FULL PAPER

BJC

British Journal of Cancer (2017) 117, 367–375 | doi: 10.1038/bjc.2017.171

BJC OPEN

Keywords: heterogeneity; DNA ploidy; PTEN; prostate cancer; Gleason score; active surveillance

Tumour heterogeneity poses a significant challenge to cancer biomarker research

Karolina Cyll[1,2,9], Elin Ersvær[1,2,9], Ljiljana Vlatkovic[3], Manohar Pradhan[1,2], Wanja Kildal[1,2], Marte Avranden Kjær[1,2,✠], Andreas Kleppe[1,2,4], Tarjei S Hveem[1,2,4], Birgitte Carlsen[5], Silje Gill[6], Sven Löffeler[6], Erik Skaaheim Haug[6], Håkon Wæhre[1,2], Prasanna Sooriakumaran[7] and Håvard E Danielsen*[,1,2,4,8]

[1]Institute for Cancer Genetics and Informatics, Oslo University Hospital, Oslo NO-0424, Norway; [2]Centre for Cancer Biomedicine, University of Oslo, Oslo, NO-0424, Norway; [3]Department of Pathology, Oslo University Hospital, Oslo, NO-0424, Norway; [4]Department of Informatics, University of Oslo, Oslo, NO-0316, Norway; [5]Department of Pathology, Vestfold Hospital Trust, NO-3103 Tønsberg, Norway; [6]Department of Urology, Vestfold Hospital Trust, Tønsberg, NO-3103, Norway; [7]Surgical Intervention Trials Unit, Nuffield Department of Surgical Sciences, University of Oxford, Oxford OX3 9DU, UK and [8]Nuffield Division of Clinical Laboratory Sciences, University of Oxford, Oxford OX3 9DU, UK

Background: The high degree of genomic diversity in cancer represents a challenge for identifying objective prognostic markers. We aimed to examine the extent of tumour heterogeneity and its effect on the evaluation of a selected prognostic marker using prostate cancer as a model.

Methods: We assessed Gleason Score (GS), DNA ploidy status and phosphatase and tensin homologue (PTEN) expression in radical prostatectomy specimens (RP) from 304 patients followed for a median of 10 years (interquartile range 6–12). GS was assessed for every tumour-containing block and DNA ploidy for a median of four samples for each RP. In a subgroup of 40 patients we assessed DNA ploidy and PTEN status in every tumour-containing block. In 102 patients assigned to active surveillance (AS), GS and DNA ploidy were studied in needle biopsies.

Results: Extensive heterogeneity was observed for GS (89% of the patients) and DNA ploidy (40% of the patients) in the cohort, and DNA ploidy (60% of the patients) and PTEN expression (75% of the patients) in the subgroup. DNA ploidy was a significant prognostic marker when heterogeneity was taken into consideration. In the AS cohort we found heterogeneity in GS (24%) and in DNA ploidy (25%) specimens.

Conclusions: Multi-sample analysis should be performed to support clinical treatment decisions.

Tumour heterogeneity is a well-known trait of most cancers (Lawrence *et al*, 2014; Andor *et al*, 2015). Heterogeneity originates from genomic instability (McGranahan *et al*, 2012), one of the hallmarks of cancer (Hanahan and Weinberg, 2011). Genomic instability comprises a variety of alterations ranging from small to large structural and numerical alterations in the genome, and may give rise to diverse populations of cells (Burrell *et al*, 2013). In the presence of multiple clones of tumour cells with different characteristics, a single sample from one small tumour region might not be optimal for predicting the tumour's aggressiveness.

Prostate cancer is an ideal model to study tumour heterogeneity as the whole gland is extracted during radical prostatectomy and all the tumour material is available for analysis. Furthermore, it is a complex disease that can follow multiple paths, from indolent disease causing little or no harm, to aggressive disease that may progress to metastases and death. Generally, patients with indolent

*Correspondence: Professor HE Danielsen; E-mail: hdaniels@labmed.uio.no
[9]These authors contributed equally to this work.
✠Deceased.

Received 2 February 2017; revised 10 May 2017; accepted 22 May 2017; published online 15 June 2017

© The Author(s) named above

tumours are advised for active surveillance, whereas patients with aggressive tumours are often treated with prostatectomy and adjuvant therapy, when necessary (Attard *et al*, 2016). To date, prostate cancer is one of the most common malignancies and causes of cancer related mortality among men (Center *et al*, 2012). Therefore, the correct estimation of prognostic markers is crucial for proper management of the patients.

Our aim was to systematically review the extent of tumour heterogeneity using a large cohort of post-operative prostatectomy specimens (RP) and in needle biopsies from patients with low risk prostate cancer in all available tumour samples. This is the most extensive study of tumour heterogeneity performed on such a large cohort of patients with the use of three independent methods:

- Histological heterogeneity was evaluated by Gleason scoring (GS), generally considered as the strongest single prognostic marker in prostate cancer (Epstein *et al*, 2016).
- Heterogeneity in genome size was evaluated using high-resolution DNA image cytometry (DNA ploidy analysis) an established prognostic marker for various cancer types, including prostate cancer (Danielsen *et al*, 2016).
- Heterogeneity in expression of a tumour suppressor gene, phosphatase and tensin homolog (PTEN), was evaluated by immunohistochemistry. PTEN is one of the most commonly lost tumour suppressors, and has been suggested as a reliable prognostic marker for prostate cancer (Cuzick *et al*, 2013).

To investigate the practical importance of intratumour heterogeneity, we compared the strength of DNA ploidy as a predictor of recurrence following radical prostatectomy. This was compared in analyses of one sample versus multiple samples. To our knowledge, we are the first to explore the effect of heterogeneity on the evaluation of a prognostic marker.

PATIENTS AND METHODS

Patients and specimens. The main part of this study was based on material from 317 radical retro-pubic prostatectomies performed by one of the authors (HW) at the Norwegian Radium Hospital, Oslo, Norway, between 1987 and 2005 (Figure 1). The study was approved by the Regional Committees for Medical and Health Research Ethics (REK) in Norway (REK no. S-07443a and 2012/1679). Each of these resected prostates was prepared as a series of 3 mm formalin fixed, paraffin embedded tissue blocks. For the purpose of Gleason grading, a 3 μm hematoxylin and eosin (H&E) stained section was made from the top surface of each tissue block, and visually examined. Slides containing at least one tumour area $\geqslant 4\,mm^2$ (measured manually) were Gleason graded ($n = 4469$). For each RP specimen, further 50 μm sections for DNA ploidy analyses, were taken from the block(s) with highest Gleason score, and/or the greatest tumour area, or a randomly selected block with a tumour area $\geqslant 4\,mm^2$ ($n = 1271$). In order to study intrafocal heterogeneity in GS and DNA ploidy, we compared samples from 244 tissue blocks sectioned at two different depths.

