

Transcriptome and DNA methylation analysis of DKK3 in prostate cancer initiation and progression

Zhenwu Dai¹, Robert M. Kypta^{1,2*}

¹ Department of Surgery and Cancer, Imperial College London, Hammersmith Hospital, London, United Kingdom

² Centre for Cooperative Research in Biosciences, CIC bioGUNE, Derio, Spain

*Corresponding author: r.kypta@imperial.ac.uk

Abstract

Background. Prostate cancer (PCa) is a heterogeneous disease triggered by microenvironmental factors. Features in the PCa tumour microenvironment (TME) include modified extracellular matrix (ECM) and disruption of signalling between stromal and epithelial components. DKK3 is a secreted glycoprotein that promotes stromal remodelling and epithelial morphogenesis. DKK3 is downregulated in PCa and its overexpression leads to tumour cell-specific apoptosis. The diverse functions attributed to DKK3 have been proposed to be mediated through effects on the TME. Here, an informatics approach was used to identify potential functions of DKK3 in the TME by analysing gene expression in epithelial, stromal and tumour cell line models and patient samples.

Methods. The effect DKK3 gene induction in PCa cells and DKK3 gene silencing in RWPE1 prostate epithelial and WPMY1 prostate stromal cells was determined by analysis of differentially expressed genes (DEGs) using over-representative analysis (ORA) and gene set enrichment analysis (GSEA) of gene ontology (GO). Analysis of clinical samples used data from the TGCA-PRAD, MSKCC and GSE97284 cohorts. Single sample GSEA (ssGSEA) scores for ECM-related gene signatures were calculated for the GSE97284 cohort, which includes cancer epithelial and stromal material. Oncogenic and prognostic markers were filtered through Cox regression and analysis of variance (ANOVA). Co-expression analysis was performed for favourable prognostic markers and DKK3. Reduced-representation bisulfite sequencing (RRBS) was used to determine the effects of DKK3 on gene methylation. Differentially methylated genes (DMGs) from RRBS were compared with DEGs.

Results. DKK3 expression in patient samples was negatively associated with tumorigenesis and cancer progression. Analysis of DGEs found enrichment of the GO term “Collagen-containing ECM” in all three models. The ssGSEA score was significantly higher in the DKK3 high group in all cell types. Fourteen genes in the signature were prognostic in the MSKCC and TGCA-PRAD datasets. FBLN1, ANGPT1, DPT and FREM2 were potential protective prognostic markers whose expression strongly correlated with DKK3. In the cell line models, most DEGs did not show corresponding changes in methylation. However, the overlap of hypomethylated, upregulated genes was significant in DKK3-silenced RWPE-1 cells and not in DKK3-silenced WPMY-1 cells.

Conclusion. DKK3 may play an important role in the PCa TME through its effects on ECM gene expression by both prostate epithelial and prostate stromal cells.

Background

The prostate gland consists of epithelial and stromal constituents, embedded in an extracellular matrix (ECM) (1). Structurally, the adult prostate can be divided into the central, transition and peripheral zones (PZ). The most likely precursor of prostate cancer is the formation of prostatic intraepithelial neoplasia lesions, proliferative areas in glandular epithelium that also display severe cytological atypia within prostatic ducts (2). Prostate adenocarcinoma (PRAD) originates from high-grade lesions in the PZ and is the main type of prostate cancer (PCa) (3).

PCa represents a major clinical problem, with around 1.3 million diagnoses and 400,000 deaths worldwide annually (1). Generally, nearly all primary PCa are sensitive to androgens, and therefore treatment of PCa includes the use of androgen deprivation therapies (ADT) (1). However, they can develop resistance to ADT accompanied by metastasis, representing more than 90% of cancer related deaths. The prostate-specific antigen (PSA) blood test is the

most effective method for diagnosis and staging, playing a critical role in PCa management (1). In clinical applications, the PSA test is considered together with the Gleason score (GS), a grading system of PCa patients according to the glandular architecture of the prostate (4).

The tumorigenic features of PCa include germline susceptibility, somatic gene alterations, and microenvironmental factors (5). Integrative genome profiling studies of patient tumours, for example from the Memorial Sloan Kettering Cancer Center (MSKCC) and The Cancer Genome Atlas (TCGA) databases, has identified several specific gene fusions (e.g. androgen-regulated TMPRSS2-ERG fusion), mutations (e.g. SPOP, FOXA1), copy number variations (e.g. ETV1), epigenetic heterogeneity (e.g. IDH1) in primary PCa, and frequent alterations in androgen receptor (AR) pathways in metastatic PCa (6, 7).

In the prostate gland, the ECM facilitates cell adhesion and communication (3). In the tumour microenvironment (TME), cellular phenotypical changes are usually

accompanied by remodelling of the ECM. Early in tumorigenesis, changes in TME include disruption of bidirectional signalling between epithelial cells and stromal components that maintain prostate homeostasis, and transformation of normal fibroblasts into cancer-associated fibroblasts (CAFs) (8). Alterations of the molecular characteristics in the TME also complicate assessment of the disease and implementation of precise treatment (5).

The Dickkopfs (DKKs) form a family of secreted glycoproteins, consisting of DKK1-4 (9). DKK1, DKK2 and DKK4 have been shown to negatively regulate the Wnt signalling pathway but the effect of DKK3 in Wnt signalling is less clear. DKK3 is aberrantly downregulated in many types of cancer, including prostate, lung, breast and stomach, but is also remarkably upregulated in oesophageal adenocarcinoma and oral squamous cell carcinoma (9), suggesting that mechanism of action of DKK3 may be context dependent. Early in 2005, researchers showed that ectopic expression of DKK3 in PRAD induces tumour cell-specific apoptosis (10). In the normal prostate, DKK3 is recognized as a component of the prostate epithelium that is involved in stromal remodelling and blood vessel stabilisation (11). In the prostate microenvironment, DKK3 has been shown to maintain structural integrity by inhibiting TGF- β /Smad signalling (12). *In vitro* studies indicate that DKK3 silencing promotes the expression of TGF- β -induced protein (TGFB1) in prostate stromal and epithelial cells, while it has opposite effects on extracellular matrix protein-1 (ECM-1), suggesting that the effects of DKK3 on the ECM may depend on the cell type that expresses it (12).

In this study, we hypothesized that DKK3 is not the passenger, but a key regulator of TME during PCa initiation and progression. We integrated the expression profiles of PRAD bulk-seq and different micro-dissected tissue samples from public databases. In addition, enrichment, correlation and prognostic analysis of our transcriptome profiling data from cell lines in which DKK3 was induced in prostate cancer and silenced in prostate epithelial and stromal cells was conducted to identify the potential molecular mechanisms of action of DKK3 in the TME.

Material and Methods

The study workflow is shown in Figure 1.

Data collection

MSKCC gene expression profiling data pre-stored in R package prostateCancerTaylor (v1.24.0, <https://bioconductor.org/packages/release/data/experiment/html/prostateCancerTaylor.html>) available on Gene Expression Omnibus database (GEO, www.ncbi.nlm.nih.gov/geo/ GSE21032), included 138 primary or metastatic prostate cancer and 29 matched non-cancerous cases (6). RNA-seq datasets, curated survival data and clinicopathological features of TCGA-PRAD were obtained through UCSC Xena (<https://xenabrowser.net>), comprised 494 primary or metastatic prostate cancer and 56 matched non-cancerous samples. Gene expression profiling of laser captured micro-dissected prostate specimens were accessed using GEO accession code GSE97284, together with the GS and

regions of interest (ROI), including benign epithelium (B), PIN (P), tumour (T) and stroma adjacent to benign epithelium (sB), prostatic intraepithelial neoplasia (sP) and tumour (sT) (3).

Data pre-processing

For each ensemble gene id in the TCGA-PRAD cohort, matched gene symbols and gene lengths were queried through R package biomaRt (v2.50.3, <https://cran.r-project.org/web/packages/biomartr/>). Expression values were transferred from log2(counts +1) to transcripts per million (TPM) and the data matrix normalized by genes. As the MSKCC and GSE97284 datasets are from microarrays, normalization was conducted directly by each probe. The probe with highest DKK3 expression intensity was kept, while other duplicate probes were removed. For other gene symbols with duplicate probes, only the first result in the datasheet was kept for analysis. A Multidimensional scaling (MDS) plot for GSE97284 samples was made using plotMDS function in R package limma v3.50.3 (13).

Expression analysis

For the MSKCC, TCGA-PRAD and GSE97284 samples, diagrams of normalized DKK3 expression patterns were displayed through R package ggplot2 (v3.3.6, <https://cran.r-project.org/web/packages/ggplot2/>), with themes customized using ggprism (v1.0.3, <https://cran.r-project.org/web/packages/ggprism/>) and ggbbeeswarm (v0.6.0, <https://cran.r-project.org/web/packages/ggbbeeswarm/>). To identify DKK3 changes during PCa initiation and progression, paired student t-tests were performed in the MSKCC and TCGA-PRAD cohorts, comparing 1. non-cancerous with GS 6, 2. GS 6 and 7 with GS 8 and 9. Analysis of variance (ANOVA) was also implemented across GS. In the GSE97284 samples, paired student t-tests were conducted, comparing GS 6 and GS \geq 8 samples in each ROI, and was also conducted for B and T, and sB and sT.

Cell culture, RNA extraction, cDNA synthesis, and q-RT-PCR

The details of cell culture, RNA extraction, cDNA synthesis, and q-RT-PCR were described previously (12, 14, 15). Briefly, PC3 cells, which are derived from a bone metastasis of a PRAD patient, were stably transfected with the expression vector pRP-Puro-U6 (Cyagen Biosciences) to express guide RNAs (gRNAs) targeting the DKK3 transcriptional start site (TSS) or a scrambled gRNA (gRNA-control) and dCas9-VPR (Addgene) (14). RWPE1 (ATCC) and WPMY1 (ATCC) are derived from epithelial and stromal constituents of the PZ of the same histologically normal adult human prostate, immortalized with human papilloma virus 18 (HPV-18) and SV40-large-T antigen, respectively. These cells were transfected with DKK3 or control shRNA expression plasmids and selected with 1.5 ug/ml puromycin to generate the cell clones named SH6 and NS11 (15) and PSM3 and WSH8 (12). Total RNA was extracted using PureLink RNA Mini Kit (Life technologies), and cDNA was synthesized from total RNA using Quantitect Reverse Transcription Kit (Qiagen). DKK3 expression level was determined by q-PCR using SYBR Green PCR Master Mix (Bio-Rad) and a 7900HT Fast Real-Time PCR thermal cycler (Applied Biosystems).

