Journal of Pathology
J Pathol 2012; **227:** 286–297
Published online in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1002/path.4047

ORIGINAL PAPER

From sequence to molecular pathology, and a mechanism driving the neuroendocrine phenotype in prostate cancer

Anna V Lapuk,[1†] Chunxiao Wu,[1†] Alexander W Wyatt,[1] Andrew McPherson,[2,3] Brian J McConeghy,[1] Sonal Brahmbhatt,[1] Fan Mo,[1] Amina Zoubeidi,[1] Shawn Anderson,[1] Robert H Bell,[1] Anne Haegert,[1] Robert Shukin,[1] Yuzhuo Wang,[1] Ladan Fazli,[1] Antonio Hurtado-Coll,[1] Edward C Jones,[4] Faraz Hach,[3] Fereydoun Hormozdiari,[3] Iman Hajirasouliha,[3] Paul C Boutros,[5] Robert G Bristow,[6] Yongjun Zhao,[2,7] Marco A Marra,[7] Andrea Fanjul,[8] Christopher A Maher,[9] Arul M Chinnaiyan,[9] Mark A Rubin,[10] Himisha Beltran,[11] S Cenk Sahinalp,[3] Martin E Gleave,[1] Stanislav V Volik[1] and Colin C Collins[1]*

[1] *Vancouver Prostate Centre and Department of Urologic Sciences, University of British Columbia, Vancouver, BC, Canada*
[2] *Centre for Translational and Applied Genomics, BC Cancer Agency, Vancouver, BC, Canada*
[3] *School of Computing Science, Simon Fraser University, Burnaby, BC, Canada*
[4] *Department of Pathology and Laboratory Medicine and Department of Urologic Sciences, University of British Columbia, Vancouver, BC, Canada*
[5] *Informatics and Biocomputing Platform, Ontario Institute for Cancer Research, Toronto, ON, Canada*
[6] *Division of Applied Molecular Oncology, Ontario Institute for Cancer Research, Toronto, ON, Canada*
[7] *Canada's Michael Smith Genome Sciences Centre, BC Cancer Agency, Vancouver, BC, Canada*
[8] *Pfizer Global Research and Development, La Jolla Laboratories, California, USA*
[9] *Michigan Center for Translational Pathology, Ann Arbor, Michigan, USA*
[10] *Department of Pathology and Laboratory Medicine, Weill Cornell Cancer Center, New York, NY, USA*
[11] *Department of Medicine, Weill Cornell Cancer Center, New York, NY, USA*

**Correspondence to: Colin C Collins, 2660 Oak St, Vancouver, BC, V6H 3Z6, Canada. e-mail: ccollins@prostatecentre.com*

[†]These authors contributed equally to this work.

Data availability: Illumina sequence data are available from the Vancouver Prostate Centre website at http://www.lagapc.ca/FTP_Lapuk.html

Abstract

The current paradigm of cancer care relies on predictive nomograms which integrate detailed histopathology with clinical data. However, when predictions fail, the consequences for patients are often catastrophic, especially in prostate cancer where nomograms influence the decision to therapeutically intervene. We hypothesized that the high dimensional data afforded by massively parallel sequencing (MPS) is not only capable of providing biological insights, but may aid molecular pathology of prostate tumours. We assembled a cohort of six patients with high-risk disease, and performed deep RNA and shallow DNA sequencing in primary tumours and matched metastases where available. Our analysis identified copy number abnormalities, accurately profiled gene expression levels, and detected both differential splicing and expressed fusion genes. We revealed occult and potentially dormant metastases, unambiguously supporting the patients' clinical history, and implicated the REST transcriptional complex in the development of neuroendocrine prostate cancer, validating this finding in a large independent cohort. We massively expand on the number of novel fusion genes described in prostate cancer; provide fresh evidence for the growing link between fusion gene aetiology and gene expression profiles; and show the utility of fusion genes for molecular pathology. Finally, we identified chromothripsis in a patient with chronic prostatitis. Our results provide a strong foundation for further development of MPS-based molecular pathology.

Keywords: molecular pathology; massively parallel sequencing; neuroendocrine prostate cancer; REST repressor; chromothripsis

Received 27 February 2012; Revised 24 April 2012; Accepted 24 April 2012

Conflict of interest statement: There is a possible conflict of interest as Andrea Fanjul is a Pfizer employee.

Introduction

Prostate cancer remains the second most common cause of male cancer-related death [1]. Currently, patient stratification and therapy selection are based on histopathology and clinical history. Arguably, the pinnacles of this approach are the pre- and post-operative nomograms [2–5], which predict patient prognosis with high accuracy [2]. However, for many patients nomograms fail, usually with catastrophic consequences, underscoring the need for the development of technologies to aid risk stratification.

Although most of the advanced prostate cancer patients respond to the initial androgen deprivation treatment with surgical or chemical castration, the progression to the fatal castrate-resistant prostate cancer

(CRPC) is inevitable [6]. Although neuroendocrine cancers (small cell carcinoma of the prostate) originating from neuroendocrine cells are rare at initial diagnosis, 40–100% of CRPCs demonstrate evidence of neuroendocrine differentiation (NED) [7], a process whereby adenocarcinoma cells acquire a neuroendocrine phenotype. The neuroendocrine phenotype is intrinsically resistant to androgen deprivation therapies and therefore may confer a poor prognosis; however, the mechanisms underlying its acquisition are only starting to be characterized [8].

Adenocarcinomas can be classified on the molecular level by their genomic landscape and gene fusions involving *ETS* family members [9]. For example, *TMPRSS2–ERG* occurs early in 15–50% of tumours and persists through disease progression [9–11]. While transcriptome complexity beyond *ETS* fusions and gene expression signatures is largely unexplored, recent studies point to the importance of transcriptome structural variation arising from fusions [12] and alternative splicing (AS) [13] in prostate cancer. Androgen receptor (AR) splice variants in CRPC provide a sentinel example [14]. Indeed, splice signatures show great promise for improving tumour classification and providing a novel source for candidate biomarkers and therapeutic targets [15].

The combined massively parallel sequencing (MPS) of genomes (DNA-Seq) and transcriptomes (RNA-Seq) allows comprehensive discovery of deleterious cancer variants. We hypothesized that this approach would not only be valuable for the systematic molecular characterization of prostate cancer, but also serve as an effective molecular diagnostic tool. This report provides compelling support for the validity of this hypothesis.

Materials and methods

Patients' samples

All patients signed a consent form approved by the Ethics Board (UBC Ethics Board No: H09-01628; VCHRI Nos: V09-0320 and V07-0058). Surgical samples were collected and snap-frozen (FF) at the Vancouver General Hospital (VGH). The remaining tissue was fixed in formalin and used for histopathological evaluation by three independent pathologists from the VGH pathology department and Vancouver Prostate Centre (VPC). Snap-frozen and formalin-fixed blocks are stored at the VPC Tissue Bank. Tumour blocks with tumour cellularity of at least 30% were used for molecular profiling (for details see the Supporting information, Supplementary methods).

Tumour cell lines

LNCaP and C4-2 cells were kindly provided by Dr Leland WK Chung (1992, MDACC, Houston, TX, USA). The human PCa cell lines PC-3 and DU145 were purchased from the American Type Culture Collection (Manassas, VA, USA; 2008, ATCC authentication by isoenzyme analysis). LNCaP.AI cells used for sequencing were provided by Pfizer (La Jolla, CA, USA). These are hormone refractory cells engineered to express three- to five-fold higher levels of the human AR by stably transducing the cells using viral infection with a cDNA encoding for the hAR. All cell lines were authenticated using aCGH, expression microarrays, and whole-genome and whole-transcriptome sequencing on an Illumina Genome Analyzer IIx platform (Illumina, San Diego, CA, USA) at the time of data collection for the current study.

Molecular profiling

For DNA and RNA isolation, tumour cell lines and FF tumour sections were processed as before [16]. Initially, we performed genome copy number (CN) profiling on an Agilent HD-CGH Microarray (Design ID 014698; Agilent Technologies, Santa Clara, CA, USA) to assess DNA quality prior to sequencing. 0.5 µg of genomic DNA was used for hybridization according to the manufacturer's standard protocols, as previously described [16]. Biodiscovery Nexus Copy Number v5.1 was used for CGH data quality assessment, visualization, and analysis. As we gained confidence in the robustness and sensitivity of DNA-Seq of prostate tumours, this practice was abandoned in order to minimize tissue expenditure.

