bioRxiv preprint doi: https://doi.org/10.1101/818260; this version posted October 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Single-cell dissection of a rare human prostate basal cell carcinoma

Summary heading：Single-cell analysis of prostate basal cell carcinoma

Abstract

As a rare subtype of prostate carcinoma, basal cell carcinoma (BCC) has not been studied extensively and thus lacks systematic molecular characterization. Here we applied single-cell genomic amplification and RNA-Seq to a specimen of human prostate BCC ($CK34\beta E12^{+}/P63^{+}/PAP^{-}/PSA^{-}$). The mutational landscape was obtained via whole exome sequencing of the amplification mixture of 49 single cells, and the 5 putative driver genes mutated are *CASC5*, *NUTM1*, *PTPRC*, *KMT2C* and *TBX3*. The top 3 nucleotide substitutions are C>T, T>C and C>A, similar to common prostate cancer. The distribution of the variant allele frequency values indicated these single cells are from the same tumor clone. The transcriptomes of 69 single cells were obtained, and they were clustered into tumor, stromal and immune cells based on their global transcriptomic profiles. The tumor cells specifically express basal cell markers like *KRT5*, *KRT14* and *KRT23*, and epithelial markers *EPCAM*, *CDH1* and *CD24*. The transcription factor (TF) co-variance network analysis showed that the BCC tumor cells have distinct regulatory networks. By comparison with current prostate cancer datasets, we found that some of the bulk samples exhibit basal-cell signatures. Interestingly, at single-cell resolution the gene expression patterns of prostate BCC tumor cells show uniqueness compared with that of common prostate cancer-derived circulating tumor cells. This study, for the first time, discloses the comprehensive mutational and transcriptomic landscapes of prostate BCC, which lays a foundation

bioRxiv preprint doi: https://doi.org/10.1101/818260; this version posted October 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

for the understanding of its tumorigenesis mechanism and provides new insights into prostate cancers in general.

Xianbin Su[1,2,#], Qi Long[3,#], Juanjie Bo[1,#], Yi Shi[2,#], Li-Nan Zhao[2], Yingxin Lin[4], Qing Luo[2], Shila Ghazanfar[4,5], Chao Zhang[6], Qiang Liu[7], Lan Wang[2], Kun-Yan He[2], Jian He[2], Xiao-Fang Cui[2], Jean Y. H. Yang[4], Ze-Guang Han[2,*], Jian-Jun Sha[1,*], Guoliang Yang[1,*]

[1]Department of Urology and Key Laboratory of Systems Biomedicine (Ministry of Education), Renji Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai, China

[2]Key Laboratory of Systems Biomedicine (Ministry of Education) and Collaborative Innovation Center of Systems Biomedicine, Shanghai Center for Systems Biomedicine, Shanghai Jiao Tong University, Shanghai, China

[3]Key Laboratory for Regenerative Medicine (Ministry of Education), School of Biomedical Sciences, Faculty of Medicine, The Chinese University of Hong Kong, Hong Kong SAR, China

[4]Charles Perkins Center and School of Mathematics and Statistics, The University of Sydney, Sydney, Australia

[5]Cancer Research UK Cambridge Institute, University of Cambridge, Li Ka Shing Centre, Robinson Way, Cambridge CB2 0RE, United Kingdom

[6]Department of Urology, Shanghai Changhai Hospital, Second Military Medical

bioRxiv preprint doi: https://doi.org/10.1101/818260; this version posted October 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

University, Shanghai, China

[7]Department of Pathology, Renji Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai, China

[#]These authors contributed equally.

[*]To whom correspondence should be addressed:

Guoliang Yang, Department of Urology, Renji Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai 200127, China. Tel: +86-21-6373 0455; Fax: +86-21-6373 0455; Email: ygl0511@126.com

Jian-Jun Sha, Department of Urology, Renji Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai 200127, China. Tel: +86-21-6373 0455; Fax: +86-21-6373 0455; Email: shanjianjunrj@126.com

Ze-Guang Han, Key Laboratory of Systems Biomedicine (Ministry of Education), Shanghai Center of Systems Biomedicine, Shanghai Jiao Tong University, Shanghai 200240, China. Tel: +86-21-3420 7304; Fax: +86-21-3420 6059; Email: hanzg@sjtu.edu.cn

bioRxiv preprint doi: https://doi.org/10.1101/818260; this version posted October 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Running title

Single-cell analysis of prostate basal cell carcinoma

Abbreviations

BCC, basal cell carcinoma; CTCs, circulating tumor cells; MDA, multiple displacement amplification; PAP, prostatic acid phosphatase; PSA, prostate-specific antigen; TF, transcription factor; VAF, variant allele frequency.

Keywords

prostate, basal cell carcinoma, single-cell sequencing, gene expression, transcription factor, mutation

Conflict of Interest

The authors declare that they have no conflicts of interest with the contents of this article.

bioRxiv preprint doi: https://doi.org/10.1101/818260; this version posted October 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Introduction

Prostate cancer is a heterogeneous disease with complex subtypes and different cell origins [1-3], and basal cell carcinoma (BCC) is an extremely rare histological subtype comprising < 0.01% of prostate cancer [4]. Despite being largely regarded as pursuing an indolent clinical course, aggressive behaviors like recurrence or metastasis have been observed and deaths also reported, highlighting the complex and yet poorly understood mechanism [5-9]. There have been extensive reports on the genomic features of common prostate cancer [10-17], and the molecular features of normal human prostate basal cells, prostate basal cell hyperplasia, and basal populations from human prostate cancer have also been reported [18-20], but the molecular profile of prostate BCC is still lacking. The mutational and transcriptomic features of this tumor sub-type will thus be valuable for understanding of its tumorigenesis mechanism and development of future clinical treatments.

Single-cell sequencing has evolved to be a powerful tool that provides unprecedented resolution of clinical specimens especially with limited amounts, and marker-free decomposition of the constituent cell types based on single-cell transcriptional profiles allows precise dissection of complex systems such as various tissues or tumors [21-23]. Here we applied single-cell genomic amplification to a human specimen of prostate BCC and used the mixture of 49 single cells to provide mutational landscape of prostate BCC via whole exome sequencing. We then used single-cell RNA-Seq to obtain the transcriptomes of 69 cells from the same specimen. Three types of cells, namely tumor, stromal and immune cells, were identified with

bioRxiv preprint doi: https://doi.org/10.1101/818260; this version posted October 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

their molecular features revealed at single-cell resolution, providing a useful resource for not only this prostate cancer subtype but prostate cancer in general.

Results

Histological and immunohistochemical features of human prostate BCC

The tumor exhibited a widespread infiltrative growth pattern under microscopic examination, and individual cells had large pleomorphic nuclei and scant cytoplasm (Fig. 1). The tumor was negative for two known prostatic markers, prostatic acid phosphatase (PAP) and prostate-specific antigen (PSA), but strongly expressed CK34βE12 and P63, suggestive of basal cell origin (Fig. 1). Based on the cellular morphology and immunohistochemical staining, we diagnosed it as prostate BCC. The patient with BCC was resistant to chemotherapy but sensitive to radiotherapy, and followed-up for 32 months without recurrence or metastasis (Supporting Information, Fig. S1).

Whole exome sequencing reveals the mutational landscape of prostate BCC

As genetic variations are generally believed to be the causes of tumors, here we tried to reveal the mutational features of prostate BCC. Due to the limited amount of tumor specimens collected, after dissociation of the specimens into single-cell suspension, we first conducted single-cell whole genome multiple displacement amplification (MDA) on a microfluidic-chip. We then used the mixture of amplified products from 49 single cells for whole exome sequencing (Fig. 2A). A total of 91 functional exonic mutations were obtained, and 5 of the mutated genes, *CASC5*, *NUTM1*, *PTPRC*,

bioRxiv preprint doi: https://doi.org/10.1101/818260; this version posted October 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

KMT2C and *TBX3* are putative driver genes catalogued in the COSMIC Cancer Gene Census (Supporting Information, Fig. S2) [24]. Mutations in DNA repair genes are related to poor outcomes of prostate cancers [12, 25], but few are mutated in this prostate BCC sample such as *BRCA2*, consistent with good survival of this patient.