A subgroup of 40 RP specimens was selected for detailed analysis of the intratumour heterogeneity (Figure 1). The RP subgroup consisted of 23 resections with multifocal tumours from before the PSA era, and 17 resections randomly chosen from the RP procedures performed in 2001. In this subgroup, in addition to the Gleason scoring of all tumour areas $\geqslant 4\,mm^2$, we also performed DNA ploidy and PTEN analysis on all available blocks. If multiple tumour areas were present in a block, they were analysed separately if they were $\geqslant 3$ mm apart.

In addition, we investigated a set of diagnostic needle biopsies from a cohort of 102 prostate cancer patients assigned to active surveillance at Vestfold Hospital Trust between 2008 and 2013 (the AS Cohort, Figure 1). The analysis included Gleason score and DNA ploidy for all patients with Gleason score $\leqslant 3+4$ and at least two positive cores that had sufficient tumour material (408 needle biopsy cores).

Methods. In every RP case, tumour areas $\geqslant 4\,mm^2$ were chosen for analyses. Subsequently, relevant RP tumour areas were Gleason graded by one of the authors (LV) according to the updated 2005 International Society of Urological Pathology (ISUP) Consensus guidelines (Epstein, 2010) (RP Cohort) and 2014 ISUP Consensus guidelines (RP Subgroup) [5]. The biopsy cores from the AS group were Gleason graded according to the updated 2005 ISUP (Epstein, 2010) Consensus guidelines by another author (BC). Gleason grade groups (GGGs) were categorised as: 1(6); 2($3+4=7$a); 3($4+3=7$b); 4(8); 5(9–10).

Image cytometry was performed according to a modified Hedley's method (Cyll *et al*, 2015) on 50 μm thick, dissected tumour areas. DNA content histograms were classified as diploid, tetraploid or aneuploid as described by Danielsen *et al.* (2016), with the exception of the threshold for tetraploid samples that was set to 15% of nuclei in the 4c peak in the DNA ploidy histogram.

Tumour volume was estimated using computer-assisted image analysis (CAIA), H&E stained sections with marked tumour areas were scanned (NanoZoomer HT, Hamamatsu Photonics, Hamamatsu, Japan) using a 20x objective. The tumour area was then marked in NDP.view2 (NanoZoomer HT, Hamamatsu Photonics, Hamamatsu, Japan) and the area was calculated. The sum of all tumour areas for the patient was multiplied by the block thickness (3 mm) and the tissue shrinkage factor (1.12) (Perera *et al*, 2014).

Immunohistochemistry was done using the Dako Envison FLEX+ system (K8002; Dako, Glostrup, Denmark) and Dako Autostainer Link 48. Deparaffinisation and unmasking of epitopes were performed using PTlink (Dako) and EnVisionTM Flex Target Retrieval Solution at 97 °C for 20 min. Endogenous peroxidase was blocked by treating the sections with FLEX peroxidase-blocking reagent for 5 min. Further, the sections were incubated for 120 min with the rabbit monoclonal anti-body (anti-human) PTEN (138G6, Cell Signaling technology, Danvers, MA) at 1:400 dilution, followed by EnVision FLEX+ Rabbit (linker) for 15 min, EnVisionTM Flex/HRP enzyme for 20 min and 3′3-diaminobenzidine tetrahydrochloride for 10 min. Finally, the sections were dehydrated and counterstained with hematoxylin for 10 seconds.

The percentage of tumour cells with PTEN staining was scored independently by two of the authors (KC and WK). Discordant cases were reviewed by a pathologist (MP). Tumour cells not showing cytoplasmic staining were scored as negative for PTEN expression. Stromal fibroblasts were used as an internal positive control. Slides without positively stained fibroblasts were considered to have a technical failure and were excluded from the analysis. Tumour areas showing loss of PTEN staining in $<10\%$ of the cells were classified as positive, loss of PTEN staining in $>90\%$ of the cells as negative, and loss of PTEN staining in 10–90% of the cells as a mixed PTEN expression.

Follow-up and patient survival. After exclusion of 13 RP specimens because of missing clinical data or indeterminate DNA ploidy classification (Supplementary Table S1), 304 radical prostatectomy patients were followed for a median of 10 years (IQR 6–12). Time to recurrence, where the endpoint was defined as locoregional recurrence, distant metastasis or death from prostate cancer (Punt *et al*, 2007), was calculated from primary treatment to recurrence of disease or 31 December 2008. At the end of the study, 213 (70%) patients were still alive and the mean TTR for the whole group was 106 months (IQR 64–153). None of the patients received any neoadjuvant therapy before surgery. Adjuvant therapies like radiation or androgen deprivation were not used in any patients, and these secondary treatments were only instigated

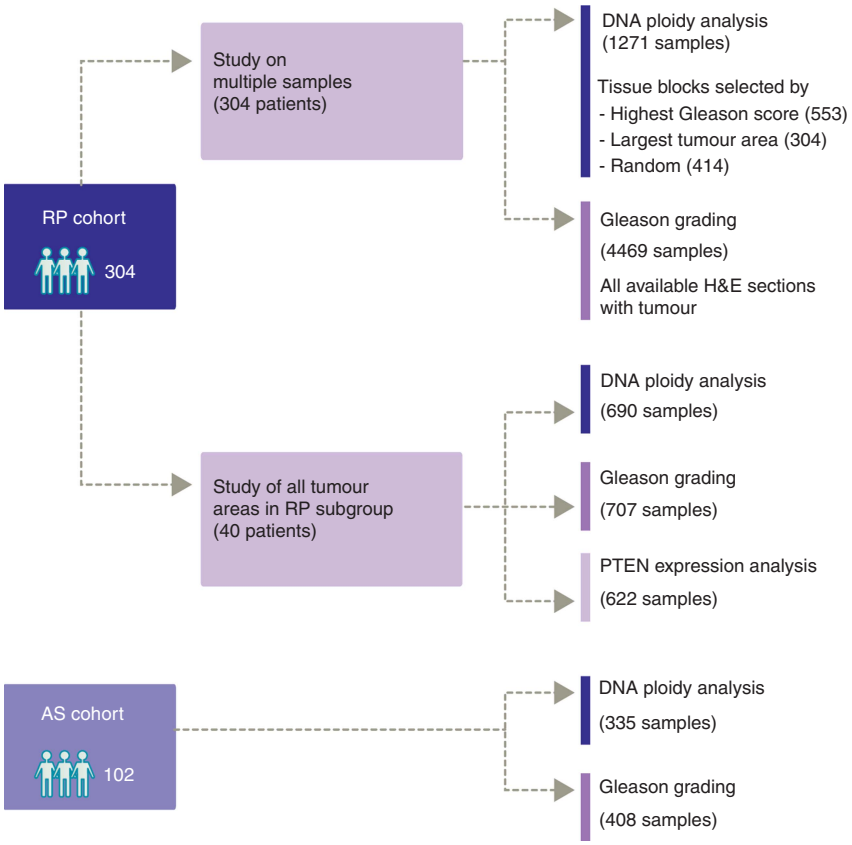

Figure 1. Overview of included patients' material from radical prostatectomy (RP) and active surveillance (AS) cohort and methods used in analysis of each cohort.

at or after the point of recurrence as previously defined. Table 1 gives clinicopathological data for both the RP and AS cohorts.