Genome and transcriptome sequencing was performed at BCCA Michael Smith Genome Sciences Centre, Vancouver, BC according to established protocols [17]. For DNA-Seq analyses, reads were mapped to the NCBI 36.1 (hg18) human genome reference sequence using MAQ 0.7.174. Genomic CN changes were derived as described in ref 18 and are shown in Supplementary Figure 3 of the Supporting information.

For RNA-Seq analyses, we used the ALEXA reference sequence database [19], containing sequences of all human exons and exon junctions (Ensembl release 54). RNA-Seq reads for each transcriptome library were aligned to the ALEXA using MAQ 0.7.174. Genes and exon expression were quantified as described in ref 20, transformed into $\log_2$ space, and quantile-normalized [21]. Alternative splicing detection was performed using the strategy described in ref 20 (for details see the Supporting information, Supplementary methods). All Illumina sequence mapping data are available from the Vancouver Prostate Centre website at http://www.lagapc.ca/FTP_Lapuk.html.

Differential gene and splice variants expression analyses

Genes were ranked by their s.d. of expression across tumours and the top 1000 were used for unsupervised hierarchical clustering (GenePattern [22]). To detect genes and splice variants differentially expressed between subsets of tumour samples, we applied a two-tailed Student's *t*-test with equal variances with a

significance level cut-off of 0.05 (Supporting information, Supplementary methods). Pathway and function enrichment analysis was performed using the Ingenuity (IPA) Knowledge Base 9 (Ingenuity® Systems; http://www.ingenuity.com) (Supporting information, Supplementary methods).

Identification of fusion genes

Gene fusions were identified from matched DNA-Seq and RNA-Seq profiles using Comrad [23] with the default parameters, the NCBI 36 (hg18) human reference genome, and Ensembl release 54 gene annotations.

Experimental validation

To validate predicted fusion transcripts and alternative splice events, we amplified the site of fusion or exon inclusion by RT-PCR from cDNA and genomic breakpoints by PCR from genomic DNA using standard techniques (PCR primers provided in the Supporting information, Supplementary Tables 2 and 3). All amplification products were validated by sequencing on an ABI PRISM® 310 Genetic Analyzer (Applied Biosystems, Foster City, CA, USA) using standard techniques.

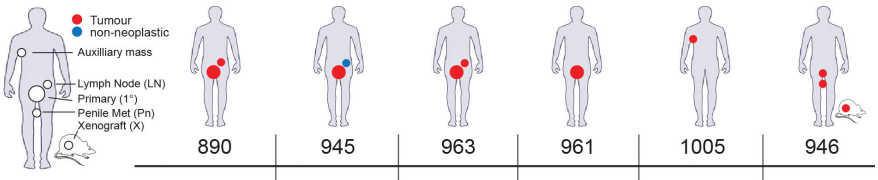
siRNA experiments

REST knockdown was performed using LNCaP cells and two independent siRNAs. For the first experiment, LNCaP cells were transfected twice with 10 nM scrambled (Scr) or REST siRNA (Dharmacon, Lafayette, CO, USA). For the second experiment, LNCaP cells were transfected twice with 20 nM Scr siRNA (Dharmacon) or REST siRNA (Santa Cruz Biotechnology, Santa Cruz, CA, USA). RNA was isolated from cells pre- and post-transfection and used for RT-PCR. Immunoblotting was performed as previously described [24]. Forty-eight hours post-transfection, total proteins were extracted and 30 µg was submitted to western blot using anti-REST antibody (Sigma-Aldrich, St. Louis, MO, USA). The staining with vinculin antibody (Sigma-Aldrich) was used as a loading control.

Results

Tumour cohort

We assembled a cohort of six high-risk patients harbouring tumours of pathological stage T3B or greater, Gleason score ≥7, with different serum PSA levels and outcome (Figure 1). Patient 945 was diagnosed with T3BN0 prostate cancer with PSA of 25 ng/ml and had undergone neoadjuvant treatment. The PSA



	890	945	963	961	1005	946
Diagnosis	PCa	PCa	hPCa	PCa	?	NEPCa
Age at Dx	64	64	46	70	61	74
PSA at Dx(ng/ml)	1.37	25	6.3	15.07	12.3	N/A
Gleason score(H&E)	4+5	5+4	5+5	4+3	4+5	5+4
Pathological staging	T3BN1	T3BN0#	T3BN1	T3BN1#	N/A	T4
BCR free time (months)	30.55	N/A*	2.9	N/A*	77.84	N/A
LN status	+	-	+	+	N/A	N/A
Hormone therapy (months)	-	NHT (5)	-	-	IAS	CHT
CRPC	-	-	-	-	-	+
Tumour cellularity	1°=80% LN=60%	1°=30% LN=0%	1°=90% LN=100%	1°=60%	100%	U=90% Pn=100% X=95%
RNA-Seq (Gbp)	1°=12.6 LN=13.8	1°=14.3 LN=13.0	1°=7.9 LN=14.8	1°=14.3	14.2	U=11.6 Pn=8.9 X=6.0

Figure 1. Summary of the prostate tumour cohort. Sample origin indicated in the top panel with histopathology status (red = malignant; blue = benign). The depth of RNA-Seq is presented as the sum of read lengths mapped to the human genome (Gb = gigabases). The tumour cellularity was estimated by histopathology. PCa = prostate adenocarcinoma; ? = unknown primary origin; hPCa = dual signature adenocarcinoma–neuroendocrine prostate cancer; NEPCa = neuroendocrine prostate cancer (small cell carcinoma of the prostate); Dx = diagnosis; 1° = primary tumour; LN = lymph node; U = urethral metastasis; Pn = penile metastasis; X = xenograft derived from urethral metastasis. Therapy prior to sample collection: NHT = neoadjuvant hormone therapy; IAS = intermittent androgen suppression (four cycles); CHT = continuous hormone therapy. CPRC = castrate = resistant prostate cancer. *The patient did not reach nadir post-surgery (PSA did not drop to 0.2 ng/dl). # Negative pathological surgical margins.

level dropped to 0.47 ng/ml 6 months post-surgery, after which the patient suffered rapid recurrence with a PSA doubling time of less than 3 months. Together with high Gleason grade and negative surgical margins, the clinical features of this patient were consistent with the metastatic recurrence. Patient 890 with T3BN1 disease and a history of chronic prostatitis and low diagnostic PSA responded well to treatment and remained recurrence-free for 30 months. Patient 963 was diagnosed with a clinically localized adenocarcinoma that rapidly progressed to metastatic disease with low serum PSA (see details in ref 25). The patient was put on the maximum androgen blockade therapy showing a rapid response, but the disease recurred in 13 months, as indicated by rising PSA. Patient 946 was diagnosed with neuroendocrine prostate cancer (NEPCa) with small-cell carcinoma morphology, and had previously undergone a salvage cystoprostatectomy. We collected two metastatic samples from the urethral and penile sites. The urethral specimen was used to establish a patient-derived mouse xenograft as described in refs 26 and 27. It was profiled together with the other tumours to determine if it maintained the salient features of the donor tumour. From patient 961 who was diagnosed with LN-positive disease, we collected a primary tumour sample. The sample from patient 1005, previously treated for prostate cancer with radiation therapy, was collected from the underarm area. Histopathology evaluation indicated an architecturally poorly differentiated non-small cell carcinoma. Immunohistochemistry showed equivocal staining for ACPP and negative staining for PSA, breast carcinoma markers (eg GCDFP, ER), KRT7, KRT20, TTF1, and p63. It was therefore classified as carcinoma of unknown primary (CUP). In total, we collected 11 tumour specimens from primary and metastatic sites for molecular profiling, including three matched lymph nodes (Figure 1).