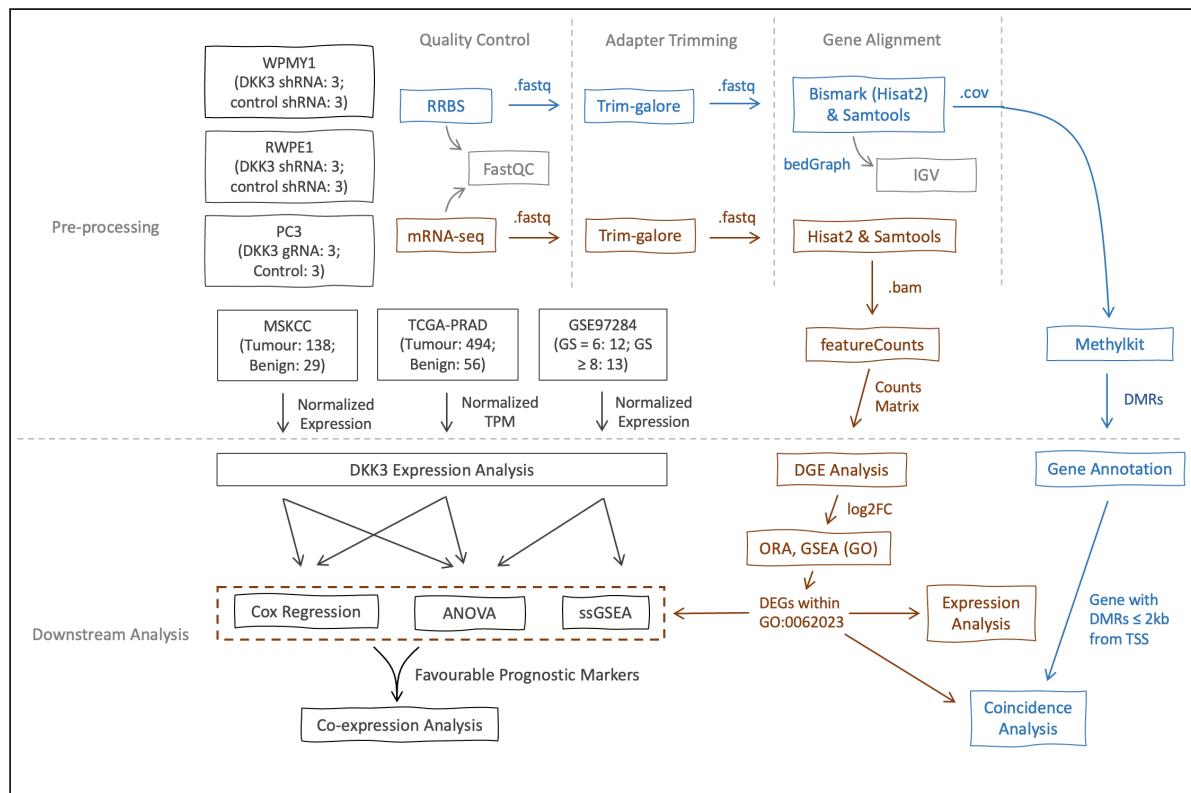


Figure 1. Flowchart of the study. External data sources were retrieved from the TCGA-PRAD and GEO databases. The MSKCC dataset (GSE21032) was pre-stored in R package prostateCancerTaylor, while gene expression in micro-dissected prostate specimens was queried using R package GEOquery with accession code GSE97284. Internal mRNA-seq and reduced-representation bisulfite sequencing (RRBS) were conducted using RNA and DNA from WPMY1, RWPE1 and PC3 cells. Brown and blue route indicates the analysis of expression profiling and methylation calls, respectively. Three replicates of treated and untreated cell clones were used for library preparation and sequencing. Sequencing reads were saved in fastq format. Pre-processing for mRNA-seq and RRBS included quality control (FastQC), adaptor trimming (Trim-galore) and gene alignment (mRNA-seq: Hisat2; RRBS: Bismark with Hisat2). Samtools for mRNA-seq processed input from Hisat2 and created sorted bam files for aggregation of counting gene reads using featureCounts, while bedGraph and coverage file were created for visualization through IGV and filtering differentially methylated regions (DMRs) using Methylkit. Downstream analysis for mRNA-seq includes differential gene expression (DGE) analysis, over-representation analysis (ORA), gene set enrichment analysis (GSEA) and visualizing the expression profile of ECM-related differentially expressed genes (DEGs) from the GO:0062023 gene set (collagen-containing extracellular matrix) using heatmaps. These ECM-related DEGs were also analyzed in three external datasets. Analysis included single sample ssgSEA, Cox regression and analysis of variance (ANOVA). Four favourable prognostic markers filtered through Cox regression and ANOVA were used for co-expression analysis. DMRs from Methylkit were annotated using R package genomat. Coincidence analysis integrated DEGs and genes with DMRs that were located within 2 kb of the TSS.

High-throughput sequencing and pre-processing

The details of High-throughput sequencing and their pre-analysis, including mRNA-seq and reduced-representation bisulfite sequencing (RRBS), are in the Supplementary methods.

Differential gene expression analysis

Gene count matrix were inputted into R package DESeq2 (v1.34.0, <https://bioconductor.org/packages/release/bioc/html/DESeq2.html>) for graphing principal component analysis (PCA) plots. Differential gene expression analysis was conducted using edgeR v3.36.0 (16). Matched gene symbols and gene lengths were queried through biomaRt. EdgeR models count data using a negative binomial distribution and uses an empirical Bayes procedure to moderate the degree of overdispersion across genes. Prior to analysis, low-expressed genes (raw counts < 10 reads) were removed. The significance of overdispersion Genes with p-value < 0.05 (adjusted through Benjamini-Hochberg (BH) method) and absolute log2 fold-change (log2FC) > 1 were considered to be significantly differentially expressed genes (DEGs). Volcano plots were made for all reference genes

involved in Differential gene expression analysis. Four Venn maps were made comparing up- and down-regulated genes in RWPE1 and WPMY1 cells, and their overlaps were evaluated using hypergeometric test.

Over representative analysis (ORA) and gene set enrichment analysis (GSEA)

ORA was performed using matched Entrez gene IDs for DEGs (combining up- and down-regulated genes) through R package clusterProfiler v4.2.2 (17), which detects over-represented genes in the Gene Ontology (GO) database, including molecular function (MF), biological process (BP) and cellular component (CC). Dot plots were graphed through Enrichplot v.1.11.2 (<https://github.com/YuLab-SMU/enrichplot/>). GSEA was also conducted using Entrez gene ID and log2FC from edgeR output and displaying enrichment results through Enrichplot, with enrichment scores and adjusted p values through BH correction (18).

Visualization of ECM-related DEGs

The expression profiles of the genes enriched in GO:0062023 in at least one cell line were collected and normalized. A heatmap was graphed using R package

ComplexHeatmap (v2.10.0, www.bioconductor.org/packages/release/bioc/html/ComplexHeatmap.html), with annotation of up- or down-regulation. Dot plots below the heatmaps show the average value for each column.

Calculation of single-sample GSEA (ssGSEA) signature scores

ssGSEA scores of the collected gene sets were computed using R package GSVA (v1.42.0, www.bioconductor.org/packages/release/bioc/html/GSVA.html). The ssGSEA algorithm assigns an enrichment score for each sample that characterizes up- or down-regulation of gene sets compared to the remaining genes measured in the sample (3). Student's *t*-test was performed between cases with low and high DKK3 expression levels, and also performed between B and T, sB and sT.

Identification of oncogenic / prognostic signatures

As more than 90% of patients in the TCGA-PRAD cohort were still alive at last census, recurrence-free interval (RFI) in the MSKCC cohort and progression-free interval in the TCGA-PRAD cohort were used for univariable Cox regression. In the MSKCC cohort, biochemical recurrence was defined as PSA ≥ 0.2 ng/ml on two occasions. Cox regression was implemented through R package survival (v3.4-0, <https://cran.r-project.org/web/packages/survival/>). Hazard ratios (HRs) and 95% confidence intervals (CI) calculated from Cox regression were displayed through forest plots using R package forestplot (v2.0.1, <https://cran.r-project.org/web/packages/forestplot/>). ANOVA was also performed for filtering genes that changed across GS. A Venn diagram was plotted to visualize the overlap between the gene signatures selected from ANOVA and Cox Regression.

Integrative analysis of favourable prognostic markers

Four key signatures with HR < 1 were selected for further analysis. Diagrams of normalized gene expression was made using MSKCC cohort and GSE97284 samples in the way that mention above. Co-expression profiling for key genes and DKK3 was visualized through correlation matrix using GGally (v2.1.2, <https://cran.r-project.org/web/packages/GGally/>). To determine if selected DEGs were altered at the epigenetic level, DKK3, FBLN1, ANGPT1, DPT and FREM2 were uploaded to the UALCAN server to determine their promoter methylation status (indicated by the beta-values) in cancer ($n = 502$) and matched normal cases ($n = 50$) in the TCGA-PRAD cohort.

Identification of the differentially-methylated bases/regions (DMRs)

Pre-analysis of RRBS results for the three cell lines (saved in .cov.gz) were read to R package methylKit v1.20.0 (19). A dendrogram for 18 samples was made based on hierarchical clustering using ward.D method. Low read (< 10) and high read (bases with more than 99.9th percentile coverage) coverages in each sample were excluded. Differential methylation was detected using "calculateDiffMeth" function with overdispersion = "MN" and test = "Chisq", which models the numbers of methylated cytosines in binomial distribution and corrects for overdispersion with Chi-squared test. R package genomation (v1.28.0, <https://bioconductor.org/packages/release/bioc/html/genomation.html>) was used for annotation and visualization of methylation data.

Gene annotation with GRCh38 Reference Sequence (RefSeq) BED file (downloaded from https://sourceforge.net/projects/rseqc/files/BED/Human_Homo_sapiens_merge_transcripts/) and CpG island annotation with BED file (from <http://genome.ucsc.edu/cgi-bin/hgTables>) were conducted and visualized using pie charts. The distance to the transcription start site (TSS) was acquired using the "getAssociationWithTSS" function. biomaRt was then used for calling matched gene symbols with Refseq sequence numbers (e.g. NM_199260). Refseq numbers without matched gene symbols were discarded in the subsequent analysis.

Integrative analysis of the methylation information

Genes with at least one DMR within 2 kilobase (kb) upstream or downstream of the TSS were considered differentially methylated. Hypergeometric test was analyzed comparing hypermethylated and downregulated genes and comparing hypomethylated and upregulated genes. Methylation information in the promoter regions of FBLN1, ANGPT1 and DPT was visualized through IGV. Bead diagrams were used to represent the methylation profiles in these regions.

Statistical analysis

All pre-analysis of the mRNA-seq and RRBS was implemented on zsh 5.8.1 (x86_64-apple-darwin21.0). Statistical analysis based on R version 4.1.3 (www.R-project.org/) was run on RStudio (www.rstudio.com/). Datasets were mainly manipulated through R package dplyr (v1.0.9, <http://cran.r-project.org/package=dplyr/>). Student's *t*-tests were conducted using R package rstatix (v0.7.0, <https://cran.r-project.org/web/packages/rstatix/>) and corrected with BH method. ANOVA and hypergeometric tests were conducted through R basic package stats v4.1.2, with p-values < 0.05 considered statistically significant throughout the study.

Results

DKK3 expression is negatively associated with tumorigenesis and cancer progression

Using tumour stage plots of the MSKCC cohort data, we found a reduction in DKK3 gene expression with increasing GS (ANOVA $p < 0.001$) (Figure 2A). Also illustrated, is decline in expression when comparing non-cancer and GS 6 groups (t-test $p < 0.001$) and comparing GS 6-7 and GS 8-9 (t-test $p < 0.001$). The MDS plot of GSE97284 samples

showed a similar expression pattern in epithelial (B, P and T) and stromal (sB, sP, sT) constituents (Figure S1), while there was clear separation between the two constituents. The expression diagram of tissue specimens also revealed that the level of DKK3 expression was lower in T and sT, compared to B and sB (t -test $p < 0.05$). Interestingly, DKK3 expression was found to be higher in GS ≥ 8 samples compared to GS 6 samples in sT (t -test $p < 0.05$) (Figure 2B).

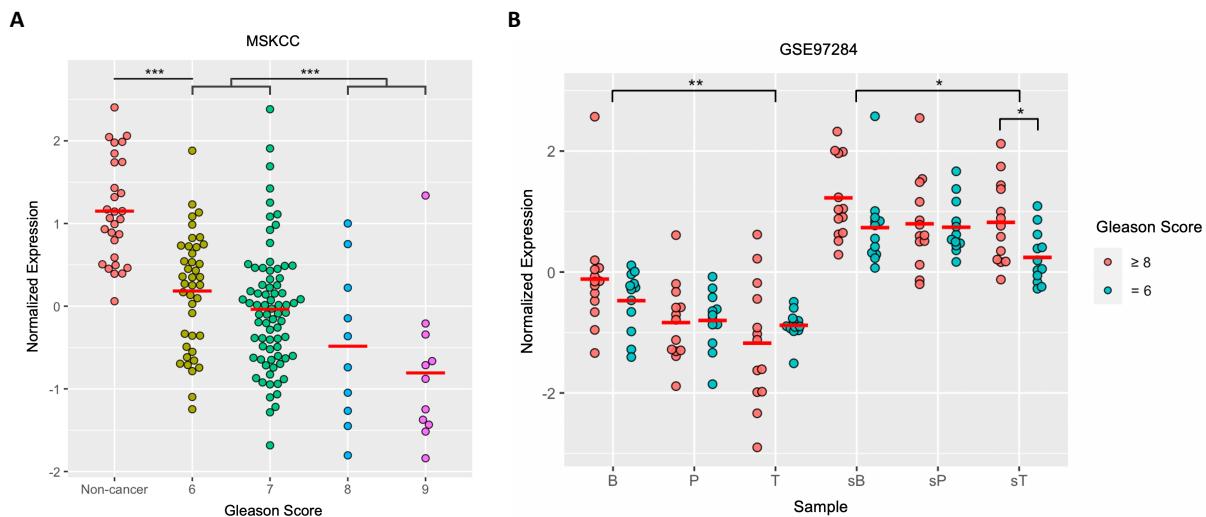


Figure 2. Analysis of the relationship of DKK3 expression, Gleason score (GS) and tissue sample type. (A) Progressive decrease in DKK3 gene expression with increasing GS ($n = 138$). T-test was performed between non-cancer and GS 6 group, and between GS 6-7 and GS 8-9 group. Red lines highlight the average expression value. (B) Comparison of DKK3 expression levels in GS 6 samples and GS 8-10 samples ($n = 150$). T-test was performed comparing low and high GS samples and comparing benign epithelium (B) with tumour (T) and benign stroma (sB) with tumour stroma (sT). $p < 0.05$ indicates statistical significance; only t-tests with statistical significance are shown. P, prostatic intraepithelial neoplasia; sP, stroma adjacent to prostatic intraepithelial neoplasia.

