


RESEARCH

Open Access



Decoding the evolutionary response to prostate cancer therapy by plasma genome sequencing

Naveen Ramesh^{1,2}, Emi Sei¹, Pei Ching Tsai³, Shanshan Bai³, Yuehui Zhao¹, Patricia Troncoso⁴, Paul G. Corn^{3,5}, Christopher Logothetis^{3,5}, Amado J. Zurita^{3,5**†} and Nicholas E. Navin^{1,2,5,6**†} 

* Correspondence: azurita@mdanderson.org; nnavin@mdanderson.org

[†]Amado J. Zurita and Nicholas E. Navin contributed equally to this work.

³Department of Genitourinary Medical Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA

¹Department of Genetics, The University of Texas MD Anderson Cancer Center, Houston, TX, USA
Full list of author information is available at the end of the article

Abstract

Background: Investigating genome evolution in response to therapy is difficult in human tissue samples. To address this challenge, we develop an unbiased whole-genome plasma DNA sequencing approach that concurrently measures genomic copy number and exome mutations from archival cryostored plasma samples. This approach is applied to study longitudinal blood plasma samples from prostate cancer patients, where longitudinal tissue biopsies from the bone and other metastatic sites have been challenging to collect.

Results: A molecular characterization of archival plasma DNA from 233 patients and genomic profiling of 101 patients identifies clinical correlations of aneuploid plasma DNA profiles with poor survival, increased plasma DNA concentrations, and lower plasma DNA size distributions. Deep-exome sequencing and genomic copy number profiling are performed on 23 patients, including 9 patients with matched metastatic tissues and 12 patients with serial plasma samples. These data show a high concordance in genomic alterations between the plasma DNA and metastatic tissue samples, suggesting the plasma DNA is highly representative of the tissue alterations. Longitudinal sequencing of 12 patients with 2–5 serial plasma samples reveals clonal dynamics and genome evolution in response to hormonal and chemotherapy. By performing an integrated evolutionary analysis, minor subclones are identified in 9 patients that expanded in response to therapy and harbored mutations associated with resistance.

Conclusions: This study provides an unbiased evolutionary approach to non-invasively delineate clonal dynamics and identify clones with mutations associated with resistance in prostate cancer.

Keywords: Tumor evolution, Liquid biopsies, Non-invasive



Background

Clonal evolution has been challenging to study in tumor tissue obtained from cancer patients, particularly in specimens collected from a single point in time [1–3]. While a small number of studies were able to collect serial core biopsies for genomic profiling [4], those studies still suffered from small amounts of tumor tissue availability and spatial sampling bias. Moreover, tumor biopsies are invasive clinical procedures with potential for complications and have significant costs [5, 6]. An alternative non-invasive approach involves using liquid biopsies, including circulating tumor cells [7–9] and cell-free DNA (cfDNA) [10–13]. Blood samples collected over time during the course of treatment provide a unique opportunity to infer tumor evolution [14]. cfDNA is particularly useful for clinical applications due to the logistical advantage of straightforward processing and the ability to cryostore materials for future analysis [15–17]. However, a major limitation has been that most cfDNA assays were developed to analyze targeted cancer gene panels and have therefore measured a limited number of CNAs and mutations [18–21]. Resolving intratumor heterogeneity and inferring clonal evolution requires measuring a large number of unbiased genomic markers, which targeted panels cannot provide.

A few studies have made initial progress towards performing unbiased exome sequencing of cfDNA in gastrointestinal cancers [22] and prostate cancer (PC) [10], or whole-genome copy number profiling of triple-negative breast cancer patients [23]. Building on those early studies, we have developed a method called PEGASUS (Plasma Exome and Genome Analysis by Size-selection Unbiased sequencing) to profile both genome-wide CNA and exome-wide mutations (~ 25,000 genes) simultaneously from archival cryostored plasma or serum DNA samples. In contrast to targeted methods [24–26], PEGASUS involves selection of small DNA fragments that contain higher tumor content relative to high molecular weight DNA from WBCs present in cryostored blood fractions. These characteristics make PEGASUS ideally suited for unbiased discovery of genomic markers and for the investigation of clonal evolution in response to therapy, when applied to longitudinal blood collections.

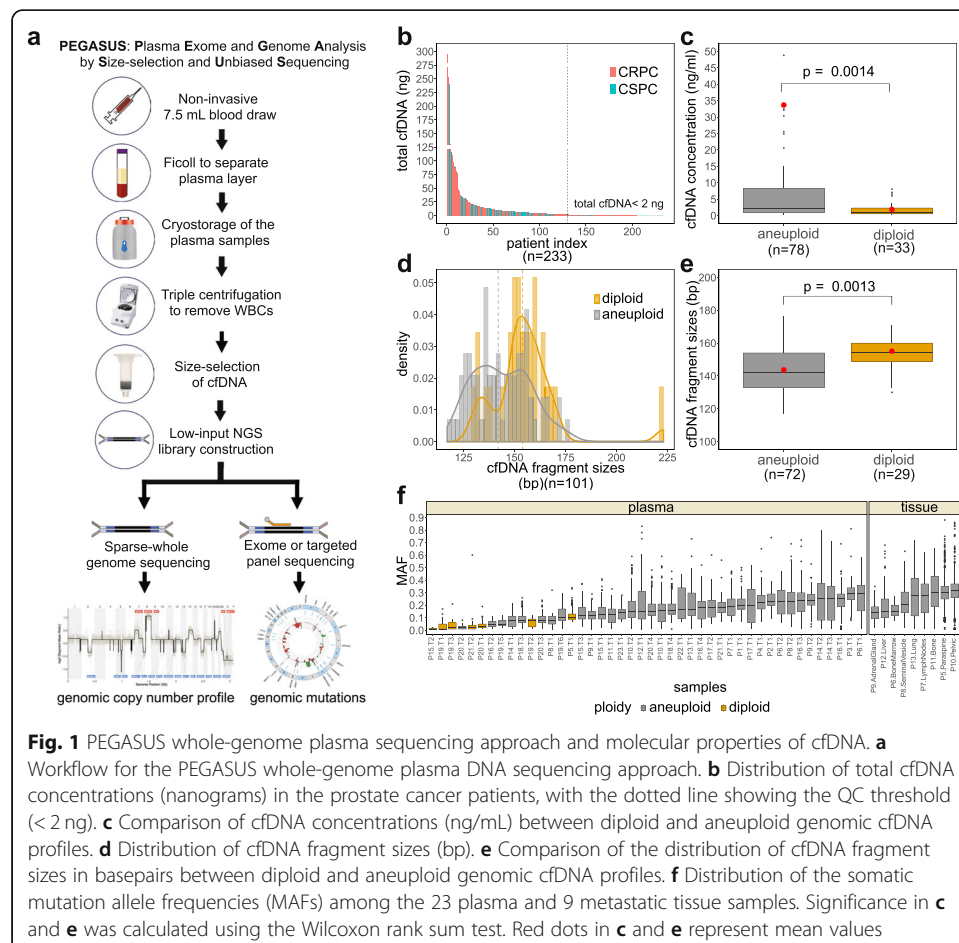
Comprehensive genomic studies have identified multiple recurrently altered genes (*AR*, *TP53*, *RB*, *SPOP*, *PTEN*, and others) and signaling pathways in metastatic prostate cancer (mPC) [27], but have also revealed the existence of many low-frequency variants and significant inter-patient heterogeneity. Moreover, PC is unique among solid tumors due to its dependence on androgen receptor-regulated pathways for progression and its high propensity to metastasize to the bone (often the only site of progression). The vast majority of previously untreated or castration-sensitive prostate cancer (CSPC) respond to initial androgen deprivation therapy (ADT), but the disease invariably adapts and progresses to lethal castration-resistance prostate cancer (CRPC) [28]. Treatment strategies for CRPC patients are limited and include advanced hormonal therapies and taxane-based chemotherapy, but no molecular markers have yet been established to guide their application and timing of treatment. Because of the bone dominance of metastatic prostate cancer (mPC), bone biopsies are most frequently needed for pathological and molecular characterization, but these are uncomfortable and difficult to perform procedures that often result in minute amounts of tumor tissue that is not suitable for genomic analysis.

As a result, mPC presents a unique opportunity for liquid biopsy genomics to investigate heterogeneity and clonal evolution over time in the context of therapy.

Results

PEGASUS method

While prospective studies typically use Streck tubes to collect plasma, most archival plasma samples have historically been collected in EDTA tubes or subjected to a gradient treatment such as Ficoll prior to cryostorage. In contrast to Streck tubes (which include a fixative), the usage of EDTA/Ficoll medium often leads to contamination by high molecular weight DNA fragments that are released from the WBC during sample processing. To address this problem, PEGASUS was designed to isolate low molecular weight DNA fragments (< 1000 bp), which is done prior to the construction of low-input libraries for whole-genome sequencing (WGS) of copy number alterations (CNAs) and exome mutation profiling (Fig. 1a, the “Methods” section). After three rounds of high-speed centrifugation to remove residual WBC, the archival plasma samples are subjected to size selection by column purification. The remaining low molecular weight DNA is used for quality control (QC) to determine the plasma DNA concentration and molecular size of the DNA fragments. Samples that pass QC are used to construct low-input NGS libraries, which are split into two parallel reactions:



(1) copy number profiling by WGS at sparse (0.1X) coverage depth and (2) exome sequencing at high coverage depth (150X tumor, 60X normal) to detect somatic mutations and indels (the “Methods” section). NGS libraries are generated in parallel from WBCs for exome capture to serve as a matched normal reference of germline variants to distinguish somatic mutations. This protocol can also be performed on fresh blood samples collected in Streck tubes.

Molecular properties of cfDNA and clinical correlations

We performed QC of cfDNA in 233 cryostored archival plasma samples collected from PC patients at MD Anderson by measuring the size and concentration, including 79 patients with CSPC and 154 patients with CRPC (Fig. 1b). QC indicated that 130 of the 233 patients (55.8%) had sufficient cfDNA (≥ 2 ng total) for construction of NGS libraries using PEGASUS. The QC data showed a wide range in the total amounts of low molecular weight cfDNA (0–5280 ng) across the patients, including a subset of 29 patients (12.4% of the total) with very high amounts of total cfDNA (≥ 20 ng) (left side, Fig. 1b). The median cfDNA concentrations in CSPC (1.25 ng/mL) and CRPC patients (0.84 ng/mL) were not significantly different ($p = 0.29$, Wilcoxon test). However, 24 CSPC patients (30.4%) had < 0.5 ng of detectable plasma DNA compared to only 5 (3.2%) of the CRPC patients (Fig. 1b and Additional file 1: Fig. S2a), suggesting that a larger proportion of CRPC patients could potentially be used for genomic profiling using PEGASUS. We performed low-pass WGS of 101 plasma samples to understand how diploid and aneuploid genomic copy number profiles correlate with cfDNA properties. This data showed that cfDNA concentrations were significantly higher in patients with aneuploid genomes compared to patients with diploid plasma copy number ($p = 0.00014$, Wilcoxon rank sum test) (Fig. 1c).

