Zeyuan Dong

TRGN -Capstone Project (Enrique I. Velazquez-Villarreal)

November 25, 2021

Title：Differences in gene expression before and after treatment in African American prostate cancer patients

1. Introduction：

Prostate cancer is one of the most commonly diagnosed cancers in men in the United States, and it is a leading cause of cancer-related deaths [1]. Now, many studies have shown that race plays a vital role in prostate cancer. African Americans have higher rates of prostate cancer morbidity and mortality compared to other racial groups.

Prostate cancer grade is an effective way to determine the risk of prostate cancer. Pathologist Dr. Donald Gleason uses the Gleason Score to grade prostate cancer. In 2014, the International Society of Uropathology revised and supplemented Gleason Score, using the classification group as the standard(from Prostate Cancer Foundation www.pcf.org).

With the development of next-generation sequencing technology, RNA-Seq has become an essential tool for transcriptome analysis and quantification. RNA-seq primarily helps researchers identify differences in gene expression [2]. The RNA-seq approach could help

researchers gain insight into the development of prostate cancer and identify potential therapeutic targets [2]. In this study, data were analyzed mainly through Bioinductor [3].

Biomarkers can guide clinical diagnosis and treatment decisions [4], so this study focused on investigating the differences in gene expression in African American prostate cancer patients before and after treatment in order to look for potential biomarkers, which will play an important role in evaluating the outcome and effectiveness of prostate cancer before and after treatment in African Americans in the future and provide more possibilities for precision medicine.

2. Method：

2.1 Data information

Data were obtained through HPC (CARE2 Bioinformatics, Statistical & Methodological Shared Resources BSMSR CORE). All patients are African Americans. The processed counts file was imported into R for gene differential expression analysis.

```
##        Type Gleason Treatment
## 34C Control       9       Pre
## 35C Control       8       Pre
## 36C Control       9      Post
## 34T   Tumor       9       Pre
## 35T   Tumor       8       Pre
## 36T   Tumor       9      Post
```

Table1.Sample information(the result of coldata inR)

There are six groups of data in the raw file, including the ID of the patient and Ensembl. They also had information before and after treatment and Gleason score. Three of the six data sets are C, and the other three are T. In ID, "C" is for Control, and" T" is for Tumor.

2.2.1 Preparing Quantification input to DESeq2

The DESeq2 package from Bioconductor was used in this project.

I first set up the working path and download the required packages, including reading the raw data into R and then processing the original Counts table. I removed the decimal point after the "ENS" number on the left side of the count table, changed it to an integer, and then assigned each "ENS" number to a human gene. I'm going to get this new table, but some genes are duplicated or not expressed, so I'm going to delete the duplicated genes and the NA from the table. To make it easier to perform the next differential expression analysis, I changed each column in the count table to "34C", "35C", "36C", "34T", "35T", "36T". In coldata, I added pre - and post-treatment information and Gleason score for the sample.

I create a DeseqDataSet by using DESeqDataSetFromMatrix function, and using the count matrix counts_filtered and the sample information coldata. Before running the DESeq2 function, I pre-filter the low count genes. On the one hand, pre-filtering can help the efficiency of DESeq2; On the other hand, rows that are rarely read can be deleted. In this project, I performed minimal pre-filtering, and setting the factor level.

2.2.2 Differential expression analysis

I will show my workflow for each step in detail in R. In this section, I perform the standard differential expression analysis steps. I build the results table, which is generated using the function results. The Result table clearly shows basemean, log2 multiples, p-values, and adjusted P-values.

2.2.3 Exploring results and visualization

The visualized results help me understand the data information and data structure more clearly. In many studies, heatmaps are usually an intuitive way to explore enumerate matrices. In DESeq2, two transformation methods are proposed. The first is the variance stable transformation (VST), and the other is the regular log transformation (Rlog). Because the running time of VST is shorter than that of Rlog, I chose VST (FPKM) for all the colors in the heat map.

2.2.4 Obtain top200 genes and find shared genes

Top200 genes of 34CT&36CT(contrasting pre- and post-treatment) and 35CT&36CT(contrasting pre- and post-treatment) by adjusted p-value were obtained respectively. Then, a list of 48 shared genes from those 2 Top200 genes was generated, and subsequently analyzed by Ingenuity Pathway Analysis (IPA). After testing, only 19 candidate genes were shared in the TOP100, and 48 candidate genes were shared in the TOP200. The maximum p-value of shared candidate genes in TOP200 was less than 10 to the power of minus 5, which was meaningful. Therefore, genes shared in TOP200 were selected and analyzed.

