

# Molecular analysis of archival diagnostic prostate cancer biopsies identifies genomic similarities in cases with progression post-radiotherapy, and those with de novo metastatic disease

Philip Vincent Charlton MBBS<sup>1,2</sup> | Dawn O'Reilly PhD<sup>1</sup> | Yiannis Philippou PhD<sup>3</sup> | Srinivasa Rao Rao PhD<sup>4</sup> | Alastair David Gordon Lamb PhD<sup>3,4</sup> | Ian Geoffrey Mills PhD<sup>4</sup>  | Geoff Stuart Higgins PhD<sup>1,2</sup> | Freddie Charles Hamdy FMedSci<sup>3,4</sup> | Clare Verrill BM<sup>4,5</sup> | Francesca Meteora Buffa PhD<sup>1</sup> | Richard John Bryant PhD<sup>3,4</sup> 

<sup>1</sup>Department of Oncology, University of Oxford, Oxford, UK

<sup>2</sup>Department of Oncology, Oxford University Hospitals NHS Foundation Trust, Oxford, UK

<sup>3</sup>Department of Urology, Oxford University Hospitals NHS Foundation Trust, Oxford, UK

<sup>4</sup>Nuffield Department of Surgical Sciences, University of Oxford, Oxford, UK

<sup>5</sup>Department of Pathology, Oxford University Hospitals NHS Foundation Trust, Oxford, UK

## Correspondence

Francesca Meteora Buffa, PhD, Department of Oncology, University of Oxford, Oxford, UK.

Email: [francesca.buffa@oncology.ox.ac.uk](mailto:francesca.buffa@oncology.ox.ac.uk)

Richard John Bryant, PhD, Nuffield Department of Surgical Sciences, University of Oxford, Oxford, UK.

Email: [richard.bryant@nds.ox.ac.uk](mailto:richard.bryant@nds.ox.ac.uk)

## Funding information

Cancer Research UK DPhil studentship; Urology Foundation, John Black Charitable Foundation, Rosetrees Trust, Prostate Cancer UK, and NIHR-HTA; Cancer Research UK, Grant/Award Numbers: C39297/A22748, C57899/A25812; European Research Council (ERC) Consolidator Award, Grant/Award Number: MICROC:772970; Cancer Research UK RadNet Oxford Center,

## Abstract

**Background:** It is important to identify molecular features that improve prostate cancer (PCa) risk stratification before radical treatment with curative intent. Molecular analysis of historical diagnostic formalin-fixed paraffin-embedded (FFPE) prostate biopsies from cohorts with post-radiotherapy (RT) long-term clinical follow-up has been limited. Utilizing parallel sequencing modalities, we performed a proof-of-principle sequencing analysis of historical diagnostic FFPE prostate biopsies. We compared patients with (i) stable PCa (sPCa) postprimary or salvage RT, (ii) progressing PCa (pPCa) post-RT, and (iii) de novo metastatic PCa (mPCa).

**Methods:** A cohort of 19 patients with diagnostic prostate biopsies ( $n = 6$  sPCa,  $n = 5$  pPCa,  $n = 8$  mPCa) and mean 4 years 10 months follow-up (diagnosed 2009–2016) underwent nucleic acid extraction from demarcated malignancy. Samples underwent 3' RNA sequencing (3'RNAseq) ( $n = 19$ ), nanoString analysis ( $n = 12$ ), and Illumina 850k methylation ( $n = 8$ ) sequencing. Bioinformatic analysis was performed to coherently identify differentially expressed genes and methylated genomic regions (MGRs).

**Results:** Eighteen of 19 samples provided useable 3'RNAseq data. Principal component analysis (PCA) demonstrated similar expression profiles between pPCa and mPCa cases, versus sPCa. Coherently differentially methylated probes between these groups identified ~600 differentially MGRs. The top 50 genes with increased expression in pPCa patients were associated with reduced progression-free survival post-RT ( $p < 0.0001$ ) in an external cohort.

Francesca Meteora Buffa and Richard John Bryant senior authors of this study.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2024 The Authors. *The Prostate* published by Wiley Periodicals LLC.

Grant/Award Number: C6078/A28736;  
Oxford NIHR BRC Surgical Innovation &  
Evaluation theme; John Black Charitable  
Foundation; Rosetrees Trust; Prostate Cancer  
UK; NIHR Oxford Biomedical Research  
Center; CRUK Oxford Cancer Center; Cancer  
Research UK Development Fund

**Conclusions:** 3'RNAseq, nanoString and 850k-methylation analyses are each achievable from historical FFPE diagnostic pretreatment prostate biopsies, unlocking the potential to utilize large cohorts of historic clinical samples. Profiling similarities between individuals with pPCa and mPCa suggests biological similarities and historical radiological staging limitations, which warrant further investigation.

#### KEYWORDS

diagnostic biopsies, metastasis, molecular analysis, prostate cancer, radiotherapy

## 1 | INTRODUCTION

Localized prostate cancer (PCa) may be treated with curative intent by radical surgery, or radiotherapy (RT) with concomitant androgen-deprivation therapy, with equivalent rates of posttreatment disease progression, but differences in side effect profiles, at 15 years clinical follow-up.<sup>1,2</sup> In the clinical oncology setting, it is recognized that at long-term follow-up a group of patients will experience disease progression following RT.<sup>3</sup> In addition to use of clinical risk scores based on prostate specific antigen (PSA), tumor grade and stage,<sup>3,4</sup> baseline molecular characterization of diagnostic biopsies offers the potential to identify patients at high risk of post-RT relapse. This approach may facilitate more accurate risk-stratification in the immediate postdiagnosis pretreatment space, to enable more appropriate treatment selection.

Several prognostic and predictive transcriptomic classifiers have been developed for PCa. However, none are routinely used in the clinic in the pre-RT setting. Currently available classifiers include the Oncotype, Decipher, Prolaris, metastatic, and hypoxia signatures.<sup>5–10</sup> The Decipher, Oncotype, and cell cycle progression (CCP) signatures have demonstrated clinical utility in predicting disease recurrence after both radical surgery and radical RT,<sup>8,11–15</sup> despite these classifiers being derived from surgically treated patients. A prostate-specific molecular signature of hypoxia has been demonstrated to predict biochemical recurrence in the salvage RT setting for local disease recurrence postradical prostatectomy and is an independent prognostic indicator for patients with localized PCa receiving RT.<sup>10</sup> Methylation-based molecular features of PCa have also been found to be associated with clinical outcome after salvage RT.<sup>15–17</sup>

Use of archival formalin-fixed paraffin-embedded (FFPE) diagnostic prostate biopsy samples for RNA sequencing is technically challenging given the small volume of available tissue, and the degradation and cross-linking of RNA over time. 3'RNAseq is potentially well suited for sequencing of degraded RNA, utilizing only the 3' ends of RNA fragments, and providing a single read per gene transcript. NanoString analysis provides an alternative strategy for molecular analysis of clinical samples, as this technique uses reporter probes to hybridize mRNA, and reports fewer genes (~600 for nanoString, vs. ~20,000 for 3'RNAseq), thus significantly reducing computational resources and analysis time. Illumina 850k methylation

analysis has greater power over previous methylation arrays due to the increased number of probes. Utilizing these distinct molecular analysis technologies in an orthogonal manner may identify biologically relevant genes, versus each single genomic technique in isolation.

Previous studies in the field have analyzed the molecular features of clinical samples from patients with localized PCa treated with RT, to derive a signature associated with clinical outcome, such as those described as being hypoxic or metastatic.<sup>9,10</sup> In this study, we investigated whether molecular features associated with subsequent development of metastatic disease could be identified in the historical diagnostic FFPE prostate biopsy samples from patients with long-term clinical follow-up after primary radical or salvage RT. We used orthogonal analyses of a carefully curated small number of samples and included cases with baseline *de novo* metastatic PCa (mPCa) at presentation as a comparator. Using this approach, we aimed to provide proof-of-concept that this technique could unlock the door to future larger scale studies from mature cohorts, such as those from the ProtecT (Prostate testing for cancer and Treatment) study,<sup>1</sup> with 15 years clinical follow-up. Such studies have the potential to provide added-value in future risk-stratification for men newly diagnosed with intermediate- or high-risk localized or locally advanced PCa undergoing radical RT, based on a personalized medicine genomic analysis of baseline molecular features of each malignancy, alongside conventional risk parameters such as PSA, tumor grade and stage.<sup>3,18–20</sup>

## 2 | MATERIALS AND METHODS

### 2.1 | Patient identification

Baseline diagnostic prostate biopsy samples from patients in the ProMPT (Prostate cancer: Mechanisms of Progression and Treatment) cohort, obtained with consent from individual patients for the use of tissue in research within ProMPT, and with appropriate institutional ethics approval (UK MREC number 01/4/61), along with carefully curated clinicopathological features and long-term clinical follow-up following primary or salvage RT, were identified. To investigate baseline molecular features of development of mPCa after RT, three groups of patients were identified for this proof-of-concept

study: individuals with (i) stable PCa (sPCa), (ii) progressing PCa (pPCa), and (iii) mPCa (Figure 1 and Supporting Information: Figure S1).