The top 3 nucleotide substitutions are C>T, T>C and C>A, which are exactly the same top 3 substitutions in recently reported localized, non-indolent prostate tumors (Fig. 2B) [13]. This suggests the etiology of prostate BCC is likely the same as common prostate tumors. The major composite COSMIC signatures are Signature 4, 5, 6 and 12 (Fig. 2C). While the etiologies of Signature 5 and 12 are still unknown, Signature 4 and 6 are associated with smoking and defective DNA mismatch repair, respectively [26].

By checking the distribution of the variant allele frequency (VAF) of the exonic mutations of prostate BCC, it is clear that most of the variations have VAF values of around 0.5, thus supporting that most of the single cells share a similar mutational background (Fig. 2D). The VAF values of the 5 driver mutations are also approximately 0.5 (Fig. 2E), indicating these single cells are probably originated from the same ancestor and form the major tumor clone. However, there are still some variations with lower VAF values which suggest sub-clonal mutations may have existed in some cells that are acquired during later diversified evolution of the tumor.

To understand whether the prostate BCC is showing similarities to other reported prostate cancer cases, we checked the mutational frequency of the prostate BCC mutated genes in prostate cancer samples collected in cBioPortal [27]. For the 88

bioRxiv preprint doi: https://doi.org/10.1101/818260; this version posted October 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

genes enquired, only 10 genes are not mutated, and 27 genes are mutated in more than 0.5% of the ~4400 prostate cancer samples (Fig. 2F). Interestingly, 3 of the driver genes from this study, *KMT2C*, *PTPRC*, and *TBX3* are also among the top mutated genes, especially *KMT2C* with mutation rate of ~6%. GSEA/MSigDB enrichment analysis [28] of the mutated genes shows that the top enriched terms are related to cell cycle, cytoskeleton and others (Fig. 2G), consistent with its basal cell origin.

Single-cell RNA-Seq reveals the transcriptomic features of prostate BCC

To understand the transcriptional profiles of prostate BCC, we then randomly selected 69 single cells for single-cell RNA-Seq (Fig. 3A). With a median unique mapping rate of 58.4% and 1.08 million mapped reads, we obtained a median of ~2800 genes per cell (Supporting Information, Fig. S3A). We used ERCC spike-ins as positive controls, and the high correlation efficiency between single-cells based on the 92 spike-ins confirmed the high quality of the data (Supporting Information, Fig. S3B-D). After filtering 5 outliers, the remaining 64 cells were classified into three clusters based on their global transcriptomic profiles (Fig. 3B-C). GSEA/MSigDB enrichment analysis identified them as tumor cells (51 cells), stromal cells (7 cells) and immune cells (6 cells) (Fig. 3D).

The three clusters of cells can also be clearly separated in tSNE plot based on the top genes expressed in each cluster (Fig. 4A). The tumor cells specifically expressed genes encoding cytokeratins including *KRT5*, *KRT14* and *KRT23*, which are reliable basal cell markers (Fig. 4B). Interestingly, epithelial markers *EPCAM*, *CDH1* and

bioRxiv preprint doi: https://doi.org/10.1101/818260; this version posted October 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

CD24 are also highly expressed in tumor cells. Stromal cells specifically express *COL1A2*, *THY1* and *BGN*, and immune cells specifically express monocytic marker *CD14*, *CXCL8* and myeloid cell nuclear differentiation antigen *MNDA* (Fig. 4B), suggestive of possible innate immune cell infiltration. Immunohistochemical staining confirmed the expression of CK14 (KRT14) at protein level, further proving the reliability of the sequencing data (Fig. 4C).

Transcription factors (TFs) covariance network analysis showed that the three clusters demonstrate different transcriptional regulatory status (Fig. 4D). Significantly, *MYC*, *ARID1A*, *TBX3* and *SOX11* are highly expressed in tumor cells, where, besides the well-known *MYC* that is recently reported to be over-expressed in human prostatic basal cells [19], the oncogenic role of *ARID1A*, a key component of chromatin remodeling complex SWI/SNF, has been recently demonstrated in liver cancer [29], the development-related gene *SOX11* was reported to be elevated in basal-like breast cancer [30], and *TBX3* plays important roles in development, differentiation and tumorigenesis [31]. The TFs co-variance network provides clues to future dissection of the regulatory mechanism of BCC.

Single-cell analysis of prostate BCC provides new insights to prostate cancer

We then checked the expression of prostate BCC tumor cell specific genes in previously reported luminal and basal cells sorted from normal and cancerous prostates by flow cytometry [18]. The prostate basal cells exhibit higher expression levels of *KRT5*, *KRT14* and *KRT23* compared with luminal cells, similarly as our

bioRxiv preprint doi: https://doi.org/10.1101/818260; this version posted October 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

prostate BCC tumor cells, further confirming their basal cell origin (Fig. 5A). Interestingly, the genes highly expressed in prostate BCC tumor cells fall into two groups based on their expression correlation patterns, with one group including *KRT5*, *KRT14*, *KRT23*, *DSC3*, *FGFR2*, etc., while the other group including *EPCAM*, *CDH1*, *CD24*, etc. (Fig. 5A). It seems the former group includes the basal cell markers while the latter group includes genes expressed in common prostate epithelial cells.

We further checked the expression patterns of BCC tumor cell specific genes in other bulk samples of human prostate cancers [10, 11, 32]. It is interesting to identify a small group of samples exhibiting basal cell gene expression features in two separate prostate cancer datasets (shaded branches in the dendrograms of Fig. 5B and Supporting Information, Fig. S4A). Likely explanations are that these tumor samples were derived from BCC patients or the samples contained basal cells. For TCGA PRAD dataset, the expressions of basal markers are detected in most samples, possibility due to the high number of genes detected by deep sequencing (Supporting Information, Fig. S4B). The split of the genes highly expressed in BCC tumor cells into 2 groups is further confirmed by these three datasets, with *KRT5*, *KRT14*, *KRT23* clustering together with *DSC3*, *FGFR2* and also possibly *ACTG2*. Our single-cell analysis of prostate BCC reveals the existence of basal cell features in current bulk samples of prostate cancer.

We then compared our results with one single-cell resolution dataset of prostate cancer [33]. Compared with circulating tumor cells (CTCs) from prostate cancer patients, the prostate BCC tumor cells exhibit distinct gene expression features with

bioRxiv preprint doi: https://doi.org/10.1101/818260; this version posted October 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

high levels of *IGF2*, *KRT5*, *KRT14*, *KRT23*, *AQP1*, *MIA*, *FGFR2* and *ACTG2*, although they all expressed *EPCAM*, *CDH1* and *CD24* (Fig. 5C). The results demonstrated the unique gene expression patterns of the prostate BCC tumor cells, which exhibit the molecular features of basal cells with activated genes like *IGF2* and *FGFR2*, consistent with report on the regulation of pluripotency by IGF and FGF pathways [34]. Another interesting finding is that the BCC immune cells clustered together with some single cells annotated as CTC-candidate, which were likely to be contaminating leukocytes to the CTCs (Fig. 5C). The single-cell level data also support the above mentioned split of the genes highly expressed in prostate BCC tumor cells, with one group representing genuine BCC markers while the other genes commonly expressed in prostate cancer.

Detection of exonic mutations in single-cell RNA-Seq reads

As our mutational data and single-cell RNA-Seq data were generated from the same prostate BCC specimen, we checked whether the exonic mutations could be detected in RNA-Seq reads. Interestingly, we indeed found mutations in RNA-Seq reads from some single cells for *GABPB2*, *ASTE1*, *COPS3* and *BEX2* (Supporting Information, Fig. S5). The variations are only detected in some of the single cells, which are likely because the mutated sites are not always covered by the relatively shallow single-cell RNA-Seq. Another phenomenon is that single cells in RNA-Seq analysis only contained either reference or altered variant, which is probably caused by allelic

bioRxiv preprint doi: https://doi.org/10.1101/818260; this version posted October 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

specific expression or allele dropout in RNA-Seq. The consistency of scRNA-Seq data and exome-seq data further suggested the reliability of our prostate BCC data.

