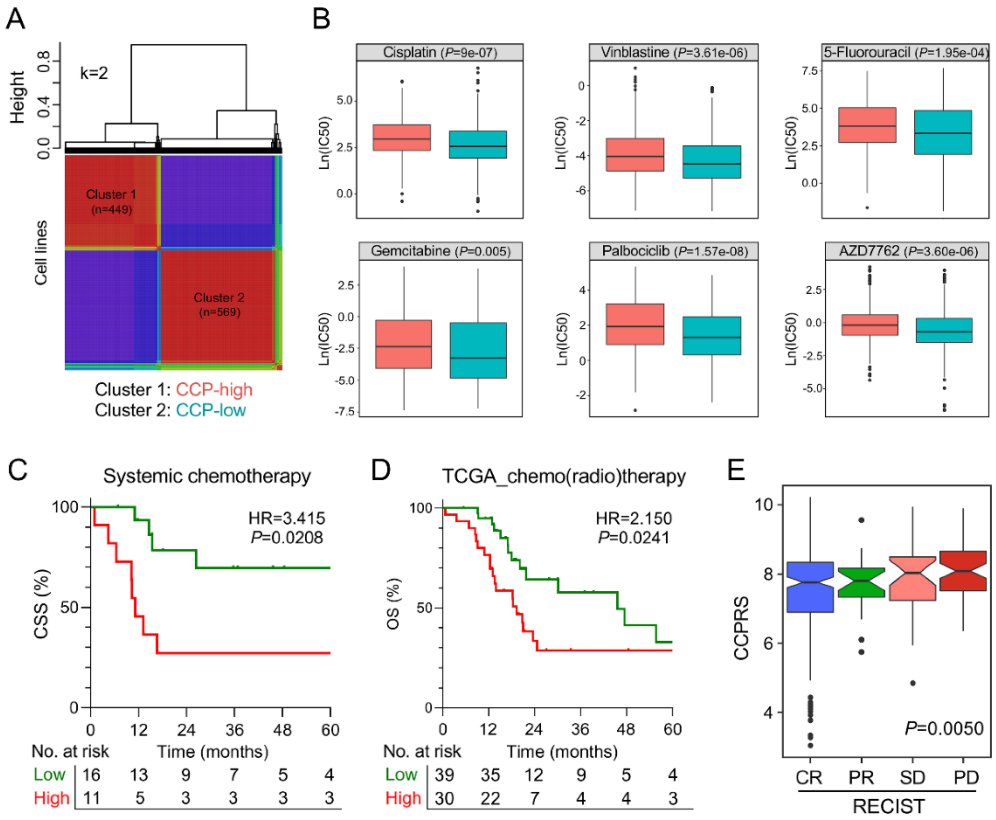


Figure 7. The gene signature serves as a promising marker for therapeutic resistance. (**A**) NMF consensus clustering was performed to divide 1018 cell lines into two clusters. (**B**) IC50 values of different routine chemotherapeutic drugs and cell cycle-targeting drugs were significantly elevated in the NMF-identified CCP-high cluster. (**C**) Among BCa patients who received systemic chemotherapy in the training cohort, those with higher CCPRS exhibited worse CSS. (**D**) Among TCGA MIBC patients who received adjuvant therapies including chemo- or/and radiotherapy, those with higher CCPRS exhibited worse overall survival. (**E**) After initial treatment, CCPRS was progressively and significantly elevated in groups with worse outcomes (p = 0.0050).

Considering MIBC accounts for a considerable part and contributes to the major mortality in BCa, we sought to optimize the risk stratification for MIBC patients by integrating CCPRS with traditional prognostic parameters. Various parameters including age, gender, LNM, grade, pT, and CCPRS were submitted for recursive partitioning analysis. Finally, CCPRS along with LNM and pT remained in the decision tree for CSS, and three different risk subgroups were defined (Figure 8D). Patients in the high-risk group exhibited worst CSS compared to other groups ($p < 0.0001$; Figure 8E). Furthermore, other MIBC patients with full-scale information of CCPRS, LNM, and pT were used to validate the classifying capacity of the decision tree. Significant differences of CSS from GEO (p = 0.0090; Figure 8F) and overall survival from TCGA (p = 0.0002; Figure S5) were observed among different risk subgroups. These results demonstrated the excellent classifying capacity of the decision tree.