The GO term “Collagen-related extracellular matrix” predominantly enriched in DEGs from three cell lines

Volcano plots of gene expression in PC3, RWPE1 and WPMY1 cells highlighted the distribution of DEGs (Figure S2). PCA plots of the expression profiles of PC3, RWPE1 and WPMY1 cells showed the cell groups of DKK3-induced or DKK3-silenced groups separated from control groups (Figure 3A). Figure 3B-D shows the most significantly involved GO terms in DEG lists from three cell lines. The term “collagen-containing extracellular matrix” (GO:0062023) in cellular component was highlighted in ORA of PC3 ($p.adjust = 0.002$), RWPE1 ($p.adjust < 0.001$) and WPMY1 ($p.adjust < 0.001$) cells. The result of GSEA in log2FC from DGE analysis revealed gene under-expression in DKK3-silenced WPMY1 cells (enrichment score = -0.452, $p.adjust = 0.047$) and overexpression in DKK3-silenced RWPE1 cells (enrichment score = 0.655, $p.adjust < 0.001$), while these genes were no enrichment in DKK3-induced

PC3 cells (enrichment score = 0.280, $p.adjust = 0.440$) (Figure 3E).

Expression profile of ECM-related genes in three cell lines

All genes that were found in GO:0062023 in at least one cell line were collected for further analysis. The overlapping genes among the cell lines are shown in Figure S3. Heatmaps of normalized gene expression in PC3, RWPE1 and WPMY1 cells highlights their relative expression levels (Figure 4). Generally, WPMY1 cells expressed more ECM-related genes compared to RWPE1 and PC3 cells and are reduced upon DKK3 silencing. By comparison, RWPE1 cells showed higher expression of the genes in the DKK3-silenced group, while in PC3 cells they were almost unchanged, overall. Interestingly, many of the DEGs that were significantly upregulated in RWPE1 cells were downregulated in WPMY1 cells.

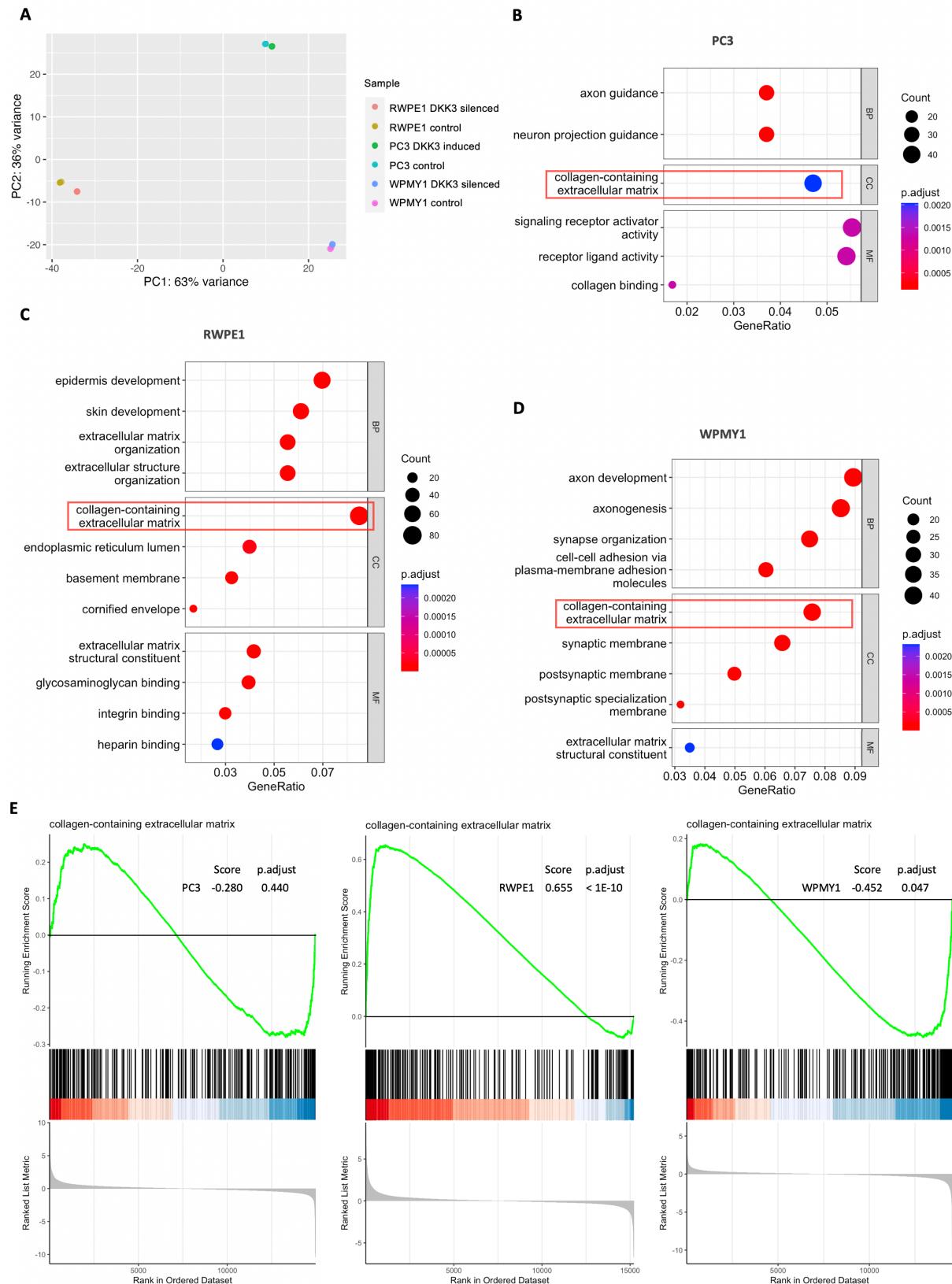


Figure 3. Functional analysis of DKK3 alterations in WPMY1, RWPE1 and PC3 cells. (A) Principal component analysis (PCA) illustrating genetic relatedness among 18 cell clones under 6 conditions. (B-D) Dot plots visualizing the top listed gene ontology (GO) terms by over representation analysis (ORA) using differentially expressed gene (DEG) lists from PC3 cells (B), RWPE1 cells (C) and WPMY1 cells (D). Dot size, distance from left and colours reflect overlapped gene counts, ratio of total DEGs and significance of enrichment (adjusted p-values). The GO term collagen-containing extracellular matrix (GO:0062023) is highlighted in red. (E) GSEA enrichment plots for the three cell line models. The final enrichment score (based on the maximum of absolute value) and adjusted p-values are shown in the plots. $p < 0.05$ was considered significant enrichment.

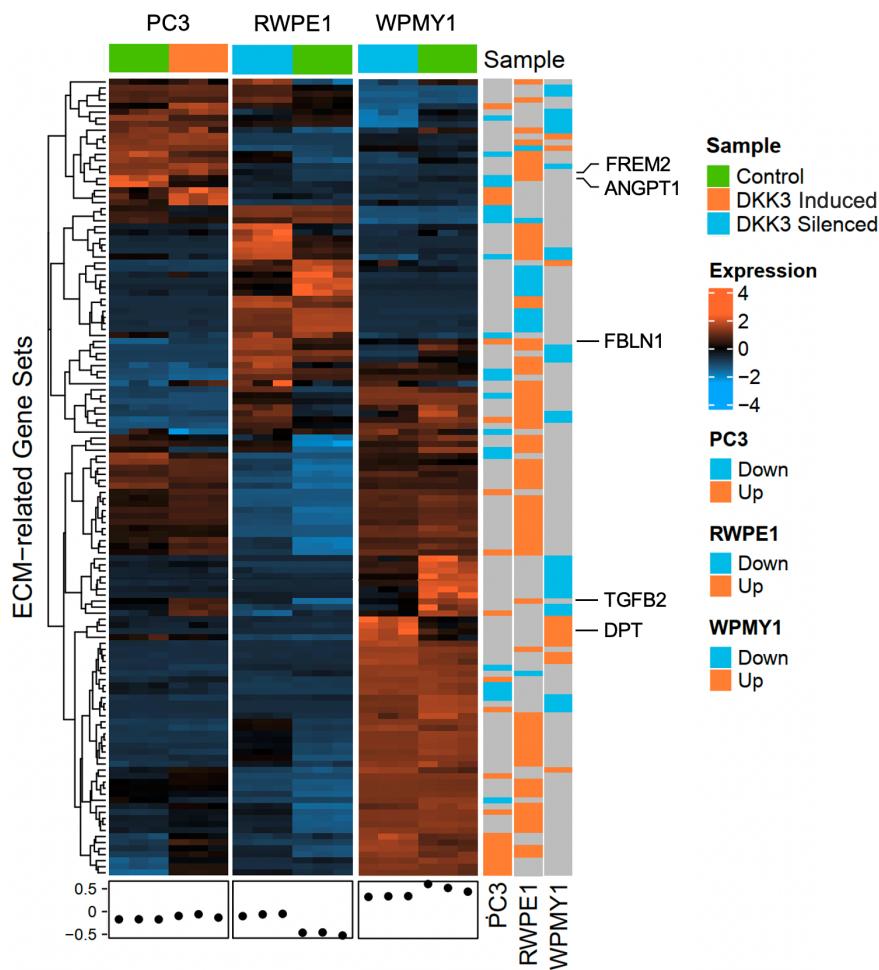


Figure 4. Heatmap of normalized expression profiles of DEGs in PC3, RWPE1 and WPMY1 cells. Bars on the right indicated up- and down-regulation of DEGs in the three cell line models. Dot plots below the heatmap showed the average value of each column (i.e. one cell replicate). Annotations on the right highlight the five genes selected for further analysis and discussion.

Expression profiles of ECM-related genes in microdissected prostate samples

To further identify the expression correlations of DKK3 with the ECM-related genes, ssGSEA enrichment scores of ECM-related gene signatures were calculated using the GSE97284 dataset (Figure 5). Our results showed that the ssGSEA enrichment scores in stromal specimens (sB, sP and sT) were generally greater than zero, while epithelial components (B, P and T) were around or below zero, except for one outlier in B. Importantly, the ssGSEA score in the DKK3 high group was significantly higher than that in the DKK3 low group in B ($p = 0.004$), P ($p = 0.044$), T ($p = 0.006$), sB ($p < 0.001$), sP ($p = 0.003$) and sT ($p < 0.001$). B and sB also have a higher ssGSEA scores than T and sT, respectively ($p = 0.005$ and $p < 0.001$).

Selecting genes signatures involved in PCa oncogenesis and progression

Univariable Cox regression and ANOVA of ECM-related genes was conducted both for samples from the TCGA-PRAD and MSKCC databases. A total of fourteen genes were found to be significant in all analyses (Figure 6A). These genes were considered as oncogenic and prognostic signatures. The results of Cox regression from the TCGA-PRAD and MSKCC cohorts are displayed in Figure 6B. Of these genes, FBLN1, ANGPT1, DPT and FREM2 were shown to be associated with favourable prognosis. The expression profiles of FBLN1, ANGPT1, DPT and FREM2 are shown in Figure 6C. The expression levels were significantly lower with increasing GS (ANOVA $p < 0.001$ in FBLN1, ANGPT1 and DPT and $p = 0.001$ in FREM2) and associated with PCa progression ($p < 0.01$ in all genes). ANGPT1 and FREM2 expression were also associated with PCa initiation ($p < 0.001$). Expression analysis of the GSE97284 dataset found higher expression in stromal than in epithelial components (Figure S4).

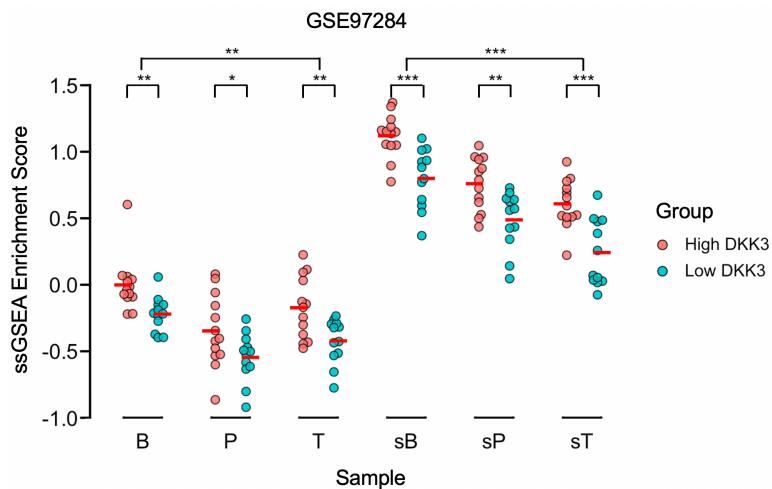


Figure 5. ssGSEA result of ECM-related DEGs in benign epithelium (B), prostatic intraepithelial neoplasia (P), tumour (T) and their adjacent stroma (sB, sP, sT) in GSE97284 dataset classified by different DKK3 expression. There were 13 high DKK3 and 12 low DKK3 in each tissue subgroup.