Discussion

As a rare malignancy, there has been little research into the molecular characteristics of prostate BCC and no consensus on its treatment. Mostly, such data exist as sporadic reports on immunohistochemical expression in a small number of cases [8, 35-37]. This study represents the first genomic and transcriptomic profiling of this prostate cancer subtype, and the single-cell resolution transcriptomic data is especially valuable for understanding of not only this subtype but prostate cancer in general.

The molecular features of normal human prostate basal cells [19], prostate basal cell hyperplasia [20], and human prostate cancer derived basal populations [18] have been reported recently, and the single-cell transcriptomic profiles of prostate BCC is a good complement to these data for comparison studies and better understanding of prostate cancer cell origin. Most of the current genomics data for prostate cancer are generated using bulk samples [10, 11, 13], and the single-cell transcriptional profiles of prostate BCC provide an opportunity for deeper utilization of these data. For example, analysis using genes highly expressed in prostate BCC reveals the existence of basal cell features in some bulk samples of prostate cancer, which is consistent with recent classification of prostate cancer samples into basal-like and luminal-like subtypes using the PAM50 classifier [1, 38]. The single-cell resolution profiles of constitutional cell types from prostate BCC provide further advantage for future

bioRxiv preprint doi: https://doi.org/10.1101/818260; this version posted October 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

de-convolution of bulk samples [39], which will maximize the values of current genomic data of prostate cancer.

The prostate BCC patient is still surviving now, indicating this case is an indolent malignancy. Our genomics and transcriptomic analysis of this chemotherapy-resistant but radiotherapy-sensitive prostate BCC provides a useful resource for future dissection of the molecular mechanisms related to the radiotherapy responsiveness and treatment guidance. Mutations that disrupt the function of DNA damage repair genes have been shown to be associated with the aggressive clinical behavior of localized prostate cancer and with cancer-specific mortality [12, 40-42], and it has also been reported that men with metastatic prostate cancer and DNA-repair gene mutations have sustained responses to poly-ADP ribose polymerase (PARP) inhibitors and platinum-based chemotherapy [43, 44]. This study showed that the prostate BCC patient with few DNA-repair gene mutations was resistant to chemotherapy, suggesting that we may directly adopt radiotherapy rather than chemotherapy for such conditions. The results indicated that precision treatment of BCC could be more than just a histopathological triviality but be based on association of tumor phenotypes with molecular features.

In summary, single-cell analysis revealed the genomic and transcriptomic landscapes of the rare prostate BCC for the first time. This provides clues for elucidation of the tumorigenesis mechanism, further discovery of biomarkers and therapeutic targets, and better understanding of not only prostate BCC but prostate cancer in general.

bioRxiv preprint doi: https://doi.org/10.1101/818260; this version posted October 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Methods

Clinical specimens

This study was approved by the Ethnical Review Board of Renji Hospital, School of Medicine, Shanghai Jiao Tong University. Specimens were obtained by prostatic needle biopsy from a 55-year-old man with pelvic pain and irritative urinary symptoms. The specimens underwent pathological evaluation. After diagnosis as prostate BCC, the patient was treated with chemotherapy and radiotherapy and then followed-up for 32 months without evidence of local recurrence or distant metastases (Supporting Information, Fig. S1). The specimens were mechanically dissociated and digested with collagenase IV and DNase I, and the single-cell suspensions were used for both single-cell genome amplification and single-cell RNA-Seq.

Immunohistochemical staining

Immunohistochemical staining was done as previously described [45], and we used the following primary antibodies: P63 (abcam, ab124762), CK34βE12 (Dako, M0630), CK14 (abcam, ab7800), PAP (Dako, M0792), PSA (Dako, A056201). We used biotinylated universal link antibody as the secondary antibody.

Single-cell genome amplification and whole exome sequencing

Single-cell capture, lysis, reverse-transcription, and whole genome MDA were done in a microfluidic-based C1 DNA-Seq IFC (10~17 μm, Fluidigm) according to its

bioRxiv preprint doi: https://doi.org/10.1101/818260; this version posted October 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

protocol using illustra GenomiPhi V2 DNA Amplification Kit (GE Healthcare, 25660031). The amplified products from 49 single cells were mixed for exonic region capture and library preparation using Agilent SureSelect Human All Exon v7 Kit (Agilent, 5191-4005). The library was sequenced using illumina HiSeq platform with 2 × 150 bp sequencing mode. We used GATK [46] for variant calling and ANNOVAR [47] for functional annotation of the mutations. As there is no para-tumor tissue for this patient, we filtered SNPs using dbSNP141 [48] and 1,000 Genomes Project (v3) database [49]. We used MutationalPatterns [50] to decipher the mutational signature composition. For genes mutated in this prostate BCC sample, we checked their mutational frequencies in prostate cancer samples collected in cBioPortal [27]. GSEA/MSigDB enrichment analysis was also conducted for the genes mutated in this prostate BCC.

Single-cell RNA-Seq and data analysis

Single-cell RNA-Seq experiment and analysis were conducted as previously described [45]. Single-cell capture, lysis, reverse-transcription, and cDNA amplification were done in a microfluidic-based C1 RNA-Seq IFC (10~17 μm, Fluidigm). After library preparation with Nextera XT Kit and Index Kit (illumina), the single-cell libraries were pooled and sequenced using NextSeq 500 (illumina) with 2 × 76 bp sequencing mode. After generation of FPKM data, SC3 [51] was used for identification of cell outliers and single-cell clustering. The genes differently expressed between clusters were then identified via ANOVA analysis ($p < 0.05$) using

bioRxiv preprint doi: https://doi.org/10.1101/818260; this version posted October 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

SINGuLAR™ (Fluidigm). GSEA/MSigDB enrichment analysis was conducted for the genes specifically expressed in each cluster to facilitate cell type identification. Transcription factors (TFs) covariance network analysis was conducted to reveal the relationship of the TFs specifically expressed in each cell cluster as previously described [45, 52]. Gene expression files were retrieved from five other prostate cancer studies [10, 11, 18, 32, 33] for comparison studies.

Acknowledgement

This work is supported in part by the National Natural Science Foundation of China (81802806, 81472621, 81402329 and 81902561), National Program on Key Research Project of China (2016YFC0902701, Precision Medicine), Shanghai Health Bureau (20164Y0124), Medical and Engineering Crossover Fund of SJTU (YG2016QN71, YG2017MS67), University of Sydney – Shanghai Jiao Tong University Joint Research Alliance (USYD-SJTU JRA) grant, funding from Key Laboratory of Systems Biomedicine (Ministry of Education) (KLSB2017QN-03), and Incubating Program for Clinical Research and Innovation of Renji Hospital SJTU School of Medicine (PYII-17-010).

Authors' contributions

XS and GY designed the study. XS, JB, LZ, CZ, QL3, LW, KH, JH, XC, WX and GY carried out experiments. XS, QL1, YS, YL, QL2, SG and JY analyzed data. XS, GY and ZH drafted the manuscript. GY, JS and ZH supervised the study. All authors read and approved final version of the manuscript.

bioRxiv preprint doi: https://doi.org/10.1101/818260; this version posted October 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

References

1. Zhao SG, Chang SL, Erho N, Yu M, Lehrer J, Alshalalfa M, Speers C, Cooperberg MR, Kim W, Ryan CJ, et al. (2017) Associations of Luminal and Basal Subtyping of Prostate Cancer With Prognosis and Response to Androgen Deprivation Therapy. *JAMA Oncol* **3**, 1663-1672.

2. Goldstein AS, Huang J, Guo C, Garraway IP & Witte ON (2010) Identification of a cell of origin for human prostate cancer. *Science* **329**, 568-571.

3. Choi N, Zhang B, Zhang L, Ittmann M & Xin L (2012) Adult murine prostate basal and luminal cells are self-sustained lineages that can both serve as targets for prostate cancer initiation. *Cancer Cell* **21**, 253-265.