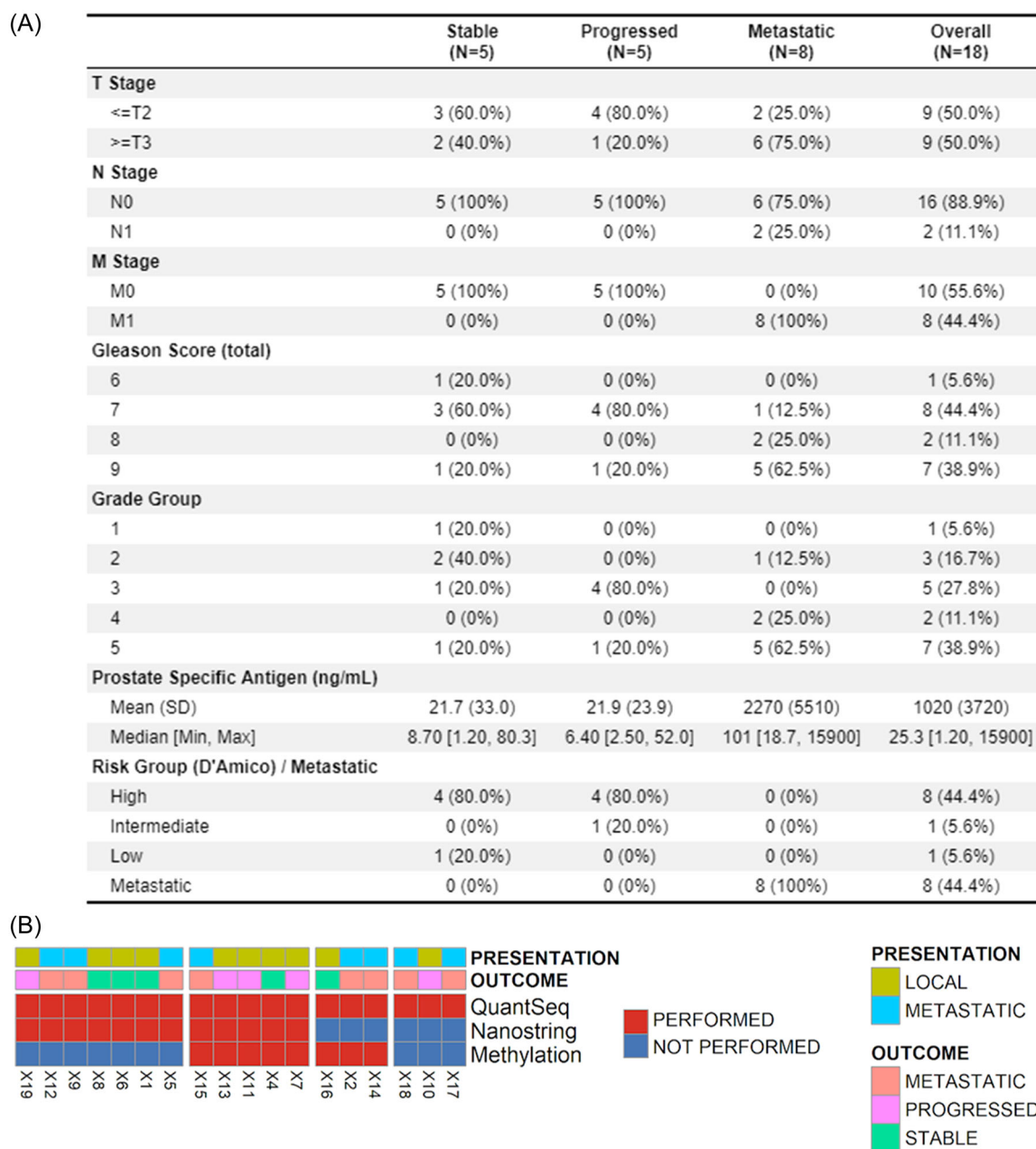
## 2.2 | Pathology and nucleic acid extraction

Pathology sections were reviewed by a specialist uropathologist for tissue availability, Gleason Grade Group, and percentage tumor content. PCa was annotated in prostate biopsy samples on a single slide, and four 5 µm serial sections per patient were macro-dissected

using a sterile scalpel. Samples were pooled per patient, and RNA extracted using the Roche HighPure FFPE kit. Eight patients samples contained sufficient material for four further slides to be utilized for DNA extraction using the ROCHE FFPE high-pure DNA kit protocol.

## 2.3 | Sequencing techniques

3'RNAseq sequencing library preparation was performed with QuantSeq. 3'-mRNA-Seq Library Prep Kit (Lexogen), and libraries were sequenced on Illumina NextSeq flow cells (75 bp fragments) at



**FIGURE 1** Summary of clinical characteristics of the cohort (A). Heatmap summary of the sequencing performed on the cohort (B). Quantseq, 3'RNA sequencing was performed for all cases; nanoString analysis was performed for 12 of the cases; DNA methylation analysis was performed for 8 of the cases. [Color figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com/terms-and-conditions)]

the Wellcome Trust Center for Human Genetics (University of Oxford).

For nanoString analysis, 12 samples with optimal spectrophotometric characteristics (determined by NanoDrop) were selected (from 4 individuals with each of sPCa, pPCa, and mPCa). Samples underwent nCounter<sup>®</sup> Human PanCancer Panel Gene Expression profiling according to the manufacturer's instructions, and data was acquired using the nCounter<sup>®</sup> SPRINT profiler. For DNA methylation analysis, bisulphite conversion was performed using the EZ-96 DNA Bisulphite Zymo Research conversion protocol. The Illumina Infinium HD Methylation Assay protocol was followed, and samples were hybridized to Human Methylation EPIC beadchips (Genomics Birmingham).

## 2.4 | Gene expression and methylation data analysis

3'RNAseq FastQ files were concatenated for each patient, and polyA and Illumina adapter trimming (AGATCGGAAGAGC) was performed using trimmomatic (v0.25), before alignment with STAR (2.7.7a) to Hg38 with featurecounts used to generate counts (Supporting Information: Methods).<sup>21</sup> Filtering of lowly expressed genes was performed, followed by differential expression analysis using DESeq. 2 (v1.26.0).<sup>22</sup> Overrepresentation analysis was performed using ClusterProfiler, enrichr and gprofiler. The gene set variation analysis GSVA (v1.34.0) package was used to perform single sample gene set enrichment analysis (ssGSEA).<sup>23</sup>

## 2.5 | Quality control (QC)

3'RNAseq QC was performed with FastQC pre- and posttrimming. One sample failed sequencing, with <100,000 counts, and was not used for further analysis. Although not strictly required for DESeq. 2 analysis, counts were filtered to only include genes with >1 count in ≥5 samples (minimum group size). For nanoString analysis, data were imported into nSolver™ analysis software v2.5, and QC performed according to nanoString guidelines, with gene transcripts normalized to housekeeping genes. For DNA methylation analysis, QC, filtering of poor performing probes, cross-reactive probes and normalization was performed before differential methylation analysis using minfi package (v1.32.0).

As a measure of the performance characteristics of the data set, the SigQC protocol<sup>24</sup> was used to generate QC metrics for previously validated signatures (DECIPHER, Prolaris, Oncotype, and prostate hypoxia, each obtained from the published literature<sup>5–10</sup> and these were compared to two previously derived “gold standard” data sets (the TCGA and Jain RT data sets). Sigcheck (package v2.18.0) was used to assess the performance of these gene sets compared to random genes and known signatures with 1000 iterations.<sup>25</sup> KMunicate and survival packages were used for survival analysis.

## 2.6 | Statistical survival analysis

Sequencing and clinical data from the external data set GSE116918 were downloaded from Gene Expression Omnibus (GEO) repository (<https://www.ncbi.nlm.nih.gov/geo/>) and survival analysis performed using a Cox proportional hazards model comprising clinicopathological features and expression of selected genes, followed by estimation of time-dependent receiver operator characteristics (ROC) using TimeROC (package 0.4)<sup>26</sup> (Supporting Information: Methods).