Statistical analysis. Statistical calculations were performed using SPSS V22.0 statistical software (IBM Corporation, Armonk, NY, USA). The χ^2-test was used to evaluate association between categorical variables. Survival analyses were performed using the Kaplan–Meier method and survival distributions were compared with the log-rank test. Associations between tumour volume and heterogeneous DNA ploidy classification in the RP subgroup were investigated with an independent samples *t*-test using log-transformed data. For each patient from the RP subgroup with at least one non-diploid sample, the expected number of samples required to detect a non-diploid sample was calculated using the definition of the expected value and under the assumption that samples were selected randomly. Two-sided *P*-values <0.05 were considered statistically significant. The interobserver agreement for the PTEN score was compared using kappa statistics.

RESULTS

Intraprostatic heterogeneity in prostatectomy specimens. Gleason grading was performed for 4469 tumour areas from 304 prostate specimens in the full RP cohort, including 707 tumour areas from the 40 specimens in the RP subgroup (Supplementary Table S1). All of the 40 prostate resections (100%) of the RP subgroup revealed more than one GGG. Whereas in the full RP cohort, 270 (89%) of the 304 RPs were heterogeneous in GGG classification (Table 2 and Figure 2A).

Within the full RP cohort, we classified 1271 DNA ploidy histograms (median 4/prostate, IQR 3–5), and discordance was observed in 121/304 (40%) cases (Supplementary Table S1). For the 40 prostate glands in the RP subgroup, a total of 690 valid DNA ploidy histograms were classified (median 15/prostate, IQR 7–28). Discordant DNA ploidy classification was found in 24/40 (60%) cases, (Table 2, Figure 2A and Supplementary Figure S1). For the RP subgroup, the median expected number of samples required to identify one of the non-diploid samples from a patient with heterogeneous DNA ploidy classifications was five.

For the patients with heterogeneous DNA ploidy classification in the RP subgroup ($n=24$), we have reviewed how often the worst DNA ploidy was present in the tumour areas with the highest GGG classification. Four patients (17%) had the worst DNA ploidy classification in all analysed samples with the highest GGG, whereas ten patients (42%) had the worst DNA ploidy identified in lower GGGs. For the last ten patients (42%), the worst DNA ploidy class was identified in some, but not all of the samples, in the highest GGG (Supplementary Table S2).

PTEN expression was scored in 622 tumour areas from the 40 prostate glands of the RP subgroup (median 13.5/prostate, IQR 6–24), (Supplementary Table S1). Positive expression was found in 328 (53%), negative in 194 (31%), and both positive and negative in 100 (16%) tumour areas. Ten prostates were entirely positive in all sections, whereas the remaining 30 prostates (75%) had heterogeneous expression (Table 2 and Figure 2A). The interobserver agreement for the PTEN score was substantial ($\kappa=0.8$). The spatial allocations of GGG, PTEN expressions and DNA ploidy classifications are exemplified in a prostate gland from the RP subgroup, which was reconstructed in 3D (Figure 2B).

Intrafocal heterogeneity in prostatectomy specimens. Intrafocal heterogeneity was investigated by comparing DNA ploidy classification and GGGs in paired samples prepared from different depths within the same tissue blocks. GGGs were compared in 244 such pairs of tissue sections, of which 99 (41%) had discordant

Table 1. Clinical data for patients from the radical prostatectomy (RP) and active surveillance (AS) cohort included in the heterogeneity study

	RP, *n* (%)		AS, *n* (%)
	Whole cohort (*n* = 304)	Subgroup[a] (*n* = 40)	Cohort (*n* = 102)
Age			
<60	102 (34)	14 (35)	24 (24)
60–70	178 (59)	23 (58)	51 (50)
>70	24 (8)	3 (8)	27 (26)
PSA (ng/ml)[b]			
<10	129 (42)	13 (33)	77 (75)
10–20	102 (34)	14 (35)	17 (17)
>20	70 (23)	13 (33)	4 (4)
Unknown	3 (1)	0	4 (4)
pT/cT[c]			
T1	0	0	69 (68)
T2	72 (24)	11 (28)	18 (18)
T3	185 (61)	21 (53)	1 (1)
T4	30 (10)	5 (13)	0
Unknown	17 (6)	3 (8)	14 (14)
Gleason score			
⩽6	17 (6)	2 (5)	51 (50)
7a	118 (39)	13 (33)	51 (50)
7b	86 (28)	11 (28)	NA
8	52 (17)	9 (23)	NA
⩾9	31 (10)	5 (13)	NA
Surgical margins			
Negative	118 (39)	11 (28)	NA
Positive	186 (61)	29 (73)	NA
Recurrence			
No	204 (67)	21 (53)	NA
Yes	100 (33)	19 (48)	NA
<60 months	51 (51)	12 (63)	NA
>60 months	49 (49)	7 (37)	NA
Vital status			
Alive	213 (70)	25 (63)	NA
Dead	91 (30)	15 (48)	NA
Due to PCa	34 (37)	9 (60)	NA
Not due to PCa	57 (63)	6 (40)	NA

Abbreviations: NA = not applicable; PCa = prostate cancer. Due to rounding the numbers may not sum to 100%.
[a]Data from the RP subgroup were collected from the RP cohort.
[b]Preoperative PSA values are used for the RP cohort and subgroups, PSA at the time of biopsy is used for the AS cohort.
[c]Pathological T stage (pT) is summarised for the RP cohort and subgroup, clinical T stage (cT) for the AS cohort.

Table 2. Intraprostatic heterogeneity in classifications of DNA ploidy, Gleason grade groups and PTEN expression for patients from radical prostatectomy (RP) and active surveillance (AS) cohort

	RP, *n* (%)		AS, *n* (%)
	Whole cohort (*n* = 304)	Subgroup (*n* = 40)	Cohort (*n* = 102)
DNA Ploidy			
Homogeneous DNA ploidy (%)	183 (60)	16 (40)	77 (76)
Diploid	174 (95)	16 (100)	75 (97)
Tetraploid	7 (4)	0	1 (1)
Aneuploid	2 (1)	0	1 (1)
Heterogeneous DNA ploidy (%)	121 (40)	24 (60)	25 (25)
Diploid/aneuploid	21 (17)	4 (17)	5 (20)
Diploid/tetraploid	72 (60)	14 (58)	18 (72)
Tetraploid/aneuploid	10 (8)	0	2 (8)
Diploid/tetraploid/aneuploid	18 (15)	6 (25)	0
Gleason grade groups			
One grade group	34 (11)	0	78 (76)
Two grade groups	82 (27)	15 (38)	24 (24)
Three grade groups	79 (26)	6 (15)	NA
Four grade groups	80 (26)	11 (28)	NA
Five grade groups	29 (10)	8 (20)	NA
PTEN expression			
Homogenous expression (%)	NA	10 (25)	NA
PTEN only positive expression	NA	8 (100)	NA
Heterogeneous expression (%)	NA	30 (75)	NA
Positive/negative	NA	3 (10)	NA
Positive/mixed	NA	14 (47)	NA
Mixed/negative	NA	1 (3)	NA
Positive/negative/mixed	NA	12 (40)	NA

Abbreviations: NA = not applicable; PTEN = phosphatase and tensin homolog. Due to rounding the numbers might not sum to 100%.