4. Tannenbaum M (1975) Adenoid cystic or "salivary gland" carcinomas of prostate. *Urology* **6**, 238-239.

5. Halat SK & MacLennan GT (2008) Adenoid cystic/basal cell carcinoma of the prostate. *J Urol* **179**, 1576.

6. Komura K, Inamoto T, Tsuji M, Ibuki N, Koyama K, Ubai T, Azuma H & Katsuoka Y (2010) Basal cell carcinoma of the prostate: unusual subtype of prostatic carcinoma. *Int J Clin Oncol* **15**, 594-600.

7. Chang K, Dai B, Kong Y, Qu Y, Wu J, Ye D, Yao X, Zhang S, Zhang H, Zhu Y, et al. (2013) Basal cell carcinoma of the prostate: clinicopathologic analysis of three cases and a review of the literature. *World J Surg Oncol* **11**, 193.

8. Ali TZ & Epstein JI (2007) Basal cell carcinoma of the prostate: a clinicopathologic study of 29 cases. *Am J Surg Pathol* **31**, 697-705.

9. Ayyathurai R, Civantos F, Soloway MS & Manoharan M (2007) Basal cell carcinoma of the prostate: current concepts. *BJU Int* **99**, 1345-1349.

10. The Cancer Genome Atlas Research Network (2015) The Molecular Taxonomy

bioRxiv preprint doi: https://doi.org/10.1101/818260; this version posted October 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

of Primary Prostate Cancer. *Cell* **163**, 1011-1025.

11. Robinson D, Van Allen EM, Wu YM, Schultz N, Lonigro RJ, Mosquera JM, Montgomery B, Taplin ME, Pritchard CC, Attard G, et al. (2015) Integrative clinical genomics of advanced prostate cancer. *Cell* **161**, 1215-1228.

12. Pritchard CC, Mateo J, Walsh MF, De Sarkar N, Abida W, Beltran H, Garofalo A, Gulati R, Carreira S, Eeles R, et al. (2016) Inherited DNA-Repair Gene Mutations in Men with Metastatic Prostate Cancer. *N Engl J Med* **375**, 443-453.

13. Fraser M, Sabelnykova VY, Yamaguchi TN, Heisler LE, Livingstone J, Huang V, Shiah YJ, Yousif F, Lin X, Masella AP, et al. (2017) Genomic hallmarks of localized, non-indolent prostate cancer. *Nature* **541**, 359-364.

14. Grasso CS, Wu YM, Robinson DR, Cao X, Dhanasekaran SM, Khan AP, Quist MJ, Jing X, Lonigro RJ, Brenner JC, et al. (2012) The mutational landscape of lethal castration-resistant prostate cancer. *Nature* **487**, 239-243.

15. Quigley DA, Dang HX, Zhao SG, Lloyd P, Aggarwal R, Alumkal JJ, Foye A, Kothari V, Perry MD, Bailey AM, et al. (2018) Genomic Hallmarks and Structural Variation in Metastatic Prostate Cancer. *Cell* **174**, 758-769.

16. Espiritu SMG, Liu LY, Rubanova Y, Bhandari V, Holgersen EM, Szyca LM, Fox NS, Chua MLK, Yamaguchi TN, Heisler LE, et al. (2018) The Evolutionary Landscape of Localized Prostate Cancers Drives Clinical Aggression. *Cell* **173**, 1003-1013 e1015.

17. Baca SC, Prandi D, Lawrence MS, Mosquera JM, Romanel A, Drier Y, Park K, Kitabayashi N, MacDonald TY, Ghandi M, et al. (2013) Punctuated evolution of prostate cancer genomes. *Cell* **153**, 666-677.

18. Smith BA, Sokolov A, Uzunangelov V, Baertsch R, Newton Y, Graim K, Mathis C, Cheng D, Stuart JM & Witte ON (2015) A basal stem cell signature identifies

bioRxiv preprint doi: https://doi.org/10.1101/818260; this version posted October 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

aggressive prostate cancer phenotypes. *Proc Natl Acad Sci U S A* **112**, E6544-6552.

19. Zhang D, Park D, Zhong Y, Lu Y, Rycaj K, Gong S, Chen X, Liu X, Chao HP, Whitney P, et al. (2016) Stem cell and neurogenic gene-expression profiles link prostate basal cells to aggressive prostate cancer. *Nat Commun* **7**, 10798.

20. Henry G, Malewska A, Mauck R, Gahan J, Hutchinson R, Torrealba J, Francis F, Roehrborn C & Strand D (2017) Molecular pathogenesis of human prostate basal cell hyperplasia. *Prostate* **77**, 1344-1355.

21. Gao S, Yan L, Wang R, Li J, Yong J, Zhou X, Wei Y, Wu X, Wang X, Fan X, et al. (2018) Tracing the temporal-spatial transcriptome landscapes of the human fetal digestive tract using single-cell RNA-sequencing. *Nat Cell Biol* **20**, 721-734.

22. Han X, Wang R, Zhou Y, Fei L, Sun H, Lai S, Saadatpour A, Zhou Z, Chen H, Ye F, et al. (2018) Mapping the Mouse Cell Atlas by Microwell-Seq. *Cell* **172**, 1091-1107 e1017.

23. Tanay A & Regev A (2017) Scaling single-cell genomics from phenomenology to mechanism. *Nature* **541**, 331-338.

24. Sondka Z, Bamford S, Cole CG, Ward SA, Dunham I & Forbes SA (2018) The COSMIC Cancer Gene Census: describing genetic dysfunction across all human cancers. *Nat Rev Cancer* **18**, 696-705.

25. Taylor RA, Fraser M, Rebello RJ, Boutros PC, Murphy DG, Bristow RG & Risbridger GP (2019) The influence of BRCA2 mutation on localized prostate cancer. *Nat Rev Urol* **16**, 281-290.

26. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SA, Behjati S, Biankin AV, Bignell GR, Bolli N, Borg A, Borresen-Dale AL, et al. (2013) Signatures of mutational processes in human cancer. *Nature* **500**, 415-421.

27. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, Sun Y,

bioRxiv preprint doi: https://doi.org/10.1101/818260; this version posted October 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Jacobsen A, Sinha R, Larsson E, et al. (2013) Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal* **6**, pl1.

28. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, et al. (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545-15550.

29. Sun X, Wang SC, Wei Y, Luo X, Jia Y, Li L, Gopal P, Zhu M, Nassour I, Chuang JC, et al. (2017) Arid1a has context-dependent oncogenic and tumor suppressor functions in liver cancer. *Cancer Cell* **32**, 574-589 e576.

30. Shepherd JH, Uray IP, Mazumdar A, Tsimelzon A, Savage M, Hilsenbeck SG & Brown PH (2016) The SOX11 transcription factor is a critical regulator of basal-like breast cancer growth, invasion, and basal-like gene expression. *Oncotarget* **7**, 13106-13121.

31. Wansleben S, Peres J, Hare S, Goding CR & Prince S (2014) T-box transcription factors in cancer biology. *Biochim Biophys Acta* **1846**, 380-391.

32. Beltran H, Prandi D, Mosquera JM, Benelli M, Puca L, Cyrta J, Marotz C, Giannopoulou E, Chakravarthi BV, Varambally S, et al. (2016) Divergent clonal evolution of castration-resistant neuroendocrine prostate cancer. *Nat Med* **22**, 298-305.

33. Miyamoto DT, Zheng Y, Wittner BS, Lee RJ, Zhu H, Broderick KT, Desai R, Fox DB, Brannigan BW, Trautwein J, et al. (2015) RNA-Seq of single prostate CTCs implicates noncanonical Wnt signaling in antiandrogen resistance. *Science* **349**, 1351-1356.

34. Bendall SC, Stewart MH, Menendez P, George D, Vijayaragavan K, Werbowetski-Ogilvie T, Ramos-Mejia V, Rouleau A, Yang J, Bosse M, et al. (2007) IGF and FGF cooperatively establish the regulatory stem cell niche of pluripotent

bioRxiv preprint doi: https://doi.org/10.1101/818260; this version posted October 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

human cells in vitro. *Nature* **448**, 1015-1021.

