BECKMAN COULTER
Life Sciences

Transcriptomic analysis reveals differential gene expression in African American prostate cancer samples

Nallasivam Palanisamy[1], Darshan Shimoga Chandrashekar[2], Han Wei [3]

[1] Henry Ford Health, Detroit, MI, United States
2 University of Alabama at Birmingham, Birmingham, AL, United States
3 Beckman Coulter Life Sciences, Indianapolis, IN, United States

Introduction

Prostate cancer (PCa) is the most frequent cancer in men in the Western world, and its incidence is rapidly increasing[1]. In addition, population-based studies demonstrate that African Americans (AA) in the United States have an increased risk of developing prostate cancer and a 4.4% risk of death compared with a 2.4% risk of death in Caucasians (CA)[2]. Despite advances in diagnosis and treatment, the molecular mechanisms underlying prostate cancer disparity in AA are not fully understood yet.

Previous studies have identified genetic[3] and environmental factors[4] that may contribute to this disparity, but a better understanding of the molecular mechanism involved in PCa in AA is needed. Transcriptomic analysis can provide insights into molecular pathways and genes that are dysregulated in PCa and may identify potential targets for therapy.

Formalin fixation and paraffin embedding (FFPE) of solid tumor samples have been pathologists' standard sample preparation method for decades, thus offering a vast archival resource of matched disease and normal tissues paired with other clinically annotated samples and patient outcome data[5]. The ability to sequence cohorts of AA FFPE samples is crucial to understanding the molecular heterogeneity of multifocal prostate cancer, but FFPE sample preparation methodology and storage conditions create unique challenges in the utilization of this sample type. Foremost among these difficulties is obtaining sufficient quantities of high-quality nucleic acids (i.e., DNA or RNA)[6]. Formalin causes nucleic acid fragmentation, especially in RNA. This fragmentation can impair nucleic acid recovery and lead to the recognition of only short sequences (approximately 100–200 nucleotides)[6].

In this whitepaper, we conducted a proof of principle study to identify transcriptomic signatures in the AA population using in-house PCa AA FFPE and fresh frozen samples (**Figure 1**). We identified 3,936 downregulated and 9,696 upregulated genes in in-house PCa AA groups. Commonly regulated protein-coding genes (mRNA) and long non-coding (lncRNA) genes were detected in in-house PCa AA and The Cancer Genome Atlas (TCGA) PCa AA groups. Functional enrichment analysis of the differentially expressed genes linked the top biological processes and associated pathways to the in-house PCa AA group.

Materials and Methods

FFPE RNA isolation

For each FFPE sample, RNA was extracted from a single 10 µm curl input using the FormaPure XL RNA kit (Beckman Coulter Life Sciences, Indiana, U.S.) per the manufacturer's instructions. Briefly, the samples were placed into 1.5 mL microcentrifuge tubes and deparaffinized in mineral oil. After

complete deparaffinization of the pieces, the lysis buffer and proteinase K were added, and the sample was incubated for two hours on a heat block at 60°C. The lysate was then transferred into a brand-new microcentrifuge tube. The RNA was bound to magnetic beads and DNase-treated, impurities were rinsed away from the beads, and the RNA was eluted in 40 µL elution buffer.

FFPE RNA quantification and integrity

All RNA samples were quantified using the Quant-it RiboGreen RNA reagent according to the manufacturer's instructions (Life Technologies, California, U.S.), and 1 ng from each sample was run on the Agilent 2100 Bioanalyzer instrument using the RNA kit (Agilent, California, U.S.) to assess fragmentation and sample integrity. RNA integrity was evaluated using DV200 values (percent of fragments greater than 200 nucleotides); the RNA with DV200 of greater than 30% was used for subsequent processing.

Library preparation and sequencing

All libraries were prepared using the xGen Broad-Range RNA Library Preparation kit (IDT, Iowa, U.S.) following the manufacturer's protocol. Up to 600 ng of total FFPE RNA was used. Excess primers and primer dimers were removed using AMPure XP beads (Beckman Coulter Life Sciences, Indiana, U.S.). The samples were then sequenced on the NovaSeq instrument (Illumina, San Diego, U.S.) using paired-end sequencing. The protocol has a depth of approximately 80 million reads per sample. The average unique read depth is 18 million.

Sequencing data analysis

Raw sequencing reads in fastq files were processed using the Trim_galore (v0.6.6) tool [**https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/**] to remove adapter sequences and low-complexity regions. Trimmed reads were mapped to the human genome GRCh38 using STAR software[7] v2.7.9a6. The plug-and-play resource library from Trinity Cancer Transcriptome Analysis Toolkit (CTAT) was used for annotation. [**https://data.broadinstitute.org/Trinity/CTAT_RESOURCE_LIB/GRCh38_gencode_v22_CTAT_lib_Mar012021.plug-n-play.tar.gz**].

For the purpose of differential expression analysis, raw reads were mapped using the STAR aligner tool[8] using default parameters. Aligned and sorted BAM output files were processed using the analysis tool HTSeq27[9] to enumerate raw read counts for each gene.

We also downloaded raw read counts from The Cancer Genome Atlas (TCGA) data portal for 7 tumor samples from African American prostate cancer patients and 4 normal samples related to African American individuals [**https://portal.gdc.cancer.gov/**]. Differentially expressed genes (DEGs) between the in-house PCa AA cohort (n=17) and TCGA Prostate Adenocarcinoma (PRAD) cohorts were identified using the DESeq2 tool[10-11]. Analysis was performed as per standard protocol [**https://bioconductor.org/packages/release/bioc/vignettes/DESeq2/inst/doc/DESeq2.html**].