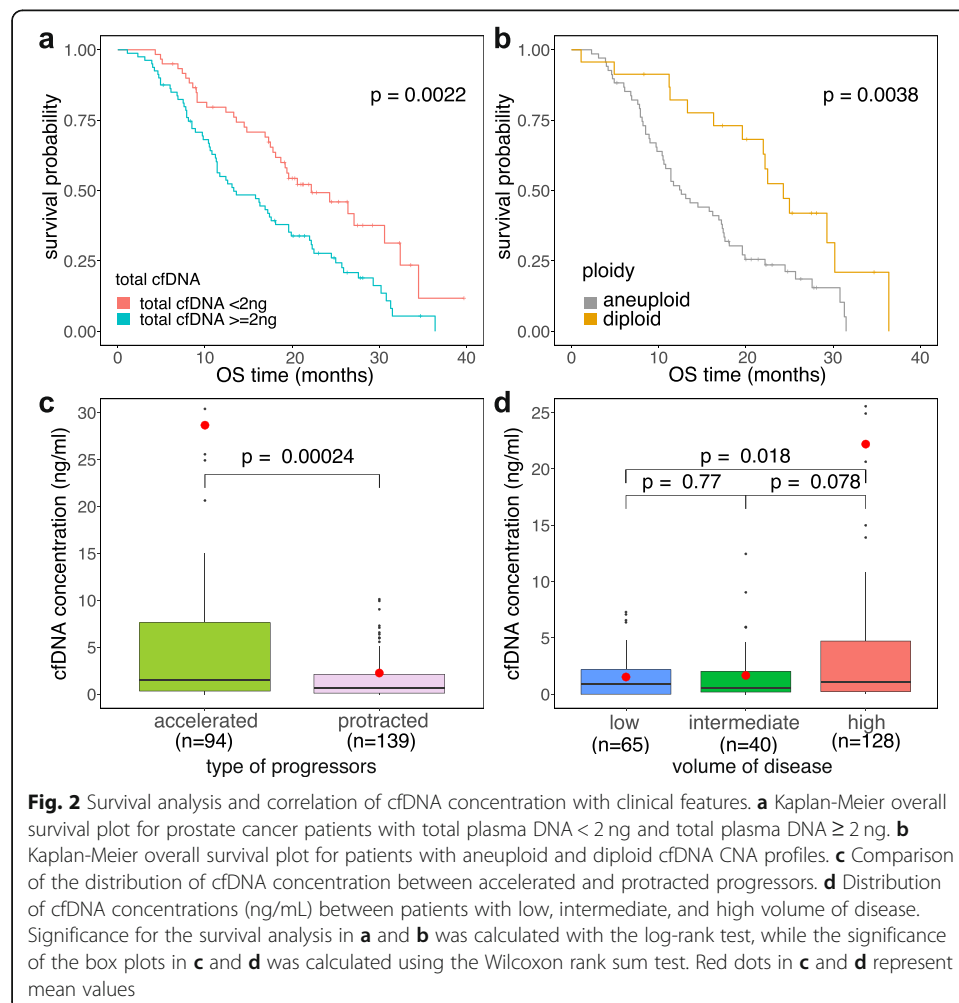
Analysis of the cfDNA fragment sizes showed a mean molecular size of 146.6 ± 1.29 bp (SEM) across the patients, with a bimodal cfDNA size distribution (Fig. 1d, Additional file 1: Fig. S1). PC patients with diploid copy number had significantly ($p = 0.0013$, Wilcoxon rank sum test) larger cfDNA fragment sizes (mean 154 ± 1.66 bp) compared to patients with aneuploid profiles (mean 142 ± 2.91 bp) (Fig. 1d, e). These numbers are consistent with previous reports on plasma DNA fragment size distributions for circulating tumor DNA (ctDNA) compared to cfDNA that was isolated from normal cells [29, 30], suggesting that the tumor cells are not shedding large amounts of DNA into the plasma. Notably, a subset of patients had additional peaks at double and triple the mean cfDNA fragment size, possibly indicating that the DNA was protected by multiple nucleosomes, rather than a single nucleosome detected in most PC patients (Additional file 1: Fig. S1a-d).

While most of the 233 patient samples used for QC were from single-timepoint blood samples, there were 9 patients with matched metastatic tissues and 12 patients with longitudinal blood samples collected during therapy that were selected for genomic CNA profiling and exome sequencing by PEGASUS. To test the variability in ctDNA concentration after isolation, we performed 4 experimental replicates from 4 test plasma samples, which showed only minor variations in the final concentrations of ctDNA (Additional file 1: Fig. S1e). Analysis of the somatic exome mutation allele frequency (MAF) showed a median MAF of 0.17 for the aneuploid cfDNA samples, which

was slightly lower than the MAF detected in matched metastatic tumor tissues (median 0.3) (Fig. 1f). Notably, plasma DNA with diploid genomic copy number profiles showed the lowest MAF (median 0.08), suggesting that the contribution of cfDNA from normal cells was high in these patients, with limited ctDNA.

Clinical correlations with plasma DNA properties

Survival analysis revealed that CRPC patients with total cfDNA < 2 ng had significantly longer overall survival (OS) than patients with total cfDNA \geq 2 ng (median 22.2 vs 13.3 months, $p = 0.0022$, log-rank test) (Fig. 2a, Additional file 1: Fig. S2). Furthermore, patients with diploid cfDNA CNA profiles had longer OS than patients with aneuploid cfDNA CNA profiles (median 21.05 vs 12.6 months, $p = 0.031$, log-rank test) (Fig. 2b, Additional file 1: Fig. S2). We next investigated associations between cfDNA concentrations and clinical parameters. Patients with accelerated PC growth had higher cfDNA levels ($n = 94$) compared to those with protracted kinetics of progression ($n = 139$; $p = 0.00024$, Wilcoxon rank sum test) (Fig. 2c). Furthermore, PC patients with high disease volume ($n = 128$) had significantly higher cfDNA levels compared to patients with low disease volume ($n = 65$; $p = 0.018$, Wilcoxon rank sum test) (Fig. 2d). Notably, cfDNA



concentrations were not significantly different ($p = 0.4642$, Wilcoxon rank sum test) in patients with histologic low grade groups 1–2 ($n = 54$) compared to high grade groups 3–5 ($n = 164$) on diagnosis (Additional file 1: Fig. S2d). By comparing local and distant metastases, PC patients with bone ($n = 132$) or visceral metastasis ($n = 41$) had respectively higher cfDNA concentrations than patients with lymph node metastasis only ($n = 30$; $p = 0.042$ and $p = 0.0099$ for bone and visceral metastasis, respectively, vs lymph node) (Additional file 1: Fig. S2e). Similarly, PSA levels had a low correlation with cfDNA concentrations ($R = 0.12$, $p = 0.073$) (Additional file 1: Fig. S2f). Finally, we utilized all the clinical and genomic factors as predictors for the OS and progression-free survival (PFS) using the univariate and multivariate Cox regression model in a subset of 70 patients that had consistent clinical variables available for all of the parameters we tested. Based on the survival analysis, we found that ploidy has significant independent and joint predictive power for OS, with significant joint predictive power for PFS (Additional file 1: Table S1). Collectively, these data suggest that higher cfDNA in the plasma and genomic aneuploidy associate with faster mPC progression and a shorter survival of PC patients.

Whole-genome profiling of single-timepoint samples

We applied PEGASUS to obtain integrated genomic CNA and exome mutation profiling of single-timepoint plasma samples from 8 CRPC patients (Fig. 3, Additional file 1: Fig. S3). These patients had a mean 34.8 ± 6.49 (SEM) CNAs and 55.6 ± 14.99 (SEM) point mutations and 16 ± 3.15 (SEM) insertion-deletions (indels). However, two of the patients (P1, P4) had a much higher mutation burden (106 and 139 mutations, respectively) (Fig. 3a). Recurrent CNAs identified included amplifications in *MYC* (8q) in all 8 patients and in *AR* (Xq) in 4 patients, as well as losses in *RBI* (13q) in 7 patients, *TP53* (17p) in 4 patients, and *APC* (5q) in 3 patients (Fig. 3c, Additional file 1: Fig. S3a). Point mutations were identified in *GNAS*, *NCOA5*, *EVL*, and *BIRC6*, while indels were found in *BRCA2*, *APC*, *TP53*, and *FOXA1* among other cancer genes. Our unbiased analysis also identified CNAs and mutations in genes that have not previously been associated with mPC progression (Fig. 3c, Additional file 1: Fig. S3b). The overall distribution of the CNAs among the 8 single-timepoint samples shows a negative binomial distribution, with a majority of the CNA genomic sizes $< 50,000$ kb (Additional file 1: Fig. S3c). On a per-patient basis, the chromosome length of the CNAs are generally $< 50,000$ kb while a subset of CNAs have larger size distributions (Additional file 1: Fig. S3d). This data demonstrates the technical feasibility of using PEGASUS to perform unbiased genomic profiling and detect both recurrent and infrequent PC aberrations in cfDNA.