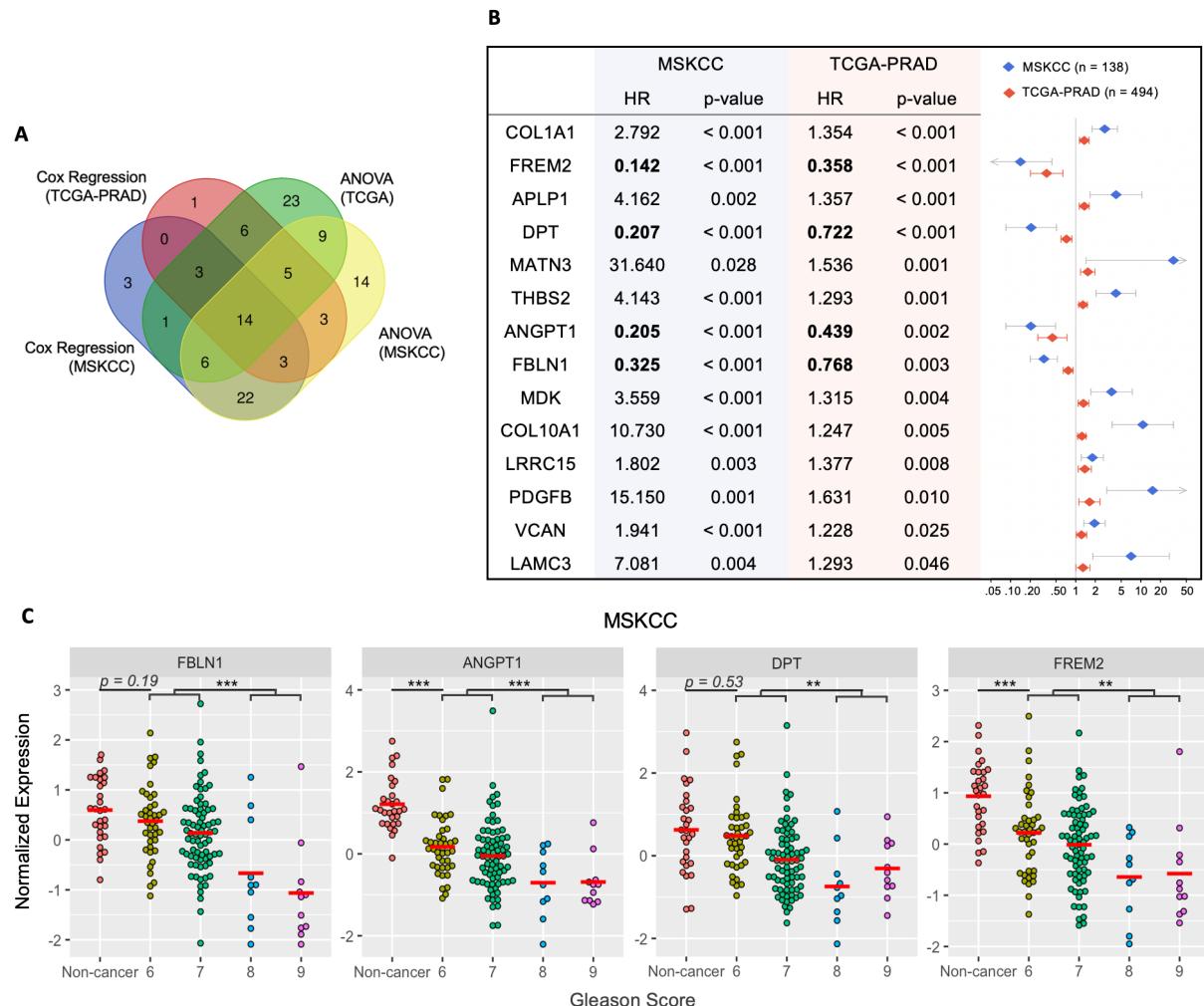


Figure 6. Selection of prognostic and oncogenic markers in cancer progression. (A) Venn diagram of significant gene signatures selected using Cox regression and analysis of variance (ANOVA) in the TCGA-PRAD and MSKCC datasets. A total of 14 genes overlapped in all gene lists. (B) Forest plots of the overlapping genes in the MSKCC (blue) and TCGA-PRAD (red) datasets. Diamonds at the midpoints of the lines show the estimated hazard ratio (HR) and the widths of the lines indicate the 95% confidence interval range. P-value prognostic significance is also displayed; $p < 0.05$ is considered statistically significant. (C) Normalized gene expression levels of FBLN1, ANGPT1, DPT and FREM2 in non-cancer and different GS cancer cases ($n = 138$). T-test was performed comparing non-cancer and GS 6 groups and GS 6-7 and GS 8-9 groups. Red lines show the average expression values.

Co-expression analysis of the four favourable prognostic genes and DKK3 reveals strongly positive correlations

We implemented a correlation analysis of the expression of any two of the four identified genes plus DKK3. To reduce the influence of GS, correlation matrices were carried out under GS classification in MSKCC, TCGA-PRAD cohort and

GSE97284 samples. The results showed strong correlations among the five genes in the MSKCC and TCGA-PRAD cohorts ($p < 0.001$), and is still significant after GS classification ($p < 0.001$) (Figure 7, Figure S5A), with the similar correlation coefficient. However, gene correlations were less significant in the microdissected GSE97284 specimens (Figure S5B-D).

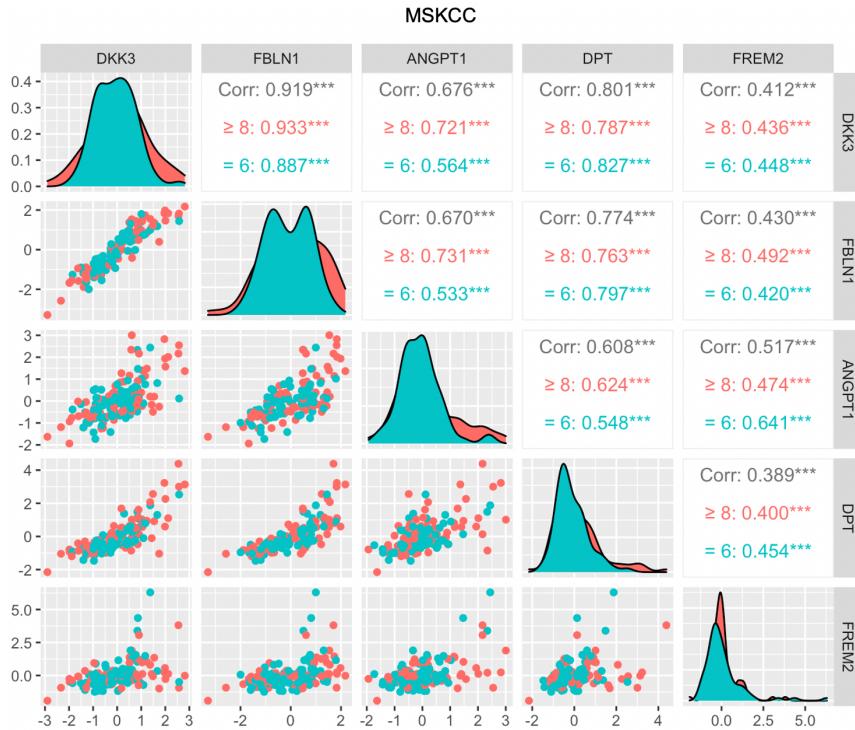


Figure 7. Co-expression analysis of favourable prognostic markers and DKK3. Correlation matrices revealed relationships between DKK3 and the four selected ECM-related genes in the MSKCC dataset. Plots on the diagonal show the distribution of each gene following a normalized expression and stratified by Gleason Score (GS) (red: GS = 6; green: GS ≥ 8). Scatterplots below the diagonal show gene correlations. The Pearson correlation coefficients for GS = 6 (red), GS ≥ 8 (green) and combined (grey) are displayed above the diagonals, of which *, **, *** indicate $p < 0.05, 0.01, 0.001$, respectively. Corr, Pearson correlation coefficient.

Methylation analysis revealed potential regulation role of DKK3 in DNA methylation

The promoter methylation level of DKK3, FBLN1, ANGPT1, DPT and FREM2 in TCGA-PRAD cohort were acquired from UALCAN server. The result showed that promoters of four genes were hypermethylated in cancer samples compared to matched non-cancer samples (t-test $p < 0.001$) except DPT ($p = 0.996$) (Figure S6).

We furtherly analysed the methylation profile of treated cell lines through RRBS. After pre-analysis, we clustered all cell samples, which suggesting a different expression profile in WPMY1 samples (Figure S7A). Further PCA diagram in WPMY cells showed a distinct methylation profile in (Figure S7B). Number of DMR was calculated in each cell lines (Figure 8A). Gene annotation and CpG island annotation file allowed us to matched DMR onto a specific region. Pie charts revealed the percentage of genes that were overlapped with Gene promoter, intron, exon or on CpG island and CpG shore (Figure S7C-E).

After matching Refseq sequences with gene symbols, we extracted DMRs within 2 kb distance of the TSS. Venn diagrams revealed that hypermethylated genes from PC3, RWPE1 and WPMY1 cells were over-represented in the downregulated gene sets (hypergeometric test $p = 0.022$, $p = 0.002$, $p = 0.009$), and hypomethylated genes from PC3 and RWPE1 cells were over-presented in the upregulated gene sets (hypergeometric test $p < 0.001$), but this was not the case in WPMY1 cells (hypergeometric test $p = 0.190$) (Figure 8B).

The methylation patterns of the genomic regions adjacent to FBLN1, ANGPT1, DPT, FREM2 and TGFB2 were displayed through IGV. No DMRs were found within 2 kb of the TSS of FBLN1, ANGPT1, DPT and FREM2 (Figure S8). However, TGFB2, a gene whose expression was found to be increased upon DKK3 silencing in RWPE1 cells, had a distinct methylation profile, including a DMR, in a region close to the TSS (Figure 8C).

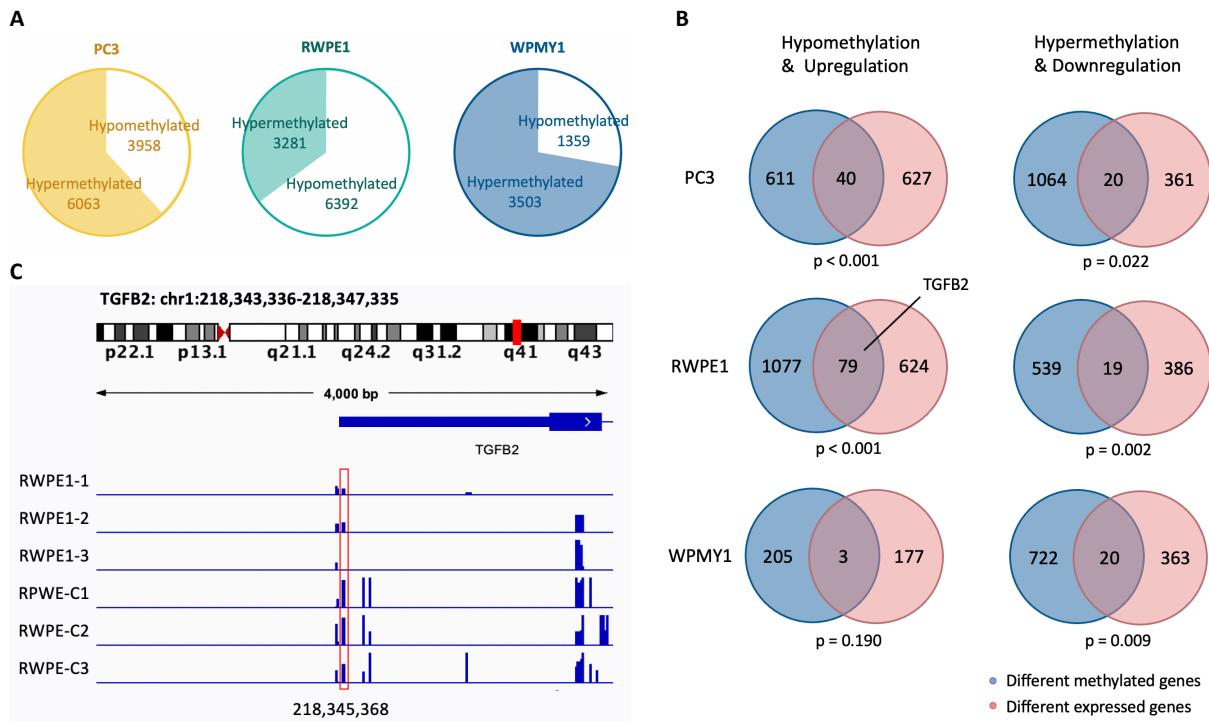


Figure 8. Identification of DNA methylation changes following DKK3 silencing or induction using reduced representation bisulfite sequencing (RRBS). (A) Number of the total differentially methylated cytosines (q -value < 0.01 and methylation difference $> 25\%$) in the three cell line models. (B) Coincidence analysis of differentially methylated gene lists (containing more than one differentially methylated CpG locus within 2 kb of the gene body (inclusive) and differentially expressed gene lists. Enrichment of overlapping genes was calculated using hypergeometric test. $p < 0.05$ indicates statistical significance. (C) Methylation profiles of the region 2 kb upstream and downstream of the TGF β 2 gene TSS in RWPE1 cell clones from Integrative Genomics Viewer (IGV). Each blue bar indicates methylation at one CpG locus, and bar height indicates methylation ratio (methylated CpGs/all CpGs * 100%). The DMR detected by Methylkit algorithms is highlighted in red.