2.2.5 Pathway Analysis and STRING protein-protein interaction prediction

In this step, I imported the selected 48 shared genes into IPA and performed simple path analysis for IPA. I submitted only genetic symbols without numbers/data/statistics. I exported the IPA results in Excel and selected the affected path that I was interested in. I made protein-protein interaction prediction through STRING.

3.Result:

I will show and explain the results of each step in detail in R.(https://rpubs.com/Zeyuan0311/840018) (https://github.com/zeyuandong/TRGN-Capstone-2021Fall)

4.Disscusion ：

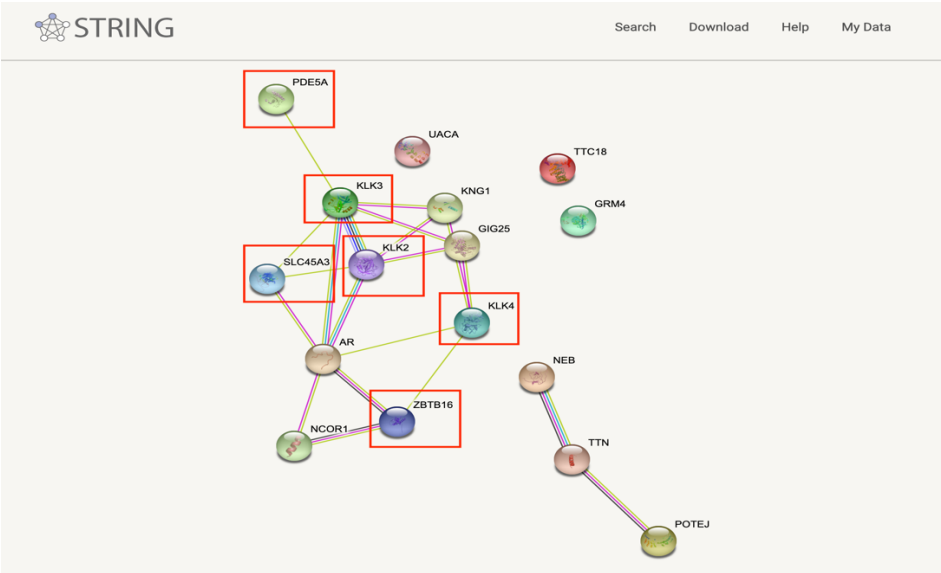

Figure.1 The protein-protein interaction prediction through STRING

KLK2, KLK3, KLK4, PDE5A, SLC45A3 and ZBTB16 were predicted to be potentially interact with Androgen Receptor(A R). NEB, TTN and POTEJ were predicted to be involved in the interaction but their functions in cancer studies are rarely reported. The rest of them are relatively independent from other proteins.

Although details about treatments remain to be updated in the future, it was assumed in this study that therapy is related to androgen because prostate is an androgen-dependent organ and androgen also plays a role in Prostate Cancer [5].

KLK3 encoding Prostate-specific antigen(PSA), and KLK2(kallikrein-related peptidase 2) are often reported together in prostate cancer studies [6]. PSA can degrade IGF binding protein 5 (IGFBP-5) [7], and cleave insulin-like growth factor (IGF) binding protein-3 (IGFBP-3)[8]. PSA testing has been advocated in prostate diagnosis and monitoring [9], but its concentration distribution varies greatly among general populations [10]. KLK3 was found to be a novel prognostic marker for prostate cancer undergoing first-line treatment with abiraterone acetate and prednisone (AA-P) at mRNA level this year by Emmy Boerrigter, et al [11]. With evaluation of in vitro, xenograft, and transgenic mice models, KLK2 functions as the protease to activate PSA [6]. In addition, both KLK2 and KLK3 polymorphism have been studied as prognostic marker in prostate cancer [12] [13]. Interestingly, the expression level of KLK3(PSA) will be reduced in malignant and continue to decrease in poorly differentiated tumors, although the level of PSA is higher in prostate cancer at protein level [14]. The enzymatic activity of PSA should also be considered and it may be associated with prostate cancer growth [14]. According to Juliana

Meola et al. [15], KLK2 mRNA expression levels in the circulating blood cells were higher in patients with prostate cancer than benign prostatic hyperplasia (BPH), while there are no difference in KLK3 expression levels. KLK3 expression in prostate cancer is always a controversial issue as mentioned above because KLK3 mRNA levels were detected lower in malignant tissues than begin tissues [14]. Although the role of KLK2 expression as a biomarker in blood needs to be improved further, we could still keep it in our gene of interest list while KLK3 will be removed.

