Prostate cancer (PCa) is the most commonly diagnosed cancer in males in the Western world. Although prostate-specific antigen (PSA) has been widely used as a biomarker for PCa diagnosis, its results can be controversial. Therefore, new biomarkers are needed to enhance the clinical management of PCs. From publicly available microarray data, differentially expressed genes (DEGs) were identified by meta-analysis with RankProd. Genetic algorithm optimized artificial neural network (GA-ANN) was introduced to establish a diagnostic prediction model and to filter candidate genes. The diagnostic and prognostic capability of the prediction model and candidate genes were investigated in both GEO and TCGA datasets. Candidate genes were further validated by qPCR, Western Blot and Tissue microarray. By RankProd meta-analyses, 2306 significantly up- and 1311 down-regulated genes were found in 133 cases and 30 controls microarray data. The overall accuracy rate of the PCa diagnostic prediction model, consisting of a 15-gene signature, reached up to 100% in both the training and test dataset. The prediction model also showed good results for the diagnosis (AUC = 0.953) and prognosis (AUC of 5 years overall survival time = 0.808) of PCa in the TCGA database. The expression levels of three genes, FABP5, C1QTNF3 and LPHN3, were validated by qPCR. C1QTNF3 high expression was further validated in PCa tissue by Western Blot and Tissue microarray. In the GEO datasets, C1QTNF3 was a good predictor for the diagnosis of PCa (GSE6956: AUC = 0.791; GSE8218: AUC = 0.868; GSE26910: AUC = 0.972). In the TCGA database, C1QTNF3 was significantly associated with PCa patient recurrence free survival (P < .001, AUC = 0.57). In this study, we have developed a diagnostic and prognostic prediction model for PCa. C1QTNF3 was revealed as a promising biomarker for PCs. This approach can be applied to other high-throughput data from different platforms for the discovery of oncogenes or biomarkers in different kinds of diseases.

© 2018 Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
2. Materials and Methods

2.1. Data and Sample Collection

Public data was collected from the Gene Expression Omnibus (GEO) dataset and The Cancer Genome Atlas (TCGA) dataset. Only microarray data that met the following criteria were included; (1) Datasets were produced by Genome-wide mRNA expression profiling by microarray, (2) The experimental platform was single-channel; (3) All cases were pathologically diagnosed to be prostate cancer tissues while the controls were identified as para-carcinoma or normal prostate tissues; (4) The minimum number of cases and controls was 3. Finally, available datasets from the following cohort were included. Wallace et al. contained gene expression profiles of primary prostate tumors resected from 33 African-American and 36 European-American patients. It also contained 18 normal prostate tissues from 7 African-American and 11 European-American patients [27]. Wang et al. contained 148 prostate samples [28]. Planche et al. [29] contained 6 prostate cancer and matched normal samples. Taylor et al. contained 218 PCa samples and 149 matched normal samples from patients treated by radical prostatectomy [30]. Ross-Adams et al. contained 99 prostate cancer samples from patients with follow-up data [31]. For the TCGA dataset, we included the TCGA-PRAD which contained 500 PCa patients.

We also analyzed 69 primary prostate cancer patients and paired adjacent normal tissues by a tissue microarray obtained from Shanghai Outdo Biotech, China (Supplementary file 1: Table S1). Another 28 independent PCa and paired adjacent normal tissues were analyzed by qPCR and western blot (Supplementary file 2: Table S2). All fresh tissues were obtained with informed consent from patients hospitalized at the Department of Urology, Longgang Central Hospital and the Department of Urology, Third Affiliated Hospital of Sun Yat-Sen University. All tissue specimens were confirmed by pathology and immediately frozen in liquid nitrogen. All experiments in this study were approved by the ethics committee of Longgang Central Hospital and Third Affiliated Hospital of Sun Yat-Sen University.

2.2. Individual Participant Data Processing

In order to integrate microarray data from different platforms, meta-analysis was carried out by RankProd. The annotation files corresponding to the types of microarrays were downloaded from the official Affymetrix website. To pre-process Affymetrix microarray data, RMAExpress 1.0.5 was introduced for background adjustments, normalization was done by Quantile and summarization by Median Polish. The output files were composed of the normalized expression values of every probe. Shared probes were extracted from different platforms using Perl 5.10 and RankProd package installed in R (v3.4.0) was run. Probe signals with percentage of false prediction (pfp) value lower than 0.05 would be considered as DEGs. GO enrichment and KEGG analysis were carried out using clusterProfiler package in R (v3.4.0) [32].

