Article

Refined Procedure to Purify and Sequence Circulating Cell-Free DNA in Prostate Cancer

Samira Rahimirad [1], Seta Derderian [1], Lucie Hamel [1], Eleonora Scarlata [1], Ginette McKercher [1], Fadi Brimo [2], Raghu Rajan [3], Alexis Rompre-Brodeur [1,4], Wassim Kassouf [1,2,4], Rafael Sanchez-Salas [4], Armen Aprikian [1,3,4] and Simone Chevalier [1,3,4,5,*]

1   Urologic Oncology Research Group, Cancer Research Program, Research Institute of the McGill University Health Center (RI-MUHC), Montreal, QC H4A 3J1, Canada; samira.rahimirad@mail.mcgill.ca (S.R.); seta.derderian@mail.mcgill.ca (S.D.); lucie.hamel@affiliate.mcgill.ca (L.H.); eleonora.scarlata@affiliate.mcgill.ca (E.S.); ginette.mckercher@affiliate.mcgill.ca (G.M.); alexis.rompre-brodeur@mcgill.ca (A.R.-B.); wassim.kassouf@mcgill.ca (W.K.); armen.aprikian@mcgill.ca (A.A.)
2   Department of Pathology, McGill University, Montreal, QC H3A 0G4, Canada; fadi.brimo@mcgill.ca
3   Department of Oncology, McGill University, Montreal, QC H3A 0G4, Canada; raghu.rajan@mcgill.ca
4   Department of Surgery (Urology Division), McGill University, Montreal, QC H3A 0G4, Canada; rafael.sanchez-salas@mcgill.ca
5   Department of Medicine, McGill University, Montreal, QC H3A 0G4, Canada
*   Correspondence: simone.chevalier@mcgill.ca

Abstract: Cell-free DNA (cfDNA), a fragmented DNA circulating in blood, is a promising biomarker for cancer diagnosis and monitoring. Standardization of cfDNA isolation to enhance the sensitivity of molecular analyses in prostate cancer (PCa) is required. Towards this goal, we optimized existing methods to obtain a high quantity and quality of cfDNA from low volumes of plasma. The protocol was applied to samples from healthy males and three patient categories: radical prostatectomy (RP), disease-free (>6 years post-RP), and metastatic castration-resistant PCa (mCRPC). The yield was significantly higher in mCRPC cases, and the size of fragments was shorter. We compared for the first time library preparation using two cfDNA inputs and low vs. high sequencing depth. Clonal events were observed irrespective of input and depth, but lower input showed more subclonal events. The clinical application of the refined protocols to cfDNA samples from an mCRPC patient showed no tumor fraction before RP, while it increased to 25% at the advanced stage. Among chromosomal changes and mutations, the androgen receptor gene amplification was detected. Altogether, this comprehensive study on improved cfDNA procedures is highly promising to enhance the quality of liquid biopsy-based research for discoveries and much-needed clinical applications.

Keywords: prostate cancer; liquid biopsies; cell-free DNA; circulating tumor DNA; whole genome sequencing

1. Introduction

Prostate cancer (PCa) is the second most common cancer in men throughout the world and is also among the leading causes of cancer death. Various risk factors are associated with PCa, including age, race, family history, lifestyle, diet, and physical activity [1]. PCa generally remains asymptomatic until it reaches an advanced stage. Therefore, early detection is crucial. Prostate tissue biopsy is currently the gold standard for diagnosing cancer and detecting histopathological anomalies, allowing the assessment of determinant

parameters for clinical decision-making [2,3]. The blood levels of prostate-specific antigen (PSA) are also commonly used for the detection of PCa [4]. Radical prostatectomy (RP) and radiation therapy (RT) are effective curative treatments and are associated with a decrease in blood PSA to undetectable or nadir levels. However, rising PSA after curative therapies is an indication of biochemical recurrence (BCR) [4,5]. Other approaches for monitoring the disease include imaging modalities, which are non-invasive ways to reveal metastases in bone and soft tissues, but only when the cancer has progressed enough for detection [3,6,7]. Obtaining additional biopsies of prostate tumors from patients who did not undergo an RP and/or biopsies of metastases is not part of current practice for disease monitoring. Moreover, biopsies of metastases are invasive and have limitations, including low sensitivity and accuracy [8]. In contrast, obtaining liquid biopsies, such as blood, is minimally invasive. Blood contains various analytes from tumors, including tumor-educated platelets, circulating tumor (ct)DNAs/RNAs in cell-free (cf)DNAs/RNAs, extracellular vesicles, circulating tumor cells, microRNAs, and proteins. Genomic content can be released into the bloodstream from both healthy and tumor cells through apoptosis, necrosis, or active secretion. Therefore, isolated ctDNAs reflect genetic changes of tumors/metastases once released in liquid biopsies [9–11]. Liquid biopsies have the potential to complement diagnosis and are promising alternatives that could offer better understanding of tumor dynamics over the course of disease [10,12]. For instance, levels of cfDNA increase significantly in cancer patients, particularly at the metastatic stage [12]. Profiling ctDNA in liquid biopsies of PCa patients has been achieved using high-throughput approaches such as targeted and shallow whole-genome sequencing (WGS), showing mutations, copy number changes (CNVs), and gene fusions [13–15]. However, deep WGS of cfDNA from PCa patients is reported only in one study [16]. There are no guidelines on the best methods to obtain sufficient cfDNA for library preparation and sequencing. In this report, we summarize the methodology of cfDNA studies in PCa and present new findings improving existing purification protocols, library preparation, and WGS.

2. Results

2.1. Optimization of cfDNA Purification Improves cfDNA Quality and Yield Using Less Plasma

Studies reporting on cfDNA isolated from plasma of PCa patients include comparative genomic hybridization arrays, quantitative real-time PCR assays, targeted sequencing, and WGS (Table S1). Although almost all authors used Qiagen kits for isolation, there is no consensus on the volume of plasma to use as input. Also, high-depth sequencing studies (targeted or WGS) routinely used at least 6 mL of plasma to obtain enough cfDNA for library preparation and sequencing [14,16–23].

Our initial modification was to add proteinaseK before thawing a low volume of plasma (~3.6 mL), followed by centrifugation. This considerably helped to reduce the contaminating genomic DNA (gDNA), as shown by the decrease in high molecular weight material at the position of the upper marker (10,380 bp) in the Bioanalyzer (Figure 1A). Moreover, the recovery of a maximum quantity of high-quality cfDNA was achieved by performing four elution steps with warm buffer (56 °C) and increasing volumes to maximize yields/percentages (%) of cfDNA eluted at the first step. The total quantity of cfDNA was defined as the sum of short fragments, between 50 and 550 bp, representing mono-, di-, and tri-nucleosomes, detected in the Bioanalyzer electropherogram for each cfDNA elution. The cfDNA concentration was expressed relative to the volume of plasma (ng/mL). Results on seven plasma samples eluted with varying volumes of AVE buffer from the Qiagen kit show different cfDNA concentrations, with recoveries of 77–100% at the first elution step (Table S2). The choice of elution buffer to elute cfDNAs varies across PCa studies, with some researchers choosing AVE [22,24], while others used water [16,19,20].