Tumor DNA concordance in plasma and metastatic tissue

To investigate the concordance of genomic events between the cfDNA and metastatic tumor tissues, we applied PEGASUS to 9 CRPC patients with matched tissue specimens obtained from different metastatic organ sites (Fig. 4a). Global genomic analysis showed that the CNA burden in plasma and metastatic tumor tissue was highly correlated (mean $r = 0.9$) in most patients (P8–P12) (Fig. 4b). However, in 4 patients (P5, P6, P7, P13), the CNA burden in the plasma was lower than the corresponding tumor tissue (mean $r = 0.6$), suggesting that some tumor clones in the metastatic sites did not

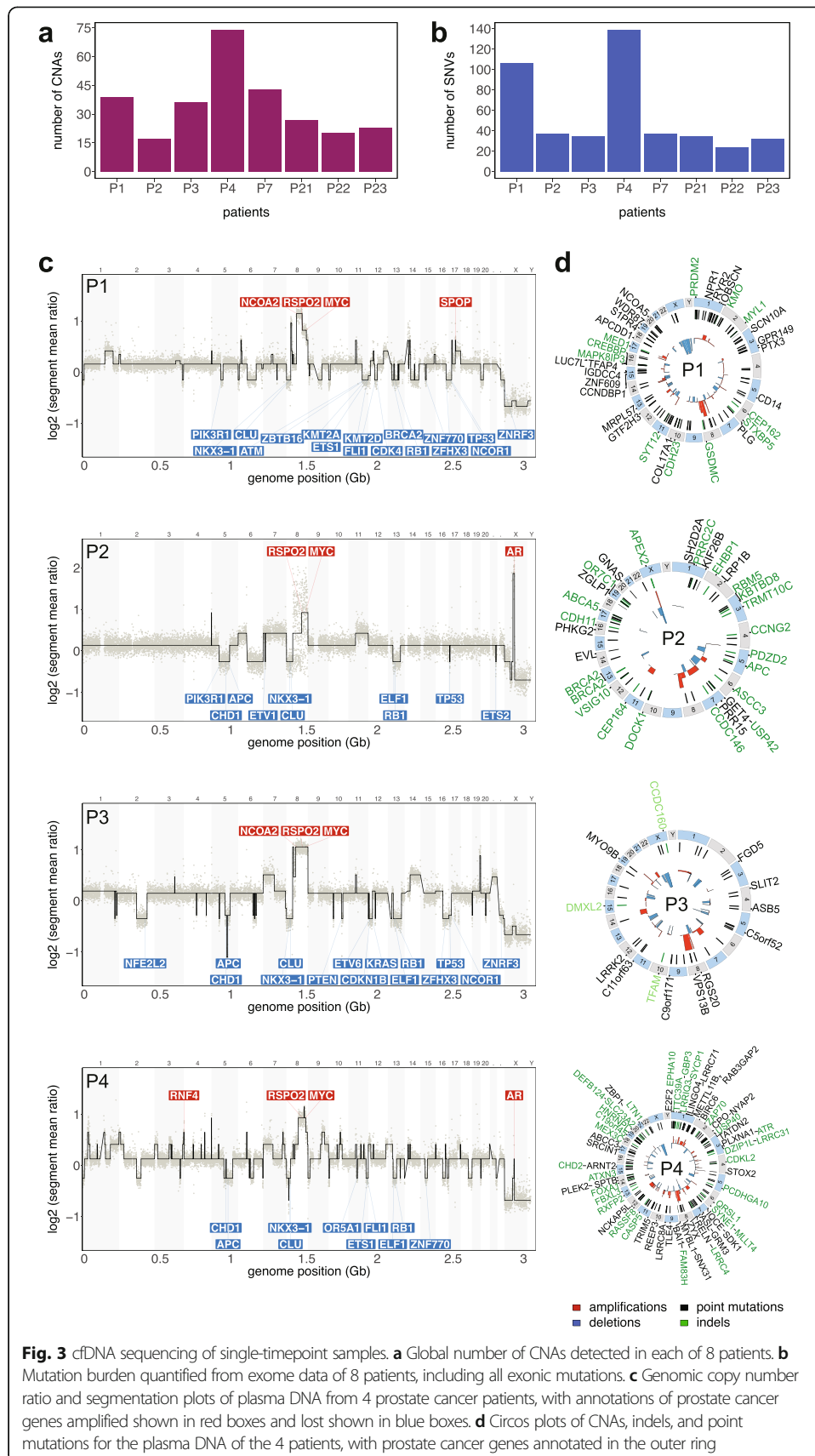
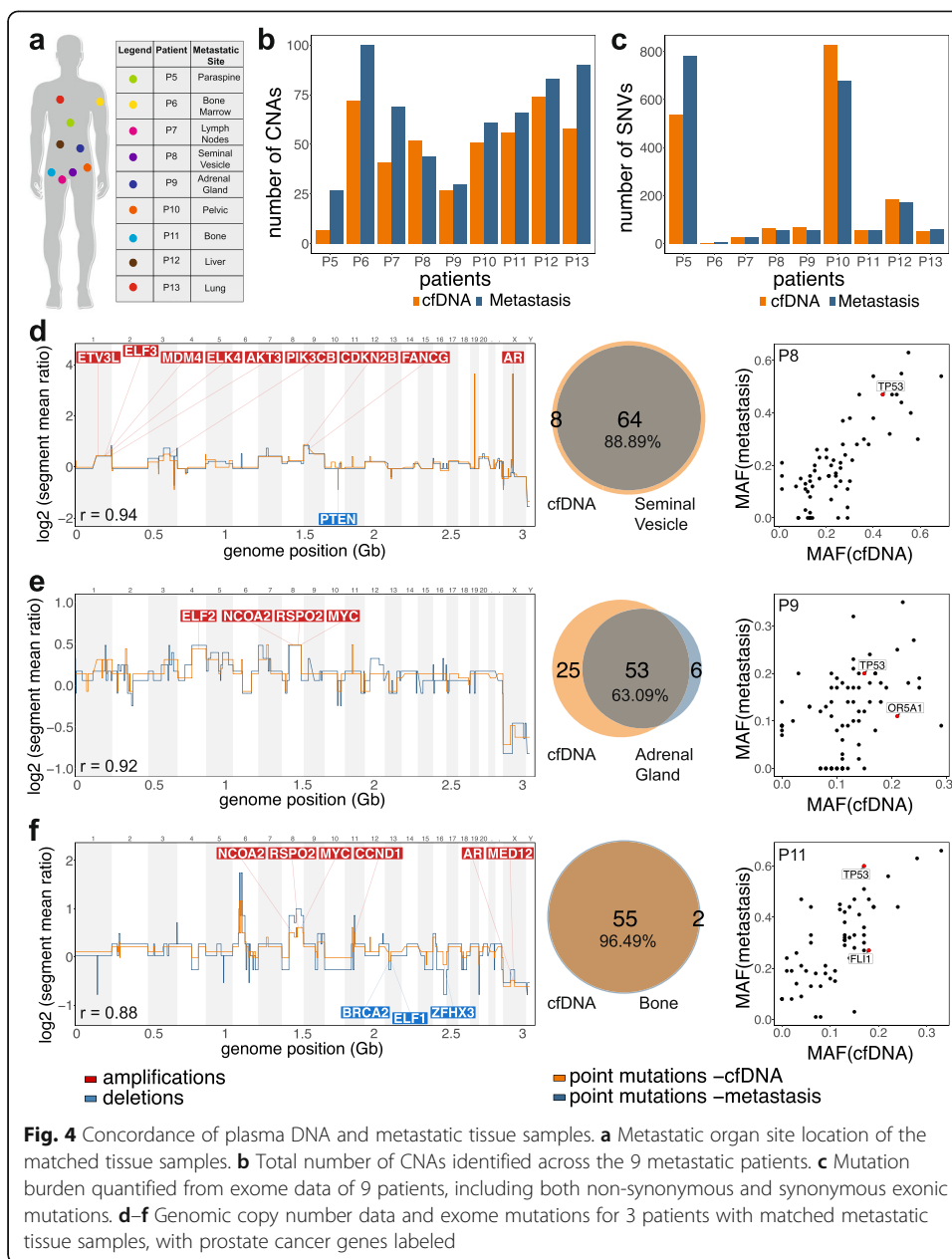


Fig. 3 cfDNA sequencing of single-timepoint samples. **a** Global number of CNAs detected in each of 8 patients. **b** Mutation burden quantified from exome data of 8 patients, including all exonic mutations. **c** Genomic copy number ratio and segmentation plots of plasma DNA from 4 prostate cancer patients, with annotations of prostate cancer genes amplified shown in red boxes and lost shown in blue boxes. **d** Circos plots of CNAs, indels, and point mutations for the plasma DNA of the 4 patients, with prostate cancer genes annotated in the outer ring



shed sufficient DNA into the blood for detection (Fig. 4b). Similarly, the total mutation burden was highly concordant between the matched cfDNA and metastatic tissues (51–96%), including two patients (P5 and P10) who had a very large number of somatic mutations (> 200) in both the blood and tissue, consistent with a hypermutator genotype [31, 32] (Fig. 4c). In P10, deletions in both *MSH2* and *MSH6* were detected, while in P5, an *MLH1* deletion was identified, which may have contributed to the high mutation burden in these patients (Additional file 1: Fig. S4a; Additional file 1: Fig. S4d).

Further analysis showed that most CNAs, including amplification in *AR* and deletion in *PTEN* in P8, amplifications in *ELF2* and *MYC* in P9, and amplifications in *MYC* and *CCND1* in P11, were concordant in the plasma and tumor tissues (Fig. 4d–f). However, we also found a small number of discordant CNAs that were exclusive to the metastatic

tumor tissue (P6) or the cfDNA (P8, P9, P12, and P13). P6 had a focal amplification in 12p21 (3.72 mb, including *KRAS*) in a bone metastasis site that was not detected in the cfDNA. In P8, a focal amplification in 19p (1.52 mb) was not detected in the matched seminal vesicle metastasis, while in P9, a focal amplification in 11p (6.7 mb) was not detected in the corresponding adrenal gland metastasis. Similarly, P12 had amplifications in 8q (*MYC*) and Xp (*AR*), while P13 had two focal amplifications in Xp (*AR* and *ELK1*), which were not detected in the matched metastatic tissues (Additional file 1: Fig. S4f).