Log2-fold change of ≥ 1 and adjusted P-value <0.05 were cut-off parameters considered to identify DEGs. Gene lists were compared using an online tool, Venny [**https://bioinfogp.cnb.csic.es/tools/venny/**]. Visualizations such as heatmaps and volcano plots were generated using R v4.2. KEGG pathways and Gene Ontology biological process enrichment in DEGs were obtained using the Bioconductor package "clusterProfiler"[12].

Results and Discussion

FFPE RNA demonstrates good yield and integrity

The integrity and quantity of RNA used in an RNA sequencing (RNA-seq) workflow can substantially impact the library yield and data quality. Because FFPE RNA is often degraded, assessing the quality

of the isolated material is important for adjusting the RNA-seq workflow to optimize results. We first examined the FFPE RNA yield from 5 PCa AA samples (**Figure 2A**). We observed that sufficient yield was obtained from all samples and exceeded the 600 ng required for RNA sequencing. We next evaluated the integrity of the isolated RNA as assessed by the DV200 value. As a valuable predictor of successful sequencing results, a DV200 value ≥ 30% (**Figure 2B**, red line) is recommended by the xGen Broad-Range RNA Library Preparation kit manufacturer for FFPE library preparations. Of these 5 in-house PCa AA FFPE samples, all had RNA of sufficient integrity (> 60%) for RNA sequencing. In summary, this data indicates that RNA obtained from FFPE samples using the FormaPure XL RNA kit is of sufficient yield and integrity for RNA-seq.

Alignment score for FFPE RNA-seq

FFPE processing results in fragmented RNA and chemically altered residues that can lead to sequence alterations and read misalignment, causing reduced aligned read counts. To ensure minimum artifacts, we examined the number and percentage of mapped reads from FFPE-derived sequences to the reference genome. Using STAR, we calculated that the average unique reads across all in-house PCa AA samples was 68 million (**Figure 3A**), or 75% (**Figure 3B**) of the reference genome. We are confident with the alignment and sequencing quality based on these results.

The transcriptomic landscape of PCa AA

In-house FFPE and frozen PCa AA and TGCA PCa AA gene expression levels were compared to TGCA AA non-pathological prostate data. Volcano plots in **Figures 4A** and **4B** show downregulated (blue nodes) and upregulated (red nodes) genes. Compared to TGCA AA PRAD normal, we identified 3,936 downregulated and 9,696 upregulated genes in in-house PCa AA groups, and 904 downregulated and 753 upregulated genes in the TGCA PRAD AA group. We also identified commonly regulated protein-coding genes (mRNA) and lncRNA-coding genes (long non-coding genes) in in-house PCa AA and TCGA PCa AA groups (**Figure 4C**). Over 300 mRNA and lncRNA-coding genes were commonly upregulated (Left), and about 60-100 genes were commonly downregulated (Right).

Furthermore, the expression profiles for the differentially expressed genes revealed key differences between in-house PCa AA samples and TGCA AA healthy samples, as shown in **Figures 5A** and **5B**. This suggests that PCa AA samples had distinct transcriptomic changes at multiple molecular levels when compared to reference samples.

Further functional enrichment analysis revealed that several well-known biological processes related to PCa were significantly enriched, such as *homophilic cell adhesion*, *regulation of hormone levels*, *cell-cell adhesion*, and *hormone metabolic processes*[13-15].

Among the enriched KEGG pathways, the Renin-angiotensin system (RAS) had the top enrichment. RAS has been shown to affect cell morphology, enhance cell proliferation and survival, and increase the risk of PCa[12]. We also noticed that the Rap1 pathway was enriched. Rap1 has been implicated in cancer tumorigenesis by affecting PCa cell migration and invasion, leading to metastatic progression[16].

Conclusion

In this study, we sought to identify novel biomarkers associated with racial disparities in PCa morbidity and mortality from FFPE and frozen samples. Despite the challenges associated with FFPE samples, we were able to isolate sufficient quantities of RNA that passed the DV200 cut-off and were suitable for downstream NGS.

We investigated the transcriptomic alterations in in-house PCa AA and TCGA reference samples. To achieve this, we identified 3,936 downregulated and 9,696 upregulated genes as distinct in the PCa AA group. Furthermore, we detected commonly regulated mRNA and lncRNA-coding genes in both PCa AA and TCGA PCa AA groups. Subsequently, a functional enrichment analysis was conducted to link the top biological processes, which included cell cycle, signal transduction, cell-cell interactions, cell adhesion, and hormone changes, to the differentially expressed genes.

Our study suggests that these genes might be involved in different networks that lead to aggressive PCa behavior in AA samples, potentially serving as biomarkers for the PCa racial disparity. However, further studies are warranted to validate our findings and to better understand the functional significance of the differentially expressed genes identified in this study.