Integrated sequence-based characterization of prostate tumours

We sequenced the genomes ($2-7\times$ coverage relative to haploid genome) and transcriptomes (to an average of 33-fold coverage assuming 200 MB transcriptome size) of each tumour using Illumina GA-IIx paired-end technology (Figure 1 and Supporting information, Supplementary Table 1 and Supplementary methods). From DNA-Seq data we derived CN profiles (Supporting information, Supplementary Figure 2), which were highly similar to aCGH profiles derived from the same tumour block (data not shown). In the instance where direct comparison was possible, the correlation between aCGH and DNA-Seq CN profiles was high (Supporting information, Supplementary Figure 3). Low genome coverage and lack of matched normal samples precluded the detection of sequence variants. However, integrative analysis of genome and transcriptome sequence data using the Comrad algorithm [23] allowed the detection of numerous novel fusions that were private to patients; 53 of those predicted with highest confidence were experimentally validated by PCR (Supporting information, Supplementary Table 2). The average fusion transcript burden was 8 per tumour with the exception of patient 961's tumour, where the absence of fusions was consistent with minimal CN changes (Supporting information, Supplementary Table 1). One of the fusions involved an ETS family gene (*ETV1*). None of sequenced samples expressed the *TMPRSS2–ERG* fusion transcript. The lack of evidence for this fusion transcript in sequencing data was supported by aCGH showing absence of the characteristic deletion in the 21q22 region that gives rise to the fusion event (Supporting information, Supplementary Figure 2) and absence of *ERG* up-regulation even in the samples with high *TMPRSS2* expression. The frequency of the *TMPRSS2–ERG* gene fusion is estimated to be between 18% and 50%, depending on the cohort [9,28,29]. Therefore, the probability of not having it represented in a cohort of six patients would range from 0.30 to 0.02 (for 18% and 50% frequency, respectively). Additionally, the *TMPRSS2–ERG* fusion is associated with a lower Gleason score (<7) [29]. The tumours in our cohort, however, were of all high risk, with Gleason scores ≥ 7.

RNA-Seq was used to estimate the expression levels of all human genes, exons, and exon junctions (see Methods). Analysis of exon and exon-junction expression resulted in the identification of approximately 1000 genes differentially spliced between the tumours, a subset of which was validated by RT-PCR (Figure 2a and Supporting information, Supplementary Table 3). Some of the observed differential splicing events are located within functional protein domains (Figure 2b); others showed a specific splice pattern in NEPCa versus PCa (Figures 2c and 2d). Overall, concordance between RNA-Seq predictions and RT-PCR was high, with an 85% validation rate (Supporting information, Supplementary Table 3), demonstrating the robustness of our approach.

High-resolution molecular pathology and detection of sub-clinical metastasis

MPS allows genome/transcriptome-wide copy number and expression profiling and the identification of fusion transcripts that may represent unique tumour-specific biomarkers [30]. We explored the global gene expression signatures of the tumours by performing unsupervised hierarchical clustering of the genes differentially expressed among samples. This resulted in sample clusters accurately reflecting the tumour subtype and/or sample composition (Figure 3a and Supporting information, Supplementary Figure 5). The cluster of primary adenocarcinomas was distinct from that of adenocarcinoma cell lines. As expected, the clustering of PCa tumours as a group was driven by high expression of prostate-specific stromal and basal cell markers and by the expression of androgen-responsive

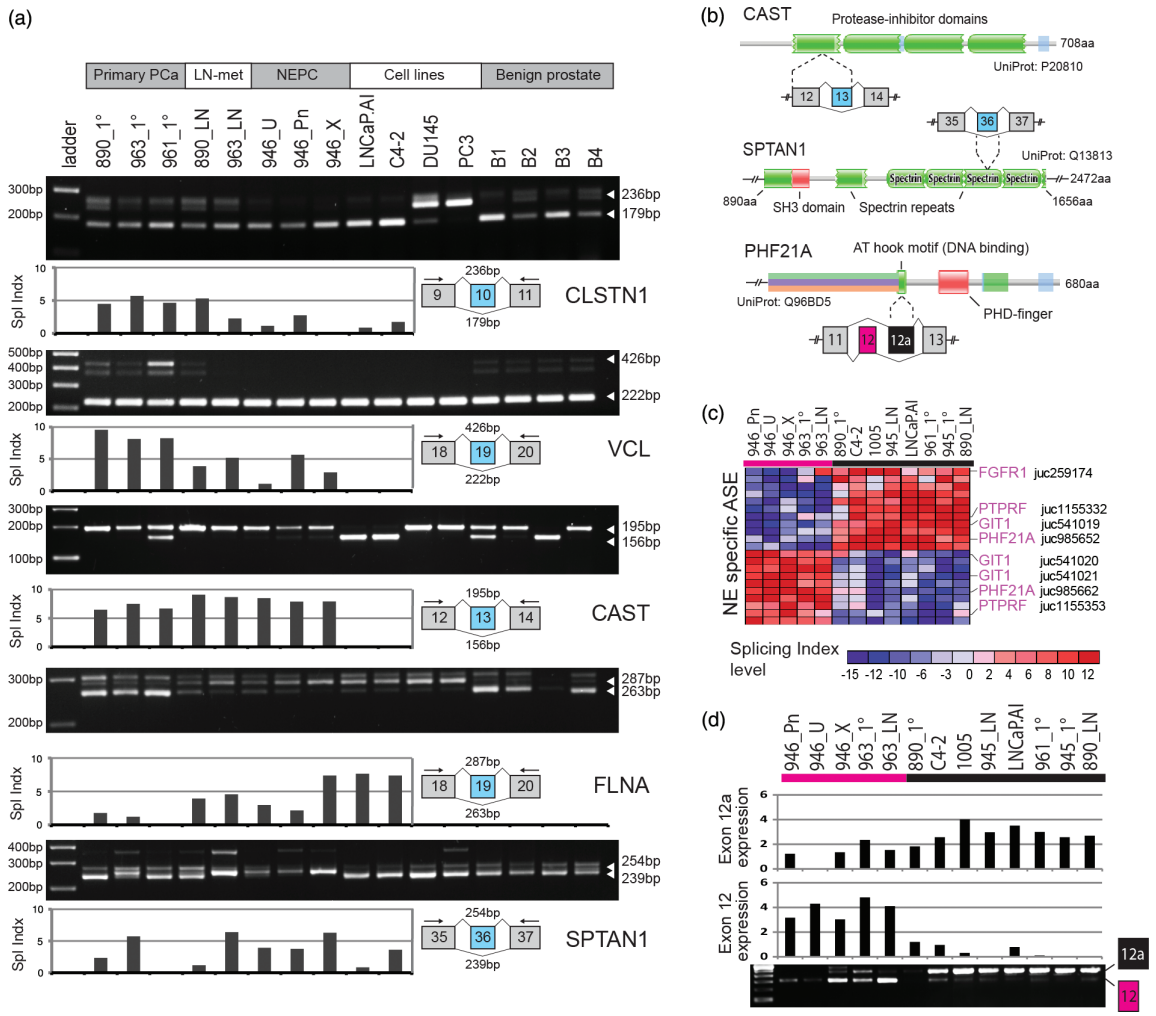


Figure 2. Alternative splicing in prostate tumours. (a) RT-PCR validation of alternatively spliced genes detected in prostate tumours and cell lines. Primers designed for the flanking constitutive exons with expected PCR product sizes are shown in the cartoons below each gel electrophoresis image. RNA-Seq-based splicing data for inclusion isoform (corresponding to the upper PCR band) are shown below each gel as splicing index bar charts for corresponding exons/junctions. (b) Possible functional consequences of alternative splicing. Protein domains (Pfam [61]) are shown for three genes, where alternatively spliced regions fall within functional domains. Alternatively spliced region of a gene is shown below the protein domain structure (blue = alternative exons; grey = constitutive exons) and its portion coding for functional domains are indicated with dashed lines. (c) Heat map of the splicing index data for exons/junctions within 12 genes showing difference in splicing in neuroendocrine samples versus primary adenocarcinoma tumours. Genes labelled on the right with magenta are involved in nervous system development [58–60]. Exon/junction IDs are shown on the right from gene symbols. (d) Alternative splicing of the *PHF21A* gene. The expression of the mutually exclusive exons of *PHF21A* by RNA-Seq is shown as a bar chart. PCR validation gel is shown underneath.

genes [eg *KLK3* (PSA), *KLK2, TMPRSS2*]. LN samples from patients 890 and 945 each had a substantial normal LN tissue component (Figure 1 and Supporting information, Supplementary Figure 1) and consequently exhibited high expression of LN-specific genes, including those associated with T cells, B cells, dendritic cells, and macrophage cells (Figure 3a). NEPCa tumours from patient 946 formed a separate cluster together with those of patient 963, driven by the expression of neuroendocrine markers. The xenograft 946_X clustered closely with the corresponding tumours from patient 946, suggesting that the xenograft retained the salient features of the donor tumour. However, unlike the donor NEPCa tumour, the xenograft expressed several androgen-responsive genes at low level. This can be explained by the fact that the xenograft was grown in intact male mice supplemented with testosterone. The mechanism of this, however, is unclear, since the xenograft does not express detectable AR protein [31]. Interestingly, patient 963's tumours exhibited a dual neuroendocrine and adenocarcinoma gene expression phenotype and represented a distinct sub-cluster with a unique signature (Figure 3a). Immunohistochemistry ruled out classical NED and adenocarcinoma, and indicated the presence of a hybrid luminal-neuroendocrine tumour in this patient, which is characterized in depth in ref 16.