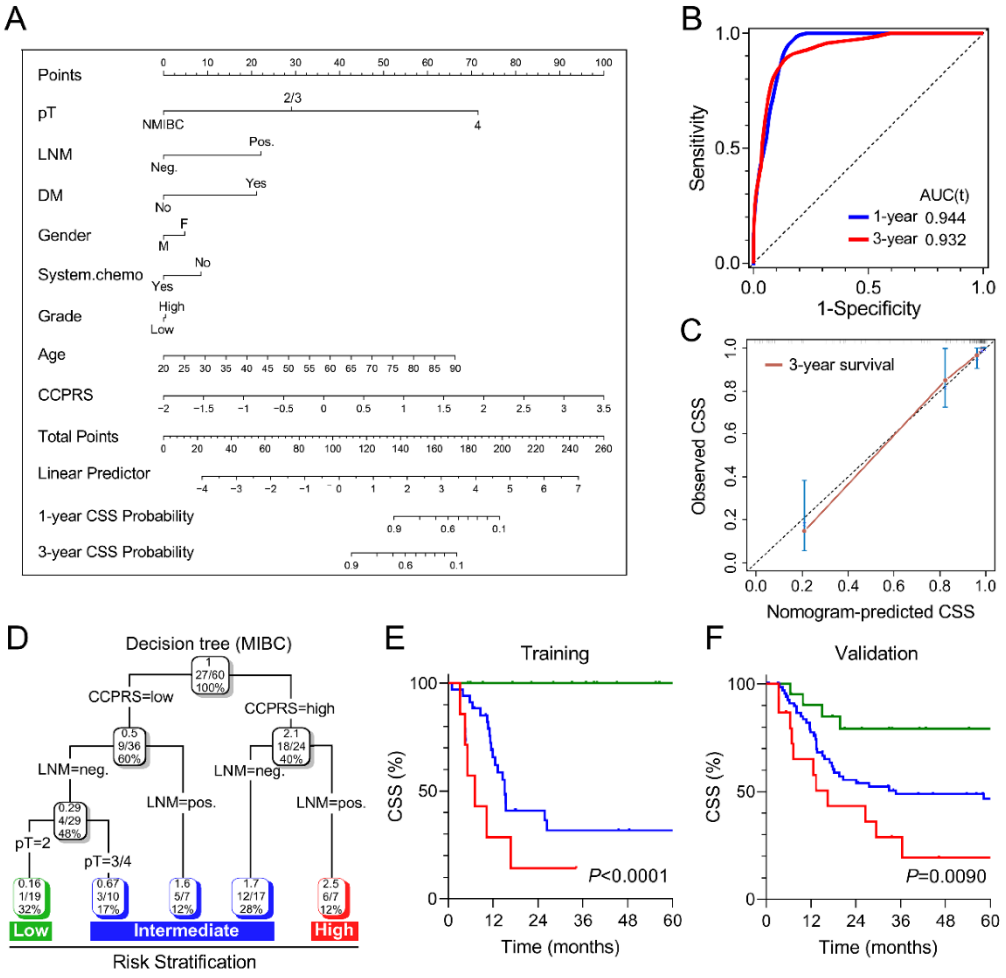


Figure 8. Combination of CCPRS and clinical variables to improve risk assessment and stratification. (**A**) A nomogram was constructed to quantify risk assessment and predict CSS probability for individual patients. (**B**) Time-dependent ROC analysis was performed to evaluate the predictive power of the nomogram for 1-year and 3-year CSS. (**C**) Calibration analysis indicated the nomogram has a high accuracy of CSS prediction. (**D**) An integrated decision tree was generated to optimize the risk stratification for MIBC patients. (**E**) In the training cohort, MIBC patients in the high-risk group exhibited worst CSS compared to other groups ($p < 0.0001$). (**F**) Significant difference of CSS was observed among different risk subgroups in the validation cohort.