We compared the performance of water with the same volumes on these samples, given that AVE contains sodium azide (NaN_3), a chemical reported to inhibit library preparation for DNA sequencing [25]. Water was as efficient as AVE and led to the collection of at least 80–100% of cfDNA at the first elution step (average 92.9% for both) (Table S2). As the recovery of cfDNA with 85 μL of water was ≥94% in the first elute for four tested samples, further experiments were carried out with 85 μL of water added to the column at the first elution step, followed by 40 μL for the second step and 20 μL for the third and fourth steps. The median percentage of cfDNA fragments collected at the first elution step with water was 82% for concentrated samples compared to 100% for samples below 1 ng/μL (Figure 1B). These findings demonstrate that repeating elution steps at least three to four times is necessary to perform a complete isolation of cfDNA and increase yields.
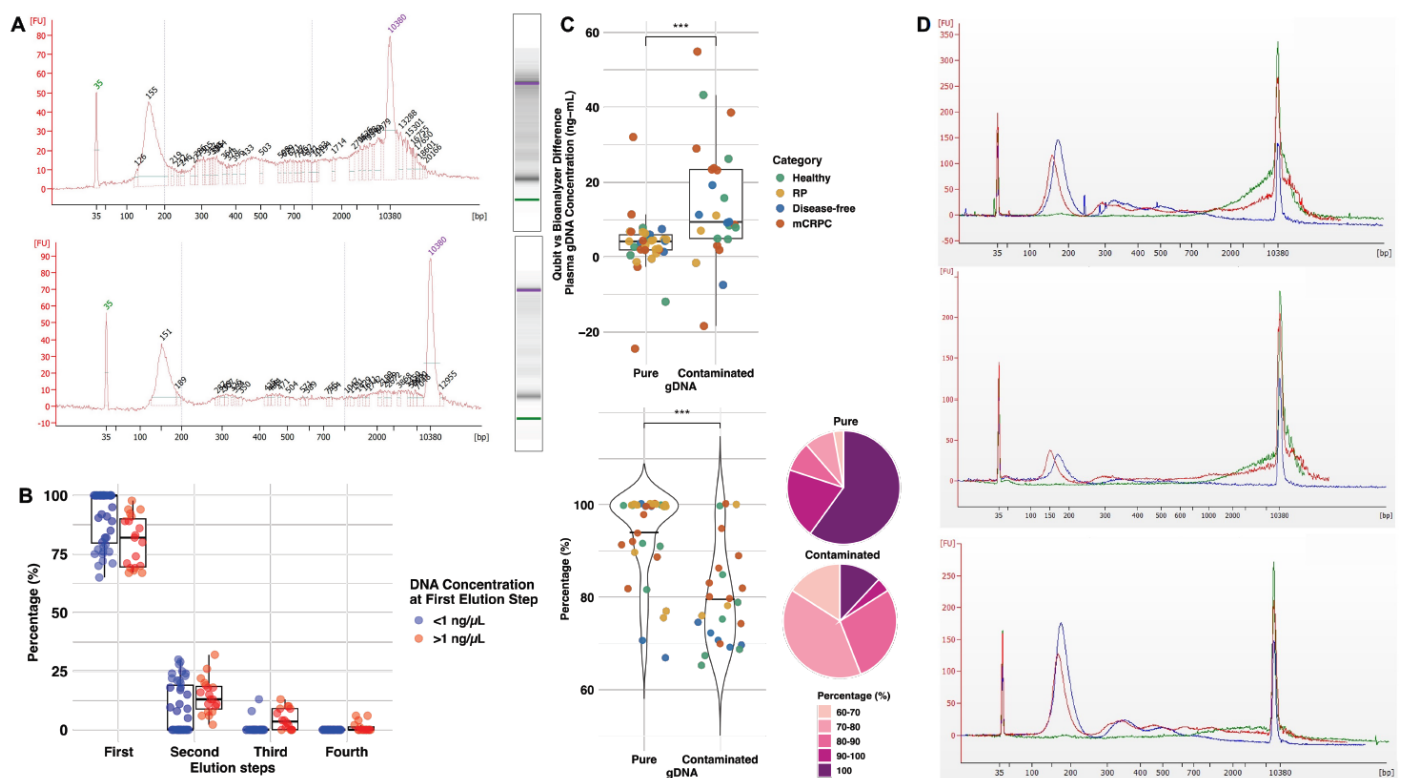

Figure 1. Purification protocol to obtain high quality cfDNA. (**A**). Bioanalyzer electropherograms and their corresponding gel electrophoresis of thawed plasma (3.6 mL) from an advanced patient, without (top panel) and with (lower panel) centrifugation. (**B**). Quantity of cfDNA collected (%) by elution step for low (blue, <1 ng/μL) vs. high (red, >1 ng/μL) DNA concentration (Qubit). (**C**). Differences between Qubit and Bioanalyzer readings were expressed in percentage and used to estimate the gDNA contamination in cfDNA samples from patients and healthy males. (Top) Bar graphs show the difference in the gDNA concentrations between samples considered "pure" vs. "contaminated" based on Bioanalyzer. (Bottom left) Violin plots show the significant difference of fragments (%) collected at first elution in pure vs. contaminated groups from the same samples. (Bottom right) Pie charts show percentages of DNA recovery at the first elution step in both pure and gDNA-contaminated samples. Statistically significant differences between groups are shown by *** $p < 0.001$. (**D**). Electropherograms showing the effects of magnetic beads to remove contaminating gDNA in plasma samples from three patients. Red line-sample before magnetic beads, blue line-pure cfDNA after magnetic beads, and green line-gDNA contamination remaining after magnets. Abbreviation: Base pair: bp; Cell-free DNA: cfDNA; Genomic DNA: gDNA; Metastatic castration-resistant prostate cancer: mCRPC; Radical prostatectomy: RP.

2.2. The Recovery of High-Purity cfDNA Reaches over 90%

Once refined, the protocol was applied to plasma samples from different categories of patients: prior to RP (*n* = 17), patients with no recurrence of disease (disease-free) for at least 6 years post-RP (*n* = 10), and patients at the mCRPC stage (*n* = 19), as well as healthy males (*n* = 14). Based on the Bioanalyzer results on gDNA (Figure 1A) and the processing of these additional samples, we observed that some were contaminated with fragments longer than 5000 bp. On the other hand, several preparations mainly contained small 50–550 bp fragments after four elution steps, reflecting a high degree of purity. We defined a threshold of >50 pg/µL of fragments longer than 5000 bp to call samples contaminated by gDNA vs. pure cfDNA. Due to the Bioanalyzer's limited ability to quantify larger fragments migrating at the upper marker position, we used values from Qubit and Bioanalyzer to quantify the gDNA concentration in plasma. The median difference between Qubit and Bioanalyzer is presented in box plots, which show 35 pure samples with very low gDNA levels at 4.3 ng/mL and 25 others with significantly higher contamination at 10.7 ng/mL (*p* < 0.0002, Figure 1C). We next expressed DNA recovery as a percentage for each elution step. For pure samples (based on Bioanalyzer), we observed that most fragments (80–100%) were recovered at the first elution, as shown in pie charts (Figure 1C). In contrast, there was a wide distribution in contaminated samples, with lower recovery in the first elution (60–80%). The pie charts also show that a small portion of samples in the contaminated group were completely recovered in the first elution (100%). They were all at a low concentration of DNA (<1 ng/µL). The converse was true for pure samples of low recovery (60–70%), whose DNA concentrations were higher than 1 ng/µL. Overall, we estimated that the probability of collecting the majority of pure cfDNA fragments at first elution was 94% (average over all 60 samples), while it was 79.6% for contaminated samples (*p* = 7.7 × 10^{-6}, Figure 1C, Violin plots), in agreement with data in Table S2. Therefore, gDNA contamination reduces the likelihood of collecting all circulating pure cfDNA fragments at the first elution step. To further reduce this contamination, a highly performant two-step wash with magnetic beads was added, allowing a complete removal of gDNA (Figure 1D). This procedure was reproducible and did not lead to noticeable losses of cfDNA (Figure S1). It was routinely used to remove contaminating gDNA when detected by Bioanalyzer. Altogether, our improved purification protocol yielded high quality and quantity of cfDNA using less plasma than previously reported (3.6 mL frozen plasma; Table S1). We highly recommend repeating elution steps four times to capture all cfDNA fragments and using magnetic beads to efficiently remove gDNA contamination and obtain pure cfDNA.