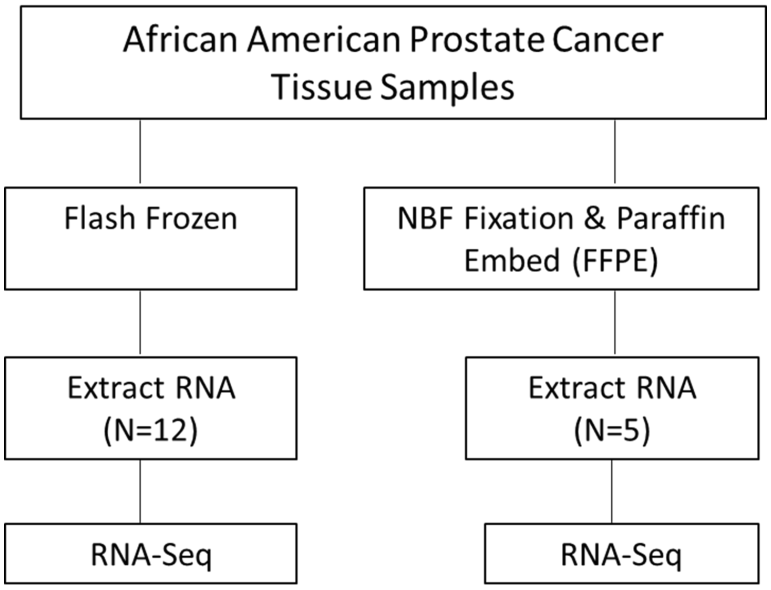


Figure 1. Experimental overview. The study aimed to identify transcriptomic signatures of PCa in AA using FFPE (N=5) and frozen (N=12) samples. TCGA AA normal prostate (N=4) and TCGA AA PCa (N=7) were used as reference data. NBF: neutral-buffered formalin.

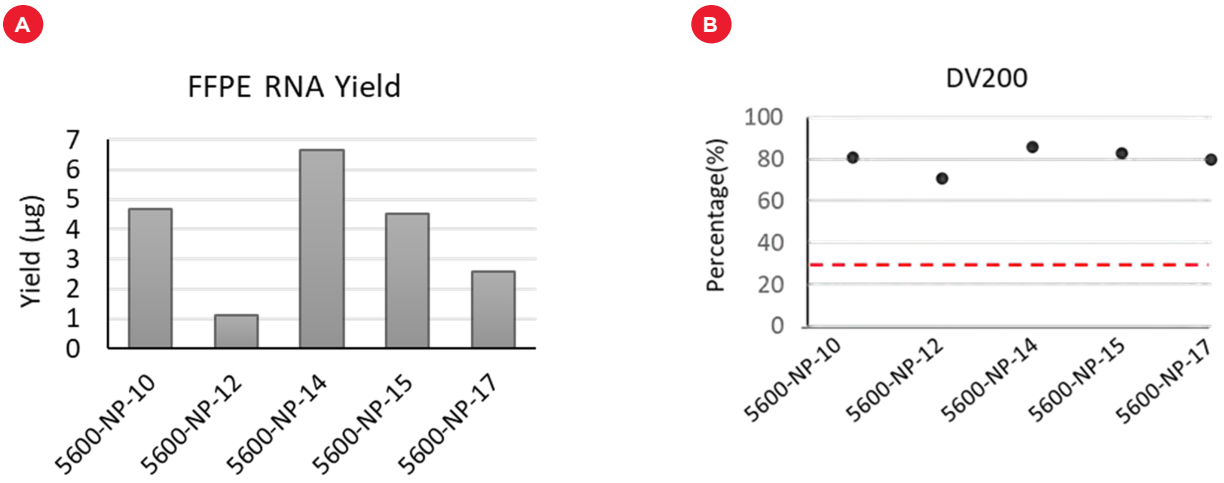
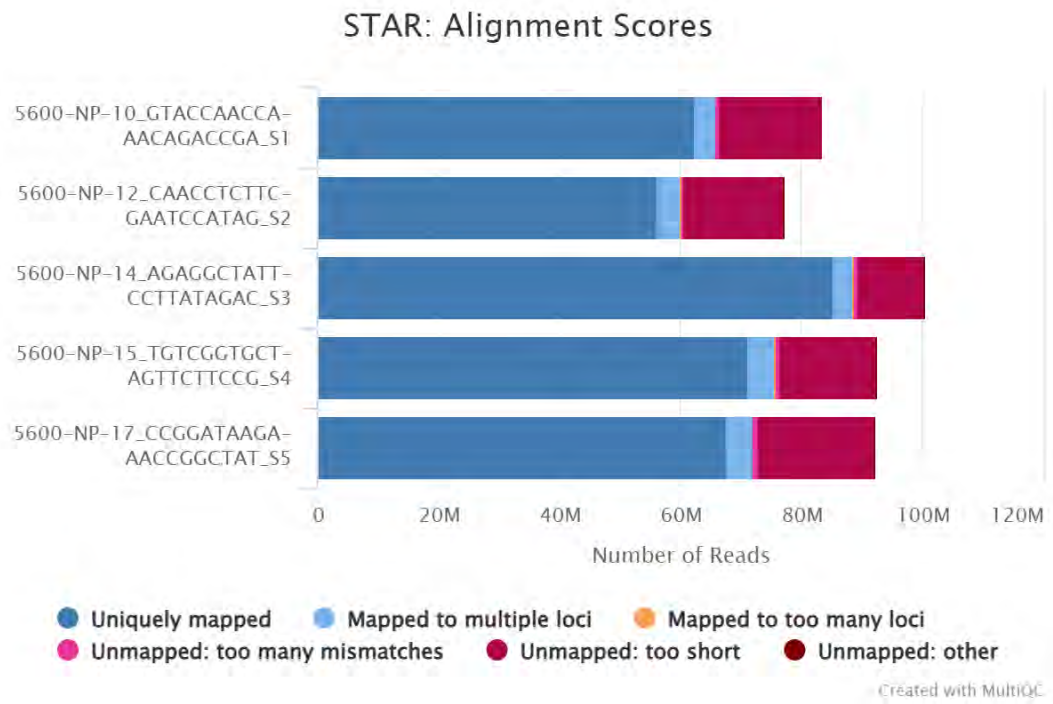


Figure 2. FFPE RNA quantification and integrity. (**A**) AA PCa samples were subjected to RNA isolation from 10 µm of tissue and evaluated for overall RNA yield and (**B**) RNA integrity determined by DV200 value. DV200 = 30 denotes recommended minimal technical limits for successful sequencing.