Discussion

To our knowledge, this is the first bioinformatic study exploring the characteristics of DKK3 gene expression and gene promoter methylation in cells of the prostate TME. Through mining public expression profiling data, we noted a reduction in the expression of DKK3 at PCa initiation (comparing benign and tumour samples) and progression (comparing expression with GS). In GSE97284 specimens, the results indicated that DKK3 expression is higher in stromal cells than in epithelial cells. Consistent with previous immunohistochemistry analysis of an Imperial college PCa cohort (12), DKK3 expression was found to be lower in tumour epithelia (T) and tumour stroma (sT) than in their benign counterparts, B and sB, respectively. Unexpectedly, DKK3 gene expression was higher in high GS tumour stroma than in low GS tumour stroma. This might reflect a unique feature of this cohort and so requires confirmation in another dataset. It is interesting to note that many of the genes more highly expressed in stroma adjacent to high-grade cancer in this cohort are genes expressed by the immune system and by osteoblasts, so may not be stromal cells (3), and DKK3 is known to be expressed by osteoblasts, where it has been proposed to induce PCa cell quiescence (20).

In this study, benign epithelium not only serve as epithelial cells interacting with tumour, but also as matched non-cancerous samples in PCa analysis. However, the following analysis suggested gene expression in benign samples (B,

sB) also differed between high and low GS (Figure 5, Figure S4). It is uncertain whether the difference is cause by germline / somatic mutations or affected by the TME. Generally, benign epithelium and stroma from high GS samples are likely to be more spatially close to tumour cells than those from low GS samples, and therefore easier to be affected. A previous study reported that PCa cells can share genomic alterations with adjacent benign tissue (5), and it has been speculated that half of the somatic alterations in a tumour-initiating cell may have taken place before neoplastic transformation, and a subset of these changes are shared with adjacent benign cells (21).

In this study, RWPE1 and WPMY1 were treated as benign epithelium and stromal components. After differential gene expression analysis between control cell clones and DKK3-induced/DKK3-silenced cell lines, ORA in DEGs showed predominant enrichment in “collagen-containing extracellular matrix” in all three cell line models, together with “collagen binding” in PC3 cells, “glycosaminoglycan binding” in RWPE1 cells and “extracellular matrix structural constituents” in WPMY1 cells. These significantly involved GO terms suggest DKK3 plays a role in the regulation of ECM turnover in the microenvironment of prostate and PCa, consistent with recent observations (22). Further analysis using GSEA highlighted the differential distribution of DEGs in the pre-ranked lists, indicating higher expression of genes induced by DKK3 silencing in RWPE1 cells and lower expression in DKK3-silenced WPMY1 cells. Importantly, the majority of collagen-containing ECM-

related DEGs are expressed at very low levels in these cell lines (e.g. average TPM < 0.5 for ANGPT1, DPT and FREM2, not shown). Therefore, an effect of DKK3 induction or silencing on the expression of these genes in the three cell line models might not be detected.

The results from GSEA contrast with those from the ssGSEA analysis of patient samples (Figure 5), which indicates that DKK3 expression is always positively associated with expression of the selected DEGs in epithelial, stromal and tumour constituents. This suggests that induction of these genes may require additional stimulation, such as by cytokines or chemokines (23). For example, pro-inflammatory cytokine interleukin-6 (IL-6) can bind to IL-6R and elicit actions through multiple pathways, including JAK/STAT and PI3K. Analysis of signalling crosstalk between PCa cells and CAFs revealed a mandatory role of carcinoma-derived IL-6 in fibroblast activation and matrix metalloproteinase secretion (24). In our study, the IL6 gene is moderately expressed by PC3 cells and at very low levels in WPMY1 cells (data not shown).

Co-expression analysis of FBLN1, ANGPT1, DPT and FREM2 and DKK3 showed highly similar expression patterns, even after GS stratification. ANGPT1 encodes angiopoietin 1, an endothelial cell marker that affects vessel stability and induces angiogenesis through binding to TIE2 (25). Previous studies found that DKK3 regulates angiopoietin-1 levels in prostate stromal cells (11). However, while ANGPT1 was increased in DKK3-silenced RWPE1 cells and decreased in DKK3-induced PC3 cells, we did not observe changes in WPMY1 stromal cells. While there are no studies linking DKK3 to FBLN1, DPT or FREM2, each of these genes has been implicated in PCa. FBLN1 encodes fibulin 1, a secreted glycoprotein incorporated into fibrillar ECM can induce PCa cell death (26) and limit cell migration and invasion (27). In contrast, DPT, which encodes dermatopontin, a tyrosine-rich acidic matrix protein, promotes the growth of PC3 tumour xenografts (28). Finally, FREM2 gene encodes FRAS1-related extracellular matrix protein 2. FREM2 with FRAS1 and FREM1 form a FRAS1/FREM complex that contributes to embryonic epithelial–mesenchymal integrity (29). In a study of cancer driver genes, FREM2 was among the 10 most recurrently mutated genes in PCa (30). Of four genes, DPT is the only gene that may induce PCa progression, but this still needs experimental validation.

Methylation might be the inner driving force of the prostate cancer (31). We previously observed that the TGF- β treatment of PC3 cells increased the expression of the DNA methyltransferase DNMT1 and that this was reduced by induction of DKK3 (14), suggesting some of the effects of DKK3 might be mediated through changes in gene promoter methylation. Analysis of the RRBS data from the cell line models found that alterations in DKK3 gene expression clearly resulted in altered DNA promoter methylation, and that the alterations largely differed in three cell line models. Moreover, there is a significant overlap of the genes that are differentially expressed and those that are differentially methylated. A deep analysis of the data was not possible for reasons of time. However, some ECM genes were examined. Among these, TGFB2 is a good example of an ECM-related DEG that was found to be remarkably demethylated at the TSS site in DKK3-silenced RWPE1 cells. TGFB2 encodes a secreted ligand of the TGF-

superfamily and triggers TGFB2/Smad/METTL3 pathway (32). Considering that DKK3 exerts some of its biological effects through TGF β /Smad signalling (12, 14), this effect of DKK3 on TGFB2 gene promoter methylation may be of relevance.

However, the Venn diagrams also indicate that differential methylation accounts for a minority of the DKK3-dependent changes in gene expression. This may be the case for the DEGs FBLN1 and ANGPT1, whose promoters are methylated at several sites but there are no clear effects of DKK3. Also, no DMRs were found for FREM2, despite a clearly lower methylation level in DKK3-silenced WPMY1 cells. More advanced methods have been developed for combining measurements of individual CpGs across large genomic regions, such as Rnbeads (33). The future use of such methods may allow for a deeper analysis and interpretation of our data. Moreover, it should also be noted that there were evident differences in methylation within the cell replicates that hamper interpretation of these findings. Multiple studies have shown that minor experimental variation can lead to a significant impact on epigenome outcome measures and data interpretation (34).

In addition, many of the effects of DKK3 on gene expression may be independent of effects on gene promoter methylation. For example, DKK3 may affect the expression of a transcription factor that drives the expression of many other genes. Indeed, we observed that DKK3 silencing altered the expression of the transcription factor GHRL3 in one of the cell line models and many of the genes affected by DKK3 in that setting are predicted GHRL3 targets (not shown).

The results from this study need to be confirmed and validated using *in vitro* (e.g. co-culture) and *in vivo* studies (e.g. DKK3 null mice). Further mining of transcriptome and epigenome data and analysis of the different groups classified by DKK3 expression levels may also be helpful. As DKK3 expression and gene methylation levels are likely to change with GS, the impact of GS may be interpreted as the impact of the effect of DKK3 in a particular GS setting. Further insight into the impact of DKK3 in PCa requires the use of studies with larger cohorts in order to exclude potential interference of GS status (i.e. using more cases with same GS).

Conclusions

Through public expression datasets, we confirmed that DKK3 expression is negatively associated with tumorigenesis and cancer progression. The expression of multiple ECM-related genes is affected by DKK3 and correlate with DKK3, suggesting DKK3 may play a coordinating role in the TME. Looking forward, more *in vitro* and *in vivo* studies will be required to validate these findings.

Acknowledgements

I would like to thank Dr Robert Kypta for his guidance throughout the project, and his powerful 17-inch MacBook with the highest performance and fastest running speed.

Also, Lakshmi M. Konduri, who guided me in upstream analysis of mRNA-seq. Special gratitude to my partner Jiannan: Although life is tough, you are the sweetest I've ever seen.

References

1. S. Sandhu, C. M. Moore, E. Chiong, H. Beltran, R. G. Bristow, S. G. Williams, Prostate cancer. *The Lancet*. **398**, 1075–1090 (2021).
2. M. Zhou, High-grade prostatic intraepithelial neoplasia, PIN-like carcinoma, ductal carcinoma, and intraductal carcinoma of the prostate. *Mod Pathol.* **31**, 71–79 (2018).
3. S. Tyekucheva, M. Bowden, C. Bango, F. Giunchi, Y. Huang, C. Zhou, A. Bondi, R. Lis, M. Van Hemelrijck, O. Andrén, S.-O. Andersson, R. W. Watson, S. Pennington, S. P. Finn, N. E. Martin, M. J. Stampfer, G. Parmigiani, K. L. Penney, M. Fiorentino, L. A. Mucci, M. Loda, Stromal and epithelial transcriptional map of initiation progression and metastatic potential of human prostate cancer. *Nat Commun.* **8**, 420 (2017).
4. J. I. Epstein, L. Egevad, J. R. Srigley, P. A. Humphrey, The 2014 International Society of Urological Pathology (ISUP) Consensus Conference on Gleason Grading of Prostatic Carcinoma. *Am J Surg Pathol.* **40**, 9 (2016).
5. M. C. Haffner, W. Zwart, M. P. Roudier, L. D. True, W. G. Nelson, J. I. Epstein, A. M. De Marzo, P. S. Nelson, S. Yegnasubramanian, Genomic and phenotypic heterogeneity in prostate cancer. *Nat Rev Urol.* **18**, 79–92 (2021).
6. B. S. Taylor, N. Schultz, H. Hieronymus, A. Gopalan, Y. Xiao, B. S. Carver, V. K. Arora, P. Kaushik, E. Cerami, B. Reva, Y. Antipin, N. Mitisades, T. Landers, I. Dolgalev, J. E. Major, M. Wilson, N. D. Soccia, A. E. Lash, A. Heguy, J. A. Eastham, H. I. Scher, V. E. Reuter, P. T. Scardino, C. Sander, C. L. Sawyers, W. L. Gerald, Integrative Genomic Profiling of Human Prostate Cancer. *Cancer Cell.* **18**, 11–22 (2010).
7. Cancer Genome Atlas Research Network, The Molecular Taxonomy of Primary Prostate Cancer. *Cell.* **163**, 1011–1025 (2015).
8. F. Bonollo, G. N. Thalmann, M. Kruithof-de Julio, S. Karkampouna, The Role of Cancer-Associated Fibroblasts in Prostate Cancer Tumorigenesis. *Cancers.* **12**, 1887 (2020).
9. E.-J. Lee, Q. T. T. Nguyen, M. Lee, Dickkopf-3 in Human Malignant Tumours: A Clinical Viewpoint. *Anticancer Res.* **40**, 5969–5979 (2020).
10. F. Abarzua, M. Sakaguchi, M. Takaishi, Y. Nasu, K. Kurose, S. Ebara, M. Miyazaki, M. Namba, H. Kumon, N. Huh, Adenovirus-Mediated Overexpression of REIC/Dkk-3 Selectively Induces Apoptosis in Human Prostate Cancer Cells through Activation of c-Jun-NH₂-Kinase. *Cancer Research*. **65**, 9617–9622 (2005).
11. C. Zenzmaier, N. Sampson, E. Plas, P. Berger, Dickkopf-related protein 3 promotes pathogenic stromal remodeling in benign prostatic hyperplasia and prostate cancer. *Prostate*. **73**, 1441–1452 (2013).
12. Z. Al Shareef, H. Kardooni, V. Murillo-Garzón, G. Domenici, E. Stylianakis, J. H. Steel, M. Rabano, I. Gorroño-Etxebarria, I. Zabalza, M. dM Vivanco, J. Waxman, R. M. Kypta, Protective effect of stromal Dickkopf-3 in prostate cancer: opposing roles for TGFBI and ECM-1. *Oncogene*. **37**, 5305–5324 (2018).
13. M. E. Ritchie, B. Phipson, D. Wu, Y. Hu, C. W. Law, W. Shi, G. K. Smyth, limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*. **43**, e47–e47 (2015).
14. H. Kardooni, E. Gonzalez-Gualda, E. Stylianakis, S. Saffaran, J. Waxman, R. Kypta, CRISPR-Mediated Reactivation of DKK3 Expression Attenuates TGF-β Signaling in Prostate Cancer. *Cancers*. **10**, 165 (2018).
15. D. Romero, Y. Kawano, N. Bengoa, M. M. Walker, N. Maltby, C. Niehrs, J. Waxman, R. Kypta, *Journal of Cell Science*, in press, doi:10.1242/jcs.119388.
16. M. D. Robinson, D. J. McCarthy, G. K. Smyth, edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. **26**, 139–140 (2010).
17. T. Wu, E. Hu, S. Xu, M. Chen, P. Guo, Z. Dai, T. Feng, L. Zhou, W. Tang, L. Zhan, X. Fu, S. Liu, X. Bo, G. Yu, clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *The Innovation*. **2**, 100141 (2021).
18. A. Subramanian, P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, A. Paulovich, S. L. Pomeroy, T. R. Golub, E. S. Lander, J. P. Mesirov, Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences*. **102**, 15545–15550 (2005).
19. A. Akalin, M. Kormaksson, S. Li, F. E. Garrett-Bakelman, M. E. Figueroa, A. Melnick, C. E. Mason, methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome Biol.* **13**, R87 (2012).
20. L.-Y. Yu-Lee, Y.-C. Lee, J. Pan, S.-C. Lin, T. Pan, G. Yu, D. H. Hawke, B.-F. Pan, S.-H. Lin, Bone secreted factors induce cellular quiescence in prostate cancer cells. *Sci Rep.* **9**, 18635 (2019).
21. C. Tomasetti, B. Vogelstein, G. Parmigiani, Half or more of the somatic mutations in cancers of self-