2.3. Development of GA-ANN PCa Prediction Model

After acquiring the DEG list, we constructed the ANN model in MATLAB (MathWorks, Massachusetts, USA) by setting the clinical phenotype of 163 microarray samples as the output variable (normal or cancer patients) and the expression values of the top 500 up- and down-regulated probes as the input variables. A training set was built with 100 randomly selected microarray samples and the other 63 microarray samples were used as a test set. The model was composed of 3 layers with 1000 nodes as the input layer (each representing an expression value of a probe) and 1 node as the output layer (the clinical phenotype). We set the maximum recursive time to 100 and the threshold of mean square error to 0.005. The weight-corrected learning rate was 0.1 and the transfer function from input layer to hidden layer was tanh while purelin was configured as the transfer function from...
hidden layer to output layer. In terms of optimization by GA, the number of initial population was 100 and the maximum evolutionary generation was 50. During each round of calculation, GA-ANN randomly selected the useful input variables keeping the computational accuracy stable. Therefore, the number of input variables could reduce nearly half every round. After 6 rounds of calculations, 15 candidate input variables (probes) were obtained.

### 2.4. Diagnosis Assay for 15-Gene Signature in Independent Dataset

The prediction accuracy was calculated both in the training and test set. We employed the genes from the TCGA cohort to assay their relative risk and capacity of diagnosis by logistic regression. This test was performed on “glm” function in R software. Then, a linear model was constructed by combining the gene expressions. A coefficient of logistic regression and index by combination was assigned to each sample. Finally, the area under curve (AUC) of the receiver operating characteristic (ROC) curves was employed to estimate the performance of the model with the “ROCR” package in R.

### 2.5. Prognostic Index of 15-Gene Signature in Prognosis of Survival of PCa

A prognostic index (PI) [33] was constructed as an integrated indicator of the 15 candidate genes selected by the ANN model for each PCa patient. The PI was calculated as a linear combination of the expression value of the genes weighted by univariate Cox regression coefficients. The standard form of PI was defined as follow:

\[
\text{Prognostic index (PI)} = \sum_i \left( \beta_i \times X_i \right)
\]

\(\beta_i\) is the regression coefficient of the ith variable and \(X_i\) is the value of the ith variable. For the form of PL, \(X_i\) is the log2-transformed expression value of each mRNA and \(\beta_i\) is the univariate Cox regression coefficient of the ith RNA.

### 2.6. Investigation of Diagnosis and Prognosis Capacity of C1QTNF3 in PCa

The capacity of C1QTNF3 to diagnose PCa was evaluated by measuring the AUC of the ROC curves using the “ROCR” package in R. To integrate and combine the results from three C1QTNF3 probes, the “aggregate” function of R was applied. Differential expression of C1QTNF3 in tumor and normal tissue was computed by the “limma” package in R. Logistic regression was measured using the “glm” function in R (Version 3.4.0). C1QTNF3 was validated by analyzing available PCa samples in the TCGA database with the cbioPortal web tool (http://www.cbioportal.org/index.do) [34]. Survival analysis was calculated automatically by this tool.

### 2.7. Quantitative Real-Time PCR Analysis (qRT-PCR)

Total RNA was extracted from patients’ tissues samples with TRizol reagent (Invitrogen, USA) and treated with DNase I (Merck, Sigma, USA). A total of 2 μg of RNA was reverse transcribed into cDNA with oligo (dT) primers using the cDNA synthesis kit (Takara, Japan). Quantitative PCR was performed in 20 μl reactions using SYBR Green qPCR Master Mix (Takara, Japan) according to manufacturer’s instruction. β-actin mRNA levels were used for normalization. The following primers were used to amplify a 110-bp PCR product for C1QTNF3: forward, 5′- CGCAACACA GTCTTACGAT-3′; reverse, 5′- ATTCGACGCAAACCTTAC-3′; a 98-bp PCR product for FABPS: forward, 5′- AGATGGTCATGGTTTTCG-3′; reverse, 3′-TACGACAACTTACGACAT-3′; an 115-bp PCR product for FABPS: forward, 5′- CACCTTTCGATTTTGTC-3′; reverse, 3′- GGCTGCT TTCTATGCTGCT-3′ and a 120-bp PCR product for ACTB: forward, 5′- ACCTTCCAGCTTCCCT-3′; reverse, 5′- CGTACAGTCTTGGGCTG-3′. The PCR amplification program was as follow: initial denaturing at 95 °C for 10 min, and then denature at 95 °C for 10 min, followed by 40 cycles of 95 °C for 15 s and 60 °C for 45 s. The mRNA level of C1QTNF3, FABPS, and LPHN3 was measured using the Applied Biosystems 7500 Real-Time PCR System (ABI, USA). Measurements were repeated 3 times and relative quantification analysis was performed using the comparative CT (2^–ΔΔCT) method.