A



B

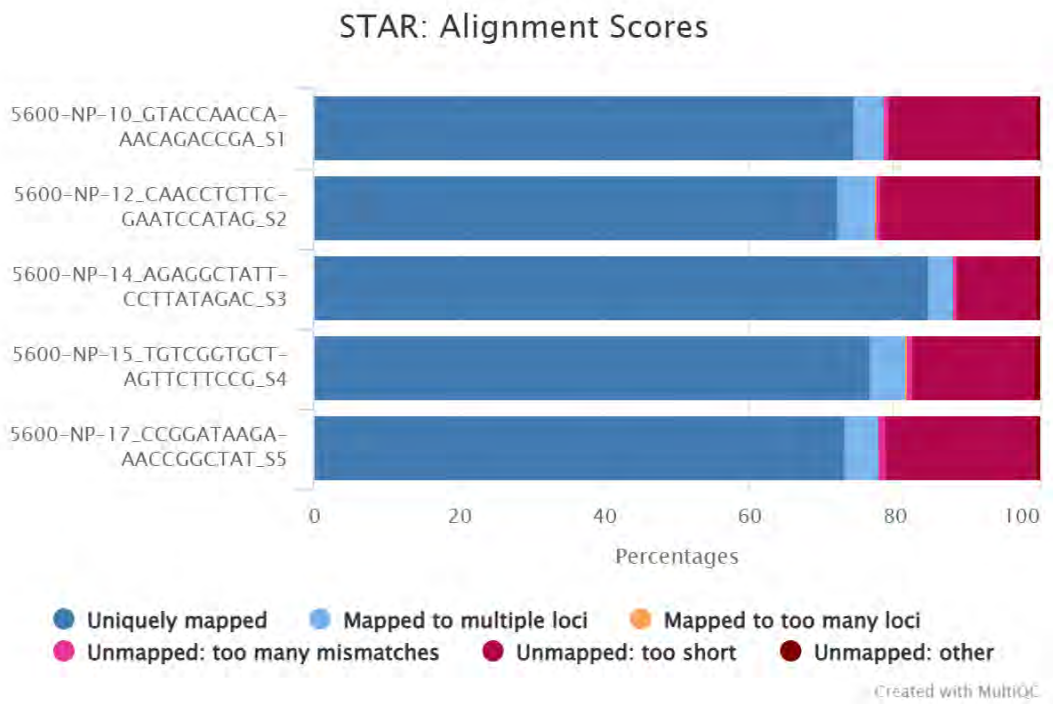


Figure 3. Read alignment across AA PCa FFPE samples. The average unique reads were 68 million (**A**), or 75% (**B**) unique reads mapping across the samples.