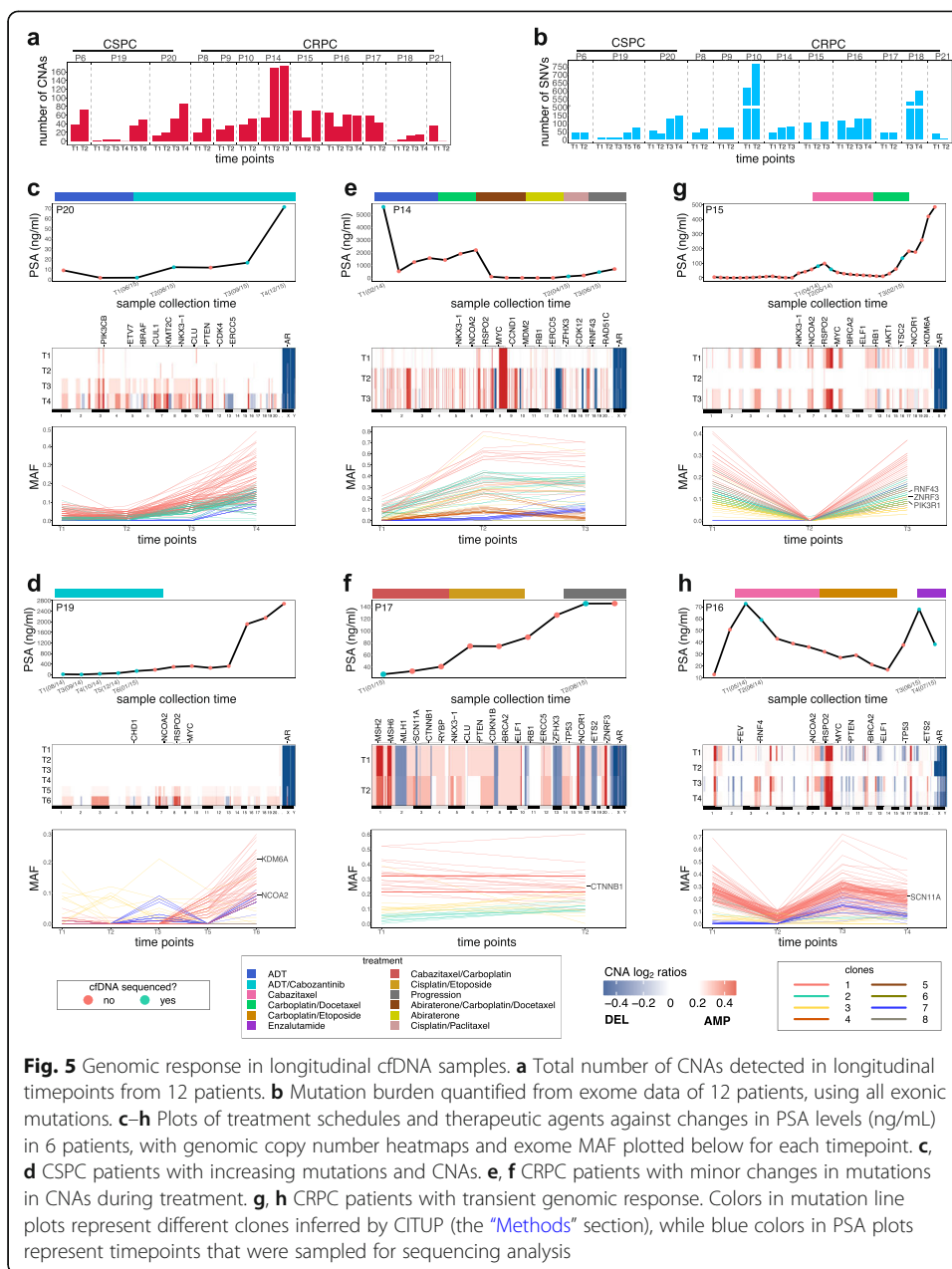
Analysis of the MAFs showed a linear correlation for most somatic mutations, but also identified mutations that were exclusive to either the plasma DNA or the metastatic sites (Fig. 4d–f, Additional file 1: Fig. S4a–S4f). However, most driver mutations including *TP53*, *AR*, *ATM*, *SPOP*, *FLII*, and *OR5A1* were detected in both the plasma DNA and tissues. Overall, the matched tissue data showed a high concordance in CNAs and point mutations between cfDNA and metastatic tissues, suggesting that the ctDNA is highly representative of many of the genomic aberrations detected in tumor tissues.

Genomic response to therapy in plasma DNA

We next applied PEGASUS to plasma DNA samples collected serially (2–6 timepoints) from 12 patients, including 9 CRPC patients (P8, P9, P10, P14, P15, P16, P17, P21) and 3 CSPC patients (P6, P19, P20), to study genomic response to therapy (Additional file 1: Table S2). The 3 CSPC patients were treated with ADT and a tyrosine kinase inhibitor (cabozantinib), while the CRPC patients received different chemotherapeutic and androgen-targeted agents (Fig. 5, Additional file 1: Table S2). From each timepoint, genomic CNA profiling and exome sequencing (mean depth 125X) were performed on the cfDNA, as well as matched normal PBMC samples (mean depth 77X) to detect germline variants.

On average, 126 somatic mutations (Fig. 5a) and 21 CNAs (Fig. 5b) were detected per patient, consistent with previously reported values in advanced prostate tumors [20, 31, 33, 34]. Integrated analysis revealed that the CNA burden (Fig. 5a) and point mutation burden (Fig. 5b) did not change substantially between the pre-treatment and post-treatment timepoints for most CRPC patients. However, in the mid-treatment timepoints, the CNA burden and mutation burden decreased substantially in several patients (P15, P16, P21), suggesting a transient genomic response to therapy. Notably, two CRPC patients (P10, P18) had high levels of somatic mutations (mean 601.5 SNVs), consistent with a hypermutator genotype [31, 32].

To investigate genomic response to therapy, we compared the plasma CNA profiles and exome MAF with the PSA levels for each patient (Fig. 5c–h, Additional file 1: Fig. S5a–S5f). In 2 CSPC patients (P19 and P20), we found increasing numbers of CNAs and mutations during treatment that corresponded to increasing PSA levels. In P6, who was hormone-naïve at the time of his first blood collection, an aneuploid profile was detected at both timepoints and accumulated CNAs at T2 as the disease progressed to CRPC. A comparable increase in mutation burden over time was found in one CRPC patient (P18) receiving chemotherapy (Additional file 1: Fig. S5d). In these patients, the increase in genomic aberrations over time was likely due to increasing tumor purity



due to increasing tumor volume and/or increasing tumor DNA shedding, rather than acquisition of new CNAs in response to treatment, since the total number of chromosome breakpoints did not change over timepoints.

In the 9 CRPC patients, complex aneuploid rearrangements and high MAF were detected at the baseline plasma sample (prior to treatment) and at the mid- or post-treatment timepoints. In P9, P10, P14, and P17, the complex aneuploid rearrangements and MAF were pre-existing before therapy and persisted through all of the timepoints analyzed, suggesting the tumors were intrinsically resistant to the therapies administered (Fig. 5e–f, Additional file 1: Fig. S5c–S5d). In contrast, P15, P16, and P21 showed transient genomic responses to therapy in the mid-treatment timepoints, in which the genomic CNA profiles approached a near-diploid state and the MAF decreased

substantially (Fig. 5g, h, Additional file 1: Fig. S5e). However, these responses were temporary, and the complex CNA profiles and high MAF returned in the later timepoints which is consistent with the change in PSA values.

In most patients, all of the CNAs were detected at the pre-treatment timepoint and did not change during treatment, suggesting the CNA had been acquired at earlier stages of tumor progression, prior to treatment (Fig. 6). This included amplifications in *AR* (10/12 patients), *MYC* (7/12 patients), and *NCOA2* (7/12 patients) and deletions in *PTEN* (5/12 patients), *RBI* (3/12), and *BRCA2* (3/12) (Fig. 5, Additional file 1: Fig. S5). The only exception was in P6 who acquired an *AR* amplification in response to ADT (Additional file 1: Fig. S5a).

In contrast to the CNAs, many mutations underwent dynamic changes in MAFs in response to therapy and the rise or drop of MAFs was fairly consistent with the rise or drop in PSA levels. In the CSPC patients, the MAFs increased (P19, P20) or remained stable (P6) during treatment timepoints. In the CRPC patients, the MAFs persisted with only minor frequency changes in patients with intrinsically resistant disease (e.g., P14, P17) or showed transient decreases in patients that were responding to therapy (e.g., P15, P16). This included point mutations in known PC driver genes, including *TP53* (P6, P9, P10), *AR* (P10), *SCN11A* (P16), *NCOA2* (P19), and other genes (Fig. 5, Additional file 1: Fig. S5). We also identified subclonal mutations that emerged during treatment, including mutations in *PTEN* (P21), *RNF43*, *PIK3R1* and *ZNRF3* (P15), *SPOP*, *RUNX1* and *AR* (P10), and *CTNBB1* (P17). Many other somatic mutations were detected in genes not associated with PC, and increased or decreased in MAFs during treatment, suggesting a potential association with sensitivity or response to the therapeutic agents.

Clonal evolution in response to therapy

To infer clonal evolution and identify subclones associated with resistance, we integrated the genomic CNA and exome mutation data from the serially collected plasma samples of the 12 patients. The confounding effects of tumor purity on CNA and MAF were normalized, and subclones were inferred across multiple timepoint samples using PyClone2 [35] and CITUP [36] (the “Methods” section, Additional file 1: Fig. S6). We identified multiple subclones (range 2–8) in the 12 patients and identified dynamic changes in clonal frequencies in response to therapy (Fig. 6). In 9 patients, minor subclones (1–21%) were identified that expanded in response to therapy (Fig. 6a), while in 3 patients (P6, P21, and P17), the most prevalent clone retained a similar frequency during treatment, suggesting that the dominant clone was intrinsically resistant to the treatment (Fig. 6b).

The clonal frequency data identified rare subclones that dynamically expanded during therapy and were therefore associated with resistance (Fig. 6a). For example, in P20, a minor subclone (clone G) with 7–11% clonal frequency in the initial timepoints (T1–T3) expanded to 75% at the final treatment timepoint (T4). This subclone harbored significant damaging mutations (SIFT < 0.05, POLYPHEN > 0.85) in two genes: *CAPN2* and *MRPL47*. In P18, a rare subclone (clone G) expanded from 6% at T3 to 53% at T4 and harbored three significant mutations: *SLC18A3*, *TNIK*, and *OR2AE1*. In P16, a rare subclone (clone G) expanded from 3 to 4% at the initial timepoints (T1–T2) to 41–57%

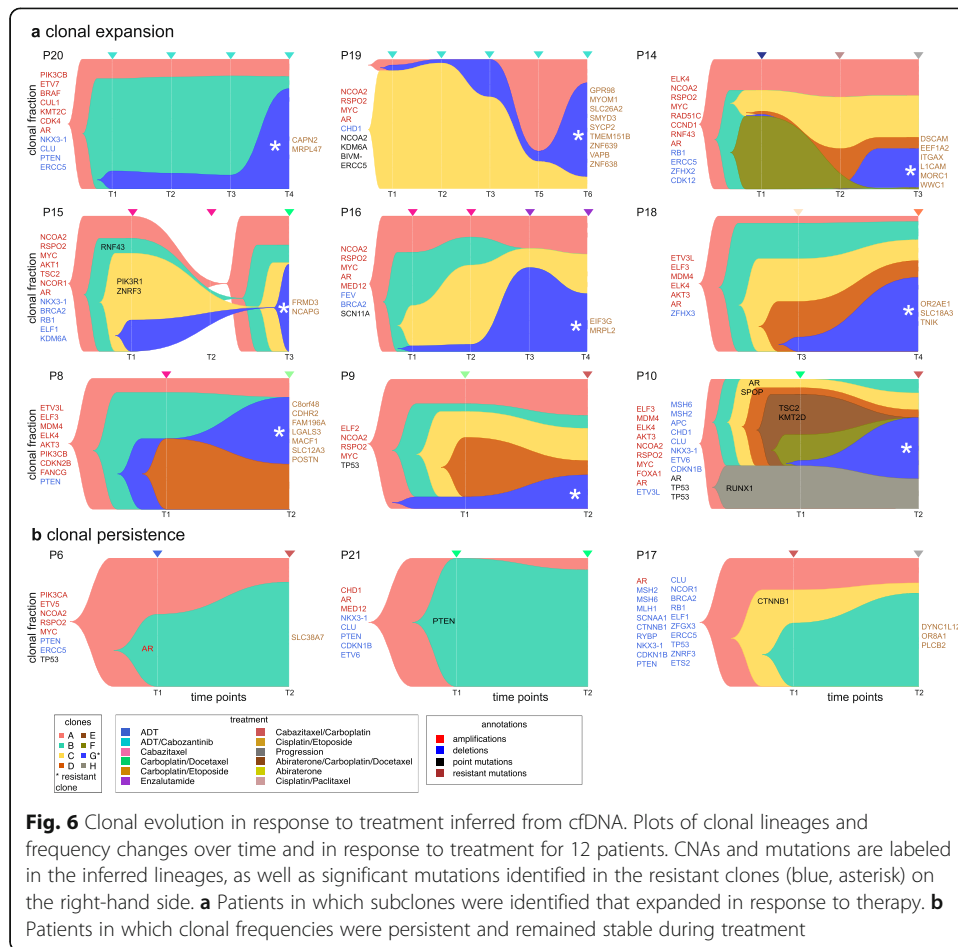


Fig. 6 Clonal evolution in response to treatment inferred from cfDNA. Plots of clonal lineages and frequency changes over time and in response to treatment for 12 patients. CNAs and mutations are labeled in the inferred lineages, as well as significant mutations identified in the resistant clones (blue, asterisk) on the right-hand side. **a** Patients in which subclones were identified that expanded in response to therapy. **b** Patients in which clonal frequencies were persistent and remained stable during treatment

at the final timepoints (T4–T5) and harbored significant mutations in *EIF3G* and *MRPL2*. In total, 9 patients were identified in which a minor subclone expanded during treatment, suggesting association with resistance.