4. Discussion

Alteration of cell cycle process (CCP) acts as a critical hallmark of cancer [21,22]. Dysregulation of CCP usually results from a series of changes in the activity of cell cycle regulators, which induces uncontrolled cell division and contributes to cancer development and progression. Abnormalities in expression and amplification of some important CCP-related genes or regulators such as cyclin D1, E1, and E2F1 were frequently observed in various cancer types. Furthermore, it was widely reported that these genomic changes usually promote malignant phenotypes and predict unfavorable outcomes [23–25]. In bladder cancer, cell cycle alteration was reported to have prognostic value and some genes involved in CCP have become attractive therapeutic targets [26].

Since Cuzick et al. reported an established cell cycle progression score in prostate cancer in 2011 [27], many studies have followed this approach and validated its prognostic value in different prostate cancer cohorts [28]. In brief, Cruzick's score is calculated upon the relative expression levels of

31 selected CCP genes normalized to 15 housekeeper genes using quantitative RT-PCR. However, some shortcomings are inevitable in this method. First, these 31 genes, selected from a set of documented CCP genes using Pearson's correlation coefficient, seem to hardly represent the hallmark of CCP comprehensively. Second, the results of quantitative RT-PCR are quite susceptible to several subjective factors such as different RNA extraction methods, inconsistent human operations, and heterogeneous samples quality in repeated experiments. In this study, ssGSEA algorithm was used to quantify CCP and other cancer-related hallmarks and pathways in BCa samples. ssGSEA calculates separate enrichment scores for each pairing of a sample and gene set, and ssGSEA score represents the degree to which the genes in a particular gene set are coordinately up- or downregulated within a sample. Obviously, this algorithm could overcome shortcomings mentioned above.

Among various cancer-related hallmarks and pathways, CCP was identified as the dominant risk factor for cancer-specific survival (CSS) in BCa. Subsequently, WGCNA was performed to identify CCP-related gene module based on transcriptome profiling data and CCP ssGSEA score. Consensus clustering was used to evaluate whether the identified "CCP module" could represent CCP. Univariate and LASSO Cox regression analyses were successively used to filter out most robust prognostic biomarkers to establish a CCP-related gene signature. The risk score derived from CCP-related gene signature is named as CCP-related risk score (CCPRS) in our study. The prognostic value of CCPRS was evaluated in the training and other independent validation cohorts across different platforms. Moreover, patients with higher CCPRS exhibited worse outcomes in the adjuvant therapy groups, suggesting CCPRS could serve as a useful marker for resistance to anti-cancer treatments. Regarding clinical application, a nomogram was constructed in combination of CCPRS and clinicopathological features to quantify risk assessment and predict CSS probability for individual patients. Considering the significant proportion and high mortality of MIBC in bladder cancer, we combined CCPRS with traditional prognostic parameters to build a decision tree to optimize risk stratification for CSS of MIBC patients.

Most biomarkers involved in the gene signature are dysregulated in BCa compared to normal tissues, and some of them have also been studied in other cancer types. For example, ADM2, one risk biomarker in the gene signature, was reported to predict poor survival in patients with pancreatic adenocarcinoma [29]. Moreover, Wang et al. reported that ADM2 could enlarge the vascular lumen by inducing the quiescent endothelial cell proliferation [30]. Deletion of RCE1 reduced the growth of fibroblasts and skin carcinoma cells [31], and overexpression of RCE1 was reported to correlate with prostate cancer progression and predict poor prognosis [32]. ZIC2, a zinc finger transcription factor, was required for the self-renewal maintenance of liver cancer stem cells, and its depletion reduced sphere formation and xenograft tumor growth in mice [33]. Recessive WDR62 mutations were identified in severe brain malformations [34], and the interaction between WDR62 and mitotic kinase AURKA is essential for drosophila brain growth [35]. What is more, overexpression of WDR62 is associated with poor prognosis in lung adenocarcinoma and gastric cancer [36,37]. In summary, evidence from current literature suggested that some biomarkers involved in the gene signature are closely related to cell proliferation or tumor growth, but the putative role in CCP still needs further investigation.