Lymph node (LN) metastases are important determinants of disease progression and are challenging to detect by conventional histopathology. NHT treatment brings an additional challenge, as tumour foci shrink and remaining deposits are even harder to detect [32]. However, three lines of evidence from DNA- and RNA-Seq data suggested the presence of an occult

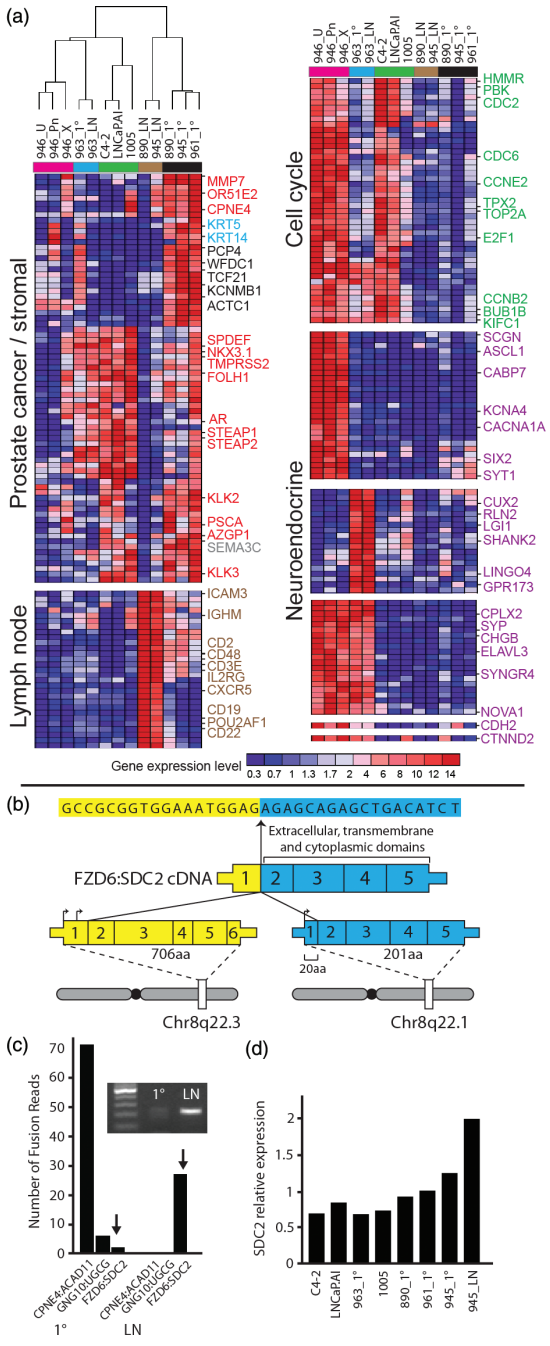


Figure 3. RNA-Seq-derived molecular pathology of prostate tumours and detection of a sub-clinical metastasis. (a) Two-way unsupervised hierarchical clustering of the gene expression profiles of the prostate cancer cohort using the top 1000 genes differentially expressed among tumours. Sample clusters colour bar is shown underneath sample labels. Representative gene clusters with cell type or tissue-specific signatures are shown (for full clusters see Supporting information, Supplementary Figure 5). Red = prostate-specific/androgen-responsive genes; blue = basal cell markers; green = luminal epithelial cell markers; magenta = neuroendocrine cell markers; black = stromal cell markers; brown = lymph node markers. (b) Schematic representation of the *FZD6:SDC2* fusion gene showing maintenance of the *SDC2* reading frame and protein domains. (c) RNA-Seq-derived expression level of the fusion genes detected in the primary tumour and matched lymph node sample from patient 945. RT-PCR validation gel image showing the enrichment of *FZD6:SDC2* is also provided. (d) Expression of *SDC2* ($\log_2$) normalized to prostate-specific genes to account for tumour cellularity, demonstrating overexpression of *SDC2* in the LN of patient 945.

metastasis in the histologically non-neoplastic 945_LN sample. First, we detected moderate expression of several prostate-specific genes including *ACPP, KLK2*, and *FOLH1*, as well as luminal epithelial cell markers, normally associated with adenocarcinoma (Figure 3a), in 945_LN. Second, the 945_LN genome CN profile had many of the genomic alterations commonly found in prostate cancers (such as *PTEN, RB1*, and 8p deletions and 8q gain) (Supporting information, Supplementary Figure 2). Finally, one of three fusion genes identified in 945_1° (*FZD6:SDC2*; see Supporting information, Supplementary Table 2) was also expressed in 945_LN (Figures 3b–3d). Interestingly, *FZD6:SDC2* was expressed at much higher levels in 945_LN than in 945_1° (Figures 3c and 3d). It is possible that the primary tumour of patient 945 contained a discrete population of cancer cells harbouring only the *FZD6:SDC2* fusion, from which the metastatic clone arose: a clonal selection model. Alternatively, differences in transcription regulation may have prevented the detection of the other fusions. Interestingly, *SDC2* was previously reported to be expressed in PCa, and was associated with PCa progression [33]. Increased expression of *SDC2* was shown to enhance collagen adhesion and the motility of rat intestinal epithelial cells [34], advantageous properties for a metastatic clone. Collectively, these three observations indicate the presence of a sub-clinical or occult metastasis, in line with the patient clinical features consistent with metastatic recurrence.

Conversely, the molecular profiling data of patient 890's tumours suggested the presence of potentially dormant metastasis in his LN [35,36]. Patient 890 was diagnosed as LN-positive (Supporting information, Supplementary Figure 1). The genomic profiles of the 890_1° and 890_LN samples were very similar and contained a number of CN changes that are not typical for prostate cancers (Supporting information, Supplementary Figure 2). Surprisingly, unlike the primary tumour that showed a strong PCa transcriptome signature (Figure 2a), the 890_LN transcriptome signature appeared benign, showing no expression of prostate-specific genes and yielding no gene fusions. This is unlikely to be due to the sensitivity issues, as the methods allowed detection of an unambiguous tumour signature in the pathologically benign 945_LN. One possible explanation is that the LN metastasis had become dormant, which is consistent with the fact that patient 890 responded well to treatment and remained recurrence-free for 30 months.

An axillary mass resected from patient 1005 previously treated for PCa was classified as CUP based on immunohistochemical analysis. The MPS analysis revealed numerous CN changes typical for prostate cancer (Supporting information, Supplementary Figure 2), and its transcriptome profile was similar to PCa cell lines based on the expression of prostate-specific and luminal epithelial cell-specific genes, including *PSA* (*KLK3*). Moreover, this tumour expressed truncated *AR* splice variants including the

oncogenic variant AR3(V7) [37,38], which, in the cohort, was present at the highest ratio relative to the full length AR (Supporting information, Supplementary Figure 7). Constitutively-active truncated isoforms of AR, including AR3(V7), are associated with poor outcome [14] and are likely to represent an adaptive response to castration [39]. These observations are consistent with the fact that patient 1005 had been exposed to several cycles of androgen deprivation therapy, and allowed unambiguous classification of patient's 1005 tumour as adenocarcinoma of prostatic origin.

Molecular signatures of castrate-resistant neuroendocrine prostate cancer

In the NEPCa tumours from patient 946, we observed a strong and distinct gene expression signature (Figure 2), with high expression of genes involved in cellular proliferation, the cell cycle, and mitosis, as well as genes important for endocrine biology (enrichment analysis, Fisher's exact test, adjusted p value ≤ 0.0001; Supporting information, Supplementary Table 3). We then compiled a targeted, literature-driven panel of up-regulated genes (compared with adenocarcinomas) which comprehensively defined the neuronal phenotype in the tumours from patient 946 (Figure 4). Detection of part of this gene signature in patient 963 led to the identification of a novel hybrid luminal-neuroendocrine cancer subtype [16].