The resistant subclones harbored a total of 36 mutations with significant SIFT (< 0.05) and POLYPHEN (> 0.85) functional impact scores (Additional file 1: Table S3). Notably, this data did not identify any recurrent genes associated with resistance; however, several recurrent gene families were identified, such as the solute carrier transporters (*SLC12A3*, *SLC38A7*, *SLC18A3*, and *SLC26A2*) and the mitochondrial ribosomal protein genes (*MRPL2* and *MRPL47*). Despite the lack of recurrence, many of the genes identified in the resistant clones have been previously implicated in PC progression or therapeutic resistance. For example, *NCAPG* detected in P15 is part of the condensin complex and has been related to CRPC pathogenesis [37, 38]. *LGALS3* detected in P8 is a member of the galectin proteins involved in apoptosis, immunity, and adhesion, and was linked to treatment resistance in PC [39, 40]. *CAPN2* detected in P20 is an intracellular cysteine protease and has been shown to promote cell proliferation and invasion in CRPC cell lines [41]. Furthermore, *WWC1*, which was identified in P14, was found to be upregulated in antiandrogen-resistant PC cell lines [42].

We further investigated if the resistance-associated mutations identified in this study were correlated with poor survival in the TCGA datasets ($N = 3811$ patients), and found significant associations with poor survival for *CAPN2*, *CDHR2*, *LGALS3*, *LICAM*, *MORC1*, *EIF3G*, *OR8A1*, *TNIK*, and *SMYD3* (adjusted p value < 0.05 , Benjamini-Hochberg correction) (Additional file 1: Table S3, the “Methods” section). Collectively, these data suggest that the evolutionary analysis of clonal dynamics in cfDNA may be useful for delineating intratumor heterogeneity and identifying resistant clones and mutations associated with therapy response and resistance in prostate cancer patients.

Discussion

Here, we report the development of an unbiased whole-genome sequencing approach for cryostored plasma DNA and its application to study clonal diversity and evolution in response to therapy in PC patients. In contrast to targeted cfDNA methods [21, 43], PEGASUS was designed to perform unbiased genome-wide profiling of CNAs and mutations, which are necessary to infer evolutionary dynamics over time and identify clones associated with response to therapy.

In several patients, we compared plasma DNA directly to matched tissue samples from metastatic organ sites. Our findings suggest that the genomic aberrations identified in cfDNA are highly representative of the metastatic tissue sites, but also that the cfDNA also contains mutations that are not present in the matched metastatic tissues. These additional mutations in the plasma may originate from other metastatic foci or micrometastases that were not profiled in this study. Based on this data, we speculate that cfDNA provides a more holistic representation of a patient’s cancer genomic aberrations, across many of the primary and metastatic tumor sites, compared to core biopsy samples that reflect a limited spatial area in a single tissue site.

We further applied PEGASUS to analyze serial plasma samples collected from PC patients that received different combinations of hormonal and chemotherapy treatments. Our data showed that complex aneuploid rearrangements remained highly stable during treatment, with few or no new CNA acquired during this time. This data suggest that CNAs are likely to have occurred early in tumor evolution and may be related to intrinsic (rather than acquired) resistance of the tumor cells. In contrast, the MAF in the cfDNA underwent dynamic changes in response to treatments. In most patients, clones present in low frequency in the pre-treatment timepoints expanded during therapy and harbored mutations associated with resistance. Many of the genomic aberrations in the genes identified ($N = 9$) in the resistance clones were correlated with poor survival in larger cohorts of PC patients in TCGA. However, most mutations identified in the resistant clones were not recurrent across patients, a possible reflection of the heterogeneous nature of the treatments or alternatively of diverse mechanisms of resistance. An exception was several recurrent mutations in solute carrier transporter genes (SLC) and mitochondrial ribosomal protein (MRP) gene families. The SLC genes are of considerable interest from a therapeutic standpoint, since they are involved in the uptake and transport of drugs into cells and therapeutic resistance [44]. Future studies will

be needed to functionally validate these mutations and understand their potential role in therapy resistance in PC patients.

Several clinical parameters identified correlated with increased plasma cfDNA concentrations, including the presence of aneuploid genomes, increased disease volume, accelerated progression, and poor OS. For cancers progressing in sites that are difficult to biopsy and/or that are unlikely to yield enough tumor cells to allow for genomic analysis (e.g., bone), PEGASUS may provide an invaluable profiling tool to discover genomic biomarkers associated with disease behavior and drug sensitivity. Global genomic features such mutation burden measured in the cfDNA may have clinical utility for identifying patients with increased neoantigens that are ideal for treatment with immune checkpoint inhibitors. Furthermore, the CNA burden or aneuploid aberrations detected in cfDNA may be used to detect tumors with homologous recombination deficiency (HRD) in patients who may benefit from agents targeting DNA damage repair defects such as PARP inhibitors [45, 46]. Indeed, our data suggest that the detection of aneuploid copy number profiles in plasma DNA is an indicator of poor survival in PC patients, as are higher concentrations of cfDNA in the blood.

The use of PEGASUS to study genomic aberrations in plasma DNA has a few limitations. Foremost, because the genomic profiling is unbiased and broad by the assay design, it requires significant sequencing coverage (e.g., 150X) and higher cost, compared to targeted panels, which also have a higher sensitivity for detection of rare mutations. Another limitation is that the approach is more suitable for patients with advanced and metastatic disease and is unlikely to have utility in the detection of early disease, where the concentration of ctDNA in the plasma is very low. This was an issue in the initial timepoints analyzed in the CSPC patients, where only diploid genomes were detected prior to progression to CRPC disease. In such cases, the use of targeted plasma DNA sequencing panels (e.g., Guardant360, OncoPrint) would be more appropriate to increase detection sensitivity.

In closing, we expect that PEGASUS will have a myriad of applications in cancer research, particularly in the discovery and identification of genomic mechanisms of response and resistance to anticancer therapy in large patient cohorts, where longitudinal plasma samples have previously been collected and cryostored for achieving purposes and long-term clinical outcome data is available. Our approach will be particularly useful in the non-invasive genomic profiling of solid tumor tissues that are challenging to biopsy (e.g., kidney, brain, lung, bone). We expect that these longitudinal genomic analyses will reveal basic mechanisms of response and disease resistance and may lead to new clinical assays that can monitor response to therapeutic agents and guide treatment decisions in cancer patients.

Conclusions

This study shows that unbiased whole-genome sequencing of plasma DNA from prostate cancer patients can detect mutations and copy number alterations that can be used to infer clonal dynamics and genome evolution longitudinally in response to treatment. By computationally integrating this data over multiple timepoints, we use an evolutionary approach to identify clones that harbor mutations associated with therapeutic resistance.

Methods

Patient clinical data and sample information

All patients in this study were treated for prostate cancer at the University of Texas MD Anderson Cancer Center (Houston, TX) and provided informed consent per an Institutional Review Board-approved prospective protocol. Two patient cohorts were included: (i) newly diagnosed metastatic and hormone-naïve (CSPC) (including patients participating in clinical trials NCT01409200 or NCT01630590) and (ii) CRPC (including patients participating in clinical trial NCT01505868). All blood plasma and tissue samples were collected before initiation of systemic treatment, while progressing on therapy by PSA or radiologic criteria, or while on systemic treatment as indicated. Patients were prospectively followed from the time of inclusion until the last visit or death. Matched metastatic tissue samples were obtained as FFPE blocks from the respective patients. Patients were classified as aneuploid or diploid based on the whole-genome copy number data from the plasma. Patients with worsening performance status, pain, or other symptoms related to tumor growth in the 6 weeks prior to the blood specimen collection, and/or with development of > 2 new metastatic lesions in a single site or new non-nodal organ site extension in the previous 3 months, were classified as “accelerated progressors”; all other patients were defined as “protracted progressors.” Patients were classified as “high disease volume” if they had > 10 focal bone metastases or equivalent and/or tumor mass > 4 cm at any site, and/or extension to at least three organ sites with one lesion at least 2 cm in diameter; “low disease volume” if patients had ≤ 4 bone metastases with or without extension to lymph nodes up to 2 cm in diameter; all others were categorized as “intermediate disease volume” patients. A sample was classified as aneuploid if the CNA profile contained a large number of segments (> 45) and had a deviation from the median segmentation value (> 0.03) that was not explained by technical noise. If the sample had an intermediate number of segments (between 25 and 45), it was considered aneuploid if it had a deviation from the median segmentation value (> 0.03) and a large segment size or high segmentation value. All other samples were classified as diploid.

Isolation of plasma DNA from blood plasma and quality control

Blood (approximately 7.5 mL) from the prostate cancer patients was collected in Ficoll tubes (catalog no. 362753). After gentle inversion, tubes were centrifuged at 1800g for 15 min at room temperature. The plasma layer was separated from the nucleated PBMC cell layer and centrifuged three times at 1500g for 10 min to remove contaminating cells. The PBMC layer was used to isolate genomic DNA, which was sequenced separately to identify germline variants (see genomic DNA isolation and quality control section). In cases where fresh blood samples were not available, frozen plasma stocks stored at -80 C were thawed and centrifuged at 16,000g three times to remove all cryoprecipitates. Low molecular weight plasma DNA was purified by size selection (< 1000 bp) from high molecular weight DNA using the QIAamp® Circulating Nucleic Acid Kit (QIAGEN Cat. no. 55114) according to the manufacturer’s instructions. Final carrier RNA concentration of 1.3 ng/μL was used to improve yields. For quality control, plasma DNA fragment size and concentrations were measured using high sensitivity (HS) Bioanalyzer chips or TapeStation HS D1000 tape cartridges. The plasma DNA

concentration and size range that was measured for QC was gated on fragments within the 100–700-bp range to exclude contaminating high molecular weight genomic DNA when present.