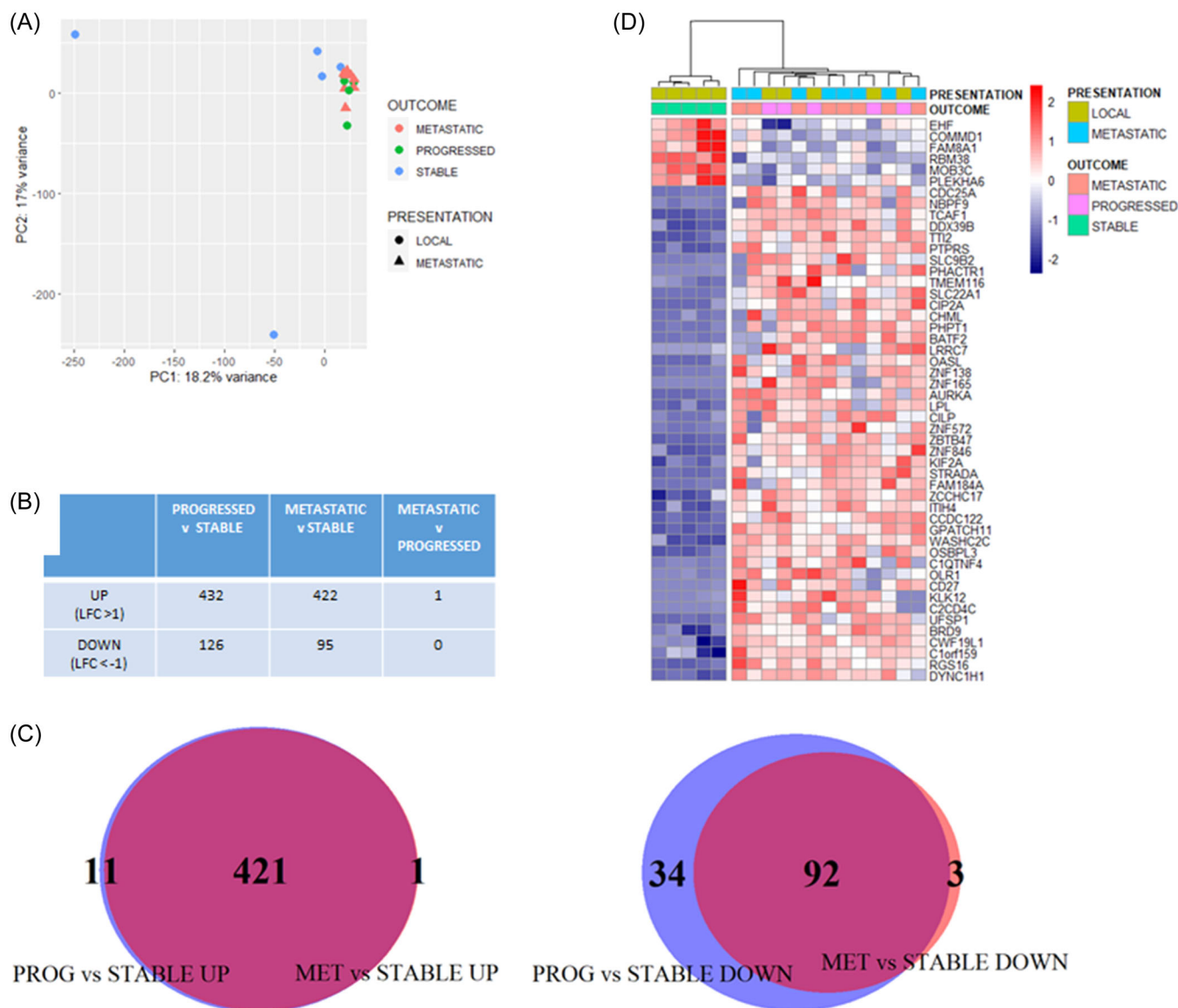
## 3 | RESULTS

Samples from 18 of 19 patients were included in this study ( $n = 6$  with sPCa,  $n = 5$  with pPCa, and  $n = 8$  with mPCa). One sample failed 3' RNAseq and was excluded from further analysis. Baseline clinicopathological features of the cohort, and a summary of the sequencing technologies ( $n = 18$  3'RNAseq;  $n = 12$  nanoString;  $n = 8$  Methylation) used per patient, are shown in Figure 1 and Supporting Information: Table S1. Baseline clinicopathological features of the two patient groups receiving RT (i.e., those with sPCa, and those with pPCa) were similar, with a similar mean PSA, and both groups contained  $n = 4$  D'Amico high-risk patients.

The sigQC protocol was used for QC to evaluate the performance of previously validated PCa signatures (prostate hypoxia, Decipher, Prolaris, and Oncotype) in the sequencing data set, compared to the “gold standard” TCGA prostate (pan-cancer) data set, and an external data set of patients with localized PCa treated with radical RT by Jain (GSE116918).<sup>11</sup> QC metrics demonstrated that these signatures perform at a comparable level in the data set compared to previously published data sets (Supporting Information: Figures S1–S3), indicating the general applicability of 3'RNAseq technology, and providing a validation of the utility of the data set.

Principal component analysis (PCA) of the 3'RNAseq data demonstrated some separation of the sPCa cases from those with pPCa or mPCa (Figure 2A). A comparison of sPCa versus pPCa cases identified 558 differentially expressed genes (DEGs) ( $p_{\text{adj}} < 0.05$ ,  $\text{LFC} > 1$  or  $< -1$ ) ( $n = 432$  increased expression,  $n = 126$  decreased expression) (Figure 2B). A similar number of DEGs were observed between mPCa and sPCa cases ( $n = 422$  increased expression,  $n = 95$  decreased expression) (Figure 2B). Only one gene was significantly differentially expressed between pPCa and mPCa cases (Figure 2B). The majority of DEGs between sPCa versus pPCa, and sPCa versus mPCa, were observed to have concordant directionality ( $n = 421$  increased expression,  $n = 92$  decreased expression) (Figure 2C).

Hierarchical clustering of the top 50 DEGs between sPCa and pPCa cases, demonstrated that pPCa and mPCa samples had a similar expression pattern, and this was distinct from sPCa cases (Figure 2D), supporting the PCA findings. Gene ontology overrepresentation analysis of increased DEGs in sPCa versus pPCa cases ( $p_{\text{adj}} < 0.01$ ), and sPCa versus mPCa ( $p_{\text{adj}} < 0.05$ ) cases, identified pathways associated with spindle pole and centrosome function, respectively (Supporting Information: Figures S4 and S5).



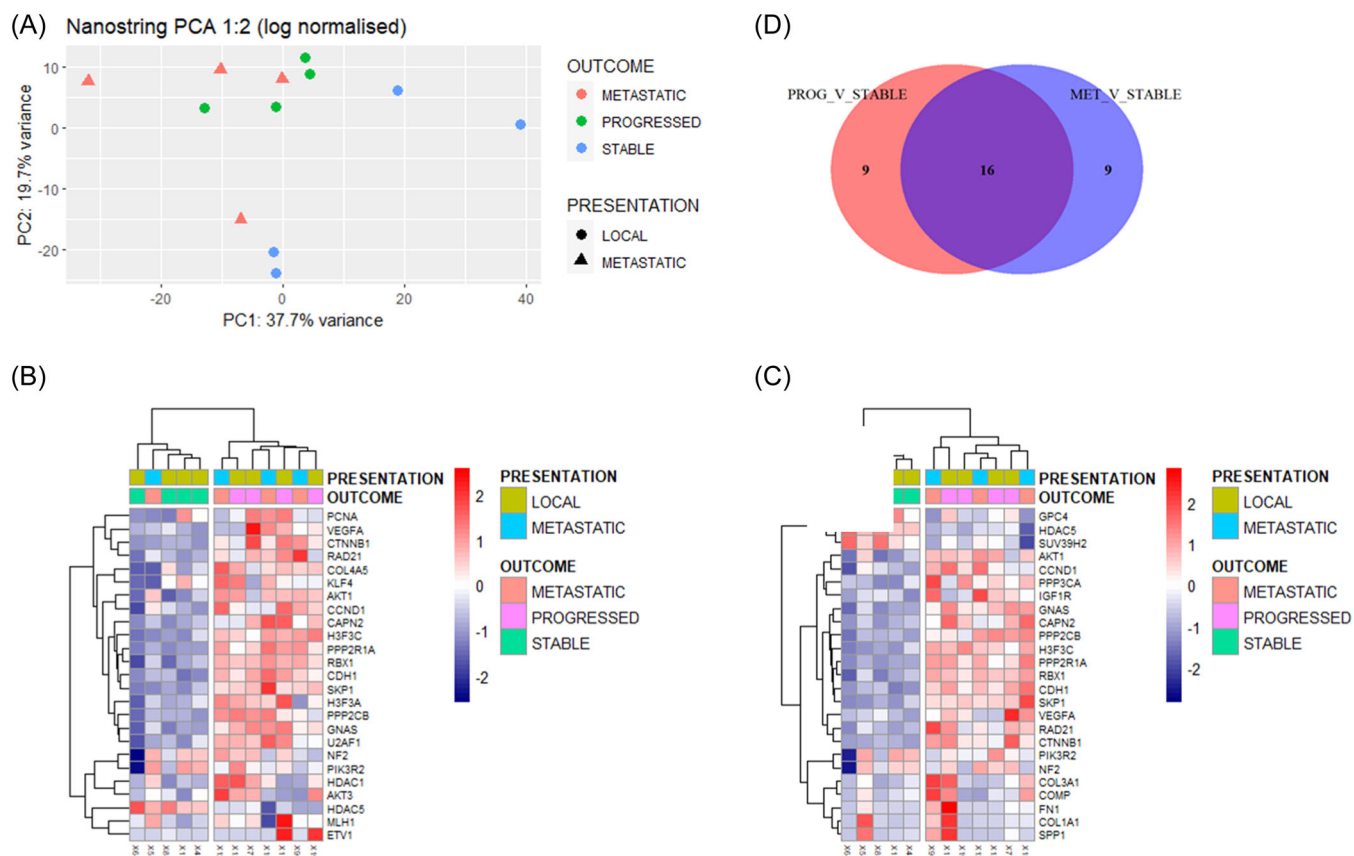
**FIGURE 2** 3'RNAseq analysis of the cohort. Principal component analysis (PCA) of normalized (rlog) gene expression data demonstrated some separation of stable cases from the progressed and de novo metastatic patients (A). The number of differentially expressed genes (DEGs) between comparisons of stable, progressed, and de novo metastatic cases is shown in (B) (adjusted  $p$ -value < 0.05, logfold change > 1/−1). A Venn diagram Venn Diagram demonstrated the overlap of DEGs in (B) by direction, increased (left panel) between progressed versus stable cases (PROG vs. STABLE UP) and de novo metastatic versus stable cases (MET vs. STABLE UP), and decreased (right panel) between progressed versus stable cases (PROG vs. STABLE DOWN) and de novo metastatic versus stable cases (MET vs. STABLE DOWN) (C). This analysis identified common DEGs genes between groups (C). Heatmap of DEGs in stable versus progressed cases (adjusted  $p$ -value < 0.05). Hierarchical clustering (ward.D2 method and euclidean distance) demonstrated clustering of the progressed and de novo metastatic cases, which have a similar gene expression pattern (D). [Color figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com/terms-and-conditions)]

NanoString analysis performed on 12 cases (4 from each of the 3 groups) demonstrated similar differences between sPCa cases and those with either pPCa or mPCa. PCA demonstrated separation of sPCa from pPCa and mPCa cases (except for one metastatic case) (Figure 3A). The top 25 DEGs by nSolver™ analysis between sPCa and pPCa cases ( $p$  < 0.05) (Figure 3B) demonstrated a similar expression pattern difference to that observed between sPCa and mPCa samples (Figure 3C), however, this result was not statistically significant on correction for multiple testing (Supporting Information:

Table S2). Nonhierarchical clustering of log-normalized nanoString expression data for these genes demonstrated that sPCa cases clustered with one mPCa case (Figure 3B). Similar findings were observed in the comparison of sPCa versus mPCa cases (Figure 3C). Sixteen of the top 25 DEGs (64%) were common between the two comparisons (Figure 3D).