35. Yang XJ, McEntee M & Epstein JI (1998) Distinction of basaloid carcinoma of the prostate from benign basal cell lesions by using immunohistochemistry for bcl-2 and Ki-67. *Hum Pathol* **29**, 1447-1450.

36. Iczkowski KA, Ferguson KL, Grier DD, Hossain D, Banerjee SS, McNeal JE & Bostwick DG (2003) Adenoid cystic/basal cell carcinoma of the prostate: clinicopathologic findings in 19 cases. *Am J Surg Pathol* **27**, 1523-1529.

37. Montironi R, Mazzucchelli R, Stramazzotti D, Scarpelli M, Lopez Beltran A & Bostwick DG (2005) Basal cell hyperplasia and basal cell carcinoma of the prostate: a comprehensive review and discussion of a case with c-erbB-2 expression. *J Clin Pathol* **58**, 290-296.

38. Zhao SG, Chen WS, Das R, Chang SL, Tomlins SA, Chou J, Quigley DA, Dang HX, Barnard TJ, Mahal BA, et al. (2019) Clinical and Genomic Implications of Luminal and Basal Subtypes Across Carcinomas. *Clin Cancer Res* **25**, 2450-2457.

39. Wang X, Park J, Susztak K, Zhang NR & Li M (2019) Bulk tissue cell type deconvolution with multi-subject single-cell expression reference. *Nat Commun* **10**, 380.

40. Leongamornlert D, Saunders E, Dadaev T, Tymrakiewicz M, Goh C, Jugurnauth-Little S, Kozarewa I, Fenwick K, Assiotis I, Barrowdale D, et al. (2014) Frequent germline deleterious mutations in DNA repair genes in familial prostate cancer cases are associated with advanced disease. *Br J Cancer* **110**, 1663-1672.

41. Castro E, Goh C, Olmos D, Saunders E, Leongamornlert D, Tymrakiewicz M, Mahmud N, Dadaev T, Govindasami K, Guy M, et al. (2013) Germline BRCA mutations are associated with higher risk of nodal involvement, distant metastasis, and poor survival outcomes in prostate cancer. *J Clin Oncol* **31**, 1748-1757.

bioRxiv preprint doi: https://doi.org/10.1101/818260; this version posted October 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

42. Gallagher DJ, Gaudet MM, Pal P, Kirchhoff T, Balistreri L, Vora K, Bhatia J, Stadler Z, Fine SW, Reuter V, et al. (2010) Germline BRCA mutations denote a clinicopathologic subset of prostate cancer. *Clin Cancer Res* **16**, 2115-2121.

43. Cheng HH, Pritchard CC, Boyd T, Nelson PS & Montgomery B (2016) Biallelic Inactivation of BRCA2 in Platinum-sensitive Metastatic Castration-resistant Prostate Cancer. *Eur Urol* **69**, 992-995.

44. Mateo J, Carreira S, Sandhu S, Miranda S, Mossop H, Perez-Lopez R, Nava Rodrigues D, Robinson D, Omlin A, Tunariu N, et al. (2015) DNA-Repair Defects and Olaparib in Metastatic Prostate Cancer. *N Engl J Med* **373**, 1697-1708.

45. Su X, Shi Y, Zou X, Lu Z-N, Xie G, Yang JYH, Wu C-C, Cui X-F, He K-Y, Luo Q, et al. (2017) Single-cell RNA-Seq analysis reveals dynamic trajectories during mouse liver development. *BMC Genomics* **18**, 946.

46. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, et al. (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**, 1297-1303.

47. Wang K, Li M & Hakonarson H (2010) ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* **38**, e164.

48. Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM & Sirotkin K (2001) dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res* **29**, 308-311.

49. Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA & Abecasis GR (2015) A global reference for human genetic variation. *Nature* **526**, 68-74.

50. Blokzijl F, Janssen R, van Boxtel R & Cuppen E (2018) MutationalPatterns:

bioRxiv preprint doi: https://doi.org/10.1101/818260; this version posted October 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

comprehensive genome-wide analysis of mutational processes. *Genome Med* **10**, 33.

51. Kiselev VY, Kirschner K, Schaub MT, Andrews T, Yiu A, Chandra T, Natarajan KN, Reik W, Barahona M, Green AR, et al. (2017) SC3: consensus clustering of single-cell RNA-seq data. *Nat Methods* **14**, 483-486.

52. Zhang HM, Liu T, Liu CJ, Song S, Zhang X, Liu W, Jia H, Xue Y & Guo AY (2015) AnimalTFDB 2.0: a resource for expression, prediction and functional study of animal transcription factors. *Nucleic Acids Res* **43**, D76-81.

bioRxiv preprint doi: https://doi.org/10.1101/818260; this version posted October 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure Legend

Fig. 1. Histological and immunohistochemical staining of the prostate BCC. H&E staining and the expressions of CK34βE12, P63, PAP and PSA are shown.

Fig. 2. Whole exome sequencing reveals the mutational landscape of prostate BCC. **(**A) General description of the exome sequencing of mixtures from genomic amplification products of prostate BCC single cells and numbers of genetic variations. (B) Mutational spectra of the prostate BCC based on the variants after SNP filter. (C) Contributions of COSMIC signatures to the prostate BCC mutational spectra. (D) Distribution of the variant allele frequency (VAF) of the exonic mutations of prostate BCC. (E) Summary of the driver (overlapped with The Cancer Gene Census) and putative sub-clonal (VAF lower than 0.35) mutations of prostate BCC. (F) Percent of samples mutated in the cBioPortal prostate cancer datasets with those above 1% shown for the exonic mutated genes in prostate BCC. The red stars indicate Census driver genes. (G) The items enriched with the exonic mutated genes in prostate BCC via GSEA/MSigDB canonical pathway enrichment analysis.

Fig. 3. Single-cell RNA-Seq reveals the constituent cell types in the prostate BCC. (A) Experimental workflow. (B) Heat-map showing the single-cell gene-expression patterns of the three clusters identified. (C) Heat-map showing the Pearson correlation coefficients between single-cells. (D) The top 10 items enriched with the gene sets specifically expressed in each cell cluster via GSEA/MSigDB canonical pathway

bioRxiv preprint doi: https://doi.org/10.1101/818260; this version posted October 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

enrichment analysis.

Fig. 4. Single-cell RNA-Seq reveals the transcriptomic features of prostate BCC. (A) Heat-map showing the expression patterns of top 20 genes specifically and highly expressed in each cluster, and t-SNE plot of the three clusters. (B) The expression of selected markers with the median values and the first and third quartiles shown. (C) Immunohistochemical staining of the specimen with CK14 antibody. (D) TFs covariance networks of the single-cell data. Each node represents a TF, and each edge represents correlation between two TFs. The TFs specifically expressed in each cluster are colored differently.

Fig. 5. Single-cell analysis of prostate BCC provide new insights to prostate cancer. (A) (left) Heat-map showing the clustering of FACS sorted luminal (CD49f Lo) and basal (CD49f Hi) cell populations from human benign and cancerous prostates based on the genes highly expressed in prostate BCC tumor cells; (right) Heat-map showing the Pearson correlation coefficients between the genes. (Data from Smith *et al*. PNAS 2015; 112: E6544-E6552). (B) Analysis of prostate cancer samples in similar approach as (A). (Data from Robinson *et al*. Cell 2015; 161: 1215–1228). (C) Comparison of the gene expression profiles of single cells from Cluster 1~3 of prostate BCC with human prostate CTCs (CTC-candidate and CTC-confirmed), bulk primary tumors, single cells from prostate cancer cell lines (Pca cell line) and white blood cells (Data from Miyamoto *et al*. *Science* 2015; 349: 1351-1356).