### 2.8. Western Blot

Cancer or paired normal tissue (0.2 g) was crushed in liquid nitrogen and lysed in RIPA lysis buffer (Cellytic, Sigma-Aldrich) in the presence of a protease inhibitor cocktail (Merck Millipore, USA). Total protein extracts were separated by SDS-PAGE and transferred to PVDF membranes (Merck Millipore, USA). Immunoblotting was done with rabbit polyclonal antibody against C1QTNF3 (ab368780, Abcam, 1:1500 dilution) in accordance with the manufacturer’s instruction. Signals were visualized using enhanced chemiluminescent substrate (ECL, BioRad, Richmond, CA, USA) and the Western Breeze chromogenic detection system (Invitrogen).

### 2.9. Tissue Microarray and Immunohistochemistry Staining

The human PCa tissue microarray (HPro-Ade180PG-02; Shanghai Outdo Biotech, China) was constructed with formalin-fixed, paraffin-embedded PCa tissues and paired adjacent normal tissues. Immunohistochemistry staining was performed by Shanghai Outdo Biotech Co., Ltd. Tissue microarray sections were blocked with goat serum, incubated with anti-C1QTNF3 (ab368780, Abcam, 1:200 dilution), deparaffinized, rehydrated, and subjected to heat-induced antigen retrieval, as previously described [35]. The expression of C1QTNF3 in each tissue was semi quantitatively graded by two independent pathologists according to staining intensity (0, negative, 1, weakly positive; 2, moderately positive; or 3, strongly positive).

## 3. Results

### 3.1. Raw Microarray Data and Overall Processing Methodology Description

The microarray assays enrolled in this study included GSE6956 [27], GSE8218 [28], GSE26910 [29], GSE21032 [30] and GSE70769 [31]. GSE6956 is composed of 69 primary prostate tumors and 18 non-tumor prostate tissues (Platform: GPL571 Affymetrix Human Genome U133A 2.0 Array). GSE8218 contained 148 prostate samples with various amount of different cell types of which 10 were normal (Platform: Affymetrix Human Genome U133A Array). By filtering out unidentified cellular component samples from GSE8218, we obtained 133 cases and 30 controls in total. GSE26910 contained 6 samples of stroma surrounding invasive prostate primary tumors and 6 matched samples of normal stroma samples (Platform: GPL570 Affymetrix Human Genome U133 plus 2.0 Array) and was used for validation. A detailed description of GEO datasets is available in Supplementary file 3: Table S3. Gene expression profile from the TCGA was used for verification. After screening the clinical data (excluding NA in survival time), 466 patients’ samples were selected. Furthermore, two independent datasets were used to validate prognostic capability of the 15-gene signature. From the Taylor dataset, gene expression and follow-up data of 140 patients with primary prostate cancer were collected from GSE21032. The Ross–Adams dataset was collected from GSE70769. After matching gene expression and clinical data samples, 92 primary prostate cancer patients were included for analysis. The flowchart in Fig. 1 shows the data analysis process.

### 3.2. DEG Identification by RankProd

RankProd was performed to detect DEGs. When we restricted the conditions to pfp=0, the number of up-regulated probes dropped to 2306 and down-regulated to 1311. The top 500 up- and down-
regulated probes are shown in a heatmap plot (Fig. 2) (Supplementary file 4: Table S4). The whole DEG lists are available in additional file (Supplementary file 5: Table S5).

3.3. Gene Ontology Enrichment and Pathway Enrichment

Differential expressed genes were annotated using the ClusterProfiler package. GO and KEGG analysis indicated that up-regulated genes enriched in pathways were obviously different from down-regulated genes (Supplementary file 6: Table S6). For example, KEGG analysis showed that the up-regulated genes related to cancer pathways included proteins that were involved in protein processing in the endoplasmic reticulum and lysosome, while downregulated genes were mainly involved in cancer pathways including focal adhesion, and complement and coagulation cascades (Fig. 3).