- renewing tissues originate prior to tumor initiation. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 1999–2004 (2013).
22. J. Kano, H. Wang, H. Zhang, M. Noguchi, *The FEBS Journal*, in press, doi:10.1111/febs.16529.
 23. T. O. Adekoya, R. M. Richardson, Cytokines and Chemokines as Mediators of Prostate Cancer Metastasis. *IJMS*. **21**, 4449 (2020).
 24. E. Giannoni, F. Bianchini, L. Masieri, S. Serni, E. Torre, L. Calorini, P. Chiarugi, Reciprocal Activation of Prostate Cancer Cells and Cancer-Associated Fibroblasts Stimulates Epithelial-Mesenchymal Transition and Cancer Stemness. *Cancer Research*. **70**, 6945–6956 (2010).
 25. G. C. Jayson, R. Kerbel, L. M. Ellis, A. L. Harris, Antiangiogenic therapy in oncology: current status and future directions. *The Lancet*. **388**, 518–529 (2016).
 26. K. Lertsuwan, L. H. Choe, I. R. Marwa, K. Lee, R. A. Sikes, Identification of Fibulin-1 as a Human Bone Marrow Stromal (HS-5) Cell-Derived Factor That Induces Human Prostate Cancer Cell Death. *The Prostate*. **77**, 729–742 (2017).
 27. P.-L. Kuo, K.-H. Shen, S.-H. Hung, Y.-L. Hsu, CXCL1/GRO α increases cell migration and invasion of prostate cancer by decreasing fibulin-1 expression through NF- κ B/HDAC1 epigenetic regulation. *Carcinogenesis*. **33**, 2477–2487 (2012).
 28. T. Takeuchi, M. Suzuki, J. Kumagai, T. Kamijo, M. Sakai, T. Kitamura, Extracellular matrix dermatopontin modulates prostate cell growth in vivo. *Journal of Endocrinology*. **190**, 351–361 (2006).
 29. E. Pavlakis, R. Chiotaki, G. Chalepakis, The role of Fras1/Frem proteins in the structure and function of basement membrane. *The International Journal of Biochemistry & Cell Biology*. **43**, 487–495 (2011).
 30. X. Zhao, Y. Lei, G. Li, Y. Cheng, H. Yang, L. Xie, H. Long, R. Jiang, Integrative analysis of cancer driver genes in prostate adenocarcinoma. *Mol Med Report* (2019), doi:10.3892/mmr.2019.9902.
 31. D. Pellacani, A. P. Droop, F. M. Frame, M. S. Simms, V. M. Mann, A. T. Collins, C. J. Eaves, N. J. Maitland, Phenotype-independent DNA methylation changes in prostate cancer. *Br J Cancer*. **119**, 1133–1143 (2018).
 32. C. Song, C. Zhou, HOXA10 mediates epithelial-mesenchymal transition to promote gastric cancer metastasis partly via modulation of TGFB2/Smad/METTL3 signaling axis. *J Exp Clin Cancer Res.* **40**, 62 (2021).
 33. Y. Assenov, F. Müller, P. Lutsik, J. Walter, T. Lengauer, C. Bock, Comprehensive analysis of DNA methylation data with RnBeads. *Nat Methods*. **11**, 1138–1140 (2014).
 34. D. Boison, S. A. Masino, F. D. Lubin, K. Guo, T. Lusardi, R. Sanchez, D. N. Ruskin, J. Ohm, J. D. Geiger, J. Hur, The impact of methodology on the reproducibility and rigor of DNA methylation data. *Sci Rep.* **12**, 380 (2022).

Abbreviations

CZ, central zone; TZ, transition zone; PZ, peripheral zone; PCa, Prostate cancer; PSA, Prostate-specific antigen; GS, Gleason score; ADT, Androgen deprivation therapy; MSKCC, Memorial Sloan Kettering Cancer Centre; TCGA, The cancer genome atlas; ECM, Extracellular matrix; PRAD, Prostate adenocarcinoma; TME, Tumour microenvironment; CAFs, Cancer-associated fibroblasts; DKK, Dickkopf; TGFB1, TGF- β -induced protein I; ECM-1, Extracellular matrix protein-1; ROI, Regions of interest; B, Benign epithelium; P, Prostate intraepithelial neoplasia; T, Tumour; sB, Stroma adjacent to benign epithelium; sP, Stroma adjacent to prostate intraepithelial neoplasia; sT, Stroma adjacent to tumour; TPM, Transcript per million; MDS, Multidimensional scaling; ANOVA, Analysis of variance; TSS, Transcriptional start site; PCA, Principal component analysis; BH, Benjamini-Hochberg; log2FC, log2 fold change; DEGs, Differentially expressed genes; ORA, Over-representative analysis; GSEA, Gene set enrichment analysis; GO, Gene Ontology; MF, Molecular function; BP, Biological process; CC, Cellular component; ssGSEA, single sample GSEA; HRs, Hazard ratios; CI, Confidence interval; DMR, Differentially methylated region; kb, kilobase; Refseq, Reference Sequence.

Supplementary Methods

mRNA-seq

RNA concentration from different samples were measured in Qubit 2.0, with Qubit RNA assay kit (Invitrogen, Cat.# Q32855). RNA integrity characterization was conducted by Agilent 2100 Bioanalyzer using an Agilent RNA 6000 Nano Chips Cat. # 5067-1511. After RNA normalization, sequencing libraries were prepared with the TruSeq Stranded mRNA Sample Preparation Guide (Illumina Inc.) Amplified library concentration was determined with Qubit fluorometer using the Qubit®dsDNA HS assay kit (Invitrogen, Cat.# Q32854). Samples were then run on Illumina platform HiSeq2500 with 51bp single-end reads. Sequencing data is converted into raw data (fastq.gz) for analysis. For quality control, fastq were checked using FastQC (v0.11.9, <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Adaptors were trimmed using Trim-galore (v0.6.6, <https://github.com/FelixKrueger/TrimGalore>) with phred, stringency and length as 33, 3 and 36. Read sequences were aligned through Hisat2 (v2.2.1, <http://daehwankimlab.github.io/hisat2/manual/>). The GRCh38 index file was downloaded from Hisat2 website and unzipped. Outputs of Hisat2 are sam files, which were transformed into sorted bam files using Samtools (v1.16, <http://www.htslib.org/download/>). featureCounts is a software program packed in Subread (v2.0.3, <https://sourceforge.net/projects/subread/>). Sorted bam files finally count reads to genes.

Reduced representation bisulfite sequencing (RRBS)

DNA concentrations were measured in Qubit 2.0, with Qubit HS DNA assay kit (Thermo Fisher Scientific, Cat. # Q32854). RRBS sequencing library preparation used

NEXTflex Bisulfite-Seq Kit™ (BioScientific) and EZ DNA Methylation Gold™ Kit (Zymo Research). Size distribution for amplified libraries was assessed using Agilent High Sensitivity DNA Chip (Agilent Technologies, Cat. 5067-4626). These samples were run on an Illumina Platform HiSeq2500 with 51 bp single end reads. Sequencing data were converted into raw data (fastq.gz) for analysis. For quality control, fastq were checked using FastQC v0.11.9. Adaptors were trimmed using Trim-galore v0.6.6, with phred, stringency and length as 33, 3 and 36. DNAs were digested using Mspl/TaqαI, with --rrbs, --non_directional and --retain_unpaired options. Bismark v0.23.0 (<https://github.com/FelixKrueger/Bismark/tree/master/Docs>) maps RRBS sequencing reads to a genome of interest and performs methylation calls. Samtools are pre-installed to support Bismark function. Index primary assembly genome was downloaded from the Ensembl website (http://ftp.ensembl.org/pub/release-107/fasta/homo_sapiens/dna/-.:text=Homo_sapiens.GRCh38.dna.primary_assembly.a.gz). --hisat2 option created bisulfite indexes for use with Hisat2. After genome indexing, Bismark alignment was conducted using Hisat2 mode. The output bam files were then operated on Bismark to extract methylation calls. --bedGraph option created bedGraph files, which were imported in IGV (v2.13.2, <https://software.broadinstitute.org/software/igv/>) for display of valued methylation data in track format. Outputs also include a coverage file (cov.gz) for downstream analysis and a text file. Text files can be used for generating a summary report of Bismark alignment, and CpG/ CHG/ CHH calls.

Supplementary Figures

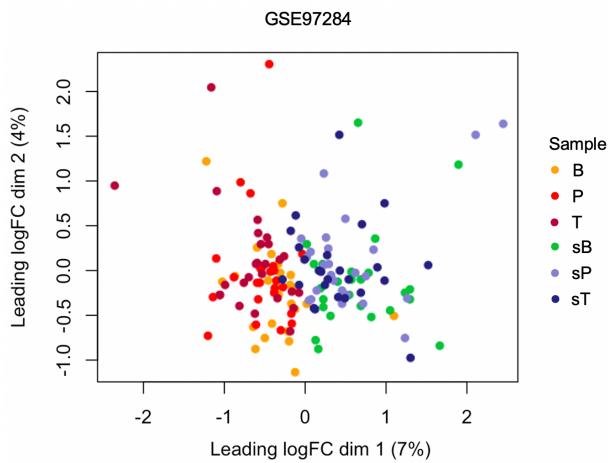


Figure S1 Multidimensional scaling (MDS) plot of gene expression profiles in benign epithelium (B), prostatic intraepithelial neoplasia (P), tumour cells (T) and their corresponding stromal tissues (sB, sP, sT). These 150 samples came from 25 prostate cancer patients (6 different samples from each patient).