2.3. The Plasma of mCRPC Patients Contains Elevated cfDNA Levels Compared to Other Categories of Patients and Healthy Males

With the optimized protocol, the plasma cfDNA concentrations of the above categories of patients and healthy men were measured by Qubit and Bioanalyzer. With both methodologies, the cfDNA concentration of mCRPC cases (median by Qubit = 34.5 and Bioanalyzer = 25.5 ng/mL) was the highest compared to other categories (Figure 2A,B). RP cases had the lowest quantity (median by Qubit = 8.6 and Bioanalyzer = 5.8 ng/mL). The overall plasma cfDNA concentration was comparable for healthy (median by Qubit = 14.6 and Bioanalyzer = 7.6 ng/mL) and disease-free cases (median by Qubit = 15.6 and Bioanalyzer = 8.3 ng/mL). This is supporting the Qubit overestimation of cfDNA quantity due to gDNA contamination in some samples (Figure 1C). Overall, the median plasma cfDNA concentration was lower in the healthy men (2.9 times, *p* = 0.02) and disease-free cases (2.6 times, not significant) than in mCRPC cases (22.2 ng/mL–without outliers).
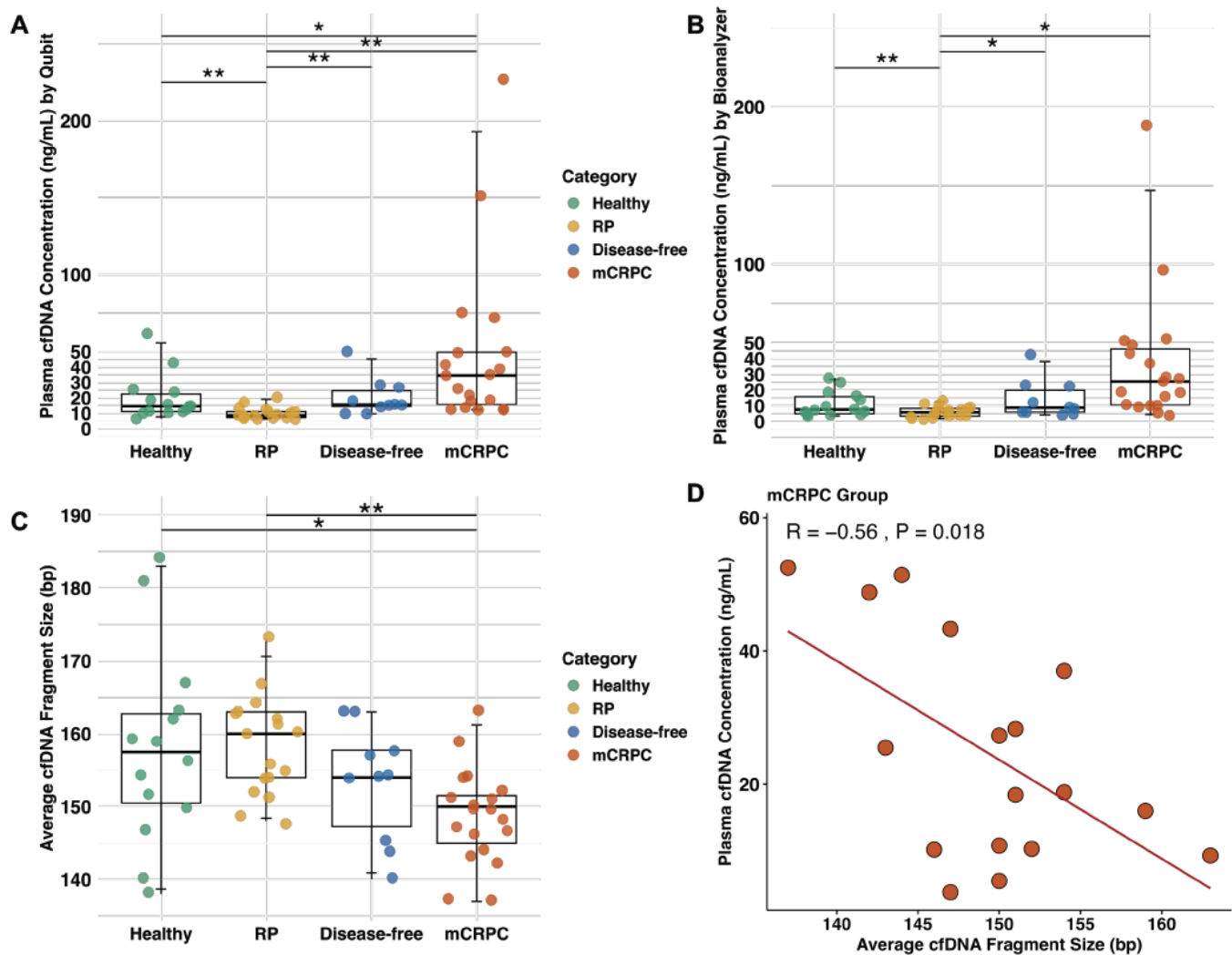
Figure 2. Comparison of cfDNA concentrations and fragment size in different categories of PCa patients and healthy males. Concentrations of cfDNA by categories of patients and controls are presented for (**A**) Qubit and (**B**) Bioanalyzer. (**C**). Average cfDNA fragment size (bp). Healthy: median = 158 bp (range = 138–184 bp); RP: median = 160 bp (range = 148–173 bp); Disease-free: median = 154 bp (range = 140–163 bp); mCRPC: median = 150 bp (range = 137–163 bp) (**C**). Statistically significant differences between groups are shown by * $p < 0.05$, ** $p < 0.01$. (**D**). Correlation between average cfDNA fragment size and concentration in mCRPC patients, with a Pearson coefficient (R) = −0.56. While two outliers were excluded in (**D**), their removal in (**A**,**B**) did not change the finding of significantly higher cfDNA concentrations in the mCRPC category. Abbreviations: Base pair: bp; Cell-free DNA: cfDNA; Metastatic Castration-Resistant Prostate Cancer: mCRPC; Radical Prostatectomy: RP.

2.4. The cfDNA Fragments from mCRPC Patients Are Smaller than Other Patient Categories and Healthy Males

The Bioanalyzer provides a global representation of cfDNA as mono-, di-, and tri-nucleosomes. The average cfDNA (mono-nucleosome) fragment size estimated for the first peak on the Bioanalyzer was ~50–250 bp. Healthy cases had a wider distribution of fragment sizes. Among PCa cases, the RP group had the largest fragments, while mCRPC had the lowest ($p = 0.0016$) (Figure 2C). A significant negative correlation was seen in the fragment size vs. plasma cfDNA concentration for the mCRPC group (Figure 2D). No association was found for the other groups (Figure S2). These findings imply that shorter cfDNA fragments are found in plasma samples containing more cfDNA.

2.5. Exploring cfDNA Concentration of Patients with Their Clinical Features

For all categories of cases and healthy males, there was no correlation of cfDNA concentration with age at time of blood draw (Figure S3). For patients, cfDNA concentration did not correlate with blood PSA prior to RP, but a positive correlation was found with blood PSA at inclusion for mCRPC patients (Figure S4). For all patients, we found no difference with tumor grade or stage (Figure S5). For patients from whom plasma was isolated before RP, no difference in cfDNA levels was observed with other pathologic features (Figure S6). For the 19 mCRPC cases, there was a trend for higher median cfDNA concentrations for the 12 cases under progression at the time of inclusion (excluding two outliers; Figure S7). There was no relationship with a history of having previous progression on Abiraterone and Docetaxel in mCRPC cases (Figure S8). Finally, overall survival was not significantly associated with high cfDNA levels (above the median at 22.2 ng/mL) (Figure S9). Further validation with more cases is required.

