

## Peer Review File

**Article information:** <http://dx.doi.org/10.21037/tcr-20-1509>

### Reviewers' comments

*Thank you for the opportunity to review this interesting paper. The manuscript is well written and methodologically sound. However, I have some questions:*

**Comment 1.** *Do authors have any information about symptoms scores, nicturia and other subjective and objective measures of LUTS (i.e. qmax, post void residual)*

**Reply 1.** Thank you for noting this. We routinely collect IPSS questionnaires for BPH patients as they were evaluated at OPD, and subsequently performed measurements of uroflow rate with post-void residual and TRUS for the patients with severe LUTS (IPSS >20). Most patients who received TURP were suffered from severe LUTS or complications of BPH. In contrast, patients who received TRUS-guided needle biopsy were owing to elevated PSA or suspected prostate cancer, and their IPSS might reveal mild to moderate LUTS (IPSS <20). In this study, these patients were divided into two groups based on averaged BPH volume, and each group contained patients who either received prostate biopsies with mild to moderate LUTS or TURP due to severe LUTS. We found the comparison of the IPSS, Qmax, and post-void residual between these two

groups insignificant in statistics, and therefore did not show these data in previous manuscript. We hereby added these data in **Table 1**.

**Changes in the text:** We added data of several parameters including IPSS, Qmax, and postvoid residual volume in Table 1, and made a short explanation in Result (see Page 14, Line 12-15).

**Comment 2.** *Why patients undergo a transitional zone biopsy? It is not the current clinical practice to perform biopsies in the TZ, instead urologist is more prone to focus on peripheric zone were more frequently is located cancer?*

**Reply 2.** Thanks for the comment. In our routine practice of TRUS-guided prostate biopsies, we usually obtain 12 specimens, including parasagittal 1-3 and peripheral 1-3 in each lobe (peripheral zone (PZ) predominant), plus lesion focus specimens if necessary. Some studies have reported certain cancer detection rate existed in TZ biopsy, however, we don't perform transitional zone (TZ) biopsy routinely unless specific hypoechoic lesion is shown by TRUS. In each case of this study underwent TRUS biopsy, after suspicious malignancy (either in TZ or PZ) was excluded preliminarily by TRUS, we would obtain 1~2 TZ biopsies along bilateral para-sagittal planes for study in addition to routine practice. Nevertheless, either the specimens from TURP or biopsy, the patients whose pathologic reports showed cancerous or precancerous lesions were

excluded from this study.

**Changes in the text:** We added some explanatory text about “**Collection of patients’ data and samples**”. (see Main Document, Materials and Methods, Page 6, line 8-9, 10-12).

**Comment 3.** *Did these patients who had specimens from TZ had a suspect of cancer?*

*In this case, it is possible that simply the biopsy missed the cancer focus. Did patients perform multiparametric mri before biopsy?*

**Reply 3.** We appreciate your insightful suggestions. First, as above mentioned, the patients whose pathologic reports revealed cancerous or precancerous lesions in biopsy or TURP were excluded from this study. Second, we do agree with that there is actually certain proportion of cancer focus missed by simply the biopsy. It is still an important issue about how to improved cancer detection rate of prostate biopsy, and some methods, such as MRI-guided biopsy had been developed and well researched. However, limited by the regulations and supervision of the National Health Insurance Administration in Taiwan, multiparametric MRI or MRI-guided biopsy performed before verified diagnosis of prostate cancer is still not encouraged and won’t be “paid”, despite several studies have reported that multiparametric MRI might be useful to avoid unnecessary biopsy and aid to elevate the cancer detection rate.

**Changes in the text:** N/A

**Comment 4.** *Did authors account for false discovery rate in the analyses of data of gene expression?*

**Reply 4.** Thank you for noting this. In this study, the false discovery rate in the analyses of data of gene expression is 25%. We will clarify this in “**Statistical analysis**”.

**Changes in the text:** We added a paragraph to account for FDR. (see Main Document, Materials and Methods, Page 13, line 6-9).

**Comment 5.** *Please specify the software used for data analysis.*

**Reply 5.** Thanks for the comment. We have rewritten the part of “**Data mining and pathway analysis**” in **Materials and Methods**. First, we used the Partek software (Partek Inc., St. Louis, MO, USA) to conduct robust multi-array average analysis, which contains background correction, quantile normalization, probe level intensity calculation, and significant gene identification. Second, after data processed, Partek software was also used to conduct principal component analysis (PCA), which reduces higher-dimensional data into a three-dimensional graph, and to evaluate the similarity of gene expression profiles for different samples. Third, we used the Genesis program, which is a platform to simultaneously visualize and analyze a whole set of gene

expression data, to perform hierarchical clustering (HC) analysis and generate a visual representation (heatmap) of the expression profiles. Finally, Ingenuity Pathway Analysis (IPA) was applied to describe gene-gene interaction networks, biological functions, and canonical pathways of differentially expressed genes.

**Changes in the text:** We have modified our text as advised to specify the use of software. (see Page 7, line 13-19 and Page 8, line 1)

**Comment 6.** *Table 1 authors may want to report data as median and iqr since they used non-parametric test to test differences between groups*

**Reply 6.** Thank you for precious suggestion. We added median and IQR in the parameters tested by non-parametric test in **Table 1**.

**Changes in the text:** Please see **Table 1**.

**Comment 7.** *How authors defined the importance of down or up-regulated genes? A lasso regression model could be used to test the importance of each gene and to reduce the number of genes that should be included in the pathway analysis.*

**Reply 7.** Thanks for your insightful question and suggestion. First, as we compared the genes between two groups of prostate volume: group 1, volume  $\leq 40$  mL; and group 2, volume  $> 40$  mL, 361 differentially expressed genes (DEGs) with a fold change of  $>$

1.4 or  $< -1.4$  and  $FDR < 25\%$  were found. Among them, 214 genes (59.2%) were down-regulated and 147 genes (40.8%) were up-regulated in BPH tissue of group 2 compared with group 1. Since we assumed these DEGs might reveal the potential mechanism of chronic inflammation leads to differential BPH growth/progression, we focused on the genes associated with immune/inflammation pathways instead of relying on gene list ordered by significance ( $p$  value) calculated with t-test method. Second, as your suggestion, we have used LASSO regression model to analyze our data. The result showed several significant genes (ordered by absolute value of weight, such as, *C7*, *STK10*, *PA2G4*, *STC1*, *MAP3K7IP1*, *PTGES2*, *RBAF600*, *SAAL1*, *LSM3*, *PTK9*, *PNPLA4*). Some of these genes (including *C7*, *MAPK3KIP1*, *PA2G4*, *RBAF600*, *SALL1*, *STC1*, *STK10*) are associated with innate immune system/inflammation, and some (such as, *PTGES2*, *PTK9*, *RBAF600*, *STC1*, *STK10*) are associated with malignancy. We will evaluate these genes to unveil potential biomarker/mechanism in the future.

**Changes in the text:** N/A

**Comment 8.** *Authors reported that they used PCA to reduce in two dimensions the dataset. However, they reported a 3d figure (figure 1). Did they used the first three principal components?*

**Reply 8.** Thank you for noting this. We are very sorry for this mistake and will correct it. The true sentence is that “we used PCA to reduce high dimension dataset into three dimension” as shown in Fig. 1A.

**Changes in the text:** We have modified our text in “**Data mining and pathway analysis**”. (see Page 7, line 16-18)