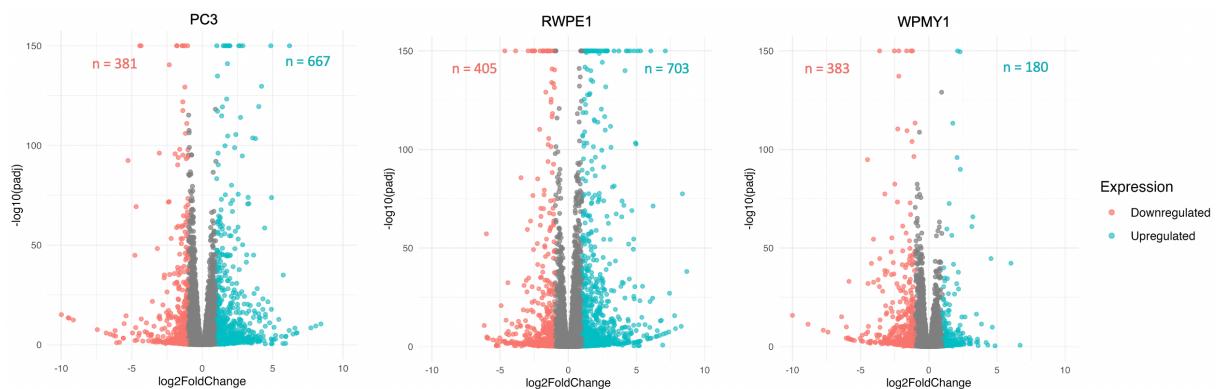


Figure S2. Integrated analysis of differentially expressed genes (DEGs) in each cell lines. (A) Volcano plots of DEGs. Green and red dots indicate significant upregulation (adjusted p-value < 0.05 and log₂ fold change ≥ 1) and downregulation (adjusted p-value < 0.05 and log₂ fold change ≤ -1), respectively. **(B)** Venn diagram of DEGs from WPMY1 versus that from RWPE1. p-values for overlapping genes were calculated using hypergeometric test.

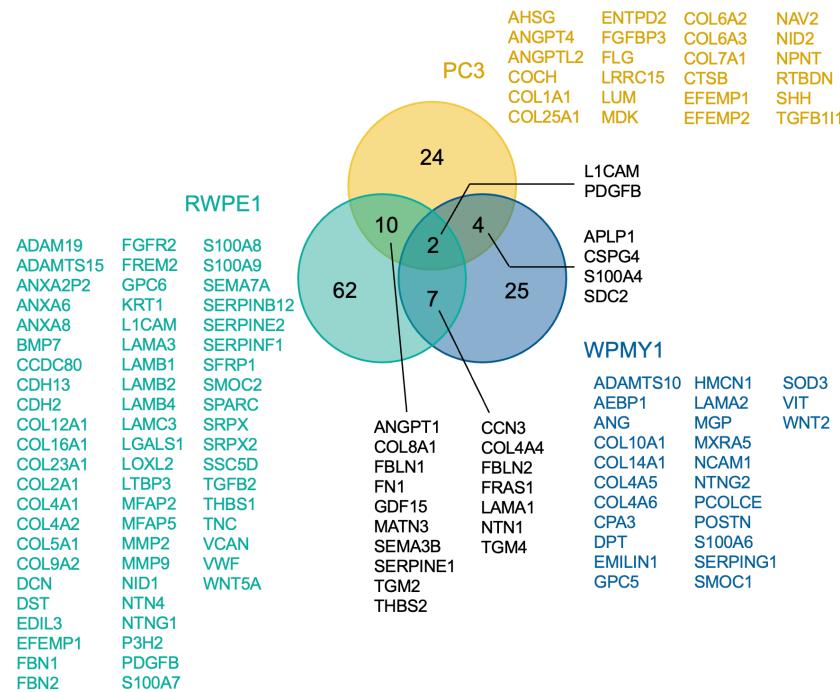


Figure S3. Venn diagram depicting overlap of genes differentially expressed in response to DKK3 in the three cell line models for the GO term “collagen-containing extracellular matrix” (n = 133).

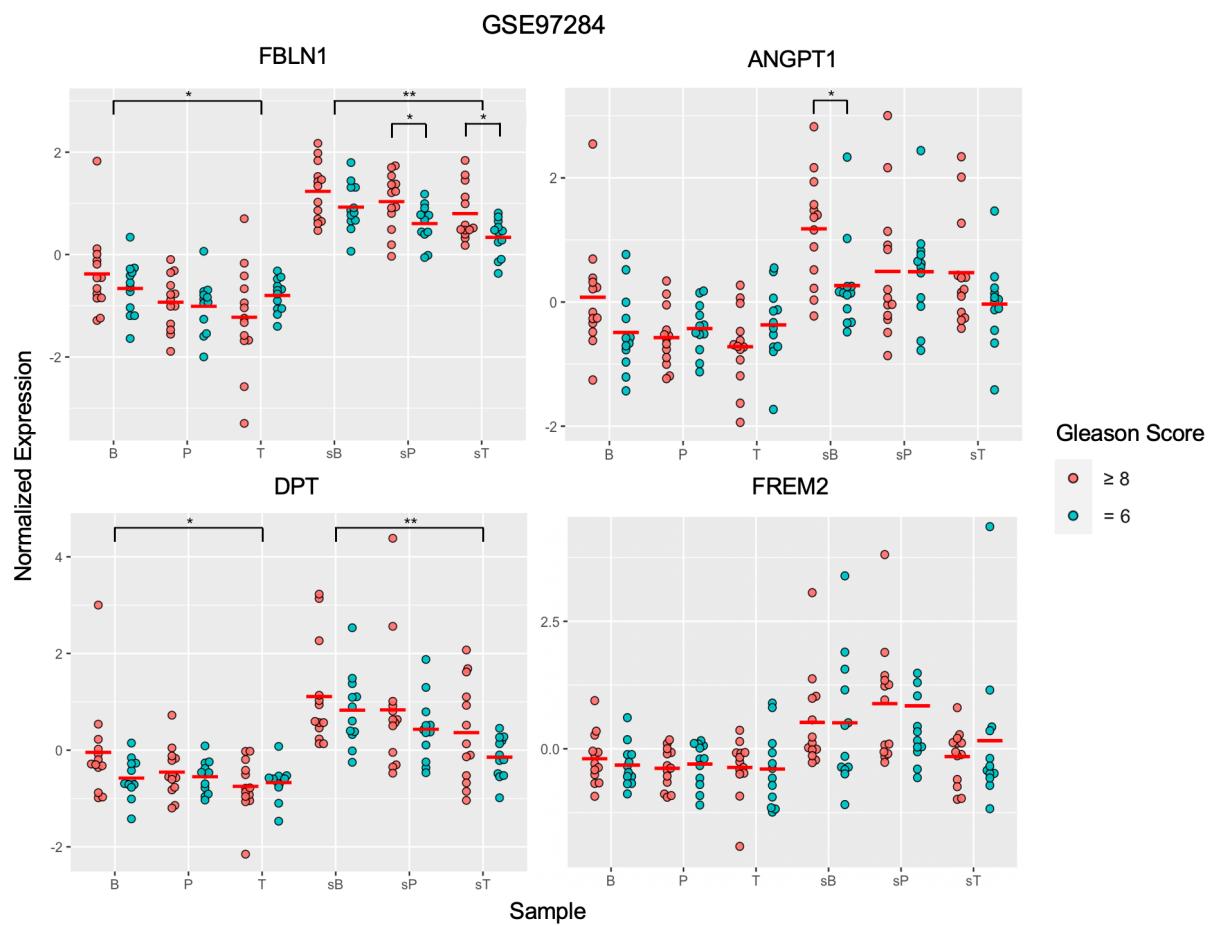


Figure S4. Comparison of DKK3 expression levels in GS 6 and GS 8-10 samples ($n = 150$). T-tests were performed comparing low and high GS samples and comparing benign epithelium (B) and tumour (T), and benign stroma (sB) and tumour stroma (sT). $p < 0.05$ indicates statistical significance; only t-tests with statistical significance are shown. P, prostatic intraepithelial neoplasia; sP, stroma adjacent to prostatic intraepithelial neoplasia.

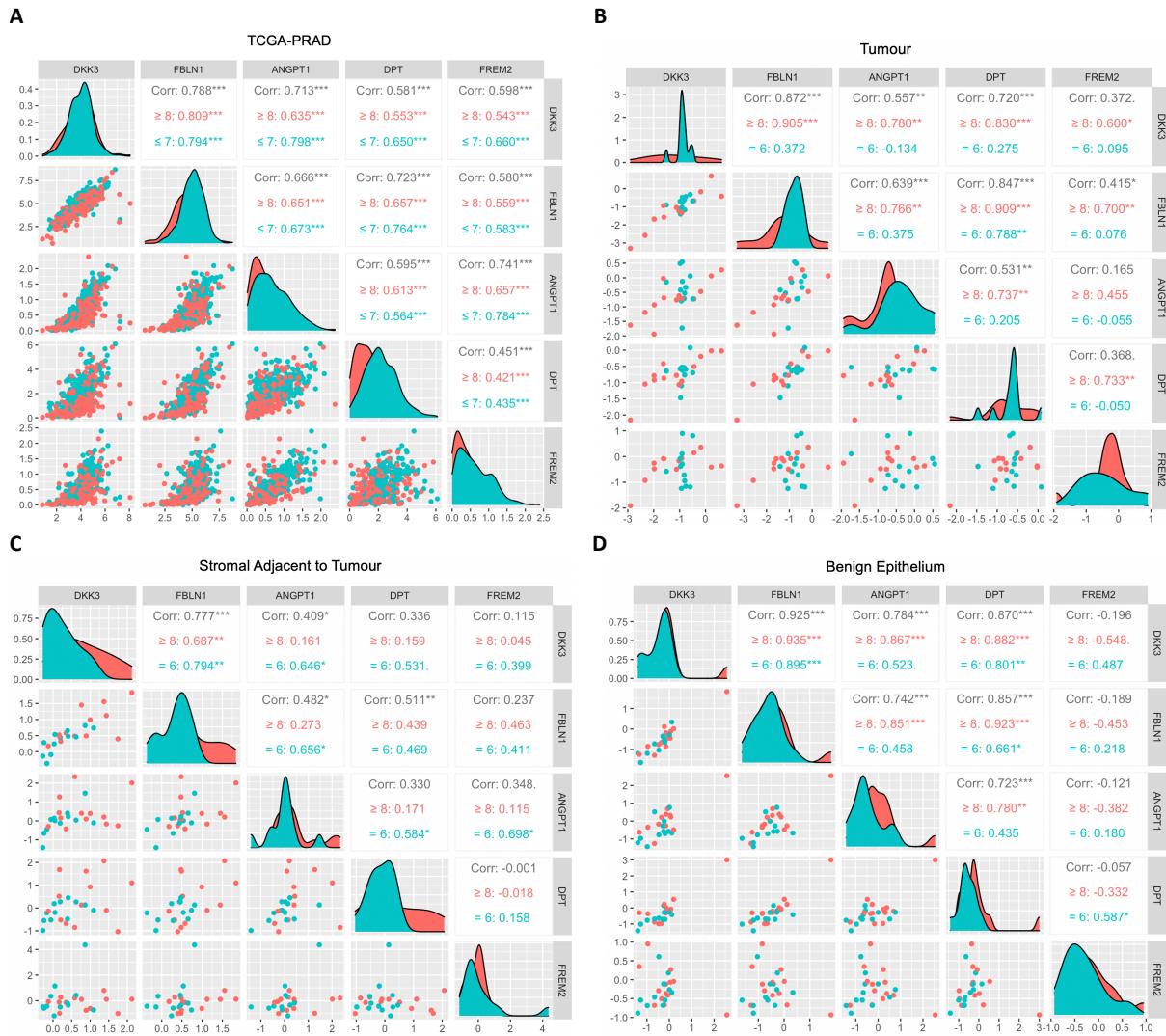


Figure S5. Correlation matrix showing the relationship between DKK3 and four selected ECM-related genes in TCGA-PRAD cohort (A), tumour cells in GSE97284 (B), benign epithelium in GSE97284 (C) and stroma adjacent to tumour in GSE97284 (D). Figures on the diagonal show the distribution of each gene following a normalized expression and stratified by Gleason score (GS) (red: GS = 6 (n = 12); green: GS ≥ 8 (n = 13)); Scatterplots below the diagonal show gene correlations. The Pearson correlation coefficients for GS = 6 (red), GS ≥ 8 (green) and combined (grey) are displayed above the diagonal, of which *, ** and * indicate p < 0.05, 0.01 and 0.001, respectively.**