Some limitations in our study should be acknowledged. First, this is a retrospective study, so the prognostic robustness and clinical usefulness of the gene signature need further validation in prospectively designed clinical trials. Second, further experimental studies are needed to reveal the regulatory role of the gene signature in BCa progression.

5. Conclusions

In summary, a novel CCP-related gene signature was established to predict CSS in BCa patients. The prognostic value of CCPRS was further validated in independent cohorts. In combination of the gene signature and clinicopathological features, a nomogram was constructed to quantify risk assessment for individual patients, and a decision tree was built to optimize risk stratification for CSS of MIBC patients. We hope the novel CCP-related gene signature could be a useful tool to select high-risk

BCa patients who may benefit from adjuvant therapies and contribute to personalized management of BCa.

Supplementary Materials: The following are available online at http://www.mdpi.com/2072-6694/12/5/1146/s1, Figure S1. A detailed screening diagram to establish a robust gene signature for CSS; Figure S2. Sample clustering showed one outlier was detected and excluded in WGCNA; Figure S3. A power of $\beta = 9$ was chosen as the optimal soft threshold to ensure a scale-free co-expression network; Figure S4. Scatter diagram showed a highly positive correlation between GS and MM in the brown module ($r = 0.87$, $p < 1 \times 10{-}200$); Figure S5. Significant difference of overall survival in TCGA MIBC patients ($p = 0.0002$) was observed among different risk subgroups defined by the decision tree.

Author Contributions: Conceptualization, C.B. and M.L.; Formal analysis, R.S. and X.B.; Validation, P.R. and C.S. (Christian Schaefer); Resources, K.U. and S.L.; Writing—original draft preparation, R.S. and M.L.; Writing—review and editing, A.B. and C.S. (Christian Stief); Investigation, N.-S.S.-H., K.U., and J.S.; Supervision, C.B., and M.L. All authors have read and agreed to the published version of the manuscript.

Funding: We greatly thank the China Scholarship Council (CSC) for supporting the work of Run Shi (No. 201708320347), Xuanwen Bao (No. 201608210186), and Jing Sun (No. 201808320327).

Conflicts of Interest: The authors declare no competing interests.

Abbreviations

BCa	Bladder cancer
CCP	Cell cycle process
CSS	Cancer-specific survival
MI	Muscle invasive
TNM	Tumor-Node-Metastasis
GEO	Gene Expression Omnibus
TCGA	The Cancer Genome Atlas
CCLE	Cancer Cell Line Encyclopedia
GDSC	Genomics of Drug Sensitivity in Cancer
ssGSEA	single-sample gene set enrichment analysis
LASSO	least absolute shrinkage and selection operator
WGCNA	Weighted correlation network analysis
GS	Gene significance
MM	Module membership
CCPRS	Cell cycle process-related risk score
LNM	Lymph node metastasis
RECIST	Response Evaluation Criteria in Solid Tumors
CR	Complete remission
PR	Partial remission
SD	Stable disease
PD	Progressive disease
NMF	Non-negative matrix factorization
tROC	Time-dependent receiver operating characteristic