Fig 1

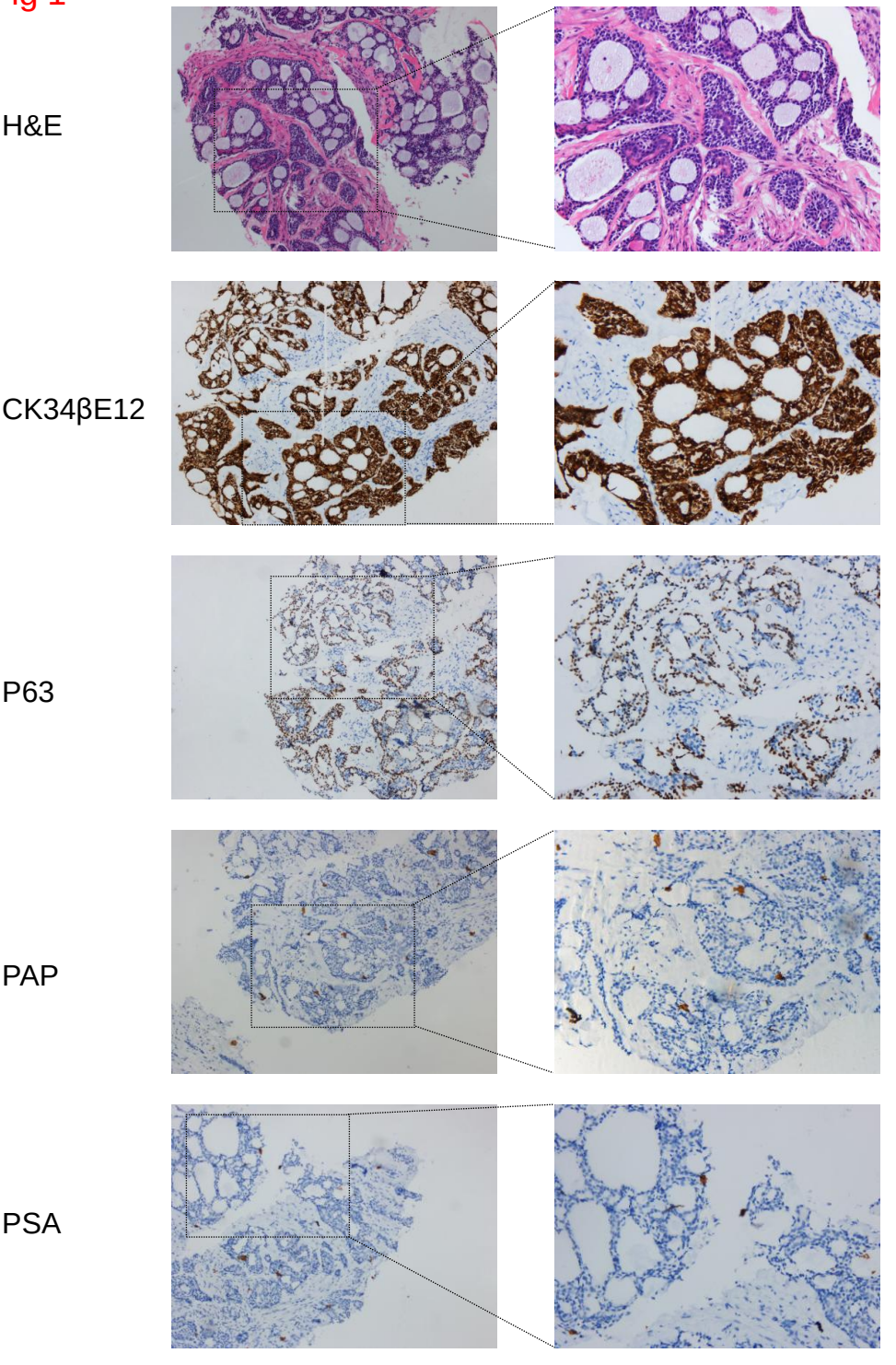
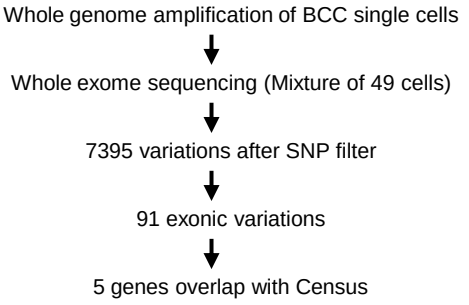
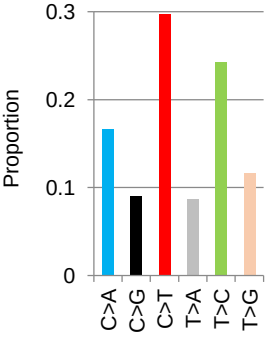


Fig 2

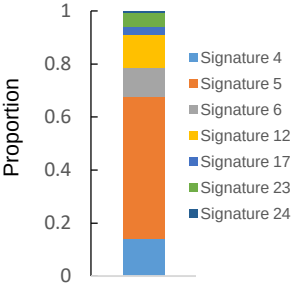
A



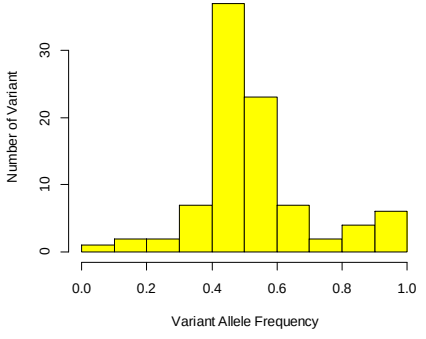
B



C



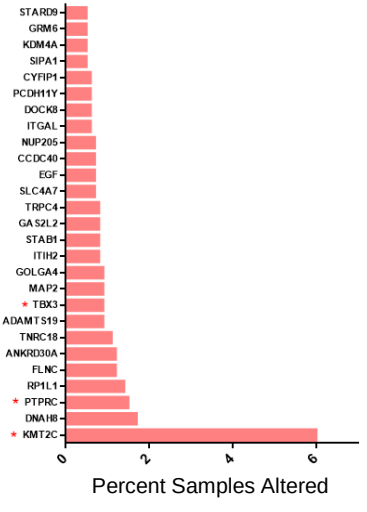
D



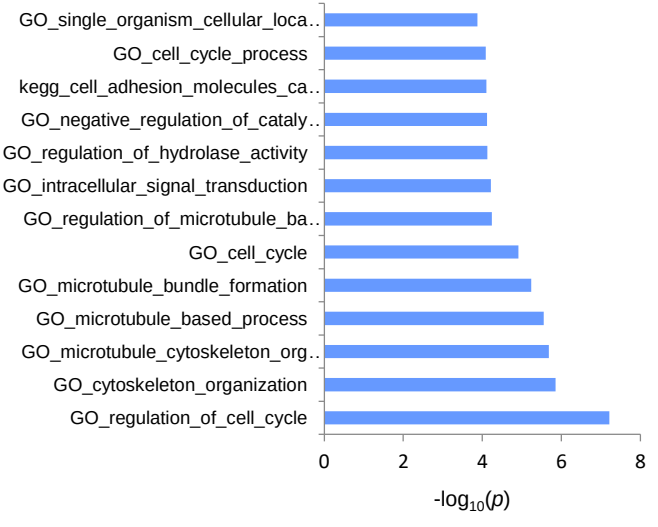
E

Group	Gene	#Chr	Position	Ref	Alt	AA change	VAF
Driver	CASC5	15	40903002	A	G	S86G	0.68
	NUTM1	15	34640367	G	C	V90L	0.63
	PTPRC	1	198713275	T	A	H769Q	0.59
	KMT2C	7	151845915	A	T	I4366N	0.54
	TBX3	12	115112422	C	G	A420P	0.42
Sub-clonal	KRT18	12	53343099	G	A	V48M	0.33
	TNRC18	7	5360200	G	A	P2200L	0.33
	KRT18	12	53343132	G	T	G59W	0.33
	SCUBE3	6	35213742	C	T	T874I	0.30
	KRT18	12	53343148	C	G	A64G	0.27
	ART5	11	3661519	T	G	E47A	0.25
	MST1L	1	17084321	A	G	C566R	0.17
	CCDC40	17	78064020	C	A	T972K	0.12
	C7orf26	7	6647724	G	C	A428P	0.09