GGGs (Table 3). Change of a primary pattern (3 or 4) was observed in 27 of 108 (25%) tissue blocks with Gleason score 7. In 33 randomly selected blocks with Gleason 7, the pathologist estimated the percentage of Gleason pattern 4 in both sections. Different percentages were reported for the two separate sections from 23 blocks, the median difference in percent Gleason 4 within a single tissue block was 10% (IQR 10–20).

Heterogeneity in DNA ploidy classification was found within 42/233 (18%) tissue blocks (Table 3). Most commonly, when two samples were sectioned at different depths of a tumour block, one sample was classified as diploid and the other as tetraploid (28/42 blocks, 67%).

Intraprostatic heterogeneity in biopsy specimens. For the AS cohort we acquired 335 DNA ploidy classifications in needle biopsy cores from 102 patients (median 3 positive cores/prostate, IQR 2–4). Heterogeneous DNA ploidy classification was observed in 25/102 (25%) prostate cancers (Table 2 and Figure 2A). 75 of the 77 prostate cancers with homogeneous DNA ploidy classification were uniformly DNA diploid.

Even though all of the patients included in the AS cohort had biopsies with Gleason scores of 6 or 7a, we observed heterogeneity in GGGs in biopsy cores from 24/102 (24%) patients.

Effect of heterogeneity on the prognostic value of DNA ploidy. We found differences in TTR based on the worst DNA ploidy classification detected among all analysed samples for each patient in both RP cohort and subgroup (mean 5.7 samples/patient, median 4, IQR 3–5, $P = 0.003$) (Figure 3A). Patients with tetraploid and aneuploid DNA ploidy status had increased risk of recurrence (HR = 1.7, 95% CI: 1.1–2.7, $P = 0.028$ and HR = 2.2, 95% CI: 1.4–3.6, $P = 0.001$) compared to patients with diploid tumours. Differences in survival were also found when DNA ploidy was based on only one sample selected as highest Gleason score ($P = 0.021$), (Figure 3B). However, a significant change in survival was only observed between patients with diploid and aneuploid samples (HR = 2.1, 95% CI: 1.2–3.7, $P = 0.007$), and not between patients with diploid and tetraploid samples (HR = 1.3, CI: 0.8–2.1, $P = 0.4$). No significant results were obtained when one block was selected randomly for each patient (HR = 1.4, 95% CI: 0.8–2.4 and HR = 1.7, CI: 0.9–3.2, $P = 0.16$), (Figure 3C).

Association of tumour heterogeneity with tumour volume. Median tumour volume in the RP subgroup was estimated as 2.5 cm^3 (IQR 0.5–5.2). The median total volume of tumours for the 24 patients in this group with a heterogeneous DNA ploidy classification was estimated as 4.4 cm^3 (IQR 1.9–6.1), significantly greater than the median total volume of the 16 tumours with homogeneous DNA ploidy classifications (0.5 cm^3, IQR 0.2–1.8, $P < 0.001$). The median volume of non-diploid samples from

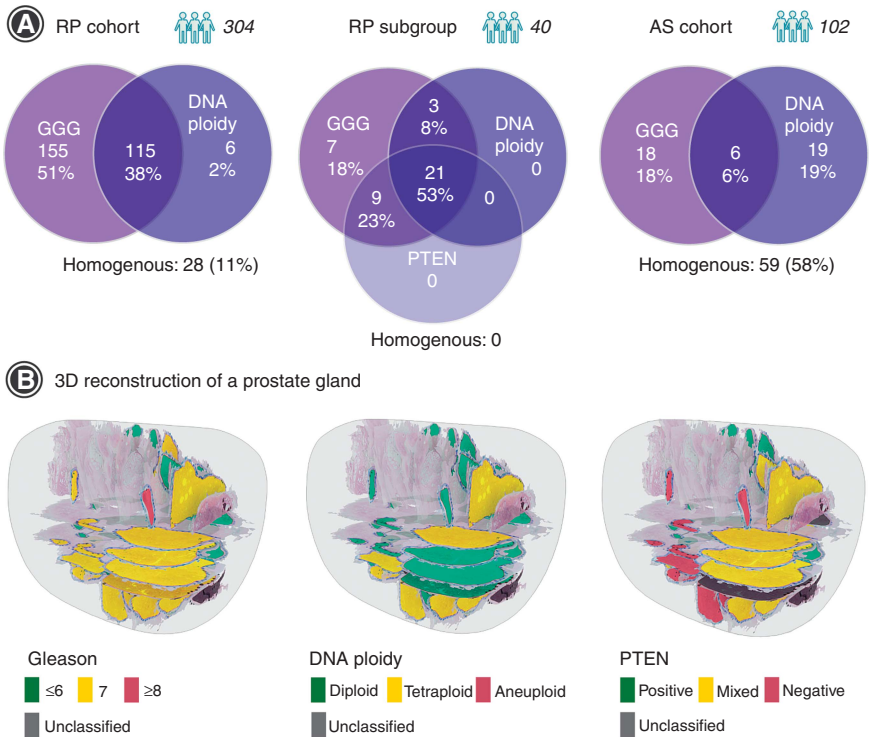

Figure 2. Visualisation of tumour heterogeneity in prostate cancer using Gleason grading, PTEN expression and DNA ploidy analysis. (**A**) Venn diagrams showing the distribution of heterogeneity of the investigated biomarkers in patients from the radical prostatectomy (RP) cohort, RP subgroup and active surveillance (AS) cohort. (**B**) A 3D reconstruction of a prostate gland from the RP subgroup using H&E stained tissue sections. The results from all three biomarkers were applied to the sampled regions.

Table 3. Intrafocal heterogeneity in DNA ploidy and Gleason grade groups classifications

	n
DNA ploidy classification (*n* = 233[a])	
Homogeneous classification	191 (82%)
Diploid	147
Tetraploid	30
Aneuploid	14
Heterogeneous classification	42 (18%)
Diploid/tetraploid	28
Diploid/aneuploid	7
Tetraploid/aneuploid	7
Gleason grade groups (*n* = 244[a])	
Homogeneous classification	145 (59%)
Grade group 1 (GS 6)	23
Grade group 2 (GS 7a)	62
Grade group 3 (GS 7b)	19
Grade group 4 (GS 8)	25
Grade group 5 (GS ⩾ 9)	16
Heterogeneous classification	99 (41%)
Different by one grade group	80
Different by two grade groups	13
Different by three grade groups	6

Abbreviation: GS = Gleason score.
[a]Two samples were made from tissue sections cut at different depths in the same block.

heterogeneous tumours in the RP subgroup was estimated as 0.6 cm^3 (IQR 0.1–2.5), which, on average, accounted for only 14% of the median tumour volume in those tumours. Tumours with heterogeneous PTEN expression ($P = 0.007$) and Gleason score ($P < 0.001$) also had larger volumes. For patients from the full RP cohort, tumours with higher pT stages were also more heterogeneous in terms of DNA ploidy ($P < 0.001$) and Gleason grading ($P < 0.001$) (Supplementary Table S3).