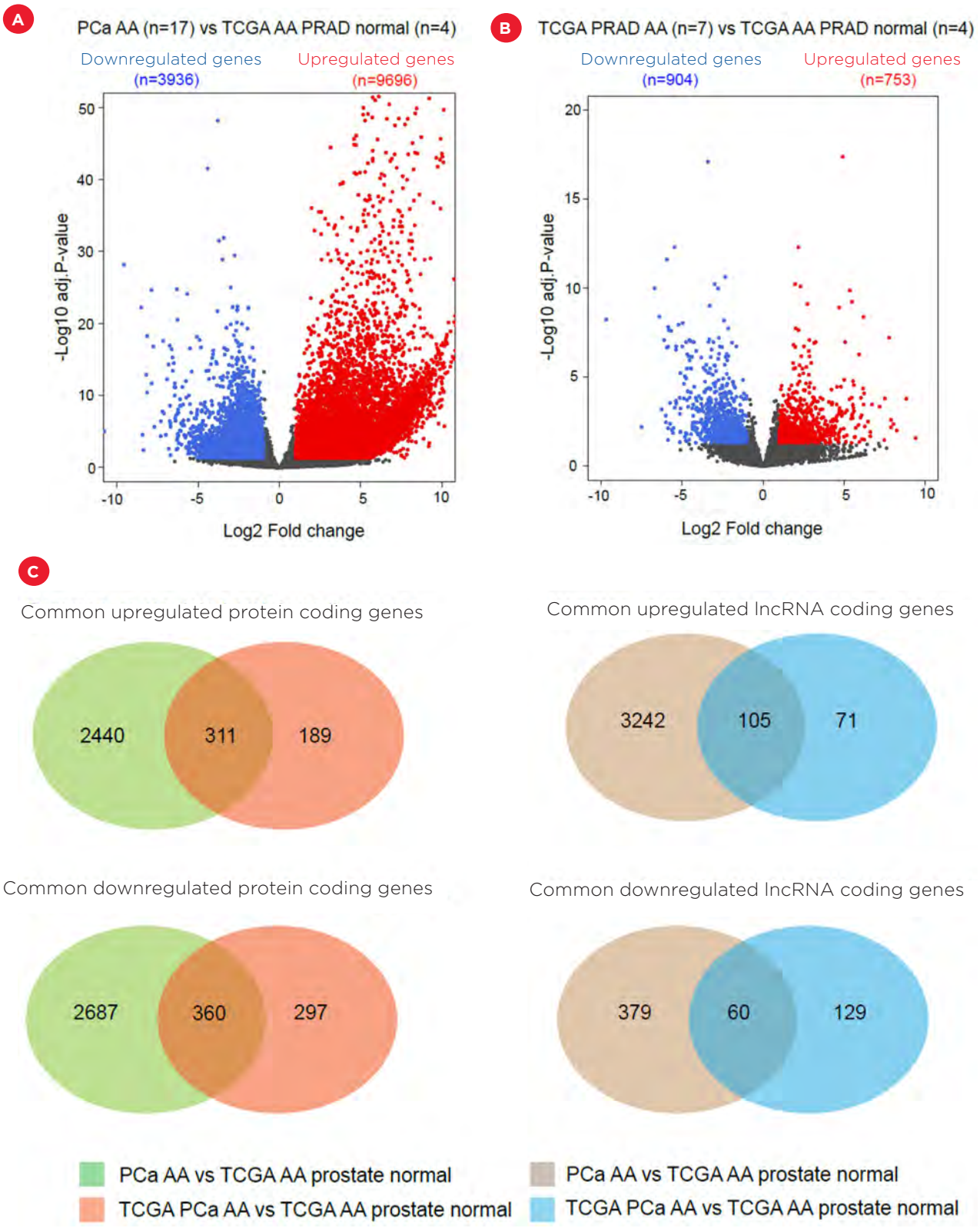
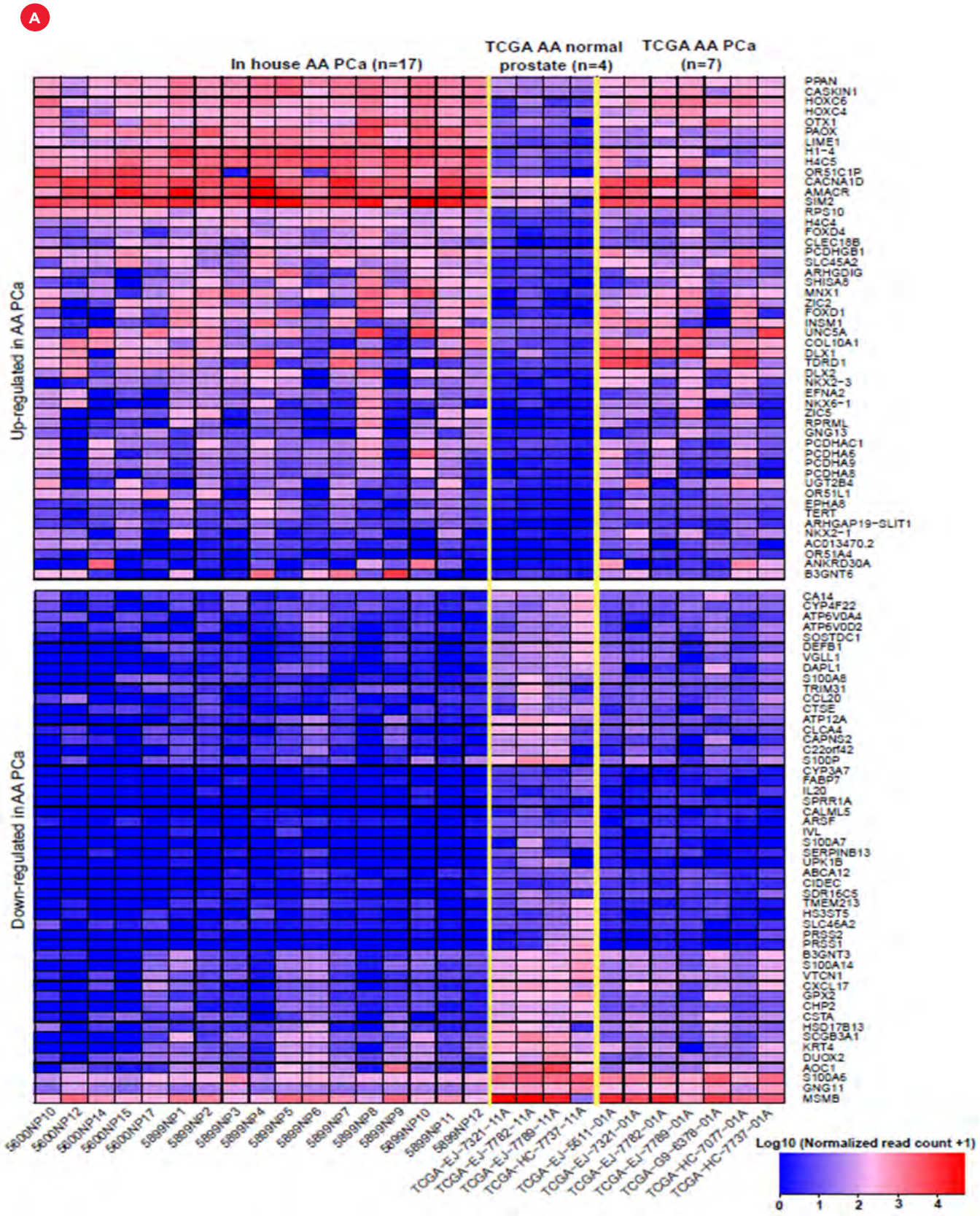


Figure 4. Gene expression level changes across AA PCa FFPE and Frozen samples compared to reference TCGA data. (**A-B**) Volcano plots showed downregulated (blue nodes) and upregulated (red nodes) genes. (**C**) The commonly regulated genes in both AA PCa and TCGA PCa AA groups. TCGA: The Cancer Genome Atlas. PRAD: Prostate Adenocarcinoma.