A comparison of nanoString expression (log-normalized) and 3' RNAseq expression (rlog-normalized) yielded an overall Spearman correlation coefficient of 0.68. We identified an overlap of





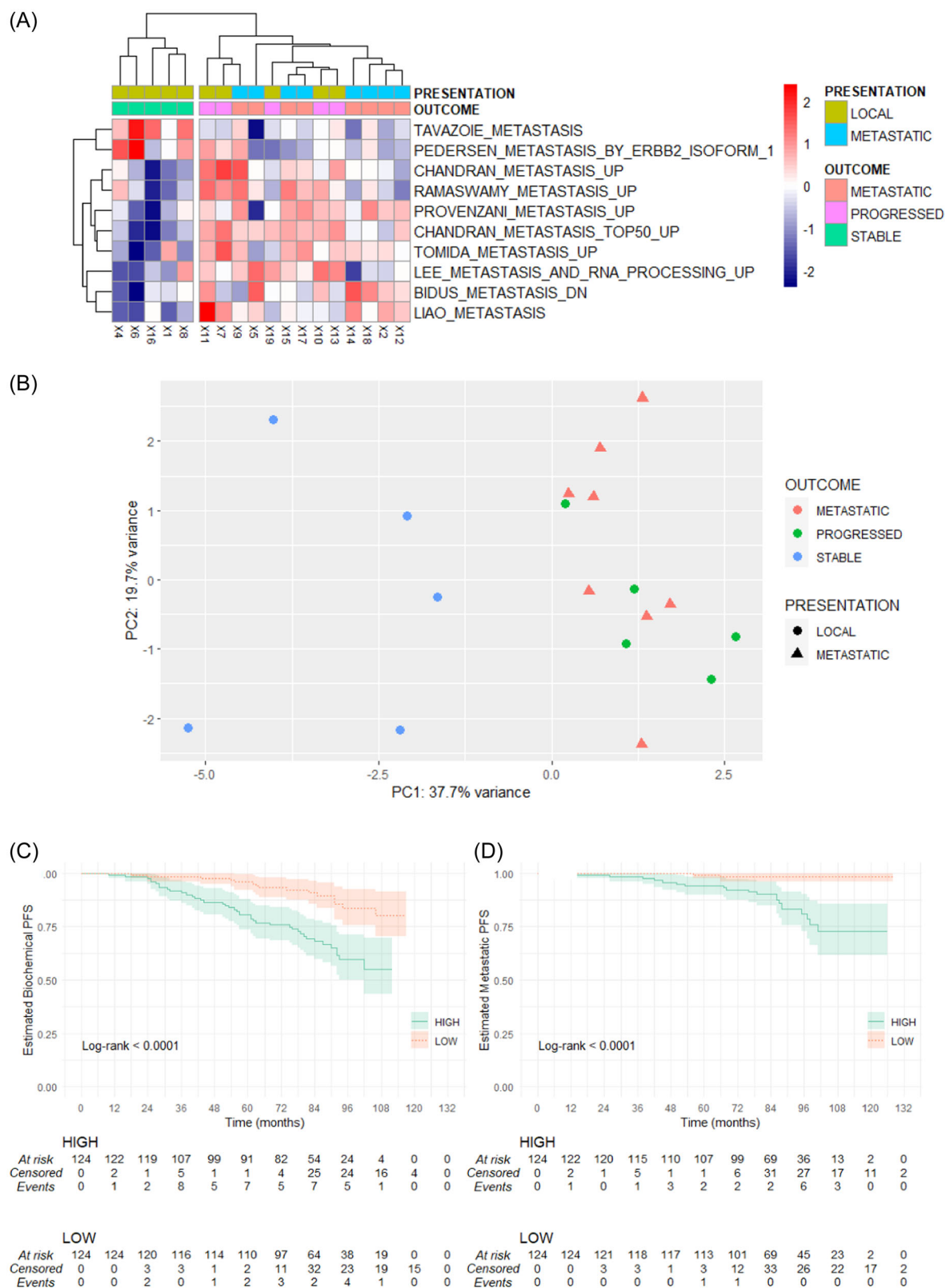
**FIGURE 3** NanoString analysis of the cohort. Principal component analysis (PCA, 1:2) of normalized (log) nanoString gene expression data demonstrated some separation of stable cases from the progressed and de novo metastatic cases (A). Heatmap of the top 25 differentially expressed genes (DEGs) identified by nanoString nSolver analysis between stable and progressed cases ( $p < 0.05$ ) (B). Nonhierarchical clustering (ward.D2 method and euclidean distance) demonstrated clustering of progressed and de novo metastatic patients with a similar expression pattern (B). A heatmap of the top 25 DEGs ( $p < 0.05$ ) between de novo metastatic and stable cases (identified in B) demonstrated clustering and similarity of the progressed and de novo metastatic cases (C). A Venn diagram demonstrated the overlap between the top 25 DEGs between progressed versus stable cases (PROG\_V\_STABLE), and de novo metastatic versus stable cases (MET\_V\_STABLE) (D). [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

4 directionally concordant DEGs between sPCa and pPCa cases (increased DEGs: *GNAS*, *ETV1*, *COL2A1*; decreased DEGs: *HDAC5*) in both the 3'RNAseq (utilizing less stringent cutoffs,  $p_{adj} < 0.2$ ) and nanoString ( $p < 0.05$ ) platforms (Supporting Information: Table S2A).

ssGSEA of Quantseq data identified DEGs in pathways associated with metastasis, centrosome, and methylation pathways in sPCa cases versus combined pPCa and mPCa cases, with only the centrosome pathways containing overlapping sets of genes (Supporting Information: Figures S6 and S7, Table S3).

Taking forward the observation that gene expression profiles were similar in pPCa and mPCa cases, ssGSEA was performed using a previously validated subset of metastatic signatures and pathways from the Molecular Signatures Database (MSigDB) website (<https://www.gsea-msigdb.org/gsea/msigdb>). Differential expression analysis was performed to compare sPCa versus pPCa, which identified 10 statistically significant signatures ( $p_{adj} < 0.1$ ) as visualized in the heatmap (Figure 4A). PCA of ssGSEA scores demonstrated separation of sPCa cases versus pPCa and mPCa cases (Figure 4B).

To explore the potential biological relevance of the DEGs from the 3'RNAseq data, the expression of the top 50 increased protein coding genes in pPCa versus sPCa cases, which were also increased in mPCa versus sPCa cases (Figure 2D), was explored in the large external Jain RT data set,<sup>11</sup> which contains data for 248 PCa patients treated with RT with clinical follow-up. Probes corresponding to the top 50 increased protein-coding genes were selected (37 genes were represented in the data set by 116 probes, see Supporting Information: Methods), and the cohort was divided into *high* and *low* expression cohorts (using the median as threshold for mean expression of all 37 genes). Biochemical recurrence-free and metastatic progression-free survival curves were observed to be significantly ( $p < 0.001$ ) different between the *high* and *low* expression cohorts (Figure 4C,D). A significantly increased hazard ratio (HR) in the *high* expression cohort for biochemical (HR: 2.5 [1.4–4.3],  $p < 0.003$ ) and metastatic (HR: 5.1 [1.7–15],  $p < 0.004$ ) progression-free survival was observed on univariable analysis in a cox proportional hazards model (Supporting Information: Figures S8 and S9). Signature performance for biochemical recurrence-free



**FIGURE 4** Analysis of findings against a previously validated subset of metastatic signatures and pathways. Guided single sample gene set analysis (ssGSEA) identified differences in metastasis pathways between stable versus progressed and de novo metastatic cases (metastatic signatures subset of C2 MsigDB curated pathways, adjusted  $p$ -value < 0.1) (A). Principle component analysis (PCA) of metastasis signature ssGSEA scores defined in (A) demonstrated separation of stable versus progressed and de novo metastatic cases (B). Thirty-six of the top 50 protein-coding differentially expressed genes (DEGs) increased in the progressed versus stable ProMPT cases were then analyzed in the external Belfast radiotherapy-treated data set (GSE116918). This demonstrated an association between the median expression of these DEGs (divided into HIGH vs. LOW by median expression of all probes) and biochemical progression-free survival (C) and metastatic progression-free survival (D). [Color figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com/doi/10.1002/pros.24715)]

survival and metastatic progression-free survival was significant compared to random signatures, cancer signatures, and permutations of survival and feature data performed using the SigCheck package with 1000 iterations (Supporting Information: Figures S10–S12). Four genes (*CDC25A*, *OLR1*, *CDON*, and *DDX39B*) were found to be independently associated with biochemical progression-free survival (lower limit of HR > 1) in a cox proportional hazards model (Supporting Information: Figure S8A). A multivariable Cox proportional hazard model sequentially incorporating clinicopathological features using clinically relevant cutoffs (T-stage, Gleason score [dichotomized as sum score 6–7 or sum score 8–10]) and PSA (dichotomized as <20, ≥20) and high versus low expression of the four genes (mean expression of all four genes with the median value used as threshold to divide cohort) demonstrated a HR of 3.47 [1.79–6.7] ( $p < 0.001$ ) for high versus low expression cohorts (Supporting Information: Table S4 and Figure S8B). Estimation of time-dependent ROC demonstrated the ability of this 4-gene signature (expressed as mean expression of all four genes), with PSA and Gleason scores, as continuous variables in a Cox proportional hazards model to predict for biochemical area under the curve (AUC: 76.8) and metastatic (AUC: 82.9) progression-free survival (package TimeROC) (Supporting Information: Figure S9). The associations of individual genes with clinical characteristics are shown in Supporting Information: Table S5.