F

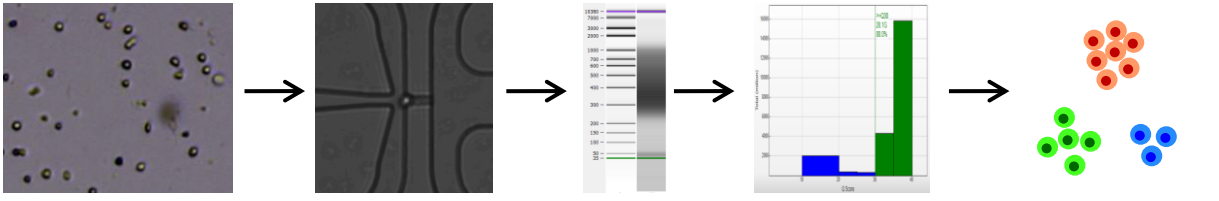


G



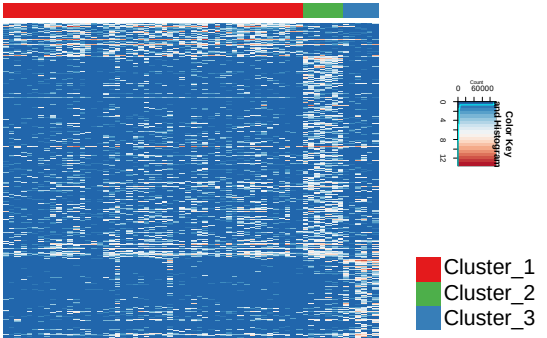
A

Fig 3

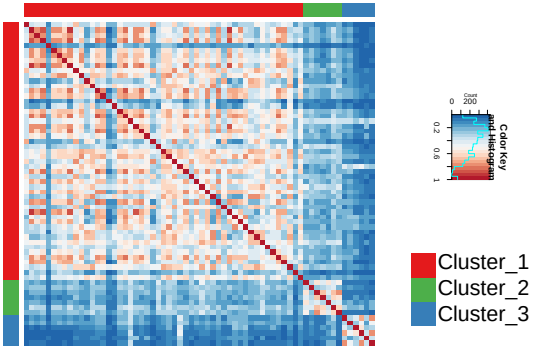


Single cells(SC) | Captured SC | Library prep | scRNA-Seq | SC transcriptome

B



C



D

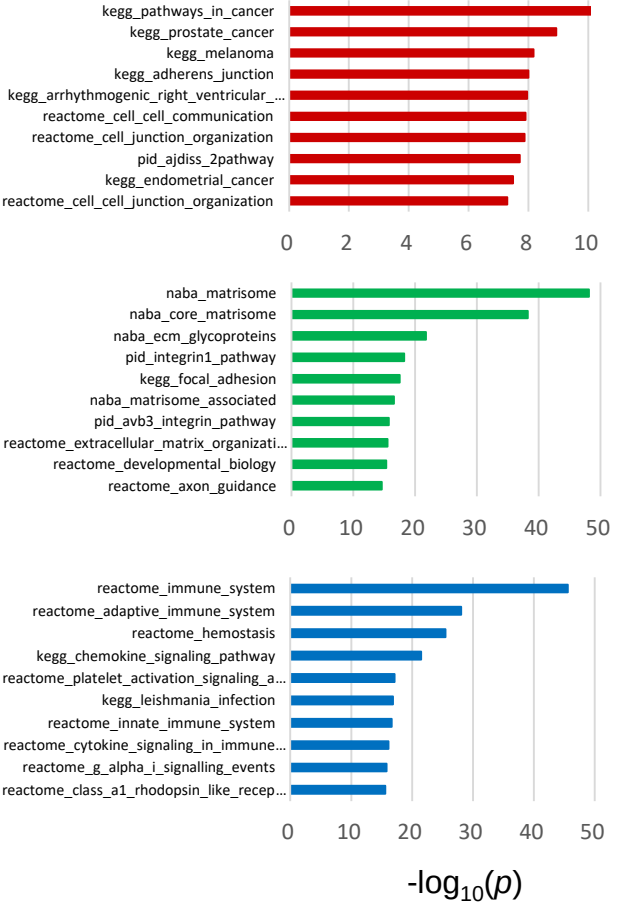
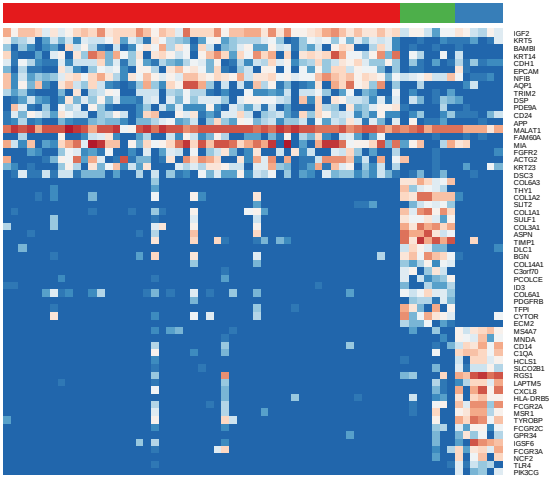
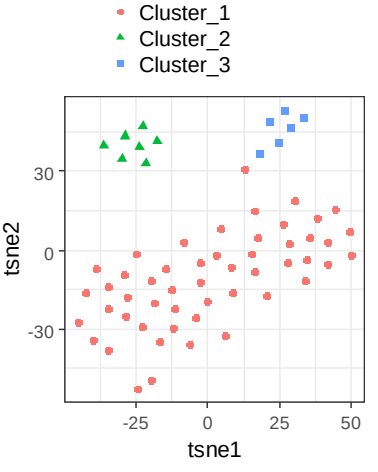


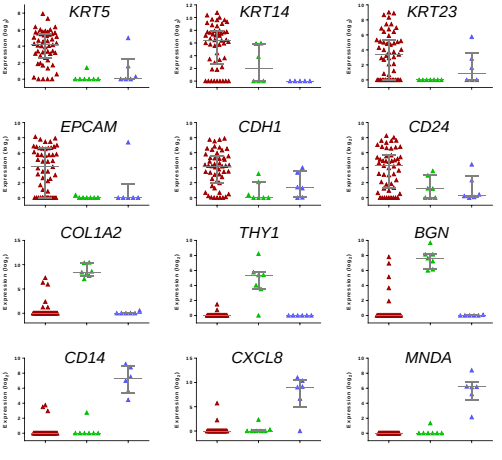
Fig 4

A

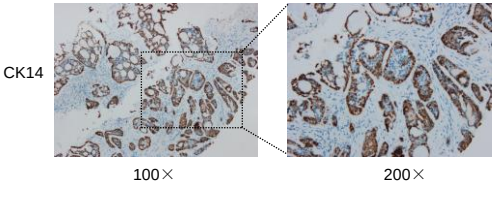

IGF2
KRT5
BAMBI
KRT14
CDH1
EPCAM
NFIB
AQP1
TRIM2
DSP
PDE8A
CD24
APP
MALAT1
FAM69A
MIA
FGFR2
ACTG2
KRT23
DSC3
COL6A3
THY1
COL1A2
SULT2
COL1A1
SULF1
COL3A1
ASPN
TIMP1
DLC1
BGN
COL14A1
C3orf70
PCOLCE
ID3
COL6A1
PDGFRB
TFPI
CYTOR
ECM2
MS4A7
MNDA
CD14
C1QA
HCLS1
SLCO2B1
RGS1
LAPTM5
CXCL8
HLA-DRB5
FCGR2A
MSR1
TYROBP
FCGR2C
GPR34
IGSF6
FCGR3A
NCF2
TLR4
PIK3CG
Cluster_1
Cluster_2
Cluster_3