Significantly, in the NEPCa and hybrid PCa/NEPCa tumours, we observed reduced expression of *REST*, a transcription factor considered to be the master repressor of neuronal differentiation [40] (Supporting information, Supplementary Figure 6). REST binds to target sites [41] within genes important for a neuronal phenotype and prevents transcription. Interestingly, 28/50 of the transcriptionally active neuroendocrine phenotype genes described in Figure 4 harbour experimentally validated *REST* binding site(s). To determine the clinical relevance of this pattern, we mined publically available data. Taylor *et al* profiled 218 prostate tumours [9], and we have found *REST* down-regulation and associated up-regulation of the neuronal signature in 50% of tumours with a NED/NEPCa component (Figure 4b). In contrast, only 3% of other tumours from that cohort exhibited this pattern. The difference in the frequency of this pattern in the two groups of tumours was highly significant ($p < 0.0001$, Fisher's exact test).

Most if not all hallmarks of cancer are associated with perturbed exon splicing [42]. To explore possible contributions of AS to the neuroendocrine phenotype, we explored NEPCa-specific AS events and found a number of genes associated with neuronal functions (Figure 2c and Supporting information, Supplementary Tables 5 and 6). Of particular interest, a component of the *REST* transcriptional complex, *PHF21A* [43], is differentially spliced in NEPCa versus PCa (Figure 2d). The *PHF21A* isoform expressed in NEPCa tumours was found to lack an exon, encoding an AT-hook protein motif necessary for DNA binding (Figure 2b).

To experimentally validate the influence of *REST* down-regulation on associated gene and exon expression changes, we performed an siRNA knockdown of *REST* in LNCaP cells (Figure 4c). We observed an up-regulation of a number of NEPCa markers along with increased exclusion of the *PHF21A* AT-hook encoding exon, consistent with observations in the clinical samples (Figure 4c).

The analysis also suggested a possible role of global splicing regulation in NED and epithelial-to-mesenchymal transition (EMT). Genes differentially spliced between NEPCa and other tumours showed moderate enrichment of cell morphology functions (Fisher's exact test, adjusted p value ≤ 0.004; Supporting information, Supplementary Table 5). Interestingly, we found alternatively spliced genes important for cell shape and invasion, which were reported by others to undergo alternative splicing in breast cancer cells upon induction of EMT [44] (Supporting information, Supplementary Table 6). NED of PCa cells was previously linked to EMT [45], a process accompanied by morphological transformation of organized epithelial cells into isolated, migratory mesenchymal cells [46]. In line with this, we noticed an up-regulation of the mesenchymal marker N-cadherin, and a number of cell adhesion and motility genes (Figures 3a and 4).

Prostatic adenocarcinomas are known to contain recurrent fusions [47], but the spectrum of NEPCa fusions is largely unexplored. We identified 11 fusion transcripts in NEPCa samples from patient 946 (Supporting information, Supplementary Table 2 and Supplementary Figure 4), which were shared by the urethra and penile metastases and the xenograft. Half of the fusions involved genes with restricted expression in neuroendocrine cells (Figure 4), thereby recapitulating the gene expression results. A similar situation was observed in the tumours of patient 963, where the fusion gene set mirrored the dual PCa/NEPCa gene expression signature [16]. This finding supports recent reports that chromosomal translocations are cell-type-specific as they occur preferentially in transcriptionally active genes [12,48].

Identification of chromothripsis in prostate cancer

While the existing paradigm dictates that chromosomal rearrangements occur gradually over time, recent evidence suggests that in at least 2–3% of cancers, tens to hundreds of genomic rearrangements involving only one or a few chromosomes can occur in a one-off cellular crisis resulting in cancer-causing lesions [49]. This phenomenon, known as chromothripsis [18], was not previously described in prostate cancer and its association with outcome is unclear. We found evidence of it in the primary tumour genome from one of our high-risk patients (890) with a history of chronic prostatitis. The 890_1° genome contained fewer amplifications than other tumours. Two chromosome arms (2p and

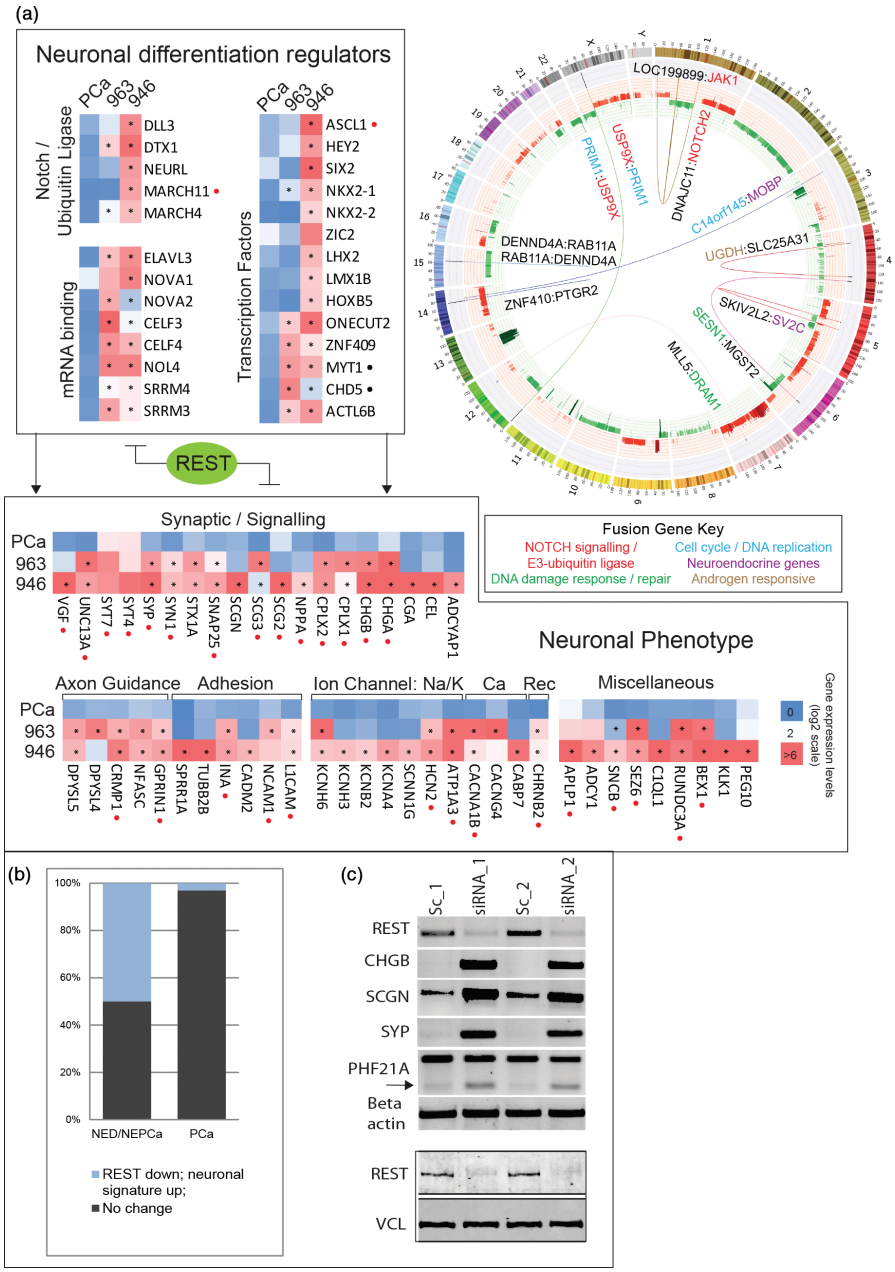


Figure 4. The acquisition and maintenance of a neuroendocrine phenotype. (a) The bottom panel shows the neuronal phenotype identified in the tumours of patients 946 and 963. For patient 946, we used the mean expression from all three samples. Stars within the heat map illustrate statistically different expression from other adenocarcinoma PCa samples. Red circles by the gene name indicate an experimentally validated REST binding site within the gene loci. Rec – receptor. The top-left panel demonstrates a panel of regulatory genes up-regulated in the same tumour samples which we hypothesize to act synergistically to maintain expression of the neuronal phenotype. The top-right panel is a Circos plot [62] of the 946_Pn sample demonstrating the CN and fusion gene profile exhibited by the tumours of patient 946. Chromosomes are arranged circularly end-to-end with each chromosome's cytobands marked in the outer ring. The inner ring displays copy number data inferred from genome sequencing, with red indicating gains and green indicating losses. Within the circle, genomic rearrangements are shown as arcs with the resultant validated fusion transcript annotated. Function of genes involved in fusions are indicated by colour and provided in the key. (b) Recurrence of the *REST* down-regulation and the neuronal signature up-regulation from a in the prostate cancer cohort published by Taylor *et al* [9] showing 50% frequency in the NED/NEPCa subset of tumours ($n = 16$) versus 3% in the PCa subset ($n = 134$). The NED/NEPCa subset was defined by overexpression of at least one of the neuroendocrine markers (*CHGA*, *CHGB* or *SYP*; Supporting information, Supplementary methods). (c) siRNA knockdown of REST in LNCaP cells. RT-PCR gel shows down-regulation of REST using two independent siRNA pools and concomitant up-regulation of neuronal signature genes as well as change in PHF21A splicing. An arrow indicates the isoform lacking the AT-hook binding domain. Sc = scrambled RNA control. Western blot using REST and vinculin antibodies is shown underneath RT-PCR, demonstrating significant down-regulation of REST protein upon siRNA transfection.