References

1. Kamat, A.M.; Hahn, N.M.; Efstathiou, J.A.; Lerner, S.P.; Malmstrom, P.U.; Choi, W.; Guo, C.C.; Lotan, Y.; Kassouf, W. Bladder cancer. *Lancet* **2016**, *388*, 2796–2810. [CrossRef]
2. Mak, R.H.; Hunt, D.; Shipley, W.U.; Efstathiou, J.A.; Tester, W.J.; Hagan, M.P.; Kaufman, D.S.; Heney, N.M.; Zietman, A.L. Long-term outcomes in patients with muscle-invasive bladder cancer after selective bladder-preserving combined-modality therapy: A pooled analysis of Radiation Therapy Oncology Group protocols 8802, 8903, 9506, 9706, 9906, and 0233. *J. Clin. Oncol.* **2014**, *32*, 3801–3809. [CrossRef]
3. Fonteyne, V.; Ost, P.; Bellmunt, J.; Droz, J.P.; Mongiat-Artus, P.; Inman, B.; Paillaud, E.; Saad, F.; Ploussard, G. Curative Treatment for Muscle Invasive Bladder Cancer in Elderly Patients: A Systematic Review. *Eur. Urol.* **2018**, *73*, 40–50. [CrossRef]

4. Kogenaru, S.; Qing, Y.; Guo, Y.; Wang, N. RNA-seq and microarray complement each other in transcriptome profiling. *BMC Genom.* **2012**, *13*, 629. [CrossRef]
5. Byron, S.A.; Van Keuren-Jensen, K.R.; Engelthaler, D.M.; Carpten, J.D.; Craig, D.W. Translating RNA sequencing into clinical diagnostics: Opportunities and challenges. *Nat. Rev. Genet.* **2016**, *17*, 257–271. [CrossRef]
6. Lee, J.S.; Leem, S.H.; Lee, S.Y.; Kim, S.C.; Park, E.S.; Kim, S.B.; Kim, S.K.; Kim, Y.J.; Kim, W.J.; Chu, I.S. Expression signature of E2F1 and its associated genes predict superficial to invasive progression of bladder tumors. *J. Clin. Oncol.* **2010**, *28*, 2660–2667. [CrossRef]
7. Van der Heijden, A.G.; Mengual, L.; Lozano, J.J.; Ingelmo-Torres, M.; Ribal, M.J.; Fernandez, P.L.; Oosterwijk, E.; Schalken, J.A.; Alcaraz, A.; Witjes, J.A. A five-gene expression signature to predict progression in T1G3 bladder cancer. *Eur. J. Cancer* **2016**, *64*, 127–136. [CrossRef] [PubMed]
8. Le Goux, C.; Vacher, S.; Pignot, G.; Sibony, M.; Barry Delongchamps, N.; Terris, B.; Piaggio, E.; Zerbib, M.; Damotte, D.; Bieche, I. mRNA Expression levels of genes involved in antitumor immunity: Identification of a 3-gene signature associated with prognosis of muscle-invasive bladder cancer. *Oncoimmunology* **2017**, *6*, e1358330. [CrossRef]
9. Ghandi, M.; Huang, F.W.; Jane-Valbuena, J.; Kryukov, G.V.; Lo, C.C.; McDonald, E.R., 3rd; Barretina, J.; Gelfand, E.T.; Bielski, C.M.; Li, H.; et al. Next-generation characterization of the Cancer Cell Line Encyclopedia. *Nature* **2019**, *569*, 503–508. [CrossRef] [PubMed]