A
In house AA PCa (n=17)
TCGA AA normal prostate (n=4)
TCGA AA PCa (n=7)
Up-regulated in AA PCa
Down-regulated in AA PCa
PRAN
CASKIN1
HOXC6
HOXC4
OTX1
PAOX
LIME1
H1-4
H4C5
OR51C1P
CACNA1D
AMACR
SIM2
RPS10
H4C4
FOXD4
CLEC18B
PCDHGB1
SLC45A2
ARHGDIG
SHISA8
MNX1
ZIC2
FOXD1
INSM1
UNC5A
COL10A1
DLX1
TDRD1
DLX2
NKX2-3
EFNA2
NKX6-1
ZIC5
RPRML
GNG13
PCDHAC1
PCDHA6
PCDHA9
PCDHA8
UGT2B4
OR51L1
EPHA6
TERT
ARHGAP19-SLIT1
NKX2-1
AC013470.2
OR51A4
ANKRD30A
B3GNT6
CA14
CYP4F22
ATP6V0A4
ATP6V0D2
SOSTDC1
DEFB1
VGLL1
DAPL1
S100A8
TRIM31
CCL20
CTSE
ATP12A
CLCA4
CAPNS2
C22orf42
S100P
CYP3A7
FABP7
IL20
SPRR1A
CALML5
ARSF
IVL
S100A7
SERPINB13
UPK1B
ABCA12
CIDEC
SDR16C5
TMEM213
HS3ST5
SLC46A2
PRSS2
PRSS1
B3GNT3
S100A14
VTCN1
CXCL17
GPX2
CHP2
CSTA
HSD17B13
SCGB3A1
KRT4
DUOX2
AOC1
S100A6
GNG11
MSMB
5600NP10
5600NP12
5600NP14
5600NP15
5600NP17
5899NP1
5899NP2
5899NP3
5899NP4
5899NP5
5899NP6
5899NP7
5899NP8
5899NP9
5899NP10
5899NP11
5899NP12
TCGA-EJ-7321-11A
TCGA-EJ-7782-11A
TCGA-EJ-7789-11A
TCGA-HC-7737-11A
TCGA-EJ-5611-01A
TCGA-EJ-7321-01A
TCGA-EJ-7782-01A
TCGA-EJ-7789-01A
TCGA-G9-6378-01A
TCGA-HC-7077-01A
TCGA-HC-7737-01A
Log10 (Normalized read count +1)
0
1
2
3
4