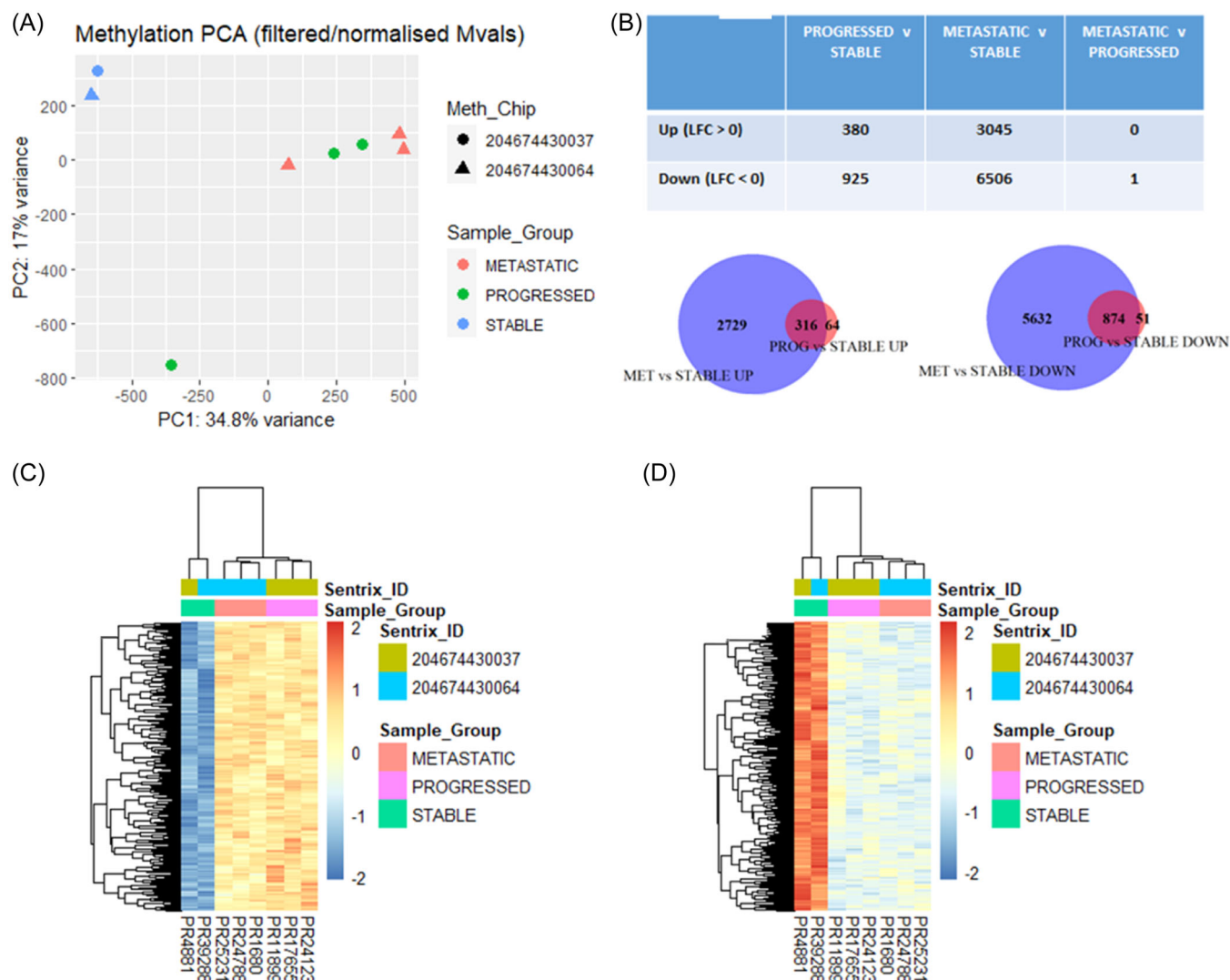
Methylation analysis was performed on eight samples (two with sPCa, three with pPCa, and three with mPCa) on two illumina 850k EPIC arrays. Following normalization and filtering of methylation data, PCA identified similar findings to those observed in the 3' RNAseq data, with separation of the sPCa cases from pPCa and mPCa cases (Figure 5A). Differential methylation analysis (performed with minfi package v1.32.0) identified 1305 probes to be significantly differentially methylated in sPCa versus pPCa cases ( $p_{\text{adj}} < 0.05$ ), and 9551 probes to be significantly differentially methylated in sPCa versus mPCa cases. Most probes (94.5%, 874 of 925 probes) hypomethylated in sPCa versus pPCa cases were also hypomethylated in sPCa versus mPCa cases. Most probes (83.1%, 316 of 380 probes) hypermethylated in sPCa versus pPCa cases were hypermethylated in pPCa versus mPCa cases (Figure 5B). This similarity is demonstrated by only one probe being significantly differentially methylated between pPCa and mPCa cases and is visualized in heatmaps of the coherently differentially methylated probes (Figure 5C,D).

Analysis of differentially methylated regions (DMRs) was performed to identify whether the differentially methylated probes correspond to known genes. Over 600 DMRs were identified in a comparison of sPCa versus pPCa or mPCa cases, including genes of biological relevance in PCa such as *GNAS* and *AR* (Supporting Information: Figures S13 and S14). The DMR plots demonstrate pPCa and mPCa cases had similar mean methylation patterns, distinct from sPCa cases. The coherently differentially methylated probes (Figure 5B) were mapped to their genomic location to identify DMRs, to explore whether they were specific to individual chromosomes. The outer track of Figure 6A demonstrates the chromosomal location

of hypomethylated (blue) and hypermethylated (red) regions, with the largest peak occurring at chromosome 19. The Rainfall plot middle track of Figure 6A shows the genomic coordinates of each region, with the y-axis corresponding to the minimum distance to neighboring regions, demonstrating clustering of differentially methylated probes at chromosome 19. To investigate whether the corresponding genes are part of specific pathways, an overrepresentation analysis was performed, which identified pathways associated ( $p_{\text{adj}} < 1 \times 10^{-14}$ ) with RNA pol II cis-regulatory region sequence-specific DNA binding, DNA binding transcription factor activity gene ontology and Herpes Simplex 1 infection kyoto encyclopedia of genes and genomes (Supporting Information: Figures S15–S18). To explore the relationship between differential methylation and gene expression, the coherently hypomethylated probes were filtered between pPCa and mPCa cases, versus sPCa cases (Figure 5B), to identify potential association with gene promoters. This data was integrated with 3'RNAseq (rlog normalized) expression data for the corresponding 84 genes to generate combined methylation and expression heatmaps and density plots. Unsupervised hierarchical clustering of the differentially methylated probes corresponding to promoter regions demonstrated clustering of sPCa cases versus pPCa and mPCa cases (Figure 6B, Top Panel). Most of the probes corresponded to open chromatin regions and DNase I hypersensitivity sites. The methylation density heatmap (Figure 6B, Second Panel) visualized the distribution of methylation (mVals), and this was similar and compact for pPCa and mPCa cases. 3'RNAseq expression of genes corresponding to the promoter-related probes (Figure 6B, Third Panel) demonstrated variation in the corresponding gene expression, particularly in the sPCa samples. Density plots of gene expression demonstrated that the expression of genes corresponding to hypomethylated promoters was more tightly distributed in the pPCa and mPCa samples, whereas gene expression in the sPCa samples (where the promoter-associated probes were comparatively hypermethylated) showed greater variation (Figure 6B, Bottom Panel). These findings were similar in the RNA expression of differentially methylated genes in the full cohort (Supporting Information: Figure S19). Overrepresentation analysis of hypo- and hypermethylated genes, performed using two separate platforms (enrichr and gprofiler), to compare sPCa cases versus pPCa and mPCa cases, identified pathways associated with histone H3 lysine 4 trimethylation, histone H3 lysine 27 trimethylation, and Polycomb Repressive Complex 2 (PRC) (Supporting Information: Figure S20). To explore the relationship between Enhancer of Zeste 2 (EZH2) expression and expression of EZH2-regulated genes in sPCa versus pPCa and mPCa cases, it was observed that a set of genes previously identified as being regulated by EZH2<sup>27</sup> were DEGs between these clinical groups of cases ( $p_{\text{adj}} < 0.05$ ) (Supporting Information: Figure S21).

While this proof-of-concept study did not specially aim to identify a gene signature prognostic for clinical outcome post-RT, a set of genes was demonstrated to be prognostic for post-RT outcome in a large external data set (Supporting Information: Figure S22). This set of genes was distinct from gene sets previously described in the literature and warrants further investigation.