10. Yang, W.; Soares, J.; Greninger, P.; Edelman, E.J.; Lightfoot, H.; Forbes, S.; Bindal, N.; Beare, D.; Smith, J.A.; Thompson, I.R.; et al. Genomics of Drug Sensitivity in Cancer (GDSC): A resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Res.* **2013**, *41*, D955–D961. [CrossRef] [PubMed]
11. Barbie, D.A.; Tamayo, P.; Boehm, J.S.; Kim, S.Y.; Moody, S.E.; Dunn, I.F.; Schinzel, A.C.; Sandy, P.; Meylan, E.; Scholl, C.; et al. Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature* **2009**, *462*, 108–112. [CrossRef] [PubMed]
12. Liberzon, A.; Birger, C.; Thorvaldsdottir, H.; Ghandi, M.; Mesirov, J.P.; Tamayo, P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst.* **2015**, *1*, 417–425. [CrossRef] [PubMed]
13. Langfelder, P.; Horvath, S. WGCNA: An R package for weighted correlation network analysis. *BMC Bioinform.* **2008**, *9*, 559. [CrossRef] [PubMed]
14. Tibshirani, R. The lasso method for variable selection in the Cox model. *Stat. Med.* **1997**, *16*, 385–395. [CrossRef]
15. Subramanian, A.; Tamayo, P.; Mootha, V.K.; Mukherjee, S.; Ebert, B.L.; Gillette, M.A.; Paulovich, A.; Pomeroy, S.L.; Golub, T.R.; Lander, E.S.; et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 15545–15550. [CrossRef]
16. Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N.S.; Wang, J.T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **2003**, *13*, 2498–2504. [CrossRef]
17. Cerami, E.; Gao, J.; Dogrusoz, U.; Gross, B.E.; Sumer, S.O.; Aksoy, B.A.; Jacobsen, A.; Byrne, C.J.; Heuer, M.L.; Larsson, E.; et al. The cBio cancer genomics portal: An open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* **2012**, *2*, 401–404. [CrossRef]
18. Wilkerson, M.D.; Hayes, D.N. ConsensusClusterPlus: A class discovery tool with confidence assessments and item tracking. *Bioinformatics* **2010**, *26*, 1572–1573. [CrossRef]
19. Lee, D.D.; Seung, H.S. Learning the parts of objects by non-negative matrix factorization. *Nature* **1999**, *401*, 788–791. [CrossRef]
20. Strobl, C.; Malley, J.; Tutz, G. An introduction to recursive partitioning: Rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychol. Methods* **2009**, *14*, 323–348. [CrossRef]
21. Hanahan, D.; Weinberg, R.A. Hallmarks of cancer: The next generation. *Cell* **2011**, *144*, 646–674. [CrossRef] [PubMed]
22. Williams, G.H.; Stoeber, K. The cell cycle and cancer. *J. Pathol.* **2012**, *226*, 352–364. [CrossRef] [PubMed]
23. Ooi, A.; Oyama, T.; Nakamura, R.; Tajiri, R.; Ikeda, H.; Fushida, S.; Dobashi, Y. Gene amplification of CCNE1, CCND1, and CDK6 in gastric cancers detected by multiplex ligation-dependent probe amplification and fluorescence in situ hybridization. *Hum. Pathol.* **2017**, *61*, 58–67. [CrossRef] [PubMed]