DNA isolation and sequencing of tissue samples

PBMC layer or plasma pellets were utilized to extract the reference genomic DNA using QIAGEN Kits (Cat. no. 51106 and 56304, respectively) according to the manufacturer's instructions. FFPE DNA from metastatic tissues was isolated using the QIAamp DNA FFPE Tissue Kit (QIAGEN Cat. no. 56404) following the manufacturer's instructions.

The isolated DNA quality was assayed through agarose gel electrophoresis or TapeStation genomic DNA tapes to determine size and concentration. The resulting gDNA was used to construct low-input DNA sequencing libraries with the Hyper Prep kit (Kapa Biosciences cat. no. KK8504) according to the manufacturer's instructions. Genomic DNA from PBMCs was sonicated to 200 bp (Covaris Peak power 175, Duty Factor 10%, cycles/burst 200, time 180 s, temp 4–7 °C) prior to end repair and a-tailing steps. Post-ligation cleanup was performed with 0.8X AMPure XP beads. PCR amplification of plasma DNA samples was performed at 11 PCR cycles. Final sequencing libraries were split into two pools for either copy number sparse WGS or exome capture. The copy number libraries were sequenced at 36 or 76 cycles single-end on the Illumina HiSeq4000 system. The exome libraries were further captured using the SeqCap EZ Exome V2 kit following the manufacturer's instructions (Nimblegen-Roche Cat. no. 05860482001). The final exome libraries were sequenced at 76 or 100 paired-end reads on the Illumina HiSeq4000 system.

Plasma DNA library construction and exome sequencing

The plasma DNA that passed QC by having ≥ 2 ng total DNA and size distributions < 1000 bp were used to construct low-input DNA sequencing libraries with the Hyper Prep kit (Kapa Biosciences cat. no. KK8504) according to the manufacturer's instructions. Post-ligation cleanup was performed with 0.8X AMPure XP beads. PCR amplification of plasma DNA samples was performed at 11 PCR cycles. Final sequencing libraries were split into two pools for either copy number sparse WGS or exome capture. The copy number libraries were sequenced directly at 36 or 76 cycles single-end on the Illumina HiSeq 4000 system. The exome libraries were captured using the SeqCap EZ Exome V2 kit following the manufacturer's instructions (Nimblegen-Roche Cat. no. 05860482001). The final exome libraries were sequenced at 76 or 100 paired-end reads on the Illumina HiSeq4000 system.

Analysis of genomic copy number data using circular binary segmentation (CBS) from plasma DNA and tissues

Reads sequenced were demultiplexed using the "bcltfastq" software (Illumina) and split into individual FASTQ files, allowing a 1 basepair mismatch for barcode edit distance. Copy number profiles were detected from read depth counting of the sequencing data using the "variable binning" pipeline as previously described [47]. This pipeline involved mapping FASTQ files to the human genome assembly NCBI Build 37 (hg19/

NCBI37) using Bowtie2 (2.1.0) alignment software [48]. The aligned reads in SAM files were converted to BAM files and sorted using SAMtools (0.1.16) [49]. PCR duplicates were marked and removed using Picard in GATK [50]. The reads were counted using variable bin sizes at an average genomic resolution of 220 kb. Unique normalized read counts were segmented using the CBS [51] method from R Bioconductor “DNAcopy” package followed by MergeLevels [52] to join adjacent segments with non-significant differences in segmented ratios. The parameters used for CBS segmentation were $\alpha = 0.0001$ and $\text{undo.prune} = 0.05$. Default parameters were used for MergeLevels, which removed erroneous chromosome breakpoints. Finally, we used ggplot packages in R to plot the segmentation and $\log_2(\text{ratio})$ values and annotate prostate cancer genes in regions of amplification and deletion. The prostate cancer gene list of 100 genes was compiled from two published papers [33, 53].

Calculation of CNA lengths and correlation between plasma and metastatic samples

The neutral copy number state was defined as the median segmentation ratio of all genomic bins. Any group of consecutive bins with the same segmentation ratio not equal to the copy neutral state within a chromosome was defined as a CNA. The CNA length was defined as a difference between the bin start position and bin end position. Every CNA contained a list of bins with the same segmentation ratio values. These segmentation values from the plasma and metastatic samples were used to calculate the correlation values using the Pearson correlation coefficient.

Detection of mutations in plasma exome sequencing data

The plasma DNA data from each patient was aligned to the human genome reference assembly (hg19) using Bowtie2 (2.1.0) [48] and converted into a binary format (BAM) with Samtools [49]. The SAM file was sorted using samtools, and PCR duplicates were marked using Picard tools [54]. Genome Analysis Toolkit [50] (GATK)’s BaseRecalibrator and PrintReads were used to obtain BAM files that have been recalibrated for base quality scores. Somatic variants were identified using the plasma and matched PBMC samples using GATK’s MuTect2 [55] to generate a variant call format (VCF) file. The filtering functionality of the MuTect2 VCF file was enabled using GATK’s FilterMutectCalls, and the somatic variants classified as “PASS” (high confidence) and “germline risk” (borderline somatic) were retained. The filtered somatic variants were split into single nucleotide variants (SNVs) and indels using GATK’s SelectVariants function. The chromosome number and position of the variants were extracted, and samtools was used to obtain the read counts across all the variant sites. Variants with less than 0.07 variant allele frequency in the plasma sample were removed from analysis. Additionally, variants were removed from analysis if they had more than 2 variant reads in the PBMC samples sequenced at less than 100X depth or more than 0.01 variant allele frequency for samples sequenced at greater than 100X depth. These read depth filters remove the mutect2 “germline risk” calls that are like to be germline mutations and are not somatic mutations. The resulting data was annotated using ANNOVAR [56] by integrating multiple databases, including dbSNP, COSMIC [57], and the Cancer Gene Census [58]. To identify mutations and copy number variations (CNV) in genes of interest, an exon coordinate file was intersected with BedTools [59]. The functional

significance of each SNV was predicted using SIFT [60] and POLYPHEN [61]. Mutations with < 0.05 SIFT scores and > 0.85 POLYPHEN scores were considered significant for impacting gene function.

Identification of concordant mutations in matched metastases

The total number of CNAs per sample was estimated above a baseline threshold for amplifications and deletions in each patient. The CNA correlation between the two sample sources was calculated using the Pearson correlation coefficient of the segmentation means, while the mutational concordance was calculated as the percentage of the ratio between the concordant mutations and the total number of mutations in the plasma and tissue sample combined.

Survival analysis by plasma DNA concentration

We utilized survival data from 140 baseline CRPC plasma samples who were part of trial NCT01505868 for survival analysis. The samples were categorized into 2 groups based on total cfDNA (> 2 ng or < 2 ng) to prepare libraries. The overall and progression-free survival months and the group information were used for survival analysis using the Kaplan-Meier estimator using the “survminer” library in R (3.5.0). The p value of the log-rank test was calculated, and the two groups were considered significantly different if the p value was $p < 0.05$.

Estimating clonal frequencies from mutation data

Non-synonymous and synonymous somatic mutations were identified in the plasma samples for 12 PC patients with 2 to 6 longitudinal samples. The plasma DNA genomic copy number profiles were estimated from the paired-end exome sequencing depth using the R package “ExomeCNV” [62]. For ExomeCNV analysis, the minimum sensitivity and specificity was set to 0.9999 while it was optimized using the AUC criteria. Tumor purity was estimated with THetA2 [63] using the tumor and normal GATK recalibrated bam files while the minimum fraction of the genome with a potential copy number event for the sample was set to 0. The variant allele frequencies from each point mutation were normalized with both exome-derived copy number profiles and estimated tumor purities using PyClone2 (v0.12.9) [35]. The copy number and purity-adjusted clonal frequencies were then used as input for CITUP [36] for the joint calculation and estimation of clonal subpopulations using the optimal trees across the longitudinal timepoints from same patient. Finally, the clonal lineages were plotted with “timescape” [64] using the CITUP tree structures and the clonal frequencies across the longitudinal timepoints for individual PC patients.

Survival analysis of resistant genes from TCGA

Selected prostate cancer studies [65–79] including the prostate cancer (MSK, 2019), prostate adenocarcinoma (TCGA, provisional), and The Metastatic Prostate Cancer Project that were available on cBioPortal (www.cbioportal.org) were used for survival analysis to determine if genomic aberrations in significant genes associated with resistance were associated with poor patient survival in the expanded patient cohorts. All the genes reported in Additional file 1: Table S3 are tested, and p values from log-rank test

were calculated. The p values from the log-rank test were then adjusted using the Benjamini-Hochberg procedure.

Cox regression model for OS and PFS using available clinical and genomic parameters

Using data from 70 prostate cancer patients with available clinical data across all clinical and genomic parameters, we performed a Cox regression model using the “survival” R package. We used the `coxph` function by utilizing one or all the factors as predictors and the survival months as the outcome variable for the univariate or multivariate analysis. Any predictor that had a p value < 0.05 was considered statistically significant. For the discrete clinical and genomic parameters, total cfDNA < 2 ng, diploid cfDNA profiles, low disease volume, protracted rate of progression, and prostate disease were used as a reference for the Cox regression models.

Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s13059-020-02045-9>.

Additional file 1. Fig. S1-S6 and Table S1-S3.

Additional file 2. Review History.

Acknowledgements

We are very grateful to the participating prostate cancer patients at MD Anderson Cancer Center for donating their blood and tumor tissues for this research work.

Peer review information

Yixin Yao was the primary editor on this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

Review history

The review history is available as Additional file 2.

Authors' contributions

NR analyzed the data, interpreted the results, and wrote the manuscript. ES, PCT, SB, and YZ performed the experiments. PGC, PT, and CL provided plasma and tissue samples from prostate cancer patients. AJZ provided plasma and tissue samples, interpreted the results, and wrote the manuscript. NN analyzed the data, interpreted the results, and wrote the manuscript. The authors read and approved the final manuscript.

Authors' information

Twitter handles:

Nicholas Navin: @nicholas_navin

Amado Zurita: @azursaa

Naveen Ramesh: @Naveen_Ramesh1

Yuehui Zhao: @LeonZhao15

Funding

This work was supported by grants to NN and AJZ from the MD Anderson Cancer Moonshot Prostate Cancer Program and to PGC from The Solon Scott III Prostate Cancer Research Fund. The work was also supported by grants to NN from the American Cancer Society (129098-RSG-16-092-01-TBG) and NIH (RO1CA240526). This study was supported by the MD Anderson Sequencing Core Facility Grant (CA016672) and the Flow Cytometry Facility grant from NIH (CA016672). The Prostate Cancer Tissue Core was supported by MD Anderson's Prostate Cancer SPORE grant P50 (CA140388). NR and YZ were supported by the CPRIT Research Training Program (RP170067).