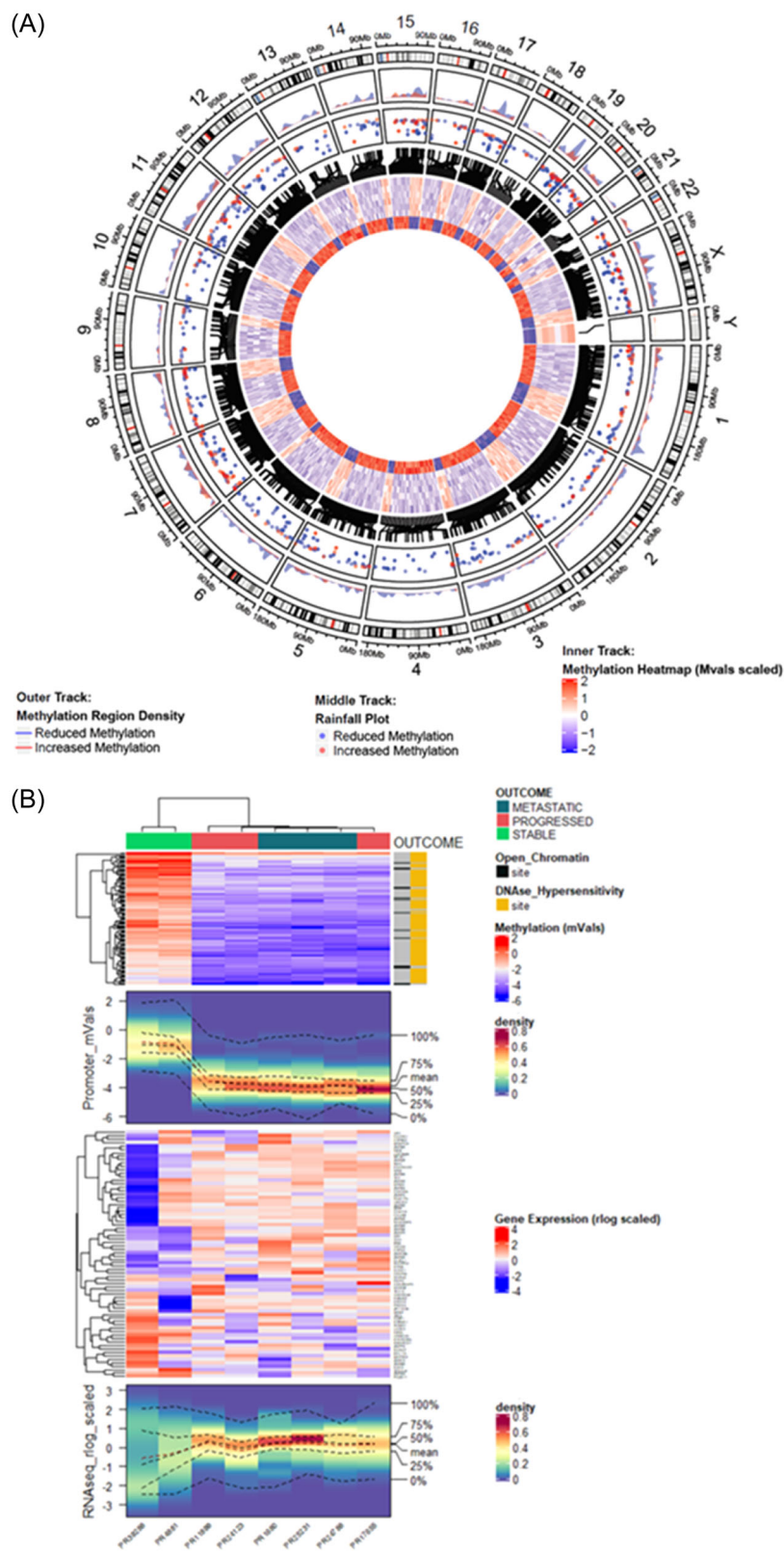
**FIGURE 5** DNA methylation analysis of the cohort. Principal component analysis (PCA, postnormalization and filtering) demonstrated some separation of stable cases from the progressed and de novo metastatic cases by Principal Component 1 (A). Analysis of differential DNA methylation (postnormalization and filtering) demonstrated minimal differential methylation (adjusted  $p$ -value < 0.05) between progressed and de novo metastatic cases, and directionally coherent overlap between progressed versus stable cases, and between de novo metastatic versus stable cases, based on the differentially methylated probes (Mvals) (B). A heatmap demonstrated overlap of differentially methylated probes from (B), increased (LFC > 1) (C) and decreased (LFC < -1) (D) in progressed and de novo metastatic versus stable cases. [Color figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com/terms-and-conditions)]

## 4 | DISCUSSION

This proof-of-concept study tested the possibility that orthogonal genomic analyses (3'RNAseq, nanoString, and DNA methylation) can identify molecular features associated with PCa progression in the baseline historical pretreatment FFPE biopsy samples with long-term clinical follow-up. This unlocks the potential to investigate powerful large clinical cohorts with long-term follow-up, such as the ProtecT cohort.<sup>1</sup> It raises the potential to utilize molecular features of prostate biopsy features in clinical decision making, to aid risk-stratification and personalized medicine approaches in the post-diagnostic space. Both retrospective studies of large historical

cohorts, and prospective studies to investigate the added value of this approach in the clinical setting, are now warranted.

Transcriptional signatures prognostic of metastatic recurrence have been described in PCa, and some of these have been validated in patients treated with RT.<sup>5-11</sup> However, most of these studies, except for the "hypoxia signature," describe signatures derived from radical prostatectomy specimens rather than from pretreatment prostate biopsies from patients undergoing RT. DNA methylation is altered in PCa development and progression, however, while prognostic biomarkers have been developed this approach has not been specifically used for RT-treated patients.<sup>16,17,28</sup>



**FIGURE 6** Analysis of coherently differentially methylated probes versus their genomic location and versus gene expression data. A differential methylation density plot of coherently differentially methylated regions between progressed and de novo metastatic cases versus stable cases (A). The outer track density plot demonstrated the fraction of the genomic window covered by differentially methylated regions in progressed and de novo metastatic cases versus stable cases (hypomethylated regions in blue, hypermethylated regions in red). The middle track rainfall plot demonstrated the genomic coordinates of each region (the y-axis corresponding to the minimum distance to the neighboring region). The inner track circular heatmap demonstrated coherently differentially methylated probes (Mvals scaled). The outer three lanes correspond to de novo metastatic cases; the middle three lanes correspond to progressed cases; the inner two lanes correspond to stable cases. A combined promoter methylation and gene expression analysis is shown in (B). The top panel methylation heatmap demonstrated promoter-associated probes hypomethylated in progressed and de novo metastatic cases versus stable cases. Open chromatin regions (OCRs) and DNase I hypersensitivity sites (DHSs) are annotated, demonstrating promoter-associated CpG sites as either DHSs or OCRs. The second panel methylation density heatmap visualizes the distribution of Mvals, with clustering as per the top panel. The third panel demonstrated the expression (Quantseq) of genes corresponding to CpGs in the upper two panels; gene expression is normalized (rlog – DESeq. 2) and scaled. The fourth and final panel is a gene expression density heatmap displayed as density of Quantseq expression values. [Color figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com/terms-and-conditions)]

Large clinical data sets, such as the DECIPHER GRID project (NCT02609269) which contains transcriptomic information from 20,000 patients, are required to reliably identify molecular features associated with disease behavior. The results of this proof-of-concept study highlight the opportunity for future larger studies to obtain

molecular data from historical diagnostic prostate biopsy samples as a method of identifying key molecular features of high-risk disease. It is demonstrated herein that 3'RNAseq, nanoString, and methylation analysis each have the capacity to achieve this, and to identify baseline molecular features associated with development of mPCa.

One sample in this analysis failed 3'RNAseq, with all other analyses achieving useable data. One sample achieved fewer counts than others (500,000 vs. ~8 million), however, this sample was included in the analysis for several reasons. First, 3'RNAseq only measures 3' transcripts and produces counts for a whole transcript, whereas standard RNA sequencing measures multiple fragments across the same gene transcript. Second, differential expression analysis tools can account for differing library sizes. Third, the analysis was also performed with this sample excluded, with minimal change to the results. In addition, to explore the impact of mapping techniques, the analysis was repeated utilizing Subread Aligner versus STAR with this sample both included and excluded, achieving similar results.

The 3'RNAseq DEGs analysis demonstrated several broad similarities in gene expression between pPCa cases and mPCa cases, both in terms of the PCA, the pattern, number and directional coherence of DEGs between groups, and the nonhierarchical clustering analysis. Two groups of samples (pPCa and mPCa) were clearly distinct from sPCa cases. This observation suggests that there is shared underlying molecular biology between pPCa and mPCa cases, that can be identified in archival FFPE pretreatment prostate biopsies using these techniques. This may potentially be explained by performance limitations of radiological staging at the time of original diagnosis of the historical cases with long-term follow-up selected for this proof-of-concept study. For example, radionuclide bone scans were used for baseline staging as part of risk-stratification, rather than the more recently developed MRI marrow or PSMA PET CT scans. It is possible that the original TNM staging of these cases, using imaging modalities available at the time, led to under-staging of some cases. The cases with pPCa may have had undetectable micro-metastases at the time of original diagnosis and treatment, thus accounting for these baseline prostate biopsy samples having similar molecular features compared to the baseline mPCa cases, with both of these two groups of patient samples being clearly distinct from sPCa cases. Nevertheless, it is interesting that such differences can be identified in the relatively small amount of genomic material available from archival FFPE prostate biopsy samples. This raises the exciting possibility that the added value of molecular analysis approach could be taken forward for evaluation in prospective clinical studies investigating the utility of this approach in treatment decision making. This would aim to improve risk stratification and clinical outcomes for patients, and warrants investigation in future studies.

The similarities in RNA expression in pPCa and mPCa cases, compared with sPCa cases, observed in the 3'RNAseq analysis, were also generally observed in the nanoString analysis using PCA and gene expression tools. However, in the nanoString analysis, there is incomplete separation of the sPCa cases, with one mPCa case clustering with the sPCa cases. This may potentially be explained by the nanoString platform being a targeted panel containing significantly fewer genes than those used for 3'RNAseq (~800 vs. >20,000). Moreover, the two techniques have different mechanisms of action, with nanoString utilizing reporter probes to hybridize mRNA, and 3'RNAseq sequencing the 3' end of transcripts, which

could be potentially affected by alternative splicing. The Spearman correlation between 3'RNAseq (rlog normalized) expression and nanoString (log normalized) expression for the subset of nanoString genes was strong at 0.68, and per-sample correlations utilizing different normalization techniques for 3'RNAseq were similar.

While the demonstrated DEGs ( $p < 0.05$ ) in the 3'RNAseq analysis did not achieve statistical significance on correction for multiple testing using the nSolver™ nanoString analysis, the results presented herein are scaled values of normalized expression, which ought to be considered to broadly support the findings of the 3' RNAseq analysis rather than be utilized in isolation. The use of these orthogonal RNA sequencing techniques demonstrates similar findings across these transcriptional analysis platforms comparing pPCa (and mPCa) cases versus sPCa cases in this proof-of-concept study.