Kallikrein-related peptidase 4 (KLK4) functions as a protease overproduced in localized prostate cancer and also participates in prostate cancer metastasis [16]. Androgen receptor(AR) and mTOR signaling pathway are very important proliferative pathway in our body, which are also involved in prostate cancer. AR and mTOR pathway are proved to be integrated by KLK4 and promyelocytic leukemia zinc finger(PLZF) [17]. PLZF is encoded by ZBTB16, which is also in our candidate gene list, and will be discussed later. PLZF can bind and inhibit AR, as well as inhibit mTOR signaling pathway. However, KLK4 can inhibit PLZF, and thus make AR and mTOR signaling pathway active [17]. This is also confirmed by the experimental fact that KLK4 siRNA could result in tumor remission in prostate cancer mice model [17]. Recently, the anti-tumor effects was also reported by Brian W-C Tse, et al.[18], but it was context-dependent and isoform-dependent, and more data and experiments should be provided to prove that in the future.Thus, we can still list KLK4 as our gene of interest because its important role in AR and mTOR signaling pathway.

SLC45A3 encoding Solute carrier family 45 member 3, is a marker for prostate cells(from Uniprot). The expression SLC45A3 is rarely reported alone, but it is frequently reported as fusion genes such as SLC45A3-ERG [19] and SLC45A3-ELK4 [20]. Little is known about its function and roles in Prostate Cancer. SLC45A3 and SLC45A2 were reported to be a good prognostic indicator for melanoma via data mining analysis study [21]. Due to the limited information about this gene, SLC45A3 was not in the list of gene of interest.

ZBTB16(Zinc Finger And BTB Domain Containing 16), also known as promyelocytic leukemia zinc finger (PLZF), is a versatile transcription factor participating in various biological processes. ZBTB16(PLZF) is an androgen-activated gene, and this has been widely confirmed by different experiments. First, its expression was confirmed in AR positive Prostate Cancer cellines LNCaP [22], VCaP and 22Rv1[23]. In addition, ZBTB16(PLZF) is not detected in AR-deficient celline DU145, and ZBTB16(PLZF) expression can be reversed by ectopic AR expression in DU145 cell [24]. ZBTB16(PLZF) is believed to be a tumor suppressor gene and a negative regulator of AR [17]. In this case, there was no obvious difference between each Tumor-Control pair group. Considering its tumor suppressor function, the reactivation of ZBTB16(PLZF) was mentioned as a potential therapy for PCa [25], but its expression was lowered after treatment in this research. The possible reason may be that activated AR can activate the expression of ZBTB16(PLZF), so the anti-androgen therapy will logically reduce ZBTB16(PLZF) expression. Although this result is reasonable, it still reminded us of the tumor suppressor function of ZBTB16(PLZF) in Prostate Cancer. Potentially, expression level of

ZBTB16(PLZF) can be regarded as a molecular marker for evaluating therapies in Prostate Cancer during designing stage because lowered expression level of ZBTB16(PLZF) is not a satisfying result to some extent. In this case, the downregulation of ZBTB16(PLZF) is acceptable.

PDE5A encoding cGMP-specific 3',5'-cyclic phosphodiesterase can play an important role in signal transduction via the regulation of the intracellular concentration of cyclic nucleotides [26]. What's more important is that PDE5A is observed mainly in glandular structures of the prostate [27], and PDE5A was identified as a drug target in prostate cancer [28]. The inhibition of PDE5A can weaken the growth of human prostate tumors [29]. PDE5A was also reported in various disease such as colon cancer [30] and cardiac hypertrophy [31]. Since PDE5A mRNA level are reported to be high in various cancers(https://www.proteinatlas.org/ENSG00000138735-PDE5A/pathology) and its role in tumor progression was studied in breast cancer models [32], PDE5A can be a factor together with other prostate cancer specific biomarkers.

Now, KLK2, KLK4, ZBTB16(PLZF) and PDE5A are in our final list of candidate genes because their functions in prostate cancer or tumor growth/progression have been supported by relatively more experiments and studies. KLK2, KLK4 and PDE5A are believed to have higher expression levels in malignant tumors in most situations. ZBTB16(PLZF) can suppress tumor so the downregulation of it would not be a ideal result to some extent, but in this case, as an AR-activating gene, ZBTB16(PLZF) will be downregulated because of anti-AR therapy.