3.4. GA-ANN Screening for Candidate PCa Biomarker Genes

After we obtained the DEGs from RankProd, we adjusted the number of nodes in the hidden layer to improve the prediction accuracy of the ANN model. As this can also lead to dramatic complexity of the neural network and an increase in modeling duration, we fixed the optimal

Fig. 2. Heatmap plot of top 1000 differentially expressed genes (DEG) from Rankprod. The blue shade represented normal tissue and red shade represented patients tumor tissue.
number of nodes in the hidden layer at 5 (Fig. 4a, b). Furthermore, we noticed a significant advantage of genetic algorithm optimized ANN over general ANN in the performance of prediction and modeling duration. The prediction accuracy of both the training and test set reached 100% with high modeling speed (1.326 s). The process of training and testing are listed in Table 1.

Finally, we obtained 15 genes (Table 2) as a minimum candidate gene list to let the ANN model predict whether a prostate sample was normal or tumor tissue (Fig. 4c, d, e).

3.5. Diagnosis and Prognosis Capacity of Candidate Genes (15-gene signature) for PCa in TCGA Datasets

The expression distribution of the 15 genes in the TCGA PCa cohort is shown by boxplot (Fig. 5a). The P values, hazard ratios (HR) and coefficients of the 15 genes in overall survival prediction model for TCGA cohort are listed (Table 2). These genes have a high AUC value (0.953), which represents the high diagnosis capacity in this model (Fig. 5b). The patients in the TCGA PCa cohort were ranked according to the PI. Using the median value of PI as the cutoff, 466 patients were divided into 2 groups: a high-risk group with 233 patients and low-risk group with 233 patients (Supplementary file 7: Fig. S1). The PI was significantly associated with PCa patient 5 years overall survival (OS) (Fig. 5c) and recurrence-free survival (RFS) (Fig. 5d). The survival rates of the high-risk group in OS and RFS were both significantly lower than that of the low-risk group (log-rank P value = .003). We used the 5 years OS and RFS survival rate to compare the prognostic capacity of the 15-gene signature model, PSA screening and the Gleason score. The Gleason score is the most popular pathology grade for PCa and is a measure of how likely the tumor will grow invasively. The results suggested that the 15-gene signature was the best index for predicting PCa in 5 years OS with an
AUC = 0.808. The AUC of PSA screening and the Gleason score were 0.631 and 0.692 respectively (Fig. 5e). As for the 5 years RFS, AUCs of the three indicators demonstrated that the 15-gene signature model (AUC = 0.614), PSA screening (AUC = 0.702) and the Gleason score (AUC = 0.740) have similar capacity, and the Gleason score performed best (Fig. 5f).

To further validate the performance of the 15-gene signature, two independent datasets of Taylor et al. and Ross-Adams et al. were employed. The results showed that the 15-gene signature performed well in both independent datasets. PI calculated from the 15-gene signature can significantly classify patients into low- and high-risk (Taylor cohort: HR = 2.893, p value = .003, AUC = 0.74; Ross-Adams cohort: HR = 1.886, p value = .03, AUC = 0.67). The Kaplan-Meier and ROC curve for the two datasets are shown in Supplementary file 8 (Fig. S2).

3.6. Validation of Candidate Genes in Prostate Cancer Tissues

From the 15 genes, we selected 3 genes (FABP5, C1QTNF3 and LPHN3) for further analysis (Table 2). To analyze mRNA levels of FABP5, C1QTNF3 and LPHN3, qRT-PCR was performed on tissues from 28 prostate cancer patients. The mRNA level of FABP5 and C1QTNF3 was in all prostate cancer tissues higher than in the paired adjacent
Three datasets (GSE6956: OR = 1.253, 95% CI: 0.872–1.202, AUC = 0.868; GSE26910: OR = 5.332, 95% CI: 2.062–14.183, AUC = 0.972, Fig. 8b). Additionally, we also tested the prognostic ability of C1QTNF3 when compared with the paired adjacent normal tissues (Fig. 7c). Overall, these data suggest that C1QTNF3 showed a higher C1QTNF3 protein expression than in the paired adjacent normal tissues (Fig. 7c). Overall, these data suggest that C1QTNF3 is constantly overexpressed in PCa.