2.6. Further Protocol Refinements Lead to Highly Reliable Sequencing Data

Recovering elevated levels of high-quality cfDNAs from small volumes of plasma is key for patients with aggressive disease donating blood for research and for downstream molecular applications. The quantity of cfDNA for library preparation and the sequencing depth have not been thoroughly investigated and are critical before initiating translational projects on plasma from PCa patients throughout the course of their disease. Pilot experiments were conducted to optimize the number of PCR cycles to prepare sequencing libraries. Skipping fragmentation in the TruSeq Nano DNA kit protocol was among the steps considered in collaboration with Illumina specialists. No further changes were applied at this point. Amplification was tested at 8 and 13 PCR cycles for the cfDNA Reference Standard (30 ng) and at 8, 10, and 13 cycles for the PCa cfDNA sample. Eight PCR cycles had the lowest off-target amplification peak (~85–105 s, equal to 500–2000 bp), as seen in both PCa samples (Figure 3A, bottom) and the Reference Standard (Figure 3A, top), compared to 10 and 13 cycles with false amplifications of larger fragments. This had no effect on library concentrations as determined by qPCR, with similar values at different numbers of cycles (Table S3). The number of PCR cycles selected for additional samples was 8 to avoid non-specific amplification of large fragments.

Most studies report on ~30 ng of cfDNA to use for library preparation. For the first time, two sequencing libraries were prepared to compare 30 and 50 ng as inputs for the same cfDNA sample from an mCRPC patient. In parallel, each of these cfDNA inputs was subjected to both low-pass ($10\times$ depth) and deep ($170\times$) WGS for higher sensitivity. At low depth, the sample with 30 ng cfDNA input had a noticeable number of false subclonal events compared to 50 ng input (Figure 3B-black arrows). Other subclonal events (Chr12, ChrX) were found as clonal events with the 50 ng input (Figure 3B-purple arrows). As expected, the 30 ng input had a lower tumor fraction, 0.18 (1st and 2nd maps), compared to 0.24–0.25 with the 50 ng input (3rd and 4th maps). At high depth, the 30 ng input sample showed results similar to low pass (1st and 2nd maps). CNVs detected with the 50 ng input at low pass were also comparable at high depth (Figure 3B-3rd and 4th maps). This infers that higher cfDNA input improves sequencing data. A more in-depth comparison of a 50 ng input sample at low-pass and deep sequencing revealed high correlations and similarities in clonal events (Figure 3C). Nonetheless, discordances were still detected for subclonal events. At low-pass WGS, some segments were detected as subclonal CNVs, while at higher depth they were neutral (Figure 3B, Chr2, ChrX). Furthermore, a clonal event at high depth was detected as a subclonal event at low pass (Figure 3B, Chr22). Also, a subclonal CNV detected at high depth was not captured at low depth (Figure 3B, Chr16).
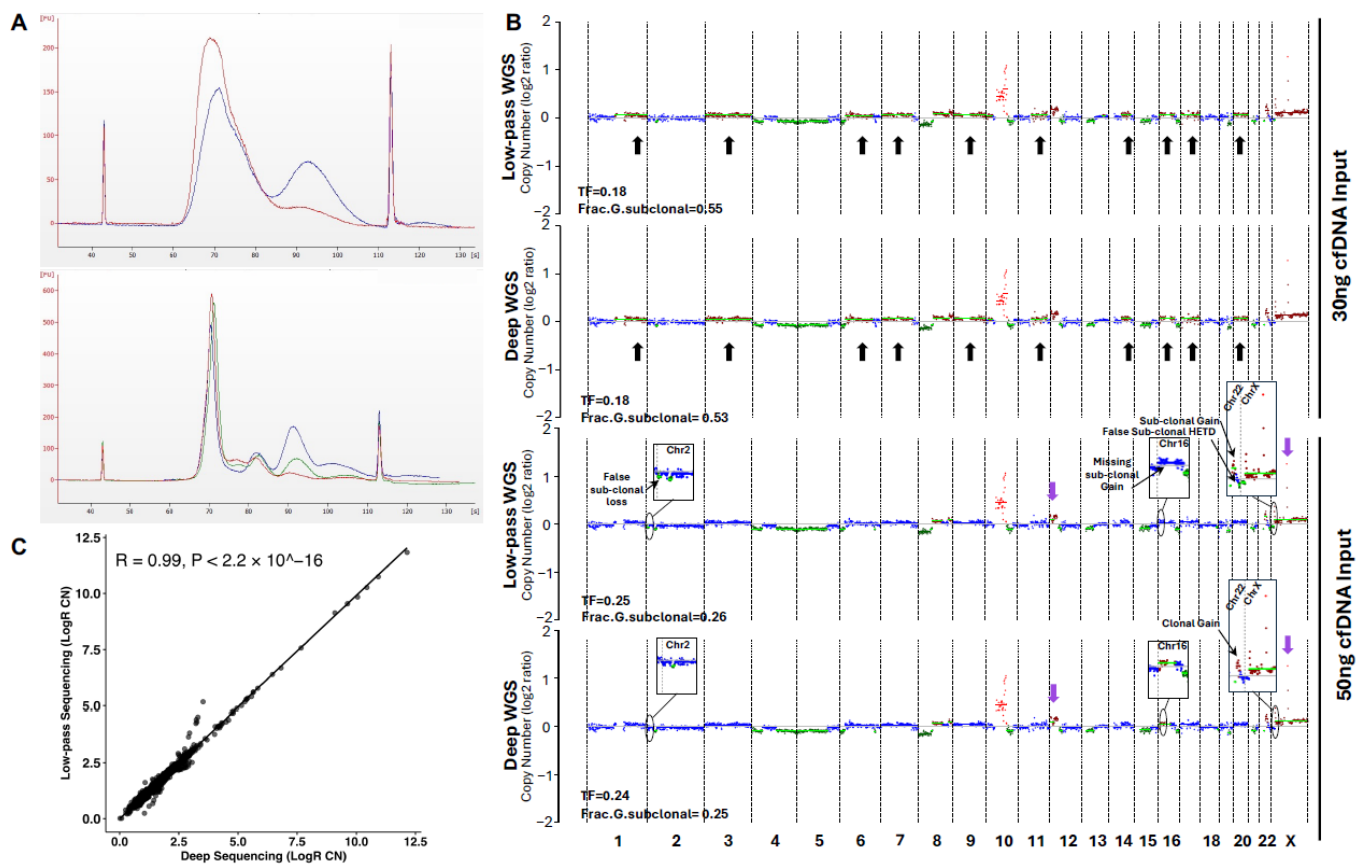
Figure 3. Optimizing conditions to prepare libraries and sequence cfDNA. cfDNA was isolated from 3.5 mL of plasma from a high-risk patient. (**A**). The cfDNAs of the Reference Standard and patient were amplified at different numbers of PCR cycles for library preparation, varying from 8 to 10 and 13. Electropherograms from the Bioanalyzer comparing, in the Top panel, the Reference Standard at 13 (blue) and 8 (red) PCR cycles; and in the Bottom panel, the patient sample at 13 (blue), 10 (green), and 8 (red) PCR cycles. (**B**). Chromosomal maps comparing the two cfDNA inputs at low-pass vs. deep WGS sequencing. The 1st and 2nd panels show CNVs (log2 ratio) for the 30 ng input at low pass (1st panel; post-alignment depth of 9.7×) and deep (2nd panel; 147×). The 3rd and 4th panels show CNVs for the 50 ng input at low-pass (3rd panel) and deep (4th panel) WGS. Black arrows indicate subclonal events. Blue dots in rectangle boxes drawn for Chr2, Chr16, Chr22, and ChrX represent neutral CN; light and dark red show amplification and clonal gain; dark green stands for loss, and light green lines are subclonal events. Purple arrows show clonal events at 30 ng input detected as subclonal at higher input. (**C**). Correlation between LogR CN for low-pass and deep WGS for the 50 ng input sample, yielding a Pearson coefficient (R) of 0.99 and $p < 2.2 \times 10^{-16}$ (^ represents the power). Abbreviations: Cell-free DNA: cfDNA; Copy number variations: CNVs; The fraction of the genome that harbors subclonal alterations: Frac.G.subclonal; Heterozygote deletion: HETD; Polymerase chain reaction: PCR; TF: Tumor fraction; WGS: Whole Genome Sequencing.