All the conclusions are established on the fact that the treatment was effective(34C,T;36C,T). For assessing the outcome of treatment in this case, all these four genes will be expected to be downregulated roughly 3~4-fold after treatment based on the vst(FPKM) value shown in heatmap. Considering the interaction of KLK4 and PLZF, it may be a good idea to compare the ratio of KLK4/ZBTB16(PLZF) expression level instead of each single value. The reduction of this ratio after treatment will be a good sign of effective treatment, but how to make it statistically significant and how to validate its relationship needs further more studies on KLK4/ZBTB16(PLZF). Certainly, all of these work is an initial trying on looking for potential biomarker at mRNA level in African male prostate cancer patients to assess the response to a specific therapy. The information such as details of treatment and effectiveness of these patients are expected to be provided in the future to help us refine this study.

Reference:

1 Daniyal, M. *et al.* Epidemiology, etiology, diagnosis and treatment of prostate cancer. *Asian Pac J Cancer Prev* **15**, 9575-9578, doi:10.7314/apjcp.2014.15.22.9575 (2014).

2 Stark, R., Grzelak, M. & Hadfield, J. RNA sequencing: the teenage years. *Nat Rev Genet* **20**, 631-656, doi:10.1038/s41576-019-0150-2 (2019).

3 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550, doi:10.1186/s13059-014-0550-8 (2014).

4 Califf, R. M. Biomarker definitions and their applications. *Exp Biol Med (Maywood)* **243**, 213-221, doi:10.1177/1535370217750088 (2018).

5 Shafi, A. A., Yen, A. E. & Weigel, N. L. Androgen receptors in hormone-dependent and castration-resistant prostate cancer. *Pharmacol Ther* **140**, 223-238, doi:10.1016/j.pharmthera.2013.07.003 (2013).

6 Williams, S. A., Xu, Y., De Marzo, A. M., Isaacs, J. T. & Denmeade, S. R. Prostate-specific antigen (PSA) is activated by KLK2 in prostate cancer ex vivo models and in prostate-targeted PSA/KLK2 double transgenic mice. *Prostate* **70**, 788-796, doi:10.1002/pros.21111 (2010).

7 Maeda, H. *et al.* Prostate-specific antigen enhances bioavailability of insulin-like growth factor by degrading insulin-like growth factor binding protein 5. *Biochem Biophys Res Commun* **381**, 311-316, doi:10.1016/j.bbrc.2009.01.096 (2009).

8 Cohen, P., Peehl, D. M., Graves, H. C. & Rosenfeld, R. G. Biological effects of prostate specific antigen as an insulin-like growth factor binding protein-3 protease. *J Endocrinol* **142**, 407-415, doi:10.1677/joe.0.1420407 (1994).

9 Lilja, H., Ulmert, D. & Vickers, A. J. Prostate-specific antigen and prostate cancer: prediction, detection and monitoring. *Nat Rev Cancer* **8**, 268-278, doi:10.1038/nrc2351 (2008).

10 Rodriguez, S. *et al.* Very low PSA concentrations and deletions of the KLK3 gene. *Clin Chem* **59**, 234-244, doi:10.1373/clinchem.2012.192815 (2013).

11 Boerrigter, E. *et al.* Liquid biopsy reveals KLK3 mRNA as a prognostic marker for progression free survival in patients with metastatic castration-resistant prostate cancer undergoing first-line abiraterone acetate and prednisone treatment. *Mol Oncol* **15**, 2453-2465, doi:10.1002/1878-0261.12933 (2021).

12 Kohli, M. *et al.* Exploratory study of a KLK2 polymorphism as a prognostic marker in prostate cancer. *Cancer Biomark* **7**, 101-108, doi:10.3233/CBM-2010-0152 (2010).

13 Li, H., Fei, X., Shen, Y. & Wu, Z. Association of gene polymorphisms of KLK3 and prostate cancer: A meta-analysis. *Adv Clin Exp Med* **29**, 1001-1009, doi:10.17219/acem/121521 (2020).

14 Mattsson, J. M., Laakkonen, P., Stenman, U. H. & Koistinen, H. Antiangiogenic properties of prostate-specific antigen (PSA). *Scand J Clin Lab Invest* **69**, 447-451, doi:10.1080/00365510903056031 (2009).

15 Meola, J. *et al.* Differential expression of the KLK2 and KLK3 genes in peripheral blood and tissues of patients with prostate cancer. *Genetics and Molecular Biology* **29**, 193-199, doi:10.1590/s1415-47572006000200001 (2006).