DISCUSSION

Tumour heterogeneity in cancer is a major obstacle in clinical practice and in the search for prognostic markers. Genomic heterogeneity in prostate cancer has been documented in studies that analysed small parts of the tumour from only a few patients (Fraser *et al*, 2014; Boutros *et al*, 2015; Cooper *et al*, 2015). There is a growing awareness that analysis of multiple samples is necessary to overcome the limiting factor of tumour heterogeneity on the discovery and validation of biomarkers (Gerlinger *et al*, 2015). Despite this, most of these studies still rely on solely one sample (Talantov *et al*, 2010; Freedland *et al*, 2013; Cooperberg *et al*, 2015; Ferraldeschi *et al*, 2015). In order to more accurately model heterogeneity, we chose to analyse a high number of samples with three independent methods applied to a large cohort of patients with long follow-up.

We decided to use prostate cancer as our model system to study heterogeneity since radical prostatectomies, which involve entire prostate gland removal, allow us to study the whole tumour and the whole organ. In addition, prostate cancer is known to be heterogeneous (Andor *et al*, 2015), which is partially explained by the fact that 56–87% of all cases of contemporary radical prostatectomies have multifocal disease (Karavitakis *et al*, 2011). A comparison of the genomic landscape in geographically distinct regions within a prostate gland has revealed that prostate cancer arises from multiple, independent clonal expansions (Karavitakis *et al*, 2011; Boutros *et al*, 2015; Cooper *et al*, 2015). The individual lesions in a prostate can also grow together and create yet another level of heterogeneity (Andreoiu and Cheng, 2010). We have found high rates of intrafocal heterogeneity in DNA ploidy classification and Gleason scoring by comparing samples taken at different depths from the same tumour area in a tissue block. These discrepancies can also be explained by the fusion of several individual foci rather than heterogeneity within one focus.

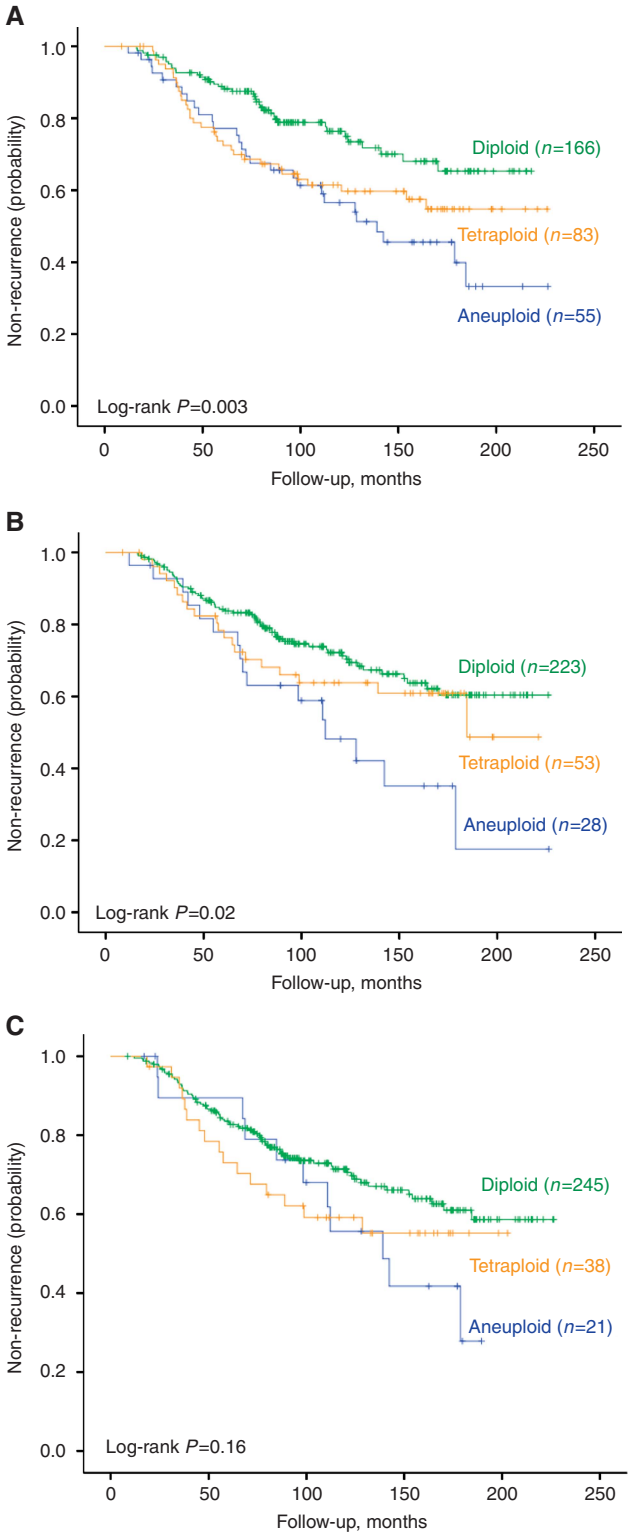

Figure 3. Time to recurrence after radical prostatectomy stratified by: (**A**) worst DNA ploidy detected among all analysed samples for each patient (**B**) DNA ploidy from one sample with highest Gleason score (**C**) DNA ploidy from one randomly selected sample.

However, prostate cancer may not be relevant as a model for less heterogeneous cancer types, for example, colorectal cancer, which do not have such multifocal growth patterns.

Gleason scoring is widely accepted as the standard for histological grading of prostate cancer. Pathologists examine and score glandular architecture of every tumour specimen reporting the Gleason score for both dominant and secondary patterns within the whole tumour. Despite that numerous prognostic tissue biomarkers have been reported in the literature during the last few decades, Gleason score remains the strongest predictor of prostate cancer progression.

Heterogeneity in Gleason scoring within multiple foci in a tumour-containing gland is a common finding. We observed extensive heterogeneity in GGGs, such that only 11% of the tumours in the resected prostates were characterised by a single GGG in all tumour-containing blocks. This is in agreement with previous studies reporting homogeneity in 10% (Aihara *et al*, 1994) and 9% of specimens (Arora *et al*, 2004), although these studies analysed an unknown number of samples from fewer patients using a Gleason scoring system that is no longer in use. In addition, while previous studies mainly focused on heterogeneity between individual foci, we report on intrafocal heterogeneity in GGGs on parallel sections from the same tumour area, which were detected in 41% of the 244 investigated tumour areas. We found a difference in prevalence of primary pattern 3 and 4 for foci with Gleason sum 7, resulting in both Gleason 7a and 7b classifications, which differ substantially from one another in terms of prognosis (Stark *et al*, 2009).