3.7. Diagnostic and Prognostic Capacity of C1QTNF3 for PCa Prediction in Various Datasets

C1QTNF3 expression was analyzed in GEO datasets GSE6956, GSE8218 and GSE26910 (Fig. 8a). The AUC of the ROC curve showed that C1QTNF3 showed good performance on diagnosis for PCa in all three datasets (GSE6956: OR = 1.253, 95% CI: 0.872–1.636, P = .001, AUC = 0.791; GSE8218: OR = 2.848, 95% CI: 1.365–4.311, P = .055, AUC = 0.868; GSE26910: OR = 5.322, 95% CI: 2.062–8.602, P = .015, AUC = 0.972, Fig. 8b). Additionally, we also tested the prognostic ability of C1QTNF3 in the TCGA dataset. The results showed that C1QTNF3 overexpression is closely associated with recurrence-free survival time (P < .001, AUC = 0.57) (Fig. 8c, d).

4. Discussion

Prediction and diagnosis is the most important step in PCa management for patients. In order to screen candidate biomarkers which may be helpful for diagnoses and prognosis for PCa, we have combined RankProd with GA-ANN to create a prediction model. This process could also provide a general framework for rational cancer gene signature discovery based on high throughput data. To datamine oncogenes, biomarkers or gene signature prediction models for prostate cancer, high throughput data from microarray or next generation sequencing is a fundamental source. Data processing approach plays a crucial part in such studies. Since the ANN model can fit any nonlinear function it has more advantages in processing high-throughput data. At the moment, depth neural networks have been applied to a variety of artificial intelligence applications. In the future, neural networks are bound to be used more in molecular medicine.

In this study, a 15-gene signature was identified by our data processing system that exhibited a great capacity for diagnosis and prognosis of PCa. The AUCs of the 15 genes signature showed a perfect diagnostic ability in PCa gene expression samples from datasets from both GEO and TCGA. Although the genes individual were not significant in the 5-year OS prognostic test, the 15-gene signature can effectively classify PCa patients into high- and low-risk groups, and showed a good prediction of the 5-year survival rate in the PCa cohort from TCGA, Taylor et al. and Ross-Adams et al. Other studies also tried to establish prediction models for diagnosis or prognosis of PCa. Cima et al. combined bioinformatic prioritization with targeted proteomics and machine learning to build predictive regression models for tissue PTEN status and diagnosis and grading of PCa [8]. Wu et al. constructed a 32-gene signature model which could predict PSA recurrence of post-radical prostatectomy patients via PCA coupling with Cox regression [36]. In 2014, Bismar et al. used a singular value decomposition (SVD) method to identify an ETS transcription factor (EGR) relative 10 genes signature for establishing a prognostic prediction model to predict patients’ clinical outcome [37]. Our work here provides a different approach to establish prediction model and select the candidate oncogenes or biomarkers.

From the 15-gene signature model, we selected 3 candidate genes (FABP5, LPHN3 and C1QTNF3) for further analysis. FABP5 has been demonstrated as a target of PCa in previous studies [38,39]. Latrophilin 3 (LPHN3) is a brain-specific member of the G-protein coupled receptor family associated to both attention-deficit/hyperactivity disorder (ADHD) genetic susceptibility and methylphenidate (MPH) pharmacogenetics [40] and was down regulated in PCa. Interestingly, we found C1q and tumor necrosis factor related protein 3 (C1QTNF3, alias as CTRP3), a highly expressed gene in PCa tissue, in the 15-gene list. In addition, the three probes (209424_s_at, 209426_s_at, 209425_s_at)
standing for C1QTNF3 respectively ranked the first, third and fourth place among the significantly up-regulated probes. These hints have inspired us to explore the role of C1QTNF3 as a susceptibility gene in PCa. Our previous studies have demonstrated that C1QTNF3 stimulated proliferation and anti-apoptosis in prostate cells through the protein kinase C signaling pathway [41]. Furthermore, C1QTNF3 regulated 14-3-3 sigma and GLRX3 which has functions in various kinds of tumors as well as in prostate cancer [41]. It suggests that C1QTNF3 may promote the transformation from prostate cells to malignant cells.

To confirm our findings, we validated C1QTNF3 by qPCR, Western Blot and tissue microarray. We were able to draw the conclusion that C1QTNF3 was significantly overexpressed in prostate tumor tissues. Besides, in the GEO dataset, the AUCs of ROC demonstrated C1QTNF3 had good performance for PCa diagnosis. The TCGA dataset showed that C1QTNF3 expression is closely associated with DFS time of prostate cancer patients. These results together indicate that C1QTNF3, as a biomarker for diagnosis and prognosis of PCa, has a high reliability and accuracy.