Availability of data and materials

The datasets supporting the conclusions of the article are available under bioproject ID: PRJNA554329 [80].

Ethics approval and consent to participate

All patients in this study were treated for prostate cancer at the University of Texas MD Anderson Cancer Center (Houston, TX) and provided informed consent per an Institutional Review Board-approved prospective protocol. Two patient cohorts that were included in this study are (i) newly diagnosed metastatic and hormone-naïve (CSPC) (including patients participating in clinical trials NCT01409200 or NCT01630590) and (ii) CRPC (including patients participating in clinical trial NCT01505868). The patient sample collection was approved by the Institutional Review Board under protocol PA13-0247. All the experimental methods comply with the Helsinki Declaration.

Consent for publication

All the samples in this study have been deidentified.

Competing interests

The authors have no financial conflicts to declare.

Author details

¹Department of Genetics, The University of Texas MD Anderson Cancer Center, Houston, TX, USA. ²MD Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences, Houston, TX, USA. ³Department of Genitourinary Medical Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA. ⁴Department of Pathology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA. ⁵David H. Koch Center for Applied Research of Genitourinary Cancers, The University of Texas MD Anderson Cancer Center, Houston, TX, USA. ⁶Department of Bioinformatics and Computational Biology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA.

Received: 3 January 2020 Accepted: 13 May 2020

Published online: 06 July 2020

References

- Davis A, Gao R, Navin N. Tumor evolution: linear, branching, neutral or punctuated? *Biochim Biophys Acta Rev Cancer*. 1867;2017:151–61.
- McGranahan N, Swanton C. Clonal heterogeneity and tumor evolution: past, present, and the future. *Cell*. 2017;168:613–28.
- Marusyk A, Almendro V, Polyak K. Intra-tumour heterogeneity: a looking glass for cancer? *Nat Rev Cancer*. 2012;12:323–34.
- Kristiansen A, Bergstrom R, Delahunt B, Samaritunga H, Guethjonsdottir J, Gronberg H, Egevad L, Lindberg J. Somatic alterations detected in diagnostic prostate biopsies provide an inadequate representation of multifocal prostate cancer. *Prostate*. 2019;79:920–8.
- Jones TA, Radtke JP, Hadaschik B, Marks LS. Optimizing safety and accuracy of prostate biopsy. *Curr Opin Urol*. 2016;26:472–80.
- Shariat SF, Roehrborn CG. Using biopsy to detect prostate cancer. *Rev Urol*. 2008;10:262–80.
- Dago AE, Stepanyk A, Carlsson A, Luttgren M, Kendall J, Baslan T, Kolatkar A, Wigler M, Bethel K, Gross ME, et al. Rapid phenotypic and genomic change in response to therapeutic pressure in prostate cancer inferred by high content analysis of single circulating tumor cells. *PLoS One*. 2014;9:e101777.
- Thiele JA, Pitule P, Hicks J, Kuhn P. Single-cell analysis of circulating tumor cells. *Methods Mol Biol*. 1908;2019:243–64.
- Haber DA, Velculescu VE. Blood-based analyses of cancer: circulating tumor cells and circulating tumor DNA. *Cancer Discov*. 2014;4:650–61.
- Adalsteinsson VA, Ha G, Freeman SS, Choudhury AD, Stover DG, Parsons HA, Gydush G, Reed SC, Rotem D, Rhoades J, et al. Scalable whole-exome sequencing of cell-free DNA reveals high concordance with metastatic tumors. *Nat Commun*. 2017;8:1324.
- De Mattos-Arruda L, Caldas C. Cell-free circulating tumour DNA as a liquid biopsy in breast cancer. *Mol Oncol*. 2016;10:464–74.
- Bettegowda C, Sausen M, Leary RJ, Kinde I, Wang Y, Agrawal N, Bartlett BR, Wang H, Luber B, Alani RM, et al. Detection of circulating tumor DNA in early- and late-stage human malignancies. *Sci Transl Med*. 2014;6:224ra224.
- Wan JCM, Massie C, Garcia-Corbacho J, Mouliere F, Brenton JD, Caldas C, Pacey S, Baird R, Rosenfeld N. Liquid biopsies come of age: towards implementation of circulating tumour DNA. *Nat Rev Cancer*. 2017;17:223–38.
- Dagogo-Jack I, Brannon AR, Ferris LA, Campbell CD, Lin JJ, Schultz KR, Ackil J, Stevens S, Dardaie L, Yoda S, et al. Tracking the evolution of resistance to ALK tyrosine kinase inhibitors through longitudinal analysis of circulating tumor DNA. *JCO Precis Oncol*. 2018. <https://doi.org/10.1200/PO.17.00160>.
- Gabriel E, Bagaria SP. Assessing the impact of circulating tumor DNA (ctDNA) in patients with colorectal cancer: separating fact from fiction. *Front Oncol*. 2018;8:297.
- Li JH, He ZQ, Lin FH, Chen ZH, Yang SY, Duan H, Jiang XB, Al-Nahari F, Zhang XH, Wang JH, et al. Assessment of ctDNA in CSF may be a more rapid means of assessing surgical outcomes than plasma ctDNA in glioblastoma. *Mol Cell Probes*. 2019;46:101411. <https://doi.org/10.1016/j.mcp.2019.06.001>.
- Markus H, Contente-Cuomo T, Farooq M, Liang WS, Borad MJ, Sivakumar S, Gollins S, Tran NL, Dhruv HD, Berens ME, et al. Evaluation of pre-analytical factors affecting plasma DNA analysis. *Sci Rep*. 2018;8:7375.
- Lanman RB, Mortimer SA, Zill OA, Sebisanoovic D, Lopez R, Blau S, Collisson EA, Divers SG, Hoon DS, Kopetz ES, et al. Analytical and clinical validation of a digital sequencing panel for quantitative, highly accurate evaluation of cell-free circulating tumor DNA. *PLoS One*. 2015;10:e0140712.
- Wyatt AW, Annala M, Aggarwal R, Beja K, Feng F, Youngren J, Foye A, Lloyd P, Nykter M, Beer TM, et al. Concordance of circulating tumor DNA and matched metastatic tissue biopsy in prostate cancer. *J Natl Cancer Inst*. 2017;109(12):d1x118. <https://doi.org/10.1093/jnci/d1x118>.
- Ulz P, Belic J, Graf R, Auer M, Lafer I, Fischereder K, Webersinke G, Pummer K, Augustin H, Pichler M, et al. Whole-genome plasma sequencing reveals focal amplifications as a driving force in metastatic prostate cancer. *Nat Commun*. 2016;7:12008.
- Xia S, Kohli M, Du M, Dittmar RL, Lee A, Nandy D, Yuan T, Guo Y, Wang Y, Tschannen MR, et al. Plasma genetic and genomic abnormalities predict treatment response and clinical outcome in advanced prostate cancer. *Oncotarget*. 2015;6:16411–21.
- Parikh AR, Leshchiner I, Elagina L, Goyal L, Levovitz C, Siravegna G, Livitz D, Rhissorakrai K, Martin EE, Van Seventer EE, et al. Liquid versus tissue biopsy for detecting acquired resistance and tumor heterogeneity in gastrointestinal cancers. *Nat Med*. 2019;25:1415–21.
- Stover DG, Parsons HA, Ha G, Freeman SS, Barry WT, Guo H, Choudhury AD, Gydush G, Reed SC, Rhoades J, et al. Association of cell-free DNA tumor fraction and somatic copy number alterations with survival in metastatic triple-negative breast cancer. *J Clin Oncol*. 2018;36:543–53.