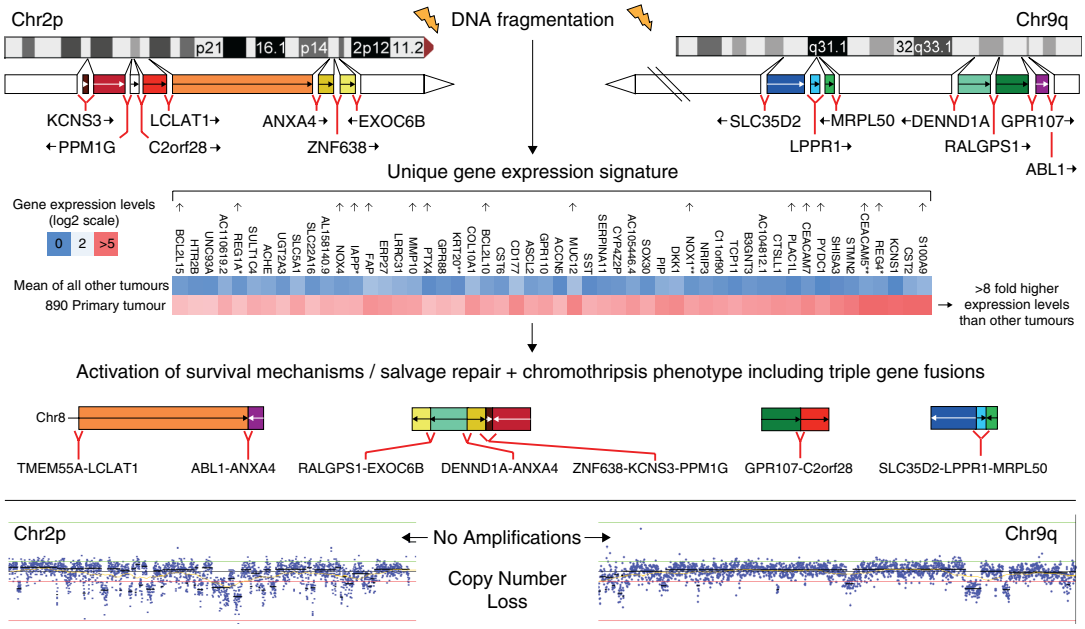


Figure 5. Chromothripsis, inflammatory response, and fusion genes in patient 890. The top panel indicates multiple putative genomic breakpoints between 2p and 9q resulting in validated fusion transcripts. The second panel demonstrates the top 50 uniquely expressed genes in the primary tumour of patient 890 compared with all other tumour samples, showing enrichment (small arrows) of genes linked to response to inflammation and apoptosis. *Pancreas-specific genes; **colon-specific genes. The third panel shows the fusion transcripts derived from genomic rearrangements between chromosomes 2 and 9, including triple gene fusions. The bottom panel provides genomic sequence-derived copy number profiles, indicating the lack of amplifications and the high number of focal deletions, a hallmark of chromothripsis [18,49].

9q) were found to harbour more deletions than other chromosome arms and to possess a more segmented CN profile (Supporting information, Supplementary Table 1 and Supplementary Figure 2). Furthermore, nine out of 11 validated fusion events in 890_1° involved genes on 2p or 9q and two fusions involved three genes, providing evidence of complex chromosomal rearrangements involving multiple breakpoints (Figure 5) (for details see the Supporting information, Supplementary data). On the transcriptome level, this tumour exhibited a unique expression signature of genes with highly restricted expression patterns in normal tissues other than prostate (Figure 5). A comprehensive literature search revealed that this panel included genes associated with resistance to bacterial infection, response to inflammation, anti-apoptosis, and cell stress (indicated by arrows in Figure 5).

Discussion

We previously used MPS to identify an aggressive hybrid luminal-neuroendocrine tumour [16] and have now expanded the study to six high-risk prostate cancer patients. We demonstrated that integrative analysis of shallow genome and deep transcriptome sequencing is a powerful tool for specific and sensitive molecular pathology, which can improve diagnosis and possibly have an impact on therapy. Moreover, we identified a mechanism that may drive the emergence of the neuroendocrine phenotype that is frequently found in CRPC cancers.

Combined DNA- and RNA-Seq-based molecular profiling of prostate tumours led to findings that have important clinical implications. Firstly, we unambiguously identified a LN metastasis not detected by standard histopathology; suggested a potentially dormant LN metastasis; classified a tumour of unknown origin; and identified a new hybrid subtype of PCa [18]. Identification of an occult LN metastasis in patient 945 is associated with a poorer outcome and may warrant more aggressive treatment [50]. The presence and higher expression of gene fusion *FZD6:SDC2* in the LN compared with the primary tumour may suggest clonal selection and supports the idea that such tumour- and patient-specific markers may play a role in molecular pathology, as well as in monitoring disease progression and response to therapy, eg by measuring their levels in patient plasma samples [30]. Conversely, the absence of a tumour transcriptome signature in the LN sample of LN-positive patient 890 potentially suggests that this patient could benefit from a less aggressive therapeutic intervention. Thus, although based only on two patients, these results provide hope that MPS may ultimately aid in differentiating quiescent from aggressive metastatic prostate cancer. Secondly, in the US alone 31 000 people will be diagnosed with cancers of unknown primary (CUP) in 2012 (approximately 2% of all cancers diagnosed), and identification of a primary site is important for therapeutic management of CUP patients [51]. Although assignment to the primary site is possible with microarrays, MPS profiling significantly broadens the spectrum of tumour-specific biomarkers, enabling

the identification of fusion transcripts, splice variants, and genomic breakpoints. Finally, the identification of a novel hybrid luminal-neuroendocrine phenotype in aggressive conventional adenocarcinoma may imply a poorer outcome for patient 963, who should be managed accordingly.

Molecular characterization of NEPCa and hybrid PCa/NEPCa tumours is a high-priority research area, given the prevalence of NED in advanced PCa and its association with resistance to current therapies [7]. Our findings indicate that the REST transcriptional complex, a master repressor of the neuronal phenotype, may be one of the important determinants of the neuroendocrine phenotype in PCa. We observed down-regulation of *REST* with associated up-regulation of neuronal genes, a pattern also found in 50% of tumours with a NED/NEPCa component in a large independent cohort [9]. Simultaneously, the REST co-factor PHF21A appears to lose the AT-hook domain important for DNA binding through alternative splicing. Additionally, NED has been previously linked to EMT [45], and our observation of splicing regulation of EMT-related functions in NEPCa supports this link. Finally, we found that the fusion genes that we identified preferentially involved transcriptionally active genes, consistent with previous reports [12,48]. Thus, we hypothesize that the fusion transcript spectrum may reflect the evolutionary history of a tumour and provide insight into cancer progression.