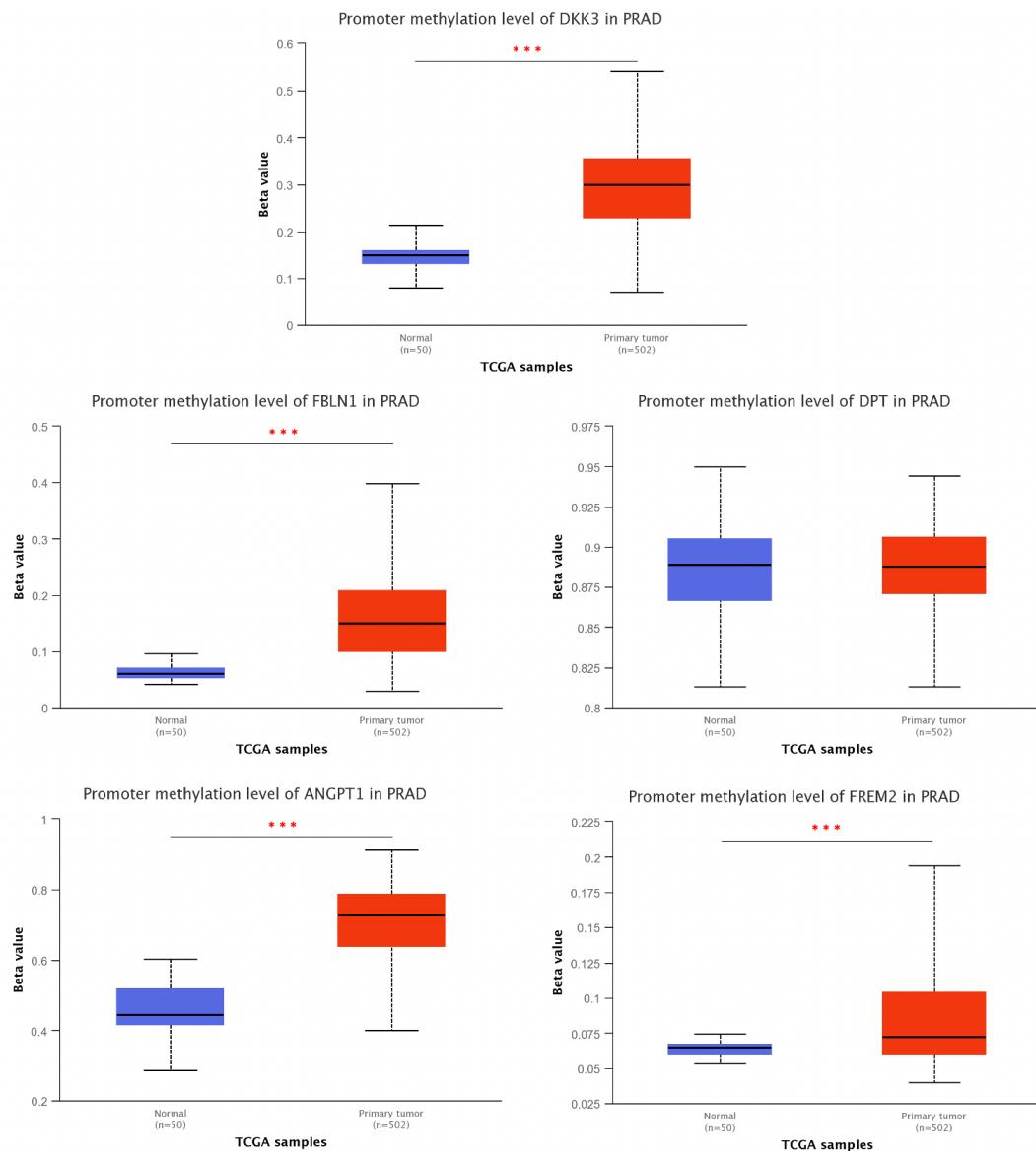


Figure S6. Promoter methylation levels in non-cancer (blue) or primary tumour (red) cases retrieved from the UALCAN server using the TCGA-PRAD dataset. The beta values indicate the DNA methylation levels ranging from 0 (unmethylated) to 1 (fully methylated). Student t-tests were performed comparing normal and primary tumour samples, with *, **, *** indicating $p < 0.05, 0.01, 0.001$, respectively.

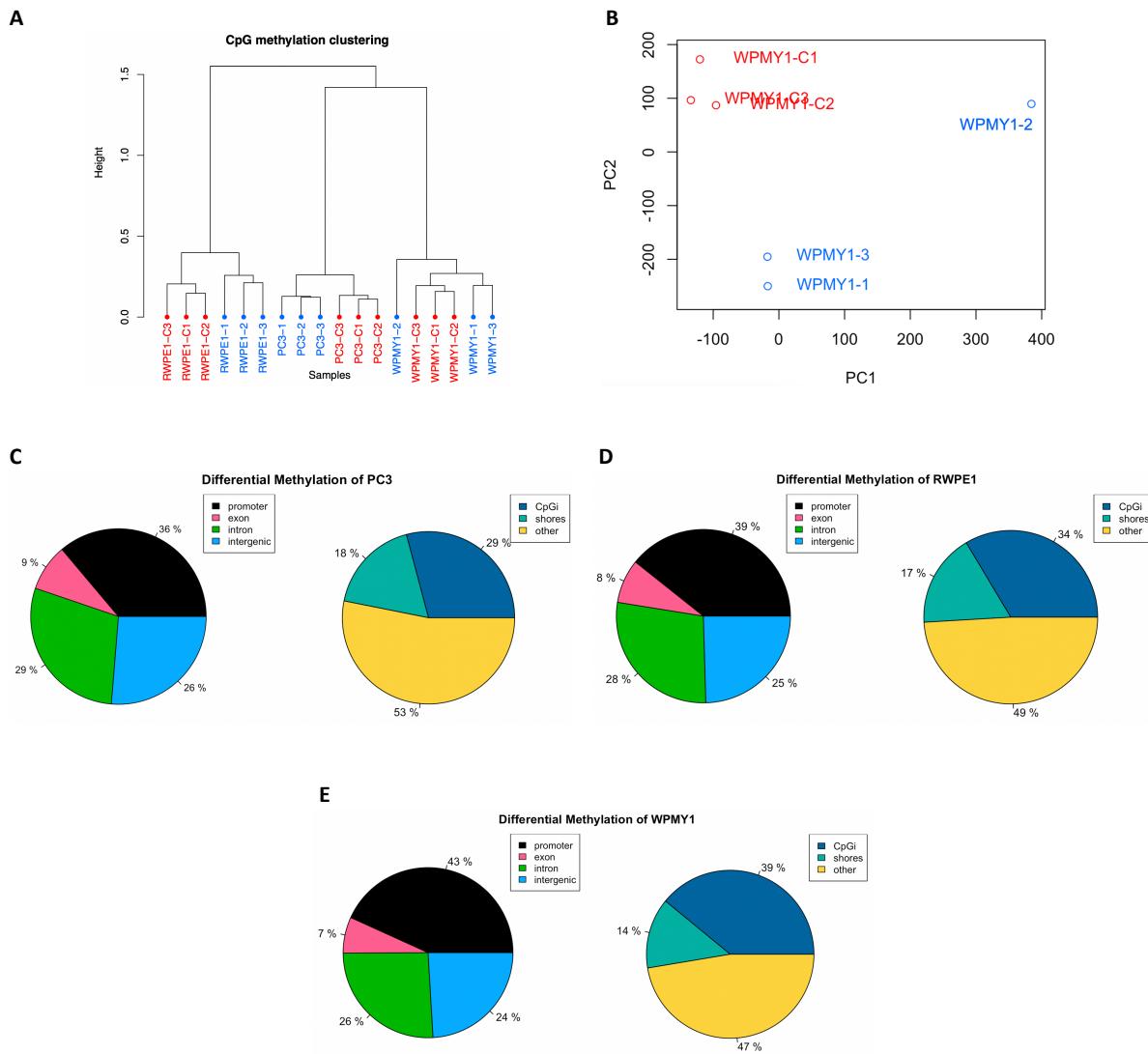


Figure S7. Analysis of DNA methylation information obtained by reduced-representation bisulfite sequencing (RRBS) using Methylkit. (A) Dendrogram of 18 samples, including three experimental samples (1/2/3) and three control samples (C1/C2/C3) for three cell line models (WPMY1, RWPE1 and PC3). **(B)** PCA diagram of methylation profile of WPMY1 cells. **(C-E)** Annotation of the distribution of differential methylation region (DMR) in PC3 **(C)**, RWPE1 **(D)** and WPMY1 **(E)** cell lines with Genome Reference Consortium Human Build 38 (GRCh38). Pie charts on the left depict the percentage of DMRs that are located in promoter, exon, intron or intergenic regions, while the charts on the right show the percentage of DMRs on CpG islands, CpG shores or other regions.

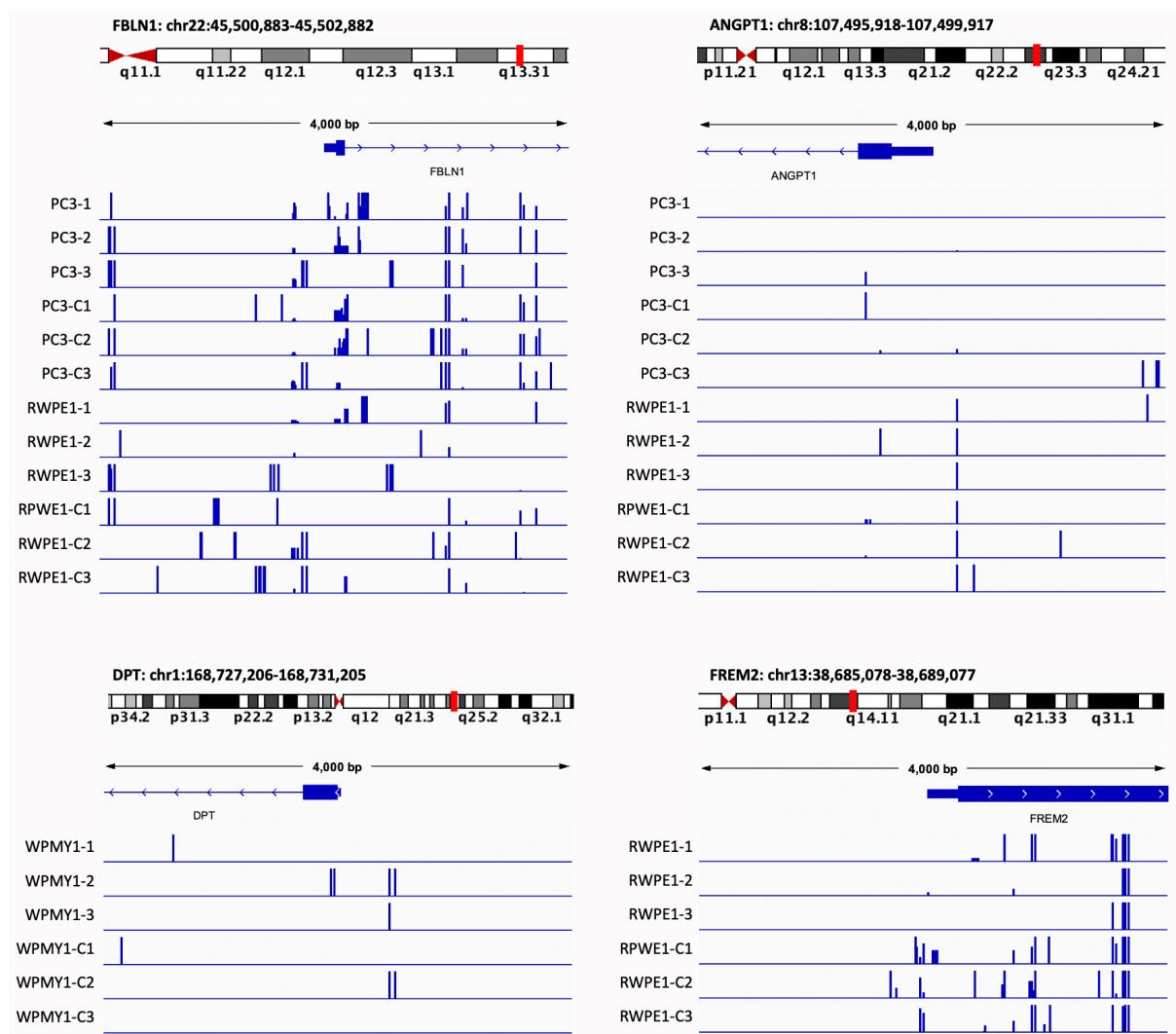


Figure S8. Representation of the methylation profile of the 2 kb regions upstream and downstream of the transcription start sites (TSS) of FBLN1, ANGPT1, DPT and FREM2 using Integrative Genomics Viewer (IGV). Each bar indicates methylation at one CpG locus, and bar height indicates methylation ratio (methylated CpGs/all CpGs * 100%). Profiles are shown only for those cell lines where the gene was differentially expressed.

Data Availability

TCGA-PRAD cohort are available at UCSC Xena brower (<https://xenabrowser.net>). MSKCC cohort and microdissected prostate specimens data could be accessed on GEO database (<https://www.ncbi.nlm.nih.gov/geo/>) through accession code GSE21032 and GSE97284. Algorithms in this study can be seen at https://github.com/realzwu/project_2. The RNA-Seq and RRBS model cell line data will be made available upon publication.