24. Abbosh C, Birkbak NJ, Wilson GA, Jamal-Hanjani M, Constantin T, Salari R, Le Quesne J, Moore DA, Veeriah S, Rosenthal R, et al. Phylogenetic ctDNA analysis depicts early-stage lung cancer evolution. *Nature*. 2017;545:446–51.
25. Page K, Guttery DS, Fernandez-Garcia D, Hills A, Hastings RK, Luo J, Goddard K, Shahin V, Woodley-Barker L, Rosales BM, et al. Next generation sequencing of circulating cell-free DNA for evaluating mutations and gene amplification in metastatic breast cancer. *Clin Chem*. 2017;63:532–41.
26. Christensen E, Nordentoft I, Vang S, Birkenkamp-Demtroder K, Jensen JB, Agerbaek M, Pedersen JS, Dyrskjot L. Optimized targeted sequencing of cell-free plasma DNA from bladder cancer patients. *Sci Rep*. 2018;8:1917.
27. Rycaj K, Tang DG. Molecular determinants of prostate cancer metastasis. *Oncotarget*. 2017;8:88211–31.
28. Davies A, Conteduca V, Zoubeidi A, Beltran H. Biological evolution of castration-resistant prostate cancer. *Eur Urol Focus*. 2019;5:147–54.
29. Moulriere F, Chandrananda D, Piskorz AM, Moore EK, Morris J, Ahlborn LB, Mair R, Goranova T, Marass F, Heider K, et al. Enhanced detection of circulating tumor DNA by fragment size analysis. *Sci Transl Med*. 2018;10(466):eaat4921. <https://doi.org/10.1126/scitranslmed.aat4921>.
30. Moulriere F, Rosenfeld N. Circulating tumor-derived DNA is shorter than somatic DNA in plasma. *Proc Natl Acad Sci U S A*. 2015;112:3178–9.
31. Kumar A, White TA, MacKenzie AP, Clegg N, Lee C, Dumpit RF, Coleman I, Ng SB, Salipante SJ, Rieder MJ, et al. Exome sequencing identifies a spectrum of mutation frequencies in advanced and lethal prostate cancers. *Proc Natl Acad Sci U S A*. 2011;108:17087–92.
32. Antonarakis ES, Shaikat F, Isaacsson Velho P, Kaur H, Shenderov E, Pardoll DM, Lotan TL. Clinical features and therapeutic outcomes in men with advanced prostate cancer and DNA mismatch repair gene mutations. *Eur Urol*. 2019;75:378–82.
33. Robinson D, Van Allen EM, Wu YM, Schultz N, Lonigro RJ, Mosquera JM, Montgomery B, Taplin ME, Pritchard CC, Attard G, et al. Integrative clinical genomics of advanced prostate cancer. *Cell*. 2015;162:454.
34. Lohr JG, Adalsteinsson VA, Cibulskis K, Choudhury AD, Rosenberg M, Cruz-Gordillo P, Francis JM, Zhang CZ, Shalek AK, Satija R, et al. Whole-exome sequencing of circulating tumor cells provides a window into metastatic prostate cancer. *Nat Biotechnol*. 2014;32:479–84.
35. Roth A, Khattra J, Yap D, Wan A, Laks E, Biele J, Ha G, Aparicio S, Bouchard-Cote A, Shah SP. PyClone: statistical inference of clonal population structure in cancer. *Nat Methods*. 2014;11:396–8.
36. Malikic S, McPherson AW, Donmez N, Sahinalp CS. Clonality inference in multiple tumor samples using phylogeny. *Bioinformatics*. 2015;31:1349–56.
37. Goto Y, Kurozumi A, Arai T, Nohata N, Kojima S, Okato A, Kato M, Yamazaki K, Ishida Y, Naya Y, et al. Impact of novel miR-145-3p regulatory networks on survival in patients with castration-resistant prostate cancer. *Br J Cancer*. 2017;117:409–20.
38. Arai T, Okato A, Yamada Y, Sugawara S, Kurozumi A, Kojima S, Yamazaki K, Naya Y, Ichikawa T, Seki N. Regulation of NCAFG by miR-99a-3p (passenger strand) inhibits cancer cell aggressiveness and is involved in CRPC. *Cancer Med*. 2018;7:1988–2002.
39. Fukumori T, Oka N, Takenaka Y, Nangia-Makker P, Elsamman E, Kasai T, Shono M, Kanayama HO, Ellerhorst J, Lotan R, Raz A. Galectin-3 regulates mitochondrial stability and antiapoptotic function in response to anticancer drug in prostate cancer. *Cancer Res*. 2006;66:3114–9.
40. Dondoo TO, Fukumori T, Daizumoto K, Fukawa T, Kohzaki M, Kowada M, Kusuvara Y, Mori H, Nakatsuji H, Takahashi M, Kanayama HO. Galectin-3 is implicated in tumor progression and resistance to anti-androgen drug through regulation of androgen receptor signaling in prostate cancer. *Anticancer Res*. 2017;37:125–34.
41. Li P, Miao C, Liang C, Shao P, Wang Z, Li J. Silencing CAPN2 expression inhibited castration-resistant prostate cancer cells proliferation and invasion via AKT/mTOR signal pathway. *Biomed Res Int*. 2017;2017:2593674.
42. Marques RB, Dits NF, Erkens-Schulze S, van Ijcken WF, van Weerden WM, Jenster G. Modulation of androgen receptor signaling in hormonal therapy-resistant prostate cancer cell lines. *PLoS One*. 2011;6:e23144.
43. Schreuer M, Meerseman G, Van Den Herrewegen S, Jansen Y, Chevolet I, Bott A, Wilgenhof S, Seremet T, Jacobs B, Buyl R, et al. Quantitative assessment of BRAF V600 mutant circulating cell-free tumor DNA as a tool for therapeutic monitoring in metastatic melanoma patients treated with BRAF/MEK inhibitors. *J Transl Med*. 2016;14:95.
44. Lin L, Yee SW, Kim RB, Giacomini KM. SLC transporters as therapeutic targets: emerging opportunities. *Nat Rev Drug Discov*. 2015;14:543–60.
45. Christenson ES, Antonarakis ES. PARP inhibitors for homologous recombination-deficient prostate cancer. *Expert Opin Emerg Drugs*. 2018;23:123–33.
46. Kurnit KC, Coleman RL, Westin SN. Using PARP inhibitors in the treatment of patients with ovarian cancer. *Curr Treat Options in Oncol*. 2018;19:1.
47. Navin N, Kendall J, Troge J, Andrews P, Rodgers L, McIndoo J, Cook K, Stepansky A, Levy D, Esposito D, et al. Tumour evolution inferred by single-cell sequencing. *Nature*. 2011;472:90–4.
48. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012;9:357–U354.
49. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Proc GPD. The sequence alignment/map format and SAMtools. *Bioinformatics*. 2009;25:2078–9.
50. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernysky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010;20:1297–303.
51. Venkatraman ES, Olshen AB. A faster circular binary segmentation algorithm for the analysis of array CGH data. *Bioinformatics*. 2007;23:657–63.
52. Dimitrov P. MergeLevels R. <https://www.rdocumentation.org/packages/aCGH/versions/1.50.0/topics/mergeLevels>.
53. Cancer Genome Atlas Research N. The molecular taxonomy of primary prostate cancer. *Cell*. 2015;163:1011–25.
54. Picard Tools. <http://broadinstitute.github.io/picard/>.
55. Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, Gabriel S, Meyerson M, Lander ES, Getz G. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol*. 2013;31:213–9.
56. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*. 2010;38:e164.

57. Forbes SA, Bindal N, Bamford S, Cole C, Kok CY, Beare D, Jia M, Shepherd R, Leung K, Menzies A, et al. COSMIC: mining complete cancer genomes in the catalogue of somatic mutations in cancer. *Nucleic Acids Res.* 2011;39:D945–50.
58. Futreal PA, Coin L, Marshall M, Down T, Hubbard T, Wooster R, Rahman N, Stratton MR. A census of human cancer genes. *Nat Rev Cancer.* 2004;4:177–83.
59. Quinlan AR. BEDTools: the Swiss-Army tool for genome feature analysis. *Curr Protoc Bioinformatics.* 2014;47:11 12 11–34.
60. Ng PC, Henikoff S. SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Res.* 2003;31:3812–4.
61. Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. *Curr Protoc Hum Genet.* 2013;7(Unit7):20.
62. Sathirapongsasuti JF, Lee H, Horst BA, Brunner G, Cochran AJ, Binder S, Quackenbush J, Nelson SF. Exome sequencing-based copy-number variation and loss of heterozygosity detection: ExomeCNV. *Bioinformatics.* 2011;27:2648–54.
63. Oesper L, Satas G, Raphael BJ. Quantifying tumor heterogeneity in whole-genome and whole-exome sequencing data. *Bioinformatics.* 2014;30:3532–40.
64. Smith M. timescape: patient clonal timescapes. R package version 1.8.0; 2019.
65. Gerhauser C, Favero F, Risch T, Simon R, Feuerbach L, Assenov Y, Heckmann D, Sidiropoulos N, Waszak SM, Hubschmann D, et al. Molecular evolution of early-onset prostate cancer identifies molecular risk markers and clinical trajectories. *Cancer Cell.* 2018;34:996–1011 e1018.
66. Grasso SC, Wu YM, Robinson DR, Cao X, Dhanasekaran SM, Khan AP, Quist MJ, Jing X, Lonigro RJ, Brenner JC, et al. The mutational landscape of lethal castration-resistant prostate cancer. *Nature.* 2012;487:239–43.
67. Abida W, Cyrta J, Heller G, Prandi D, Armenia J, Coleman I, Cieslik M, Benelli M, Robinson D, Van Allen EM, et al. Genomic correlates of clinical outcome in advanced prostate cancer. *Proc Natl Acad Sci U S A.* 2019;116:11428–36.
68. Robinson D, Van Allen EM, Wu YM, Schultz N, Lonigro RJ, Mosquera JM, Montgomery B, Taplin ME, Pritchard CC, Attard G, et al. Integrative clinical genomics of advanced prostate cancer. *Cell.* 2015;161:1215–28.
69. Beltran H, Prandi D, Mosquera JM, Benelli M, Puca L, Cyrta J, Marotz C, Giannopoulou E, Chakravarthi BV, Varambally S, et al. Divergent clonal evolution of castration-resistant neuroendocrine prostate cancer. *Nat Med.* 2016;22:298–305.
70. Baca SC, Prandi D, Lawrence MS, Mosquera JM, Romanell A, Drier Y, Park K, Kitabayashi N, MacDonald TY, Ghandi M, et al. Punctuated evolution of prostate cancer genomes. *Cell.* 2013;153:666–77.
71. Barbieri CE, Baca SC, Lawrence MS, Demichelis F, Blattner M, Theurillat JP, White TA, Stojanov P, Van Allen E, Stransky N, et al. Exome sequencing identifies recurrent SPOP, FOXA1 and MED12 mutations in prostate cancer. *Nat Genet.* 2012;44:685–9.
72. Fraser M, Sabelnykova VY, Yamaguchi TN, Heisler LE, Livingstone J, Huang V, Shiah YJ, Yousif F, Lin X, Masella AP, et al. Genomic hallmarks of localized, non-indolent prostate cancer. *Nature.* 2017;541:359–64.
73. Kumar A, Coleman I, Morrissey C, Zhang X, True LD, Gulati R, Etzioni R, Bolouri H, Montgomery B, White T, et al. Substantial interindividual and limited intraindividual genomic diversity among tumors from men with metastatic prostate cancer. *Nat Med.* 2016;22:369–78.
74. Taylor BS, Schultz N, Hieronymus H, Gopalan A, Xiao Y, Carver BS, Arora VK, Kaushik P, Cerami E, Reva B, et al. Integrative genomic profiling of human prostate cancer. *Cancer Cell.* 2010;18:11–22.
75. Hieronymus H, Schultz N, Gopalan A, Carver BS, Chang MT, Xiao Y, Heguy A, Huberman K, Bernstein M, Assel M, et al. Copy number alteration burden predicts prostate cancer relapse. *Proc Natl Acad Sci U S A.* 2014;111:11139–44.
76. Armenia J, Wankowicz SAM, Liu D, Gao J, Kundra R, Reznik E, Chatila WK, Chakravarty D, Han GC, Coleman I, et al. The long tail of oncogenic drivers in prostate cancer. *Nat Genet.* 2018;50:645–51.
77. Ren S, Wei GH, Liu D, Wang L, Hou Y, Zhu S, Peng L, Zhang Q, Cheng Y, Su H, et al. Whole-genome and transcriptome sequencing of prostate cancer identify new genetic alterations driving disease progression. *Eur Urol.* 2018;73(3):322–39. <https://doi.org/10.1016/j.eururo.2017.08.027>.
78. Gao D, Vela I, Sboner A, laquinta PJ, Karthaus WR, Gopalan A, Dowling C, Wanjala JN, Undvall EA, Arora VK, et al. Organoid cultures derived from patients with advanced prostate cancer. *Cell.* 2014;159:176–87.
79. Abida W, Armenia J, Gopalan A, Brennan R, Walsh M, Barron D, Danila D, Rathkopf D, Morris M, Slovin S, et al. Prospective genomic profiling of prostate cancer across disease states reveals germline and somatic alterations that may affect clinical decision making. *JCO Precis Oncol.* 2017. <https://doi.org/10.1200/PO.17.00029>.
80. Ramesh N, Sei E, Tsai P-C, Bai S, Zhao Y, Troncoso P, Corn P, Logothetis C, Zurita A, Navin N. Decoding the evolutionary response to prostate cancer therapy by plasma genome sequencing. *NCBI.* 2020. <https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA554329>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