DNA ploidy analysis is a robust method that detects populations of cells with alterations in DNA content, and is a prognostic marker for various cancer types, including prostate cancer (Danielsen *et al*, 2016). This method allows detection of deviations from the normal diploid state, which reflect gross chromosomal alterations and heterogeneity at the gene level (McGranahan *et al*, 2012; López-García *et al*, 2017). Even using such a method, we have shown high rates of heterogeneity in prostate cancer. When we analysed a median of four samples for each patient in the full RP cohort, we observed heterogeneity in DNA ploidy classification in 40% of the patients. Similar rates of heterogeneity of DNA ploidy have been reported previously (O'Malley *et al*, 1993; Wang and Wilkin, 2000). However, by analysing all tumour areas, we observed more than one DNA ploidy classification in 60% of the 40 patients in the RP subgroup.

Investigators often make use of Gleason grading to select a single block with highest Gleason score for analysis with other methods (Cooperberg *et al*, 2015; Freedland *et al*, 2016). However, this seemingly reasonable practice may be sub-optimal. Our findings, using selected biomarkers, clearly show that heterogeneity can affect analysis done using commercial biomarkers from just one tumour area, as confirmed in a recent publication of Wei *et al* (2017). The authors investigated the impact of genomic heterogeneity on molecular classification of 26 tissue cores, sampled from both index tumour and other lesions, from four multifocal RP specimens. Prognosticator scores were obtained using four different assays, including Prolaris and Decipher. The study demonstrated that the large extent of inter- and intratumour heterogeneity precludes the use of these platforms on a single biopsy or only on index lesion for determining appropriate treatment. The importance of sufficient sampling was also shown by Haffner and Mosbruger (Haffner and Mosbruger, 2013) who tracked the clonal relationship between primary tumour and metastases in a case study. They suggested that the lethal metastatic clone arose from a small, isolated focus with Gleason pattern 3, and not from the prevalent Gleason pattern 4 foci or a lymph node metastasis resected with the prostate. Thus, the dominant clone does not necessarily dictate the clinical outcome, further underscoring the need for sampling methods and molecular analyses that are sufficiently sensitive for the detection of the clinically important subclones. Although, several studies have shown a good correlation between high Gleason score and the presence of aneuploidy as reviewed in Danielsen *et al* (2016), we did not always find that the tumour in a block with a non-diploid DNA ploidy classification matched with the highest Gleason score and vice

versa. This discrepancy could not be explained by minor errors in Gleason scoring or DNA ploidy analysis, because we found that 64% of GGGs 4 and 5 samples were DNA diploid (Supplementary Table S4). Also, only four patients from the RP subgroup (17%) had the worst DNA ploidy classification in all analysed samples with the highest GGG (Supplementary Table S2).

Our results have clearly shown that, regardless of the selection criteria, a single sample does not provide representative information of the tumour as a whole. We compared the differences of analysing one random sample, with one sample with the highest Gleason, and multiple samples (median of four samples) for each patient in the whole RP cohort, on the efficiency of DNA ploidy as a prognostic marker. Significant results were not obtained for the randomly chosen samples. We observed significant differences in TTR for diploid and non-diploid cases when we analysed a single sample with the highest Gleason. However, when the analysis was performed on multiple samples per patient, we found significant differences in TTR between all three ploidy classifications. Importantly, such a separation between tetraploid and aneuploid tumours has never been shown before. This is likely caused by the fact that all previous studies used one sample per patient for the DNA ploidy analysis as reviewed by Böcking *et al* (2014) and Danielsen *et al* (2016). Therefore, the analysis of multiple samples for each patient has evidently improved the resolution of DNA ploidy as a prognostic marker.

A number of the specimens from the RP cohort were dated relatively early in the PSA era, when patients who went through radical prostatectomy more often had multifocal disease and larger tumour volumes. There is likely to be less heterogeneity for the average patient currently considered for radical prostatectomy than in the RP specimens from our RP cohort. However, in newer cohorts the clinical follow-up is too short to assess the ultimate outcome for the patients.

In order to document heterogeneity in a more contemporary cohort, we also included a cohort of patients enrolled in AS. However, for those patients, we examined biopsy cores where the scarce amount of tumour tissue only represents a small fraction of the tumour. Despite these shortcomings, it is encouraging that we were able to prepare satisfactory nuclear suspensions for DNA ploidy estimation for the majority of the samples. We found heterogeneity in DNA ploidy in 25% of the AS cohort of patients, which is relatively high considering that it is comprised of patients with low stage disease. For this study we have selected patients with at least two positive biopsy cores successfully analysed with DNA ploidy. It is possible that patients excluded because we were not able to analyse the required number of samples, had smaller tumours. Therefore, our results may be biased towards larger tumours which, as we have shown, are more heterogeneous.

Almost all previous estimates of DNA ploidy in needle biopsies have been made from Feulgen-processed thin histological sections by manual selection of <100 nuclei for DNA content measurement (Sengupta *et al*, 2006). For several reasons this is now recognised as a sub-optimal method (Danielsen *et al*, 2016), and our use of automated image analysis with suspensions of nuclei from the tumour areas of needle cores represents a significant step forward in this field. To the best of our knowledge there is only one other report of similar technology being applied to prostate needle biopsy cores, and this was a preliminary report that showed very promising results (Keyes *et al*, 2013). However, we were not able to make monolayers with sufficient number of nuclei for 18% of the samples because the tumour areas in the biopsy cores were very small in size.

DNA ploidy analysis can reveal sub-populations within a tumour that have a substantial alteration in genome size, but its sensitivity is usually limited to detect a gain or loss of about 2% of the total chromatin, equivalent to one whole average-sized chromosome. Smaller chromatin additions or any losses of individual genes will not be detected by DNA ploidy analysis. This is one of the reasons why we also measured the degree of heterogeneity in the expression of PTEN, a tumour-suppressor gene, which loss has been linked to poor patient prognosis in prostate cancer (Yin and Shen, 2008; Cuzick *et al*, 2013). Heterogeneous expression of PTEN was found in 75% of the RP specimens in the RP subgroup when we analysed all blocks with a tumour area $\geqslant 4\,mm^2$. Within individual tumour areas, we found heterogeneity in PTEN expression in 16% of the samples, which is high, compared to previous studies reporting heterogeneity in $<1\%$ of samples (Cuzick *et al*, 2013). These results may question the reproducibility of PTEN expression as a biomarker when only a small area of the tumour is investigated, and could also explain why it has been difficult to implement promising biomarkers in the clinic.