The presence of genome copy number aberrations and complex genome rearrangements is typical for high-risk prostate tumours [9,52]. However, we detected evidence of chromothripsis in the primary tumour of patient 890 with a history of chronic prostatitis and low PSA. Chromothripsis, recently described in other cancers, results in many genomic aberrations localized to a few genomic regions [18]. Several mechanisms were proposed to explain this phenomenon, one of which is an aborted apoptosis that may have been triggered by a noxious stimulus, such as radiation or infection [53]. Interestingly, this tumour's transcriptome exhibited a signature of non-prostate-specific genes associated with an inflammatory immune response, anti-apoptosis, and cell stress. A growing body of evidence points to a possible role of prostatic inflammation in the aetiology of prostate cancer; however, a causal link is yet to be established [54]. Our observation of chromothripsis in PCa and its potential association with chronic prostatitis warrant further investigation. Furthermore, we detected the presence of triple fusion genes, potentially arising from chromothripsis. Although the clinical relevance of this phenomenon is yet to be elucidated, it may represent a distinct mechanism of prostate carcinogenesis.

It is well established that PCa is biologically and clinically heterogeneous. Current methods of patient stratification based on integration of clinical history with tumour histopathology still fail for a significant number of cancer patients, resulting in both over- and under-treatment. The possible benefits of MPS for understanding of the biological underpinnings of prostate cancer are undisputed. Moreover, a recent pilot study demonstrated that whole-genome, targeted-exome, and transcriptome sequencing is possible within a time frame relevant for cancer patients [55]. Sequencing costs are expected to reach US $1000 or less per genome in the near future. Thus, MPS has the potential to improve on rational disease intervention and lower healthcare costs [56,57]. It is reasonable to expect that MPS will ultimately complement conventional pathology in the future, resulting in improved patient management.

Acknowledgment

This study was supported by the Canadian Institutes of Health Research (YZW, MG), Centres of Excellence for Commercialization and Research (MG), PNW Prostate Cancer SPORE P50 CA097186, Prostate Cancer Canada, Genome BC (CC), Prostate Cancer Foundation (AL, CC), and the Ontario Institute for Cancer Research (through funding provided by the Government of Ontario) to RB, PB, and CC via the CPC-GENE prostate cancer genome-network. The work in the AC laboratory was supported by NIH P50 CA69568, U01 CA111275, and R01CA132874 grants.

Author contribution statement

CCC, SVV, AL, CW, and AWW conceived experiments and analysed data. CCC funded and directed the project. SVV coordinated the project, data management, and analysis. AVL, SVV, AM, SA, and FM conducted bioinformatics analyses and RHB coordinated statistical analyses. CW, BJM, RS, AH, SB, and AZ carried out experiments. MEG, LF, AH-C, and ECJ identified tumours and performed histopathology analysis. YW provided prostate tumour mouse xenografts. YZ and MAM coordinated the next-generation sequencing. FHa, FHo, IH, SCS, CAM, and AMC assisted in bioinformatics analyses. MAR and HB provided mouse xenograft sequencing data. MEG co-directed the project and reviewed the manuscript. AZ, PCB, and RGB aided data analyses and reviewed the manuscript. AVL, AWW, SVV, and CCC wrote the manuscript.

References

Note: References 63–65 are cited in the Supporting information to this article.

1. Siegel R, Naishadham D, Jemal A. Cancer statistics, 2012. *Ca Cancer J Clin* 2012; **62**: 10–29.
2. Graefen M, Karakiewicz PI, Cagiannos I, *et al*. Validation study of the accuracy of a postoperative nomogram for recurrence after radical prostatectomy for localized prostate cancer. *J Clin Oncol* 2002; **20**: 951–956.

3. Graefen M, Karakiewicz PI, Cagiannos I, *et al*. International validation of a preoperative nomogram for prostate cancer recurrence after radical prostatectomy. *J Clin Oncol* 2002; **20**: 3206–3212.
4. Kattan MW, Eastham JA, Stapleton AM, *et al*. A preoperative nomogram for disease recurrence following radical prostatectomy for prostate cancer. *J Natl Cancer Inst* 1998; **90**: 766–771.
5. Kattan MW, Wheeler TM, Scardino PT. Postoperative nomogram for disease recurrence after radical prostatectomy for prostate cancer. *J Clin Oncol* 1999; **17**: 1499–1507.
6. Beltran H, Beer TM, Carducci MA, *et al*. New therapies for castration-resistant prostate cancer: efficacy and safety. *Eur Urol* 2011; **60**: 279–290.
7. Vashchenko N, Abrahamsson PA. Neuroendocrine differentiation in prostate cancer: implications for new treatment modalities. *Eur Urol* 2005; **47**: 147–155.
8. Beltran H, Rickman DS, Park K, *et al*. Molecular characterization of neuroendocrine prostate cancer and identification of new drug targets. *Cancer Discov* 2011; **1**: 487–495.
9. Taylor BS, Schultz N, Hieronymus H, *et al*. Integrative genomic profiling of human prostate cancer. *Cancer Cell* 2010; **18**: 11–22.
10. Tomlins SA, Laxman B, Varambally S, *et al*. Role of the *TMPRSS2–ERG* gene fusion in prostate cancer. *Neoplasia* 2008; **10**: 177–188.
11. Miyagi Y, Sasaki T, Fujinami K, *et al*. ETS family-associated gene fusions in Japanese prostate cancer: analysis of 194 radical prostatectomy samples. *Mod Pathol* 2010; **23**: 1492–1498.
12. Berger MF, Lawrence MS, Demichelis F, *et al*. The genomic complexity of primary human prostate cancer. *Nature* 2011; **470**: 214–220.
13. Friedlander TW, Roy R, Tomlins SA, *et al*. Common structural and epigenetic changes in the genome of castration resistant prostate cancer. *Cancer Res* 2012; **72**: 616.
14. Dehm SM, Tindall DJ. Alternatively spliced androgen receptor variants. *Endocr Relat Cancer* 2011; **18**: R183–R196.
15. Lee DJ, Cha EK, Dubin JM, *et al*. Novel therapeutics for the management of castration-resistant prostate cancer (CRPC). *BJU Int* 2012; **109**: 968–985.
16. Wu C, Wyatt AW, Lapuk AV, *et al*. Integrated genome and transcriptome sequencing identifies a novel form of hybrid and aggressive prostate cancer. *J Pathol* 2012; **227**: 53–61.
17. Shah SP, Morin RD, Khattra J, *et al*. Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution. *Nature* 2009; **461**: 809–813.
18. Maher CA, Wilson RK. Chromothripsis and human disease: piecing together the shattering process. *Cell* 2012; **148**: 29–32.
19. Griffith M, Griffith OL, Mwenifumbo J, *et al*. Alternative expression analysis by RNA sequencing. *Nature Methods* 2010; **7**: 843–847.
20. Lapuk A, Marr H, Jakkula L, *et al*. Exon-level microarray analyses identify alternative splicing programs in breast cancer. *Mol Cancer Res* 2010; **8**: 961–974.
21. Bullard JH, Purdom E, Hansen KD, *et al*. Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. *BMC Bioinformatics* 2010; **11**: 94.
22. Reich M, Liefeld T, Gould J, *et al*. GenePattern 2.0. *Nature Genet* 2006; **38**: 500–501.
23. McPherson A, Wu C, Hajirasouliha I, *et al*. Comrad: a novel algorithmic framework for the integrated analysis of RNA-Seq and WGSS data. *Bioinformatics* 2011; **27**: 1481–1488.
24. Zoubeidi A, Ettinger S, Beraldi E, *et al*. Clusterin facilitates COMMD1 and I-kappaB degradation to enhance NF-kappaB activity in prostate cancer cells. *Mol Cancer Res* 2010; **8**: 119–130.
25. McShane LM, Altman DG, Sauerbrei W, *et al*. REporting recommendations for tumor MARKer prognostic studies (REMARK). *J Natl Cancer Inst* 2005; **97**: 1180–1184.
26. Cutz JC, Guan J, Bayani J, *et al*. Establishment in severe combined immunodeficiency mice of subrenal capsule xenografts and transplantable tumor lines from a variety of primary human lung cancers: potential models for studying tumor progression-related changes. *Clin Cancer Res* 2006; **12**: 4043–4054.
27. Collins CC, Volik SV, Lapuk A, *et al*. Next generation sequencing of prostate cancer from a patient identifies a deficiency of methylthioadenine phosphorylase (MTAP), an exploitable tumor target. *Mol Cancer Ther* 2012; **11**: 775–783.
28. Markert EK, Mizuno H, Vazquez A, *et al*. Molecular classification of prostate cancer using curated expression signatures. *Proc Natl Acad Sci U S A* 2011; **108**: 21276–21281.
29. Fine SW, Gopalan A, Leversha MA, *et al*. *TMPRSS2–ERG* gene fusion is associated with low Gleason scores and not with high-grade morphological features. *Mod Pathol* 2010; **23**: 1325–1333.
30. Leary RJ, Kinde I, Diehl F, *et al*. Development of personalized tumor biomarkers using massively parallel sequencing. *Sci Transl Med* 2010; **2**: 20ra14.
31. Tung WL, Wang Y, Gout PW, *et al*. Use of irinotecan for treatment of small cell carcinoma of the prostate. *Prostate* 2011; **71**: 675–681.
32. Gomella LG, Singh J, Lallas C, *et al*. Hormone therapy in the management of prostate cancer: evidence-based approaches. *Ther Adv Urol* 2010; **2**: 171–181.
33. Popovic A, Demirovic A, Spajic B, *et al*. Expression and prognostic role of syndecan-2 in prostate cancer. *Prostate Cancer Prostatic Dis* 2010; **13**: 78–82.
34. Choi S, Kim Y, Park H, *et al*. Syndecan-2 overexpression regulates adhesion and migration through cooperation with integrin alpha2. *Biochem Biophys Res Commun* 2009; **384**: 231–235.
35. Aguirre-Ghiso JA. Models, mechanisms and clinical evidence for cancer dormancy. *Nature Rev Cancer* 2007; **7**: 834–846.
36. Shiozawa Y, Pedersen EA, Havens AM, *et al*. Human prostate cancer metastases target the hematopoietic stem cell niche to establish footholds in mouse bone marrow. *J Clin Invest* 2011; **121**: 1298–1312.
37. Guo Z, Yang X, Sun F, *et al*. A novel androgen receptor splice variant is up-regulated during prostate cancer progression and promotes androgen depletion-resistant growth. *Cancer Res* 2009; **69**: 2305–2313.
38. Sun S, Sprenger CC, Vessella RL, *et al*. Castration resistance in human prostate cancer is conferred by a frequently occurring androgen receptor splice variant. *J Clin Invest* 2010; **120**: 2715–2730.
39. Watson PA, Chen YF, Balbas MD, *et al*. Constitutively active androgen receptor splice variants expressed in castration-resistant prostate cancer require full-length androgen receptor. *Proc Natl Acad Sci U S A* 2010; **107**: 16759–16765.
40. Ooi L, Wood IC. Chromatin crosstalk in development and disease: lessons from REST. *Nature Rev Genet* 2007; **8**: 544–554.
41. Mortazavi A, Leeper Thompson EC, Garcia ST, *et al*. Comparative genomics modeling of the NRSF/REST repressor network: from single conserved sites to genome-wide repertoire. *Genome Res* 2006; **16**: 1208–1221.
42. Venables JP. Unbalanced alternative splicing and its significance in cancer. *BioEssays* 2006; **28**: 378–386.
43. Hakimi MA, Bochar DA, Chenoweth J, *et al*. A core-BRAF35 complex containing histone deacetylase mediates repression of neuronal-specific genes. *Proc Natl Acad Sci U S A* 2002; **99**: 7420–7425.
44. Shapiro IM, Cheng AW, Flytzanis NC, *et al*. An EMT-driven alternative splicing program occurs in human breast cancer and modulates cellular phenotype. *PLoS Genet* 2011; **7**: e1002218.