To reproduce the comparison of low-pass and high-depth sequencing, the deep sequencing data was downsampled three times to reach about 10× in depth for both cfDNA inputs (Figure 4A, only showing 10× downsampled data for one of the files for each input; other files can be seen in Figure S10). Regardless of inputs, all downsampled bam files showed the same pattern of alterations. Clonal CNVs at 10×, including chr10 amplification and chr8 loss, were similar to high-depth data. This was also true for the subclonal losses at chromosomes 4, 5, 6, 10, 12, 13, 15, 16, and 21 (Figure 4A-turquoise arrows). A great number of false subclonal gains were detected in all downsampled data (chromosomes: 1, 3, 6, 7, 9, 10, 11, 14, 17, and 20) (Figure 4A-black arrows), which were absent in deep sequencing with 50 ng input. While gains in chrX and part of chr12 were detected as subclonal at high depth,

they were annotated as clonal when downsampled at $10\times$, which is the equivalent to low depth (Figure 4A-purple arrows). This further supports that high depth of sequencing can better reflect the true clonality of alterations. Of note, downsampling the 50 ng to $10\times$ reduced the tumor fraction from 0.24 to 0.18, while the fraction of subclonal events increased in parallel (Figure 4A).
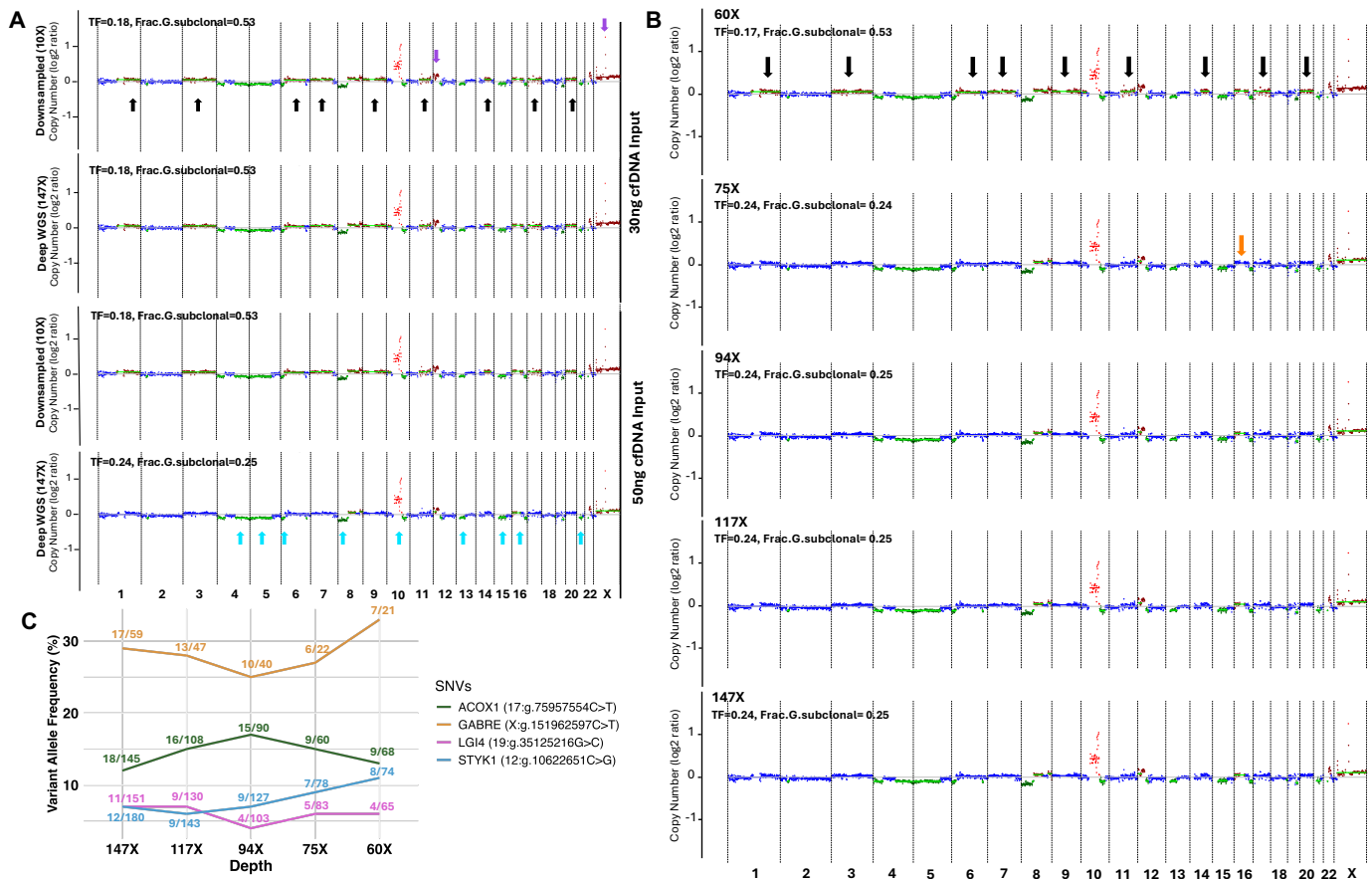
Figure 4. Gradual downsampling defines the optimum depth of sequencing. Attempt to reproduce findings at low pass and identify the minimum sequencing depth (**A**). Chromosomal maps showing CNVs after downsampling the 30 ng and 50 ng inputs deep WGS bam files (~$147\times$) to reach $10\times$ coverage. Turquoise arrows show events detected in all inputs and coverage. Black arrows show subclonal events not detected in deep WGS with 50 ng. Purple arrows show clonal gains, which were detected as subclonal events at high input and depth. (**B**). Chromosomal maps comparing CNVs detected at different depths of sequencing, including the original bam file with 50 ng input at $147\times$ depth and gradual downsampling to $117\times$, $94\times$, $75\times$, and $60\times$. Black arrows show subclonal gains detected at $60\times$ while absent at greater depth. The orange arrow represents a missing subclonal gain at chromosome 16 when downsampled at $75\times$. (**C**). Percentages of variant allele frequency of SNVs at deep sequencing ($147\times$) and gradual downsampling (~$117\times$, $94\times$, $75\times$, $60\times$). The number of reads harboring the mutations out of the total reads is indicated for each gene at different depths. Abbreviations: Cell-free DNA: cfDNA; Copy number variations: CNVs; The fraction of the genome that harbors subclonal alterations: Frac.G.subclonal; Single nucleotide variations: SNVs; Tumor Fraction: TF.

To track genomic alterations detected in deep WGS and identify the optimal depth of sequencing, we gradually downsampled the 50 ng input (20% reduction each time). The original depth of sequencing for the sample after alignment and post-alignment was $147\times$. Thus, the bam file was downsampled to reach $117\times$, $94\times$, $75\times$, and $60\times$ (Figure 4B). All clonal and subclonal events were detected at $117\times$ and $94\times$. There was one subclonal event missing on Chr16, which was not detected at $75\times$ (Figure 4B-orange arrow), while it

was found at 60×. False subclonal events were detected at 60× and the CNV pattern was similar to downsampled 10× bam files (Figure 4B-black arrows). These findings confirm the superiority of higher input and sequencing depth.