**Fig. 5.** Diagnostic and prognostic capacity of the 15-gene signature for PCa in TCGA dataset. (a) 15-gene expression value distribution in TCGA PCa cohort by boxplots. The line within the box indicates the median value; the box spans the interquartile range. (b) ROC curve for the 15-gene signature for PCa diagnosis (c) Kaplan-Meier curves for the low- and high-risk groups separated by the PI of the 15-gene signature in the TCGA PCa cohort. Significant differences in overall survival between the 2 groups were analyzed by log-rank test (P = .003). (d) Kaplan-Meier curves for the low-risk and high-risk groups of the 15-gene signature in the TCGA PCa cohort. Significant differences in DFS between the two groups were determined by the log-rank test (P = .003). (e) ROC curves for the prediction of the 5 years overall survival among the 15-gene signature model, PSA screening and the Gleason score. (f) ROC curves for the 5 years DFS among 15-gene signature, PSA screening and the Gleason score.
The data processing approach presented here provides a new view for the discovery of biomarkers with the aim of promoting diagnostic and prognostic prediction of PCa. This approach can also be applied to other high-throughput data for the discovery of oncogenes or biomarkers in different kinds of diseases and different platforms. In our study, we have established a diagnostic and prognostic prediction model and revealed C1QTNF3 as a promising biomarker for prostate cancer. However, more studies are warranted to determine the roles of the 15-gene signature prediction model and C1QTNF3 for PCa.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ebiom.2018.05.010.

Fig. 6. qPCR assay of FABP5, C1QTNF3 and LPHN3 genes in PCa and normal adjacent tissues. Scatter diagram of the gene expression and fold changed distribution of gene expression in different samples.

Fig. 7. Validation of C1QTNF3 expression in tissues microarray and Western blot. (a) Pathological sections of PCA and para-carcinoma tissue. (b) C1QTNF3 expression in tumor tissue was significantly increased when compared with para-carcinoma tissues. (c) C1QTNF3 expression assayed by Western blot.
Competing Interests

All authors declare that they have no competing interests.

Acknowledgement

This study was supported by grants from the National Post-doctoral Science Fund (2016M602451), the Natural Science Fund of Guangdong Province (2015A030310384), the Science and Technology Fund of Guangdong Province (2016A020215026), the Medical research Fund of Guangdong Province (C2015025), the Science and Technology Fund of Shenzhen City (2016A020215026), the Medical Science and Technology Program of Shenzhen Longgang District (20160606163318972). The funders had no involvement in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Authors' Contributions

QH conceptualized the project and contributed to experimental design, all data analysis and wrote the first draft of the manuscript. ZB contributed to processing, analysis, and interpretation of the data. CH, ML, ZM, XX and YL contributed to sample acquisition and experiment. KY and JL contributed to experimental design. SH and ML contributed to guide the experimental design, data analysis, and manuscript writing. All authors read and approved the final manuscript.

References

[1] Siegel RL, Miller KD, Abu-Rustum NR. Cancer statistics, 2018. CA Cancer J Clin 2018;68.
[2] Zhang W, Xiang YB, Liu ZW, Fang RR, Ruan ZX, Sun L, et al. Trends analysis of common urologic neoplasm incidence of elderly people in Shanghai, 1973–1999. Chinese J Cancer 2004;23:555.
[3] Chen W, Zheng R, Baade PD, Zhang S, Zeng H, Bray F, et al. Cancer statistics in China, 2015. CA Cancer J Clin 2016;66:115.
[4] Miller KD, Siegel RL, Lin CC, Mariotto AB, Kramer JL, Rowland JH, et al. Cancer treatment and survivorship statistics, 2016. CA Cancer J Clin 2016;66:271.
[5] Andriole GL, Grubb III RL, Baele PD, Zhang S, Zeng H, Bray F, et al. Cancer statistics in China, 2015. CA Cancer J Clin 2016;66:115.
[6] Miller KD, Siegel RL, Lin CC, Mariotto AB, Kramer JL, Rowland JH, et al. Cancer treatment and survivorship statistics, 2016. CA Cancer J Clin 2016;66:271.
[7] Andriole GL, Grubb III RL, Baele PD, Zhang S, Zeng H, Bray F, et al. Cancer statistics in China, 2015. CA Cancer J Clin 2016;66:115.
[8] Andriole GL, Grubb III RL, Baele PD, Zhang S, Zeng H, Bray F, et al. Cancer statistics in China, 2015. CA Cancer J Clin 2016;66:115.

Q. Hou et al. / EBioMedicine 32 (2018) 234–244