45. McKeithen D, Graham T, Chung LW, *et al*. Snail transcription factor regulates neuroendocrine differentiation in LNCaP prostate cancer cells. *Prostate* 2010; **70**: 982–992.
46. Kalluri R, Weinberg RA. The basics of epithelial–mesenchymal transition. *J Clin Invest* 2009; **119**: 1420–1428.
47. Clark JP, Cooper CS. ETS gene fusions in prostate cancer. *Nature Rev Urol* 2009; **6**: 429–439.
48. Lin C, Yang L, Tanasa B, *et al*. Nuclear receptor-induced chromosomal proximity and DNA breaks underlie specific translocations in cancer. *Cell* 2009; **139**: 1069–1083.
49. Stephens PJ, Greenman CD, Fu B, *et al*. Massive genomic rearrangement acquired in a single catastrophic event during cancer development. *Cell* 2011; **144**: 27–40.
50. Pagliarulo V, Hawes D, Brands FH, *et al*. Detection of occult lymph node metastases in locally advanced node-negative prostate cancer. *J Clin Oncol* 2006; **24**: 2735–2742.
51. Pavlidis N, Briasoulis E, Hainsworth J, *et al*. Diagnostic and therapeutic management of cancer of an unknown primary. *Eur J Cancer* 2003; **39**: 1990–2005.
52. Paris PL, Andaya A, Fridlyand J, *et al*. Whole genome scanning identifies genotypes associated with recurrence and metastasis in prostate tumors. *Hum Mol Genet* 2004; **13**: 1303–1313.
53. Tubio JM, Estivill X. Cancer: when catastrophe strikes a cell. *Nature* 2011; **470**: 476–477.
54. Sfanos KS, De Marzo AM. Prostate cancer and inflammation: the evidence. *Histopathology* 2012; **60**: 199–215.
55. Roychowdhury S, Iyer MK, Robinson DR, *et al*. Personalized oncology through integrative high-throughput sequencing: a pilot study. *Sci Transl Med* 2011; **3**: 111ra121.
56. Armstrong K. Can genomics bend the cost curve? *J Am Med Assoc* 2012; **307**: 1031–1032.
57. Drmanac R. The advent of personal genome sequencing. *Genet Med* 2011; **13**: 188–190.
58. Dunah AW, Hueske E, Wyszynski M, *et al*. LAR receptor protein tyrosine phosphatases in the development and maintenance of excitatory synapses. *Nature Neurosci* 2005; **8**: 458–467.
59. Maric D, Fiorio Pla A, Chang YH, *et al*. Self-renewing and differentiating properties of cortical neural stem cells are selectively regulated by basic fibroblast growth factor (FGF) signaling via specific FGF receptors. *J Neurosci* 2007; **27**: 1836–1852.
60. Zhang H, Webb DJ, Asmussen H, *et al*. Synapse formation is regulated by the signaling adaptor GIT1. *J Cell Biol* 2003; **161**: 131–142.
61. Finn RD, Mistry J, Tate J, *et al*. The Pfam protein families database. *Nucleic Acids Res* 2010; **38**: D211–D222.
62. Krzywinski M, Schein J, Birol I, *et al*. Circos: an information aesthetic for comparative genomics. *Genome Res* 2009; **19**: 1639–1645.
63. Wang XS, Prensner JR, Chen G, *et al*. An integrative approach to reveal driver gene fusions from paired-end sequencing data in cancer. *Nature Biotechnol* 2009; **27**: 1005–1011.
64. Gasi D, van der Korput HA, Douben HC, *et al*. Overexpression of full-length ETV1 transcripts in clinical prostate cancer due to gene translocation. *PLoS One* 2011; **6**: e16332.
65. Attard G, Clark J, Ambroisine L, *et al*. Duplication of the fusion of TMPRSS2 to ERG sequences identifies fatal human prostate cancer. *Oncogene* 2008; **27**: 253–263.

SUPPORTING INFORMATION ON THE INTERNET

The following supporting information may be found in the online version of this article.

Supplementary data and methods

Figure S1. Haematoxylin and eosin (H&E) staining images of patient samples from fresh frozen sections of primary tumours, axillary mass and invasion of urethra (A), and FFPE sections of lymph node samples at 2× and 20× magnification (B).

Figure S2. Whole-genome CN profile of 11 samples from the prostate tumour cohort, derived from DNA-Seq.

Figure S3. Concordance of CN profiles derived from aCGH and DNA-Seq.

Figure S4. Representative results of fusion gene validations.

Figure S5. Unsupervised hierarchical clustering of the top 1000 differentially expressed genes (by s.d. of expression across samples) in the prostate cancer cohort.

Figure S6. RNA-Seq-derived expression of the *REST* gene showing reduced expression in NEPCa compared with PCa samples.

Figure S7. Alternative splicing of the androgen receptor gene (AR) detected in the high-risk prostate tumours.

Table S1. Summary of tumour genome sequencing and CN aberrations.

Table S2. Fifty-three novel fusion genes identified in prostate tumours.

Table S3. PCR validation of predicted alternative splicing events.

Table S4. The list of genes differentially expressed between PCa and NEPCa.

Table S5. Top enriched pathways and functions identified with Ingenuity enrichment analysis.

Table S6. The list of exons/junctions alternatively spliced between PCa and NEPCa.