For SNVs, the literature reports that Mutect2 and Vardict are the most common tools used to call mutations in cfDNAs. In most cfDNA targeted sequencing studies, criteria considered to call mutations are: at least 10 supporting reads, variant allele frequency (VAF) above 1%, mutant allele fraction (MAF) 20 times higher than the background error rate (i.e., the average allele fraction across all germline DNA samples) and 3 times higher than the allele fraction in the paired germline DNA [14,18–20,26–29]. Some criteria differ in cfDNA WGS; for instance, at least 8 supporting reads, MAF $\geq$ 10%, MAF 50 times higher than the background error rate, and 10 times higher than in the matched germline sample [16]. In our study, we performed variant calling using four tools, including Mutect2, Vardict, Varscan, and Strelka2. Mutations were included if: (I) they were detected by at least two out of four tools; (II) they had a minimum depth of 10×; (III) they had a minimum VAF of 5%; (IV) they had at least 8 supporting reads in cfDNA and 0 reads in matched germline blood DNA; and (V) they had a population allele frequency below 0.01% in 1000 genomes, gnomad, kaviar, and ExAc. All variants detected as 'benign' or 'likely benign' were excluded. Of interest, SNVs were identified in the 50 ng input. They were detected in *ACOX1* (17:g.75957554C>T), *GABRE* (X:g.151962597C>T), *LGI4* (19:g.35125216G>C), and *STYK1* (12:g.10622651C>G) both in the original 147× and downsampled bam files at different depths ~117×, 94×, 75×, and 60× (Figure 4C). However, there were some fluctuations in the VAF of mutations when gradually downsampling bam files. Our findings indicate that more than 8 supporting reads require a coverage of at least 117× to detect low VAF mutations in *LGI4* (19:g.35125216G>C) and *STYK1* (12:g.10622651C>G) (Figure 4C). SNVs cannot be detected at low pass, irrespective of input quantity. Collectively, these CNV and SNV results indicate that it is preferable to use the highest quantity possible of cfDNA to prepare libraries. Furthermore, high-depth sequencing is the best option when precious samples are sequenced to ensure true clonality of alterations and to identify significant mutations.

### 2.7. Clinical Application of the Overall Refined Protocol

Given the above findings, we applied the optimized protocol to two plasma samples collected from a high-risk PCa patient. His features at diagnosis were GG 5 (Gleason score 9 (4 + 5)), pT3a staging, negative margins, and no lymph node extension. His overall trajectory was drawn based on his PSA levels over the course of the disease (Figure 5A). Briefly, he had a rapid recurrence, received standard of care based on progression, and died within 5 years.

The first blood draw (cfDNA1) was obtained prior to RP, and the second, cfDNA2, 6 months prior to death. Sample cfDNA2 had 3–4 times higher concentration than cfDNA1 (Figure 5B, top). Results by Qubit and Bioanalyzer suggest minimal gDNA contamination, as confirmed by the electropherogram (Figure 5B, bottom).

The yield of cfDNA1 from 3.6 mL of plasma was low at 30 ng compared to 130 ng for cfDNA2. To increase the quantity of cfDNA1 to prepare the sequencing library, the remaining 1.75 mL of plasma was extracted. Finally, 40 ng of cfDNA1 and 50 ng of cfDNA2 were used to generate libraries for deep WGS.

At the time of blood collection prior to surgery, no CNVs were detected in cfDNA1, which meant no tumor fraction (Figure 5C). In cfDNA2, clonal CNVs and subclonal events were detected with tumor fraction of ~0.24. As mentioned above, the main CNVs were copy number loss at chr8 as well as gains at chr10 and chr12, detected as clonal events. Noticeably, *AR* amplification was detected at chrX. The SNVs consist of four new mutations,

as described above in *ACOX1* (17:g.75957554C>T), *GABRE* (X:g.151962597C>T), *LGI4* (19:g.35125216G>C), and *STYK1* (12:g.10622651C>G). Collectively, these findings show that despite being a high-risk RP patient with a short overall survival, there was no tumor fraction in his initial cfDNA. Significant changes occurred during progression to mCRPC, which resulted in a detectable tumor fraction with both CNVs and SNVs in the sample collected late in his trajectory.
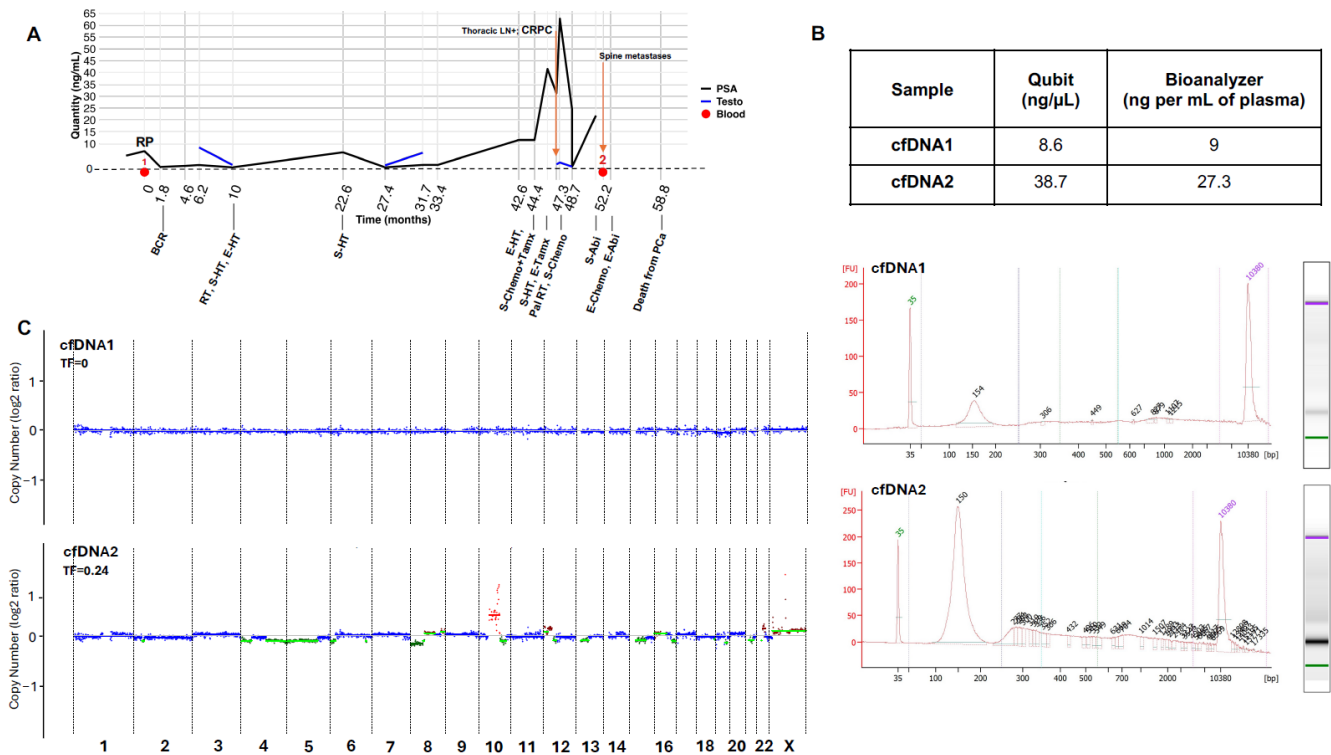
Figure 5. Refined protocol to isolate and sequence cfDNAs from plasma obtained at diagnosis and at the advanced stage of PCa. (**A**). Trajectory of high-risk PCa patient from prior to RP until death. The black and blue lines show blood levels (ng/mL) of PSA and Testosterone, respectively. Red dots indicate times of blood collections used to purify plasma cfDNA1 (prior to RP) and cfDNA2 (6 months prior to death). (**B**). Determination of cfDNA1 and cfDNA2 quantity and quality. Top panel: cfDNA concentrations assessed with Qubit and Bioanalyzer. Middle and Bottom panels: Electropherograms and corresponding gel electrophoresis from Bioanalyzer of cfDNA1 and cfDNA2, respectively. (**C**). WGS of cfDNA1 and cfDNA2 at 170×. TF is indicated on the top left. Blue dots represent neutral CN, light and dark red show amplification and clonal gain; dark green stands for loss, and light green lines are subclonal events. Abbreviations: Abiraterone: Abi; Biochemical recurrence: BCR; Cell-free DNA: cfDNA; Chemotherapy: Chemo; End: E; Palliative: Pal; Prostate Specific Antigen: PSA; Radical prostatectomy: RP; Radiotherapy: RT; Tamoxifen: Tamx; Start: S; Testosterone: Testo; Tumor Fraction: TF; Whole Genome Sequencing: WGS.