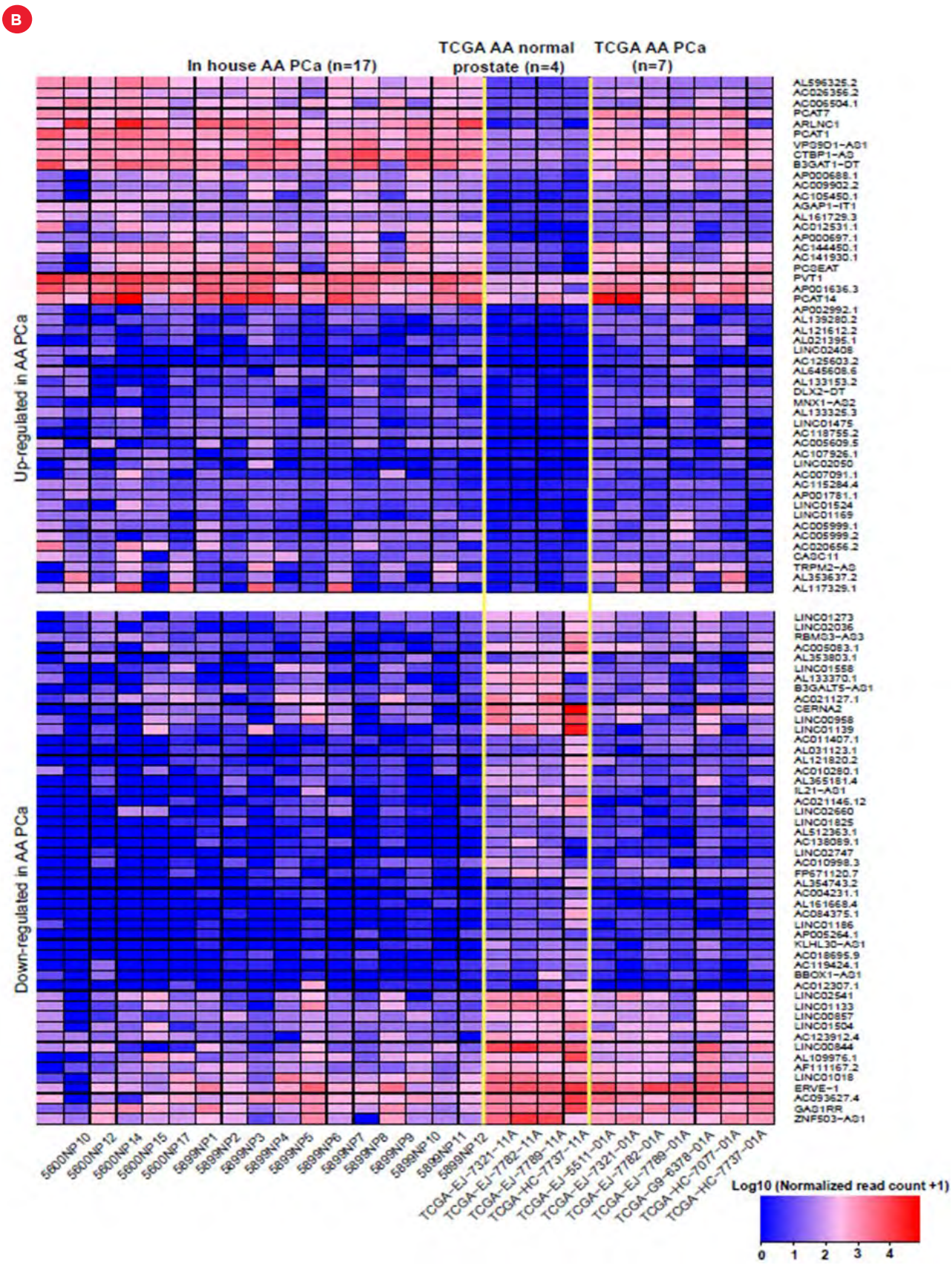


Figure 5. Heatmap plots (**A** and **B**) showing differentially-expressed gene signatures. X-axis: sample IDs starting with "56" refer to PCa AA FFPE samples, "58" refer to PCa AA frozen samples, and samples with "TGCA" are reference samples. The color-coded bars indicate expression values.

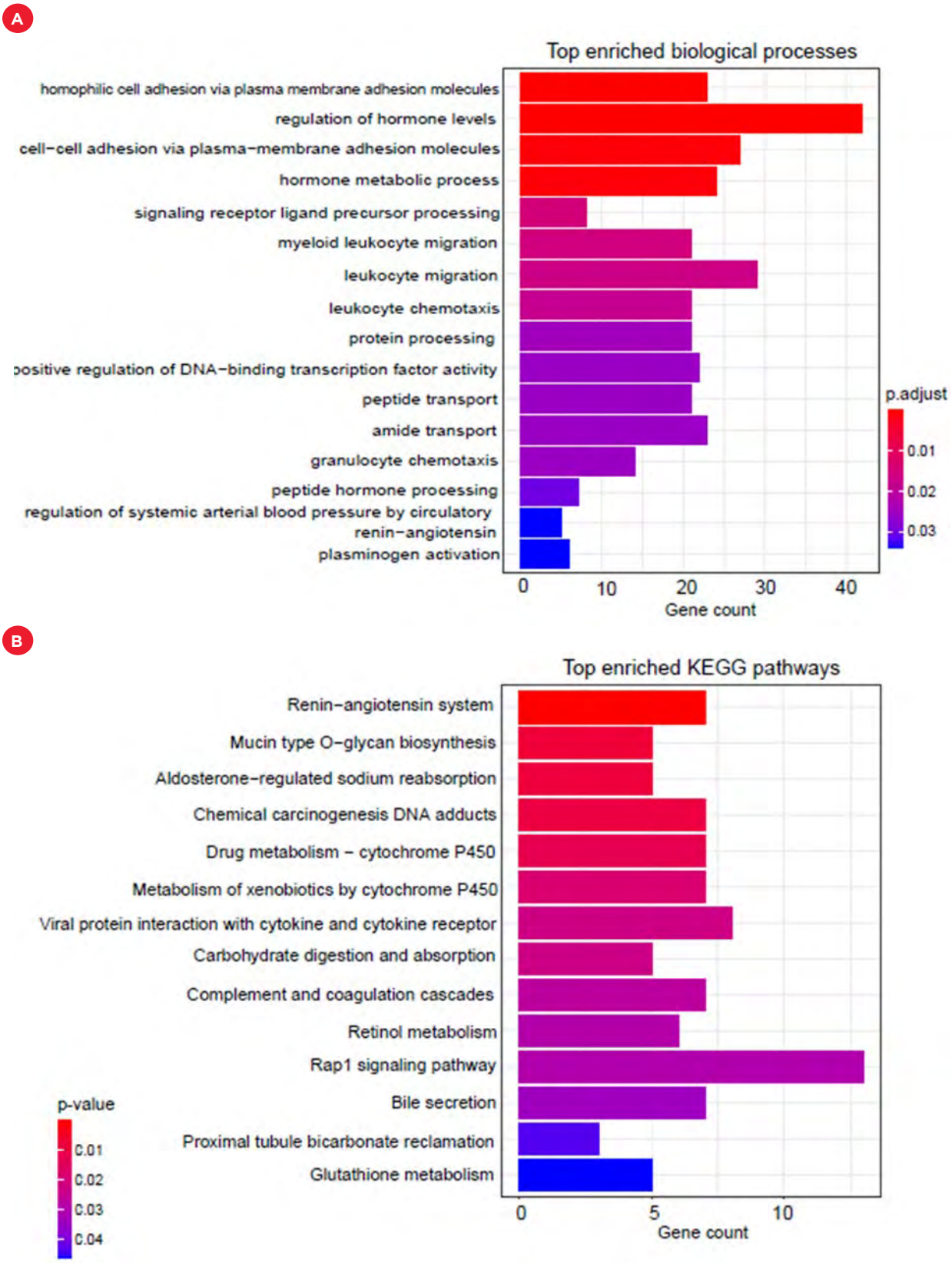


Figure 6. Functional enrichment and pathway analysis of differentially expressed genes in PCa AA. Topmost significantly enriched biological processes (**A**) and KEGG pathways (**B**). KEGG: Kyoto Encyclopedia of Genes and Genomes.