24. Lee, S.R.; Roh, Y.G.; Kim, S.K.; Lee, J.S.; Seol, S.Y.; Lee, H.H.; Kim, W.T.; Kim, W.J.; Heo, J.; Cha, H.J.; et al. Activation of EZH2 and SUZ12 Regulated by E2F1 Predicts the Disease Progression and Aggressive Characteristics of Bladder Cancer. *Clin. Cancer Res.* **2015**, *21*, 5391–5403. [CrossRef]
25. Watters, A.D.; Latif, Z.; Forsyth, A.; Dunn, I.; Underwood, M.A.; Grigor, K.M.; Bartlett, J.M. Genetic aberrations of c-myc and CCND1 in the development of invasive bladder cancer. *Br. J. Cancer* **2002**, *87*, 654–658. [CrossRef]
26. Mitra, A.P.; Hansel, D.E.; Cote, R.J. Prognostic value of cell-cycle regulation biomarkers in bladder cancer. *Semin. Oncol.* **2012**, *39*, 524–533. [CrossRef]
27. Cuzick, J.; Swanson, G.P.; Fisher, G.; Brothman, A.R.; Berney, D.M.; Reid, J.E.; Mesher, D.; Speights, V.O.; Stankiewicz, E.; Foster, C.S.; 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*, 245–255. [CrossRef]
28. Sommariva, S.; Tarricone, R.; Lazzeri, M.; Ricciardi, W.; Montorsi, F. Prognostic Value of the Cell Cycle Progression Score in Patients with Prostate Cancer: A Systematic Review and Meta-analysis. *Eur. Urol.* **2016**, *69*, 107–115. [CrossRef]
29. Hollander, L.L.; Guo, X.; Salem, R.R.; Cha, C.H. The novel tumor angiogenic factor, adrenomedullin-2, predicts survival in pancreatic adenocarcinoma. *J. Surg. Res.* **2015**, *197*, 219–224. [CrossRef]
30. Wang, L.J.; Xiao, F.; Kong, L.M.; Wang, D.N.; Li, H.Y.; Wei, Y.G.; Tan, C.; Zhao, H.; Zhang, T.; Cao, G.Q.; et al. Intermedin Enlarges the Vascular Lumen by Inducing the Quiescent Endothelial Cell Proliferation. *Arter. Thromb. Vasc. Biol.* **2018**, *38*, 398–413. [CrossRef]
31. Bergo, M.O.; Ambroziak, P.; Gregory, C.; George, A.; Otto, J.C.; Kim, E.; Nagase, H.; Casey, P.J.; Balmain, A.; Young, S.G. Absence of the CAAX endoprotease Rce1: Effects on cell growth and transformation. *Mol. Cell Biol.* **2002**, *22*, 171–181. [CrossRef] [PubMed]
32. Huang, L.; Li, M.; Wang, D.; He, J.; Wu, W.; Zeng, Q.; Li, J.; Xiao, M.; Hu, J.; He, Y.; et al. Overexpressed Rce1 is positively correlated with tumor progression and predicts poor prognosis in prostate cancer. *Hum. Pathol.* **2016**, *47*, 109–114. [CrossRef] [PubMed]
33. Zhu, P.; Wang, Y.; He, L.; Huang, G.; Du, Y.; Zhang, G.; Yan, X.; Xia, P.; Ye, B.; Wang, S.; et al. ZIC2-dependent OCT4 activation drives self-renewal of human liver cancer stem cells. *J. Clin. Investig.* **2015**, *125*, 3795–3808. [CrossRef]
34. Bilguvar, K.; Ozturk, A.K.; Louvi, A.; Kwan, K.Y.; Choi, M.; Tatli, B.; Yalnizoglu, D.; Tuysuz, B.; Caglayan, A.O.; Gokben, S.; et al. Whole-exome sequencing identifies recessive WDR62 mutations in severe brain malformations. *Nature* **2010**, *467*, 207–210. [CrossRef] [PubMed]
35. Lim, N.R.; Shohayeb, B.; Zaytseva, O.; Mitchell, N.; Millard, S.S.; Ng, D.C.H.; Quinn, L.M. Glial-Specific Functions of Microcephaly Protein WDR62 and Interaction with the Mitotic Kinase AURKA Are Essential for Drosophila Brain Growth. *Stem Cell Rep.* **2017**, *9*, 32–41. [CrossRef]
36. Zeng, S.; Tao, Y.; Huang, J.; Zhang, S.; Shen, L.; Yang, H.; Pei, H.; Zhong, M.; Zhang, G.; Liu, T.; et al. WD40 repeat-containing 62 overexpression as a novel indicator of poor prognosis for human gastric cancer. *Eur. J. Cancer* **2013**, *49*, 3752–3762. [CrossRef]
37. Shinmura, K.; Kato, H.; Kawanishi, Y.; Igarashi, H.; Inoue, Y.; Yoshimura, K.; Nakamura, S.; Fujita, H.; Funai, K.; Tanahashi, M.; et al. WDR62 overexpression is associated with a poor prognosis in patients with lung adenocarcinoma. *Mol. Carcinog.* **2017**, *56*, 1984–1991. [CrossRef]

Acknowledgements

I greatly thank Prof. Dr. Claus Belka and PD. Dr. Minglun Li for project design and guidance, and thank co-authors and colleagures for their useful comments, and thank the funding from China Scholarship Council (CSC).