Study limitations. Based on our classification of tumour foci in tissue blocks as diploid, tetraploid or aneuploid, we found that only 14% of the total median tumour volume in the radical resections was classified as non-diploid. Classifying a tissue block as non-diploid in this way is only based on the presence of aneuploid or tetraploid nuclei in the DNA histogram, which does not indicate that all tumour cells in that tissue block were non-diploid. The actual proportion of non-diploid cell nuclei was therefore substantially less than 14% of the median tumour volume. Also, we did not attempt to use Gleason grading, DNA ploidy or PTEN expression on the many small tumour foci of area $\leqslant 4\,mm^2$ for radical prostatectomies. From other studies, we know that these are usually, but not invariably, low grade and diploid (Deitch *et al*, 1993).

CONCLUSION

In this study, we demonstrated intraprostatic and intrafocal heterogeneity in Gleason score, DNA ploidy and PTEN expression in prostate cancer. Our results clearly show that for all three markers, a single sample does not provide representative information for the tumour as a whole. This high degree of genomic diversity in locally advanced prostate cancer is a major obstacle in clinical practice and research on prognostic markers. DNA ploidy status provided useful prognostic information, separating the diploid, tetraploid and aneuploid patients, only when heterogeneity was taken into account. Although many studies on biomarkers are still relying on only one sample, our findings are in line with a growing number of papers suggesting that analysis of multiple samples is necessary to compensate for heterogeneity.

ACKNOWLEDGEMENTS

We thank Marna Lill Kjæreng, Naomi Målbakken and Hege Friesgaard for technical assistance, as well as Marian Seiergren and Paul Callaghan for creating the figures. This work was supported by the Helse Sør-Øst research fund (project number 2012025 and 2013133).

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

Aihara M, Wheeler TM, Ohori M, Scardino PT (1994) Heterogeneity of prostate cancer in radical prostatectomy specimens. *Urology* **43**(1): 60–66.

Andor N, Graham TA, Jansen M, Xia LC, Aktipis CA, Petritsch C, Ji HP, Maley CC (2015) Pan-cancer analysis of the extent and consequences of intratumor heterogeneity. *Nat Med* **22**(1): 105–113.

Andreoiu M, Cheng L (2010) Multifocal prostate cancer: biologic, prognostic, and therapeutic implications. *Hum Pathol* **41**(6): 781–793.

Arora R, Koch MO, Eble JN, Ulbright TM, Li L, Cheng L (2004) Heterogeneity of Gleason grade in multifocal adenocarcinoma of the prostate. *Cancer* **100**(11): 2362–2366.

Attard G, Parker C, Eeles RA, Schröder F, Tomlins SA, Tannock I, Drake CG, De Bono JS (2016) Prostate cancer. *Lancet* **387**(10013): 70–82.

Boutros PC, Fraser M, Harding NJ, de Borja R, Trudel D, Lalonde E, Meng A, Hennings-Yeomans PH, McPherson A, Sabelnykova VY, Zia A, Fox NS, Livingstone J, Shiah Y-J, Wang J, Beck TA, Have CL, Chong T, Sam M, Johns J, Timms L, Buchner N, Wong A, Watson JD, Simmons TT, P'ng C, Zafarana G, Nguyen F, Luo X, Chu KC, Prokopec SD, Sykes J, Dal Pra A, Berlin A, Brown A, Chan-Seng-Yue MA, Yousif F, Denroche RE, Chong LC, Chen GM, Jung E, Fung C, Starmans MHW, Chen H, Govind SK, Hawley J, D'Costa A, Pintilie M, Waggott D, Hach F, Lambin P, Muthuswamy LB, Cooper C, Eeles R, Neal D, Tetu B, Sahinalp C, Stein LD, Fleshner N, Shah SP, Collins CC, Hudson TJ, McPherson JD, van der Kwast T, Bristow RG (2015) Spatial genomic heterogeneity within localized, multifocal prostate cancer. *Nat Genet* **47**(7): 736–745.

Burrell RA, McGranahan N, Bartek J, Swanton C (2013) The causes and consequences of genetic heterogeneity in cancer evolution. *Nature* **501**(7467): 338–345.

Böcking A, Tils M, Schramm M, Dietz J, Biesterfeld S (2014) DNA-cytometric grading of prostate cancer systematic review with descriptive data analysis. *Pathol Discov* **2**(1): 7.

Center MM, Jemal A, Lortet-Tieulent J, Ward E, Ferlay J, Brawley O, Bray F (2012) International variation in prostate cancer incidence and mortality rates. *Eur Urol* **61**(6): 1079–1092.

Cooper CS, Eeles R, Wedge DC, Van Loo P, Gundem G, Alexandrov LB, Kremeyer B, Butler A, Lynch AG, Camacho N, Massie CE, Kay J, Luxton HJ, Edwards S, Kote-Jarai Z, Dennis N, Merson S, Leongamornlert D, Zamora J, Corbishley C, Thomas S, Nik-Zainal S, O'Meara S, Matthews L, Clark J, Hurst R, Mithen R, Bristow RG, Boutros PC, Fraser M, Cooke S, Raine K, Jones D, Menzies A, Stebbings L, Hinton J, Teague J, McLaren S, Mudie L, Hardy C, Anderson E, Joseph O, Goody V, Robinson B, Maddison M, Gamble S, Greenman C, Berney D, Hazell S, Livni N, Fisher C, Ogden C, Kumar P, Thompson A, Woodhouse C, Nicol D, Mayer E, Dudderidge T, Shah NC, Gnanapragasam V, Voet T, Campbell P, Futreal A, Easton D, Warren AY, Foster CS, Stratton MR, Whitaker HC, McDermott U, Brewer DS, Neal DE (2015) Analysis of the genetic phylogeny of multifocal prostate cancer identifies multiple independent clonal expansions in neoplastic and morphologically normal prostate tissue. *Nat Genet* **47**(4): 367–372.

Cooperberg MR, Davicioni E, Crisan A, Jenkins RB, Ghadessi M, Karnes RJ (2015) Combined value of validated clinical and genomic risk stratification tools for predicting prostate cancer mortality in a high-risk prostatectomy cohort. *Eur Urol* **67**(2): 326–333.

Cuzick J, Yang ZH, Fisher G, Tikishvili E, Stone S, Lanchbury JS, Camacho N, Merson S, Brewer D, Cooper CS, Clark J, Berney DM, Møller H, Scardino P, Sangale Z (2013) Prognostic value of PTEN loss in men with conservatively managed localised prostate cancer. *Br J Cancer* **108**(12): 2582–2589.

Cyll K, Callaghan P, Kildal W, Danielsen HE (2015) Preparing for image based DNA ploidy (2015). [Online Video]. Available at https://www.youtube.com/watch?v=_24EkrYAwOc.

Danielsen HE, Pradhan M, Novelli M (2016) Revisiting tumour aneuploidy—the place of ploidy assessment in the molecular era. *Nat Rev Clin Oncol* **13**: 291–314.

Deitch AD, Miller GJ, deVere WR (1993) Significance of abnormal diploid DNA histograms in localized prostate cancer and adjacent benign prostatic tissue. *Cancer* **72**(5): 1692–1700.