3. Discussion

For the first time, in this present investigation, we reviewed in depth and modified the cfDNA isolation and sequencing protocols in the context of PCa to provide guidelines. The optimized methods were next applied to the plasma from several patients and to a high-risk patient.

Collection and analysis of cfDNA require specialized approaches due to the short half-life and low quantity of DNA fragments, particularly at the early stage of disease. Blood processing must be done carefully to keep blood cells intact, due to the possibility that germline DNA may contaminate cfDNA through normal blood cell lysis. We report first the modifications made to the QIAGEN cfDNA isolation kit, including adding proteinaseK

before thawing the plasma, replacing the AVE elution buffer with water, and repeating the column elution step four times with different volumes of water to ensure the complete recovery of cfDNA. This implies that significant quantities of cfDNAs likely remain on the column when the elution is carried out with only a small volume of AVE buffer and one elution step. The quantity of cfDNA obtained with the optimized procedure, requiring generally 3.6 mL of plasma from most PCa patients and healthy males, is in line with cfDNA values obtained when isolating cfDNA from 7 mL up to 55 mL, but without considering gDNA contamination [30]. This would support higher recovery of cfDNA with our procedure with consistent plasma volume. Variability in plasma volume directly impacts the cfDNA yields. Normalizing the cfDNA concentration to the volume of plasma (ng/mL) is essential to reflect biological variations rather than differences arising from sample processing. Our results also revealed that there is a limited chance of complete isolation of cfDNA at the first elution step when samples contain quantities higher than 1 ng/µL. Genomic DNA contamination reduces the possibility of collecting all cfDNA at the first elution step. Magnetic beads were reported to reduce such contamination [31,32]. We not only accounted for gDNA in cfDNA samples by comparing Qubit and Bioanalyzer data but also by removing the contaminating material with magnetic beads to obtain "pure" cfDNA preparations. We provided a reliable threshold for considering samples as pure or contaminated when they contain >50 pg/µL of fragments longer than 5000 bp in the Bioanalyzer. Accordingly, we propose a method to quantify cfDNA by summing only peaks between 50 and 550 bp, representing mono-, di-, and tri-nucleosomes.

We next isolated cfDNA from banked plasma of 60 individuals, including healthy males and PCa patients at the time of RP, disease-free, and at the mCRPC stage. There was no correlation of cfDNA concentration with clinical features at diagnosis. While blood PSA levels were not correlated with cfDNA concentration in RP, disease-free, and mCRPC patients at the time of diagnosis, an association was observed between levels of cfDNA and PSA at the time of inclusion in the study for mCRPC cases. Our findings are consistent with previous reports with a larger sample size that compared the cfDNA quantity among healthy controls, RP cases, and mCRPC patients [22,30,33]. Authors showed higher plasma cfDNA in mCRPC patients compared to RP cases and/or healthy males. It is noticeable that the lowest cfDNA quantity we detected was in the RP cases whose blood was drawn in the operating room prior to anesthesia. It is thus unlikely that lower cfDNA levels may be due to anesthetic agents reported to reduce the cfDNA quantity by affecting cell survival/death [30,34]. Therefore, cfDNA levels may be low at the early stage of PCa and at the time of surgery [30]. It was reported that the cfDNA concentration in mCRPC patients is directly affected by treatment, the amount being significantly decreased (3.89%) in responders to Abiraterone acetate and prednisone but slightly increased (0.94%) in non-responders as compared with pre-treatment levels [35]. In our study, it was not possible to state differences in cfDNA quantity based on treatments including Abiraterone, Enzalutamide, and Docetaxel, as too few patients were included. A more systematic study on mCRPC patients is required. Of interest, elevated cfDNA concentration was related to progression-free survival [36].

Among features characterizing cfDNAs is the size of fragments detected by the Bioanalyzer. When analyzed among our patients, mCRPC samples had the shortest fragments, correlating with the highest plasma concentration of cfDNA. This finding is in support of levels of cfDNA increasing and fragment size being reduced during cancer progression [37]. Our study showed no difference between fragment size in the RP vs. healthy groups, whereas a lower average cfDNA fragment size was reported in RP patients compared to healthy males [30].

In the field of PCa, the majority of cfDNA studies use targeted sequencing or low-pass WGS. Various types of library preparation kits have been used, the most common being Roche NimbleGen SeqCap EZ Choice and ThruPLEX Plasma-seq. Targeted sequencing studies focused on less than 10 genes [28] and up to about 570 genes [24]. Only one study performed cfDNA deep WGS, but for samples showing ctDNA fraction greater than 30% in targeted sequencing of a panel of 73 PCa-relevant genes [16]. To our knowledge, the best conditions to prepare libraries with an adequate number of PCR cycles, cfDNA input, and depth of sequencing have not been systematically examined. After skipping the fragmentation step in the TruSeq DNA Nano library prep protocol to be compatible with the short size of cfDNA (recommendation by Illumina), we investigated how the quantities of cfDNA affect library preparation and results of both low-pass and deep WGS. We report for the first-time that the best number of PCR cycles is 8; with 10 or 13 PCR cycles, a false positive amplification of libraries occurred at 500–2000 bp. Of note, library preparation with 8 PCR cycles was efficient irrespective of samples being treated with a magnet. Using the cfDNA sample of an mCRPC case, we showed that with a higher input, there is a lower chance of detecting false subclonal CNVs. This is in line with libraries generated for targeted sequencing with low input (below 20 ng), where there are higher levels of PCR duplicates, but excluding them resulted in unique reads correlating positively with cfDNA concentration [38]. Sequencing artifacts such as PCR amplification errors can mimic true mutations, especially at low allele frequencies, which may lead to inappropriate clinical decisions. Our two libraries (30 and 50 ng inputs) were sequenced at low-pass ($10\times$) and at high depth ($170\times$). Although all clonal CNVs were detected in both inputs at low and deep sequencing, differences were found with regard to subclonal events. There were subclonal CNVs detected at high depth but not at low depth. Also, due to low coverage, false subclonal events were found at $10\times$ depth. We thus downsampled the high-depth bam files from $147\times$ (depth after alignment) to $117\times$, $94\times$, $75\times$, and $60\times$. False subclonal events started to be detectable at $60\times$, and SNV analysis revealed that at least $117\times$ coverage is required to detect mutations with low allele frequency. These findings indicate that while deep sequencing is much more expensive than low depth, it provides more accurate CNV results as well as SNVs, which are not possible to identify at low depth. On the other hand, if the aim of a study is only clonal CNVs, low-pass WGS would be the most affordable choice.

As mentioned, collecting a sufficient quantity of cfDNA at an early stage of PCa and at the time of surgery can be challenging [30], as observed in our RP series. To move further with optimized cfDNA protocols and investigate the clinical significance of the two samples obtained for the high-risk patient, more plasma (5.3 mL) was required to isolate 40 ng of cfDNA prior to RP, whereas, as expected, 3.5 mL of his plasma yielded 130 ng of cfDNA at the late stage of disease. The estimation of tumor fraction in the plasma refers to the proportion of tumor DNA relative to total cfDNA. Library preparation and high-depth sequencing revealed no tumor fraction in the cfDNA of the first blood draw of our patient but several genomic alterations when progressing to the advanced mCRPC stage. Therefore, while ctDNA shows potential for aiding disease monitoring through analyzing genomic alterations, further studies with larger case numbers are required to validate this conclusion. The tumor fraction is mainly dependent on tumor type and disease burden and ranges from less than 1% to greater than 30% [26]. Consistent findings were reported on genomic alterations present in the ctDNA of PCa patients and DNA from metastatic biopsies, ranging between 90% and 94%, whereas none of the alterations found in ctDNA were detected in primary prostate tumors [14,29]. This is in line with the absence of ctDNA in the cfDNA1 (sample before RP), although he was a high-risk case. Of note, it was reported that up to 96.9% of somatic mutations identified in time-matched ctDNA

were shared with metastatic tissue biopsies of mCRPC patients [16]. Consistent with this finding, authors reported that truncal alterations in metastatic biopsies are detected as subclonal in ctDNA and derived from multiple metastases. Furthermore, when analyzing genomic ctDNA alterations in relation to treatment, a higher frequency was observed from before starting Abiraterone or Enzalutamide until the end of treatment in non-responders compared to responders [39].