References

1. Strand SH, Orntoft TF, Sorensen KD. Prognostic DNA methylation markers for prostate cancer. Int J Mol Sci. 2014 Sep 18;15(9):16544-76. doi: 10.3390/ijms150916544. PMID: 25238417; PMCID: PMC4200823.

2. Badal S, Aiken W, Morrison B, Valentine H, Bryan S, Gachii A, Ragin C. Disparities in prostate cancer incidence and mortality rates: Solvable or not? Prostate. 2020 Jan;80(1):3-16. doi: 10.1002/pros.23923. Epub 2019 Nov 8. PMID: 31702061; PMCID: PMC8378246.

3. Scaravilli M, Koivukoski S, Latonen L. Androgen-Driven Fusion Genes and Chimeric Transcripts in Prostate Cancer. Front Cell Dev Biol. 2021 Feb 9;9:623809. doi: 10.3389/fcell.2021.623809. PMID: 33634124; PMCID: PMC7900491.

4. Zhang W, Zhang K. Quantifying the Contributions of Environmental Factors to Prostate Cancer and Detecting Risk-Related Diet Metrics and Racial Disparities. Cancer Inform. 2023 Apr 27;22:11769351231168006. doi: 10.1177/11769351231168006. PMID: 37139178; PMCID: PMC10150431.

5. Munchel S, Hoang Y, Zhao Y, Cottrell J, Klotzle B, Godwin AK, Koestler D, Beyerlein P, Fan JB, Bibikova M, Chien J. Targeted or whole genome sequencing of formalin fixed tissue samples: potential applications in cancer genomics. Oncotarget. 2015 Sep 22;6(28):25943-61. doi: 10.18632/oncotarget.4671. PMID: 26305677; PMCID: PMC4694877.

6. Lam SY, Ioannou A, Konstanti P, Visseren T, Doukas M, Peppelenbosch MP, Belzer C, Fuhler GM. Technical challenges regarding the use of formalin-fixed paraffin embedded (FFPE) tissue specimens for the detection of bacterial alterations in colorectal cancer. BMC Microbiol. 2021 Oct 29;21(1):297. doi: 10.1186/s12866-021-02359-z. PMID: 34715774; PMCID: PMC8555202.

7. Dobin A, Gingeras TR. Mapping RNA-seq Reads with STAR. Curr Protoc Bioinformatics. 2015 Sep 3;51:11.14.1-11.14.19. doi: 10.1002/0471250953.bi1114s51. PMID: 26334920; PMCID: PMC4631051.

8. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013 Jan 1;29(1):15-21. doi: 10.1093/bioinformatics/bts635. Epub 2012 Oct 25. PMID: 23104886; PMCID: PMC3530905.

9. Putri GH, Anders S, Pyl PT, Pimanda JE, Zanini F. Analysing high-throughput sequencing data in Python with HTSeq 2.0. Bioinformatics. 2022 May 13;38(10):2943-2945. doi: 10.1093/bioinformatics/btac166. PMID: 35561197; PMCID: PMC9113351.

10. Liu S, Wang Z, Zhu R, Wang F, Cheng Y, Liu Y. Three Differential Expression Analysis Methods for RNA Sequencing: limma, EdgeR, DESeq2. J Vis Exp. 2021 Sep 18;(175). doi: 10.3791/62528. PMID: 34605806.

11. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):550. doi: 10.1186/s13059-014-0550-8. PMID: 25516281; PMCID: PMC4302049.

12. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS. 2012 May;16(5):284-7. doi: 10.1089/omi.2011.0118. Epub 2012 Mar 28. PMID: 22455463; PMCID: PMC3339379.

13. Kremer A, Kremer T, Kristiansen G, Tolkach Y. Where is the limit of prostate cancer biomarker research? Systematic investigation of potential prognostic and diagnostic biomarkers. BMC Urol. 2019 Jun 6;19(1):46. doi: 10.1186/s12894-019-0479-z. PMID: 31170942; PMCID: PMC6554887.

14. Davies G, Jiang WG, Mason MD. Cell-cell adhesion molecules and signaling intermediates and their role in the invasive potential of prostate cancer cells. J Urol. 2000 Mar;163(3):985-92. PMID: 10688036.

15. Almutlaq M, Alamro AA, Alamri HS, Alghamdi AA, Barhoumi T. The Effect of Local Renin Angiotensin System in the Common Types of Cancer. Front Endocrinol (Lausanne). 2021 Sep 3;12:736361. doi: 10.3389/fendo.2021.736361. PMID: 34539580; PMCID: PMC8446618.

16. Bailey CL, Kelly P, Casey PJ. Activation of Rap1 promotes prostate cancer metastasis. Cancer Res. 2009 Jun 15;69(12):4962-8. doi: 10.1158/0008-5472.CAN-08-4269. Epub 2009 May 26. PMID: 19470770; PMCID: PMC4195850.

For Beckman Coulter's worldwide of f ice locations and phone numbers, please visit Contact Us at beckman.com 2023-GBL-EN-101315-V4