Epstein JI (2010) An update of the gleason grading system. *J Urol* **183**(2): 433–440.

Epstein JI, Egevad L, Amin MB, Delahunt B, Srigley JR, Humphrey PA. Grading Committee (2016) The 2014 International Society of Urological Pathology (ISUP) Consensus Conference on Gleason Grading of Prostatic Carcinoma Definition of Grading Patterns and Proposal for a New Grading System. *Am J Surg Pathol* **40**(2): 244–252.

Ferraldeschi R, Nava Rodrigues D, Riisnaes R, Miranda S, Figueiredo I, Rescigno P, Ravi P, Pezaro C, Omlin A, Lorente D, Zafeiriou Z, Mateo J, Altavilla A, Sideris S, Bianchini D, Grist E, Thway K, Perez Lopez R, Tunariu N, Parker C, Dearnaley D, Reid A, Attard G, De Bono J (2015) PTEN protein loss and clinical outcome from castration-resistant prostate cancer treated with abiraterone acetate. *Eur Urol* **67**(4): 795–802.

Fraser M, Berlin A, Bristow RG, van der Kwast T (2014) Genomic, pathological, and clinical heterogeneity as drivers of personalized medicine in prostate cancer. *Urol Oncol* **33**(2): 85–94.

Freedland SJ, Choeurng V, Howard L, De Hoedt A, du Plessis M, Yousefi K, Lam LL, Buerki C, Ra S, Robbins B, Trabulsi EJ, Shah NL, Abdollah F, Feng FY, Davicioni E, Dicker AP, Karnes RJ, Den RB (2016) Utilization of a genomic classifier for prediction of metastasis following salvage radiation therapy after radical prostatectomy. *Eur Urol* **70**: 588–596.

Freedland SJ, Gerber L, Reid J, Welbourn W, Tikishvili E, Park J, Younus A, Gutin A, Sangale Z, Lanchbury JS, Salama JK, Stone S (2013) Prognostic utility of cell cycle progression score in men with prostate cancer after primary external beam radiation therapy. *Int J Radiat Oncol Biol Phys* **86**(5): 848–853.

Gerlinger M, Catto JW, Orntoft TF, Real FX, Zwarthoff EC, Swanton C (2015) Intratumour heterogeneity in urologic cancers: from molecular evidence to clinical implications. *Eur Urol* **67**(4): 729–737.

Haffner M, Mosbruger T (2013) Tracking the clonal origin of lethal prostate cancer. *J Clin Invest* **123**(11): 4918–4922.

Hanahan D, Weinberg RA (2011) Hallmarks of cancer: the next generation. *Cell* **144**(5): 646–674.

Karavitakis M, Ahmed HU, Abel PD, Hazell S, Winkler MH (2011) Tumor focality in prostate cancer: implications for focal therapy. *Nat Rev Clin Oncol* **8**(1): 48–55.

Keyes M, Macaulay C, Hayes M, Korbelik J, Morris WJ, Palcic B (2013) DNA ploidy measured on archived pretreatment biopsy material may correlate with prostate-specific antigen recurrence after prostate brachytherapy. *Int J Radiat Oncol Biol Phys* **86**(5): 829–834.

Lawrence MS, Stojanov P, Mermel CH, Robinson JT, Garraway LA, Golub TR, Meyerson M, Gabriel SB, Lander ES, Getz G (2014) Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature* **505**(7484): 495–501.

López-García C, Sansregret L, Domingo E, McGranahan N, Hobor S, Birkbak NJ, Horswell S, Grönroos E, Favero F, Rowan AJ, Matthews N, Begum S, Phillimore B, Burrell R, Oukrif D, Spencer-Dene B, Kovac M, Stamp G, Stewart A, Danielsen H, Novelli M, Tomlinson I, Swanton C (2017) BCL9L dysfunction impairs caspase-2 expression permitting aneuploidy tolerance in colorectal cancer. *Cancer Cell* **31**(1): 79–93.

McGranahan N, Burrell RA, Endesfelder D, Novelli MR, Swanton C (2012) Cancer chromosomal instability: therapeutic and diagnostic challenges. *EMBO Rep* **13**(6): 528–538.

O'Malley FP, Grignon DJ, Keeney M, Kerkvliet N, McLean C (1993) DNA Heterogeneity in prostatic adenocarcinoma. A DNA flow cytometric mapping study with whole organ sections of prostate. *Cancer* **71**(9): 2797–2802.

Perera M, Lawrentschuk N, Bolton D, Clouston D (2014) Comparison of contemporary methods for estimating prostate tumour volume in pathological specimens. *BJU Int* **113**(Suppl 2): 29–34.

Punt CJA, Buyse M, Köhne CH, Hohenberger P, Labianca R, Schmoll HJ, Påhlman L, Sobrero A, Douillard JY (2007) Endpoints in adjuvant treatment trials: A systematic review of the literature in colon cancer and proposed definitions for future trials. *J Natl Cancer Inst* **99**(13): 998–1003.

Sengupta S, Cheville JC, Lohse CM, Zincke H, Myers RP, Riehle DL, Pankratz VS, Blute ML, Sebo TJ (2006) Conventional assessment of needle biopsy specimens is more useful than digital image analysis of proliferation and DNA ploidy in prediction of positive surgical margins at radical prostatectomy. *Urology* **68**(1): 94–98.

Stark JR, Perner S, Stampfer MJ, Sinnott JA, Finn S, Eisenstein AS, Ma J, Fiorentino M, Kurth T, Loda M, Giovannucci EL, Rubin MA, Mucci LA (2009) Gleason score and lethal prostate cancer: Does $3+4=4+3$? *J Clin Oncol* **27**(21): 3459–3464.

Talantov D, Jatkoe TA, Böhm M, Zhang Y, Ferguson AM, Stricker PD, Kattan MW, Sutherland RL, Kench JG, Wang Y, Henshall SM (2010) Gene based prediction of clinically localized prostate cancer progression after radical prostatectomy. *J Urol* **184**(4): 1521–1528.

Wang N, Wilkin C (2000) Evaluation of tumor heterogeneity of prostate carcinoma by flow- and image DNA cytometry and histopathological grading. *Anal Cell Pathol* **20**(1): 49–62.

Wei L, Wang J, Lampert E, Schlanger S, DePriest AD, Hu Q, Gomez EC, Murakam M, Glenn ST, Conroy J, Morrison C, Azabdaftari G, Mohler JL, Liu S, Heemers HV (2017) Intratumoral and intertumoral genomic heterogeneity of multifocal localized prostate cancer impacts molecular classifications and genomic prognosticators. *Eur Urol* **71**(2): 183–192.

Yin Y, Shen WH (2008) PTEN: a new guardian of the genome. *Oncogene* **27**(41): 5443–5453.

This work is licensed under the Creative Commons Attribution-Non-Commercial-Share Alike 4.0 International License. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc-sa/4.0/

© The Author(s) named above 2017

Supplementary Information accompanies this paper on British Journal of Cancer website (http://www.nature.com/bjc)