Cluster_1
Cluster_2
Cluster_3
tsne2
tsne1
30
0
-30
-25
0
25
50

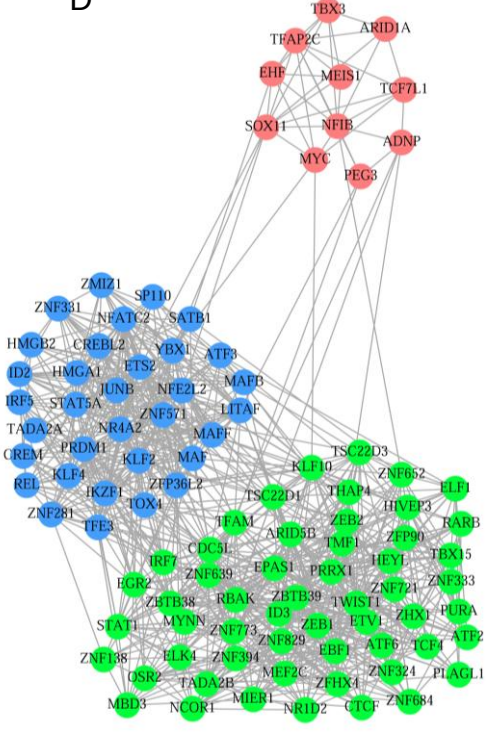
B


KRT5
KRT14
KRT23
EPCAM
CDH1
CD24
COL1A2
THY1
BGN
CD14
CXCL8
MNDA

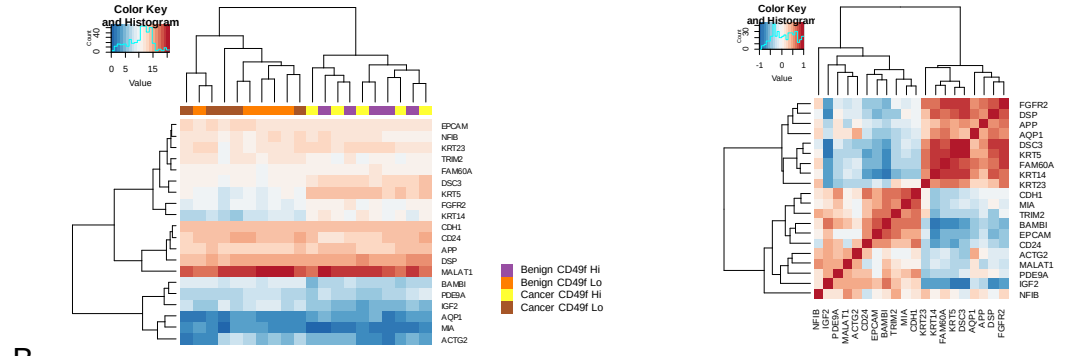
C


CK14
100×
200×

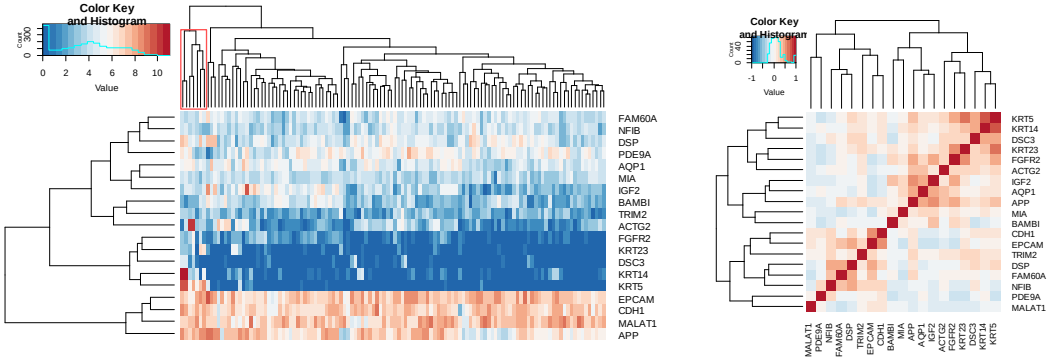
D


TBX3
ARID1A
TFAP2C
EHF
MEIS1
TCF7L1
SOX11
NFIB
ADNP
MYC
PEG3
ZMIZ1
SP110
ZNF381
NFATC2
SATB1
HMGB2
CREBL2
YBX1
ATF3
ID2
HMGA1
ETS2
JUNB
NFE2L2
MAFB
IRF5
STAT5A
ZNF571
LITAF
TADA2A
NR4A2
MAFF
CREM
PRDM1
KLF2
MAF
KLF10
TSC22D3
ZNF652
REL
KLF4
IKZF1
ZFP36L2
TSC22D1
THAP4
KLF1
ZNF281
TOX4
HIVEP3
TFE3
TFAM
ZEB2
ARID5B
RARB
CDC5L
TMF1
ZFP90
IRF7
HEYL
TBX15
EGR2
ZNF639
EPAS1
PRRX1
ZNF333
ZBTB38
RBAK
ZBTB39
TWIST1
ZNF721
STAT1
MYNN
ID3
ETV1
ZHX1
PURA
ZNF773
ZEB1
ZNF829
ATF6
TCF4
ATF2
ZNF138
ILK4
ZNF394
EBF1
OSR2
MEF2C
ZNF324
PLAGL1
TADA2B
ZFHX4
MBD3
NCOR1
MIER1
NR1D2
GTF
ZNF684

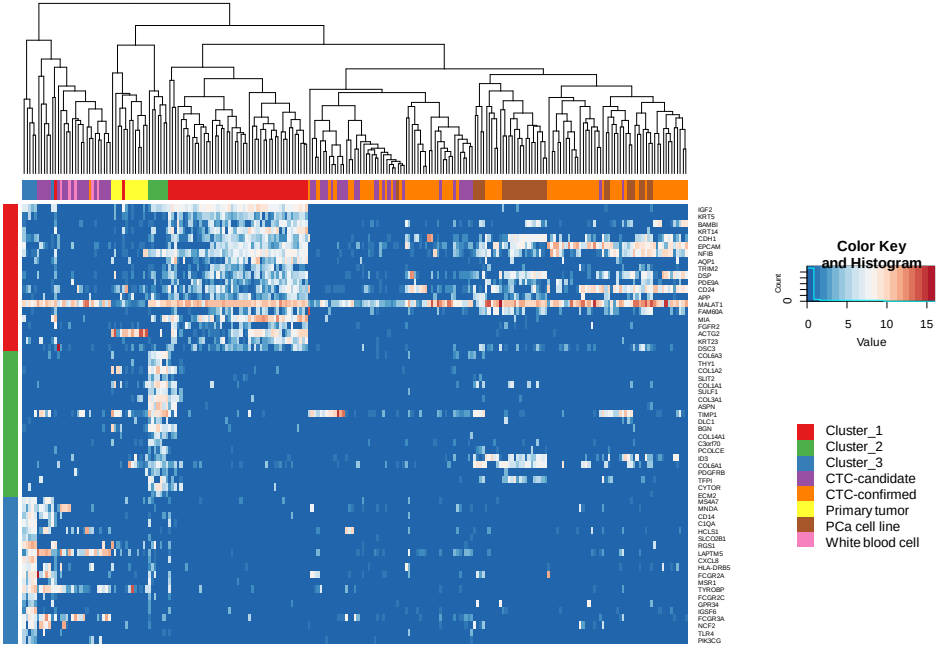
A Fig 5


Color Key and Histogram
Count
Value
EPCAM
NFIB
KRT23
TRIM2
FAM60A
DSC3
KRT5
FGFR2
KRT14
CDH1
CD24
APP
DSP
MALAT1
BAMBI
PDE9A
IGF2
AQP1
MIA
ACTG2
Benign CD49f Hi
Benign CD49f Lo
Cancer CD49f Hi
Cancer CD49f Lo

B


Color Key and Histogram
Count
Value
FAM60A
NFIB
DSP
PDE9A
AQP1
MIA
IGF2
BAMBI
TRIM2
ACTG2
FGFR2
KRT23
DSC3
KRT14
KRT5
EPCAM
CDH1
MALAT1
APP

C


Color Key and Histogram
Count
Value
Cluster_1
Cluster_2
Cluster_3
CTC-candidate
CTC-confirmed
Primary tumor
PCa cell line
White blood cell