16 Fuhrman-Luck, R. A., Stansfield, S. H., Stephens, C. R., Loessner, D. & Clements, J. A. Prostate Cancer-Associated Kallikrein-Related Peptidase 4 Activates Matrix Metalloproteinase-1 and Thrombospondin-1. *J Proteome Res* **15**, 2466-2478, doi:10.1021/acs.jproteome.5b01148 (2016).

17 Jin, Y. *et al.* Molecular circuit involving KLK4 integrates androgen and mTOR signaling in prostate cancer. *Proc Natl Acad Sci U S A* **110**, E2572-2581, doi:10.1073/pnas.1304318110 (2013).

18 Tse, B. W. *et al.* KLK4 Induces Anti-Tumor Effects in Human Xenograft Mouse Models of Orthotopic and Metastatic Prostate Cancer. *Cancers (Basel)* **12**, doi:10.3390/cancers12123501 (2020).

19 Perner, S. *et al.* Loss of SLC45A3 protein (prostein) expression in prostate cancer is associated with SLC45A3-ERG gene rearrangement and an unfavorable clinical course. *Int J Cancer* **132**, 807-812, doi:10.1002/ijc.27733 (2013).

20 Kumar-Sinha, C., Kalyana-Sundaram, S. & Chinnaiyan, A. M. SLC45A3-ELK4 chimera in prostate cancer: spotlight on cis-splicing. *Cancer Discov* **2**, 582-585, doi:10.1158/2159-8290.CD-12-0212 (2012).

21 Xie, J. *et al.* Database mining analysis revealed the role of the putative H+/sugar transporter solute carrier family 45 in skin cutaneous melanoma. *Channels* **15**, 496-506, doi:10.1080/19336950.2021.1956226 (2021).

22 Jiang, F. & Wang, Z. Identification and characterization of PLZF as a prostatic androgen-responsive gene. *Prostate* **59**, 426-435, doi:10.1002/pros.20000 (2004).

23 Hsieh, C. L. *et al.* PLZF, a tumor suppressor genetically lost in metastatic castration-resistant prostate cancer, is a mediator of resistance to androgen deprivation therapy. *Cancer Res* **75**, 1944-1948, doi:10.1158/0008-5472.CAN-14-3602 (2015).

24 Kikugawa, T. *et al.* PLZF regulates Pbx1 transcription and Pbx1-HoxC8 complex leads to androgen-independent prostate cancer proliferation. *Prostate* **66**, 1092-1099, doi:10.1002/pros.20443 (2006).

25 Jin, Y., Nenseth, H. Z. & Saatcioglu, F. Role of PLZF as a tumor suppressor in prostate cancer. *Oncotarget* **8**, 71317-71324, doi:10.18632/oncotarget.19813 (2017).

26 Loughney, K. *et al.* Isolation and characterization of cDNAs encoding PDE5A, a human cGMP-binding, cGMP-specific 3′,5′-cyclic nucleotide phosphodiesterase. *Gene* **216**, 139-147, doi:10.1016/s0378-1119(98)00303-5 (1998).

27 Uckert, S. *et al.* Immunohistochemical distribution of cAMP- and cGMP-phosphodiesterase (PDE) isoenzymes in the human prostate. *Eur Urol* **49**, 740-745, doi:10.1016/j.eururo.2005.12.050 (2006).

28 Han, Y. *et al.* Microarray analysis of copy-number variations and gene expression profiles in prostate cancer. *Medicine (Baltimore)* **96**, e7264, doi:10.1097/MD.0000000000007264 (2017).

29 Hamilton, T. K. *et al.* Potential therapeutic applications of phosphodiesterase inhibition in prostate cancer. *World J Urol* **31**, 325-330, doi:10.1007/s00345-012-0848-7 (2013).

30 Lehrer, S., Rheinstein, P. H. & Rosenzweig, K. E. Mutations of the PDE5A Gene Confer a Survival Advantage in Patients with Colon Cancer. *Cancer Prev Res (Phila)* **11**, 439-440, doi:10.1158/1940-6207.CAPR-18-0105 (2018).

31 Liu, K. *et al.* MicroRNA-19a/b-3p protect the heart from hypertension-induced pathological cardiac hypertrophy through PDE5A. *J Hypertens* **36**, 1847-1857, doi:10.1097/HJH.0000000000001769 (2018).

32 Catalano, S. *et al.* Phosphodiesterase 5 (PDE5) Is Highly Expressed in Cancer-Associated Fibroblasts and Enhances Breast Tumor Progression. *Cancers (Basel)* **11**, doi:10.3390/cancers11111740 (2019).