Overrepresentation analysis identified Gene Ontology pathways associated with spindle pole and centrosome function to be increased in pPCa and mPCa cases, versus sPCa cases, and centrosome pathways were also identified on ssGSEA, potentially due to increased mitotic activity in those samples. An RNA expression-based CCP score is independently prognostic of metastatic progression after RT, although there is minimal overlap of genes between genesets.<sup>7,18</sup> Increased expression of spindle pole genes in pPCa cases could be due to increased reliance on the spindle assembly checkpoint, due to loss of other cell checkpoints.<sup>29,30</sup> Increased expression of centrosome pathway genes could be due to centrosome amplification in primary tumor samples of patients who develop metastatic disease. Previous studies have demonstrated in situ centrosome loss in primary PCa samples,<sup>31</sup> however, in PCa cell lines, centrosome amplification has been observed in PCa cell lines with increased metastatic behavior.<sup>32</sup> Both observations may potentially be related to increased chromosomal instability and aneuploidy.

Four genes (*GNAS*, *ETV1*, *COL2A1*, and *HDAC5*) were observed to be coherently and significantly differentially expressed between sPCa and pPCa cases using both 3'RNAseq ( $p_{\text{adj}} < 0.05$ ) and nanoString platforms (less stringent  $p < 0.2$ ), however, only *COL2A1* and *ETV1* were prognostic upon application of a cox proportional hazards model. An integrative clinical genomic study previously demonstrated *GNAS* to be one of the most frequently mutated genes in advanced PCa.<sup>31</sup> *ETV1* is a transcription factor frequently overexpressed in aggressive PCa via a chromosomal translocation with androgen-responsive promoters,<sup>33</sup> and *ETV1* has been demonstrated to initiate PCa tumorigenesis in concert with the JMJD2 histone demethylase.<sup>34</sup> *COL2A1* is a candidate PCa risk gene,<sup>35</sup> and the *HDAC5* histone deacetylase gene is frequently downregulated or deleted in PCa, resulting in increased H3K27 acetylation and impaired RB-mediated repression of cell cycle-related pro-oncogenic genes.<sup>36</sup> In this proof-of-principle study data set, *HDAC5* expression is reduced in sPCa versus pPCa cases analyzed with both RNA expression techniques. Analysis of an external data set of RT-treated patients<sup>11</sup> demonstrates that these four genes (*GNAS*, *ETV1*, *COL2A1*, and *HDAC5*) are prognostic for biochemical progression-free survival, though not for metastatic progression-free survival.

Features obtained from 10 previously described metastatic signatures, from various other cancer types, were observed to be differentially expressed upon ssGSEA analysis between sPCa cases and pPCa and mPCa cases. This clear separation of samples by PCA using genes specifically associated with metastasis identifies similarities in samples from patients with pPCa and mPCa, distinct from sPCa cases. This was further demonstrated utilizing the top 50 DEGs (of which 36 are represented) from a comparison of sPCa versus pPCa cases in a large external data set.<sup>11</sup> It is noteworthy that there is no overlap between these genes and those in the Prolaris, Oncotype, Decipher, Metastatic Assay, and Prostate Hypoxia gene lists. Separation of the cohort by the median expression into HIGH and LOW groups was observed to be prognostic for both biochemical and metastatic progression-free survival, while these genes were more prognostic than would be expected from a random set of genes and selected cancer signatures. Four genes (*CDC25A*, *OLR1*, *CDON*, and *DDX39B*) were demonstrated to be independently prognostic, and each of these has been previously reported to be associated with PCa progression and/or metastasis.<sup>37–40</sup> While it is a limitation of this proof-of-concept study that the analysis is underpowered for biomarker discovery, these results demonstrate that 3'RNAseq analysis of historical FFPE prostate biopsy samples with long-term follow-up can identify biologically relevant genes which can be validated using a large external data set.

DNA hypomethylation has been described in PCa, and recent studies in advanced disease have identified specific areas of increased hypomethylation during progression from benign tissue to localized PCa to metastatic disease.<sup>41</sup> The findings in this study are consistent with this phenomenon, with similar patterns of methylation being observed in pPCa and mPCa cases, and increased methylation in mPCa versus sPCa cases, compared to pPCa versus sPCa cases.

Methylation analysis demonstrated separation of sPCa cases and pPCa/mPCa cases using both PCA and DEGs analysis, with the majority of differential gene methylation being coherent and in the same direction, accepting that a limitation of this proof-of-concept study is the relatively small number of patient samples. Previous studies have demonstrated similar methylation patterns for primary and metastatic tumor samples from the same patient,<sup>42</sup> supporting our observation that methylation characteristics of metastatic disease can be found in primary tumor biopsies. These DNA methylation results provide additional evidence beyond the 3'RNAseq and nanoString data of biological similarities in baseline samples from patients with pPCa and mPCa at both the methylomic and transcriptomic level.

The methylation technique used in this study, using 850K methylation probes, has the granularity to identify multiple areas of methylation within a single gene. We identified specific differentially methylated DNA regions which map to genes of known PCa biological relevance, including *AR* and *GNAS*. *AR* hypomethylation has been observed in mPCa,<sup>42,43</sup> and *GNAS* was hypomethylated and upregulated in pPCa/mPCa samples versus sPCa samples. The relationship between DNA methylation and gene expression is

complex, with increased gene expression being associated with hypermethylation.<sup>44</sup> Our analysis of differential methylation plots demonstrated that pPCa and mPCa samples were hypomethylated at the start of genes, and hypermethylated in other regions, compared with sPCa cases. In the case of *GNAS*, we also observed areas of differential methylation within the same gene.<sup>44</sup>

The observation that the greatest focus of hypomethylation in pPCa and mPCa samples versus sPCa samples was on chromosome 19 is interesting, given that RNA Pol II-associated chromatin interactions have been identified as determinants of transcriptional regulation in PCa.<sup>45</sup> Moreover, there is significant overrepresentation, on pathway enrichment analysis, of associated genes at the site of peak differential methylation density on chromosome 19. It is also noteworthy that RNA Pol II interactions frequently involve H3K4m3 and/or H3K27 acetylation marks.<sup>45</sup> The overrepresentation analysis of differentially methylated probes identified pathways associated with H3K4m3, H3K27m3, and PRC2, and these pathways appeared in both hyper- and hypomethylated genes, demonstrating the complexity of methylation events in different areas of the same gene. The Polycomb Group protein EZH2 in the PRC2 complex catalyzes H3K27m3 on target gene promoters, and EZH2 function has previously been associated with mPCa progression, and with metastatic progression post-RT.<sup>25,46–49</sup> We observed several genes associated with EZH2 to be significantly differentially expressed between pPCa/mPCa cases and sPCa cases.

## 5 | CONCLUSION

This study demonstrates the potential for molecular analysis of diagnostic baseline prostate biopsy samples as a tool to characterize PCa beyond the current method of risk-classification, ahead of potential curative or systemic therapy, with added value in terms of identifying patients with occult micro-metastatic disease. This warrants further investigation in both retrospective studies using larger cohorts, and prospective studies designed to investigate the use of these molecular techniques in the clinic. Taken together, the results of this proof-of-concept study demonstrate that we can now unlock the potential wealth of information that can be gained from molecular analyses of powerful large historical cohorts with baseline FFPE prostate biopsy samples and associated long-term clinical follow up. This approach may aid future risk-stratification and treatment selection in the postdiagnostic space for men with this common malignancy.<sup>50</sup>

## ACKNOWLEDGMENTS

Philip Vincent Charlton was funded by a Cancer Research UK DPhil studentship. Richard John Bryant receives research funding from the Urology Foundation, the John Black Charitable Foundation, the Rosetrees Trust, Prostate Cancer UK, and NIHR-HTA, and was funded by Cancer Research UK (C39297/A22748) during the conduct of this research. Francesca Meteora Buffa receives research funding from a European Research Council (ERC) Consolidator Award



(MICROC:772970). Geoff Stuart Higgins acknowledges research funding from the Cancer Research UK RadNet Oxford Center (C6078/A28736). Alastair David Gordon Lamb receives research funding from Cancer Research UK (C57899/A25812), the Oxford NIHR BRC Surgical Innovation & Evaluation theme, and John Black Charitable Foundation. Ian Geoffrey Mills receives research funding from the Rosetrees Trust, Prostate Cancer UK and the John Black Charitable Foundation. Freddie Charles Hamdy receives research funding from Cancer Research UK, Prostate Cancer UK, and NIHR-HTA. Clare Verrill's research time is supported by the NIHR Oxford Biomedical Research Center. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR, or the Department of Health. This project was funded by a Cancer Research UK Development Fund grant issued by the CRUK Oxford Cancer Center.

### CONFLICT OF INTEREST STATEMENT

Alastair David Gordon Lamb has received education support from Astellas, Lilly, Astrazenaca, and Ipsen, and is a stipendiary BJUI Section Editor for prostate cancer, has received honoraria for reviewing for European Urology and Lancet Oncology, and has received consulting fees from AlphaSights. The remaining authors declare no conflict of interest.

### DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

### ORCID

Ian Geoffrey Mills  <http://orcid.org/0000-0001-5347-5083>

Richard John Bryant  <http://orcid.org/0000-0002-8330-9251>

### REFERENCES

- Hamdy FC, Donovan JL, Lane JA, et al. Fifteen-year outcomes after monitoring, surgery, or radiotherapy for prostate cancer. *N Engl J Med*. 2023;388(17):1547-1558.
- Donovan JL, Hamdy FC, Lane JA, et al. Patient-reported outcomes 12 years after localized prostate cancer treatment. *NEJM Evid*. 2023;2(4):EVIDoA2300018.
- D'Amico AV. Biochemical outcome after radical prostatectomy, external beam radiation therapy, or interstitial radiation therapy for clinically localized prostate cancer. *JAMA*. 1998;280(11):969-974.
- Cooperberg MR, Pasta DJ, Elkin EP, et al. The University of California, San Francisco Cancer of the Prostate Risk Assessment score: a straightforward and reliable preoperative predictor of disease recurrence after radical prostatectomy. *J Urol*. 2005;173(6):1938-1942.
- Klein EA, Cooperberg MR, Magi-Galluzzi C, et al. A 17-gene assay to predict prostate cancer aggressiveness in the context of Gleason grade heterogeneity, tumor multifocality, and biopsy undersampling. *Eur Urol*. 2014;66:550-560.
- Nakagawa T, Kollmeyer TM, Morlan BW, et al. A tissue biomarker panel predicting systemic progression after PSA recurrence post-definitive prostate cancer therapy. *PLoS One*. 2008;3:e2318.
- Cuzick J, Swanson GP, Fisher G, et al. Prognostic value of an RNA expression signature derived from cell cycle proliferation genes in patients with prostate cancer: a retrospective study. *Lancet Oncol*. 2011;12:245-255.
- Jairath NK, Dal Pra A, Vince Jr., R, et al. A systematic review of the evidence for the decipher genomic classifier in prostate cancer. *Eur Urol*. 2021;79(3):374-383.
- Walker SM, Knight LA, McCavigan AM, et al. Molecular subgroup of primary prostate cancer presenting with metastatic biology. *Eur Urol*. 2017;72:509-518.
- Yang L, Roberts D, Takhar M, et al. Development and validation of a 28-gene hypoxia-related prognostic signature for localized prostate cancer. *EBioMedicine*. 2018;31:182-189.
- Jain S, Lyons CA, Walker SM, et al. Validation of a metastatic assay using biopsies to improve risk stratification in patients with prostate cancer treated with radical radiation therapy. *Ann Oncol*. 2018;29(1):215-222.
- Janes JL, Boyer MJ, Bennett JP, et al. The 17-gene genomic prostate score test is prognostic for outcomes after primary external beam radiation therapy in men with clinically localized prostate cancer. *Int J Radiat Oncol Biol Phys*. 2023;115(1):120-131.
- Tward J, Lenz L, Flake II, DD, et al. The clinical cell-cycle risk (CCR) score is associated with metastasis after radiation therapy and provides guidance on when to forgo combined androgen deprivation therapy with dose-escalated radiation. *Int J Radiat Oncol Biol Phys*. 2022;113(1):66-76.
- Freedland SJ, Gerber L, Reid J, et al. Prognostic utility of cell cycle progression score in men with prostate cancer after primary external beam radiation therapy. *Int J Radiat Oncol Biol Phys*. 2013;86(5):848-853.
- Nguyen PL, Martin NE, Choeurng V, et al. Utilization of biopsy-based genomic classifier to predict distant metastasis after definitive radiation and short-course ADT for intermediate and high-risk prostate cancer. *Prostate Cancer Prostatic Dis*. 2017;20(2):186-192.
- Jeyapala R, Kamdar S, Olkhov-Mitsel E, et al. An integrative DNA methylation model for improved prognostication of postsurgery recurrence and therapy in prostate cancer patients. *Urol Oncol*. 2020;38(2):39.e1-39.e9.
- Savio AJ, Kamdar S, Jeyapala R, et al. Methylation markers in prostate biopsies are prognosticators for late biochemical recurrence and therapy after surgery in prostate cancer patients. *J Mol Diagn*. 2020;22:30-39.
- Spohn SKB, Draulans C, Kishan AU, et al. Genomic classifiers in personalized prostate cancer radiation therapy approaches: a systematic review and future perspectives based on international consensus. *Int J Radiat Oncol Biol Phys*. 2023;116(3):503-520.
- Mottet N, van den Bergh RCN, Briers E, et al. EAU-EANM-ESTRO-ESUR-SIOG guidelines on prostate cancer-2020 update. Part 1: screening, diagnosis, and local treatment with curative intent. *Eur Urol*. 2021;79(2):243-262.
- Eastham JA, Auffenberg GB, Barocas DA, et al. Clinically localized prostate cancer: AUA/ASTRO guideline, part I: introduction, risk assessment, staging, and risk-based management. *J Urol*. 2022;208(1):10-18.
- Dobin A, Davis CA, Schlesinger F, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29(1):15-21.
- Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq. 2. *Genome Biol*. 2014;15(12):550.
- Hänzelmann S, Castelo R, Guinney J. GSEA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics*. 2013;14:7.
- Dhawan A, Barberis A, Cheng WC, et al. Guidelines for using sigQC for systematic evaluation of gene signatures. *Nat Protoc*. 2019;14(5):1377-1400.
- Stark R, Norden J. SigCheck: check a gene signature's prognostic performance against random signatures, known signatures, and permuted data/metadata. R package version 2.32.0; 2023.
- Blanche P, Dartigues JF, Jacqmin-Gadda H. Estimating and comparing time-dependent areas under receiver operating characteristic



- curves for censored event times with competing risks. *Stat Med*. 2013;32(30):5381-5397.
27. Xu K, Wu ZJ, Groner AC, et al. EZH2 oncogenic activity in castration-resistant prostate cancer cells is Polycomb-independent. *Science*. 2012;338(6113):1465-1469.
  28. Massie CE, Mills IG, Lynch AG. The importance of DNA methylation in prostate cancer development. *J Steroid Biochem Mol Biol*. 2017;166:1-15.
  29. Canter DJ, Freedland S, Rajamani S, et al. Analysis of the prognostic utility of the cell cycle progression (CCP) score generated from needle biopsy in men treated with definitive therapy. *Prostate Cancer Prostatic Dis*. 2020;23(1):102-107.
  30. Matthews HK, Bertoli C, de Bruin RAM. Cell cycle control in cancer. *Nat Rev Mol Cell Biol*. 2022;23(1):74-88.
  31. Robinson D, Van Allen EM, Wu YM, et al. Integrative clinical genomics of advanced prostate cancer. *Cell*. 2015;161(5):1215-1228.
  32. Wang M, Nagle RB, Knudsen BS, Cress AE, Rogers GC. Centrosome loss results in an unstable genome and malignant prostate tumors. *Oncogene*. 2020;39(2):399-413.
  33. Oh S, Shin S, Song H, Grande JP, Janknecht R. Relationship between ETS transcription factor ETV1 and TGF- $\beta$ -regulated SMAD proteins in prostate cancer. *Sci Rep*. 2019;9(1):8186.
  34. Kim TD, Jin F, Shin S, et al. Histone demethylase JMJD2A drives prostate tumorigenesis through transcription factor ETV1. *J Clin Invest*. 2016;126(2):706-720.
  35. Thibodeau SN, French AJ, McDonnell SK, et al. Identification of candidate genes for prostate cancer-risk SNPs utilizing a normal prostate tissue eQTL data set. *Nat Commun*. 2015;6:8653.
  36. Zhou Y, Jin X, Ma J, et al. HDAC5 loss impairs RB repression of pro-oncogenic genes and confers CDK4/6 inhibitor resistance in cancer. *Cancer Res*. 2021;81(6):1486-1499.
  37. Murdocca M, De Masi C, Pucci S, et al. LOX-1 and cancer: an indissoluble liaison. *Cancer Gene Ther*. 2021;28(10-11):1088-1098.
  38. González-Chavarría I, Cerro RP, Parra NP, et al. Lectin-like oxidized LDL receptor-1 is an enhancer of tumor angiogenesis in human prostate cancer cells. *PLoS One*. 2014;9(8):e106219.
  39. Hayashi T, Oue N, Sakamoto N, et al. Identification of transmembrane protein in prostate cancer by the *Escherichia coli* ampicillin secretion trap: expression of CDON is involved in tumor cell growth and invasion. *Pathobiology*. 2011;78(5):277-284.
  40. Nakata D, Nakao S, Nakayama K, et al. The RNA helicase DDX39B and its paralog DDX39A regulate androgen receptor splice variant AR-V7 generation. *Biochem Biophys Res Commun*. 2017;483(1):271-276.
  41. Zhao SG, Chen WS, Li H, et al. The DNA methylation landscape of advanced prostate cancer. *Nat Genet*. 2020;52(8):778-789.
  42. Mundbjerg K, Chopra S, Alemozaffar M, et al. Identifying aggressive prostate cancer foci using a DNA methylation classifier. *Genome Biol*. 2017;18(1):3.
  43. Ylitalo EB, Thysell E, Landfors M, et al. A novel DNA methylation signature is associated with androgen receptor activity and patient prognosis in bone metastatic prostate cancer. *Clin Epigenetics*. 2021;13(1):133.
  44. Rauluseviciute I, Drabløs F, Rye MB. DNA hypermethylation associated with upregulated gene expression in prostate cancer demonstrates the diversity of epigenetic regulation. *BMC Med Genomics*. 2020;13(1):6.
  45. Ramanand SG, Chen Y, Yuan J, et al. The landscape of RNA polymerase II-associated chromatin interactions in prostate cancer. *J Clin Invest*. 2020;130(8):3987-4005.
  46. Varambally S, Dhanasekaran SM, Zhou M, et al. The polycomb group protein EZH2 is involved in progression of prostate cancer. *Nature*. 2002;419(6907):624-629.
  47. Park SH, Fong KW, Mong E, Martin MC, Schiltz GE, Yu J. Going beyond Polycomb: EZH2 functions in prostate cancer. *Oncogene*. 2021;40(39):5788-5798.
  48. Wu X, Scott H, Carlsson SV, et al. Increased EZH2 expression in prostate cancer is associated with metastatic recurrence following external beam radiotherapy. *Prostate*. 2019;79(10):1079-1089.
  49. Zhu J, Jin L, Zhang A, et al. Coexpression analysis of the EZH2 gene using the cancer genome atlas and oncomine databases identifies coexpressed genes involved in biological networks in breast cancer, glioblastoma, and prostate cancer. *Med Sci Monit*. 2020;26:e922346.
  50. Doultosinos D, Mills IG. Derivation and application of molecular signatures to prostate cancer: opportunities and challenges. *Cancers*. 2021;13(3):495.

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Charlton PV, O'Reilly D, Philippou Y, et al. Molecular analysis of archival diagnostic prostate cancer biopsies identifies genomic similarities in cases with progression post-radiotherapy, and those with de novo metastatic disease. *Prostate*. 2024;84:977-990. doi:10.1002/pros.24715