The most interesting genomic alteration we identified is *AR* amplification, which has been reported in several ctDNA studies. For instance, it was frequently found by targeted sequencing at a high depth (500×–10,000×) in cfDNA and biopsies of metastatic tissues [40–42] and of primary prostate tumors from advanced cases [43]. Such amplification was detected in ctDNA of 55.4% of patients collected at disease progression [43]. Moreover, it was related to treatment and identified by digital droplet PCR in 14.3% of patients a month before starting first-line treatment with androgen receptor signaling inhibitors (ARSIs), like Abiraterone or Enzalutamide [21], whereas 5.3% (1 out of the 19) of patients showed *AR* amplification in cfDNA when blood samples were collected during ARSI response [43]. It is noticeable that *AR* amplification is found at higher frequency in patients progressing on Enzalutamide compared to Abiraterone or other agents and is thus associated with resistance to Enzalutamide [18,40,44]. The copy number gain varies from <1 to more than 150 copies, depending on disease stage and treatment response or non-response [16,26,43]. Shorter time to progression was reported in patients with *AR* gain that have more than 8 copies compared to patients with less than 8 copies [26]. Along with this *AR* amplification, *AR* enhancer region amplification was detected in 40% of mCRPC patients. Both progression-free survival and overall survival are significantly shorter among patients with both *AR* gene and enhancer amplifications compared to those without [41]. The survival rate in patients with baseline *AR* amplification at 10-, 20-, and 30-months post-chemotherapy was 100%, 53%, and 0%, respectively, while in patients without baseline *AR* gain, the survival rate was 96%, 80%, and 58%, respectively [45]. Overall, *AR* amplification has the potential of detecting treatment resistance over the course of the disease.

Studies also indicate that ctDNAs can reflect the landscape of clonally expanded driver mutations, even when the fraction of metastatic source is minimal. High concordance was detected in *AR* amplification with other CNV calls in genes such as *BRCA2*, *ATM*, *PTEN*, *PIK3CA*, *TP53*, and *RB1* in the ctDNA and matched solid biopsy [14,18,46–48]. These were not found in our patient, although a subclonal *PTEN* loss was observed.

The *AR* is known to be highly mutated. One study reported that 83.3% (10/12) of cases showed shared mutations in genomic DNA from prostate and ctDNA of CRPC patients [43]. In the ctDNA of our high-risk case isolated at the late stage and sequenced at high depth, there was no mutation in the *AR* gene. However, four new mutations were detected in *STYK1*, *ACOX1*, *LGI4*, and *GABRE*. The *STYK1* gene was found to be overexpressed in CRPC patients [49]. This gene was reported to promote metastasis by activating MEK/ERK and PI3K/AKT signaling in human hepatocellular carcinoma and by reducing PINT2/HAI-2 expression in non-small cell lung cancer [50,51]. Higher expression levels of *ACOX1* were detected in PC3 cells, a cell line with prostatic small cell carcinoma, compared to 22RV1 cells (a human prostate carcinoma epithelial cell line) [52]. Direct studies on *LGI4* or *GABRE* in PCa are limited. *LGI4* is known as a putative secreted protein involved in neural development and function [53]. *GABRE* encodes a subunit of the gamma-aminobutyric acid (GABA)-A receptor, which is involved in inhibitory neurotransmission in the central nervous system. GABA has been shown to promote gastrin-releasing peptide secretion in neuroendocrine-like PCa cells, contributing to tumor progression [54]. Additionally, another subunit of the GABA-A receptor, *GABRB3*, has been

identified as a potential biomarker in PCa [55]. Further research is necessary to determine whether these mutations have any significant role in PCa.

Overall, this study provides a comprehensive literature review on cfDNA methodology applicable to PCa, with step-by-step optimization of protocols for cfDNA purification, library preparation, and sequencing. Comparison of cfDNA quantity and fragment size among healthy controls and patients at the time of RP or at mCRPC is in line with previous reports. We also added one more group, disease-free patients, who had similar quantity and fragment size compared to healthy males. The highest levels and shorter cfDNA fragments were detected in mCRPC patients. The inclusion of cfDNA1 and cfDNA2 from the high-risk patient suggests the effectiveness of cfDNA levels and size in monitoring disease over time. The current work presents pros and cons with low-depth sequencing, which is more common in studying cfDNA. High depth definitely proved superiority for detecting true copy number variations and mutations if inputs are optimal. Taken together, the optimized cfDNA procedure will hopefully help in guiding the standardization of assays to better contribute to meaningful clinical applications, thereby opening the way to further experimentation on larger sample sizes.

4. Material and Methods

4.1. Established Protocols on cfDNA in PCa

A literature review was conducted to collect cfDNA studies in PCa and summarize the methods used. The screening was done between 2015 and 2025 through Pubmed, Web of Science, and Scopus databases. The following combinations of keywords were used: prostate cancer, cancer of the prostate, prostatic cancer, cancer of the prostate, prostate neoplasm, prostatic neoplasm; circulating tumor DNA, cell-free tumor DNA, cell-free DNA; liquid biopsy, plasma, serum, and blood. Abstracts and review studies were excluded, and duplicated articles were eliminated. 45 studies were kept. The complete steps in methods from isolation up to sequencing and bioinformatics analysis are presented in the Supplementary Table S1.

4.2. Ethical Review and Patient Consent

This study was approved by the Ethics Review Board of the McGill University Health Center (MUHC) (MP-37-2017-3189/MUHC biobank; MP-37-2021-6957: multicentric Quebec PROCURE PCa Biobank), in conformity with the Declaration of Helsinki. Written consent was obtained from participants donating blood for research.

4.3. Blood Germline DNA Processing and Sequencing

Whole blood was collected in EDTA tubes and centrifuged at $2500\times g$ for 15 min at room temperature within 2 h after collection. Plasma (1.8 mL aliquots) and buffy coat layers were stored at $-80\,°C$ until analysis. Germline DNA from buffy coats was extracted using a QIAamp DNA kit (Qiagen; Germantown, MD, USA), quantified by picogreen assay (Thermo Fisher Scientific; Waltham, MA, USA), and stored at $-80\,°C$. The quality of DNA was confirmed by agarose gel electrophoresis. DNA samples with high quality and sufficient quantity were sent to the McGill Genome Center (MGC) for preparing a library with the Lucigen NxSeq AmpFREE kit (Biosearch Technology; Middlesex, UK). Quality was tested using Tapestation (Agilent; Millcreek, ON, Canada). Libraries were sequenced at a depth of $40\times$ for blood DNA on the Novaseq6000-S4 using 150-bp paired-end sequencing reads (2×150 bp; Illumina).

4.4. Isolation and Quantification of cfDNA

The QIAamp circulating nucleic acid kit (Qiagen) was used for isolating cfDNA from plasma with some modifications applied to the protocol. In this study, cfDNA was isolated from 2 tubes of 1.8 mL plasma. 90 μL of proteinaseK was added to each tube before thawing at room temperature for 15 min. Each tube was gently vortexed three times during this period. Tubes were centrifuged at 12,000 rpm for 5 min at 4 °C. After centrifugation, 3.68 mL of plasma from the two tubes was transferred into a 50 mL tube to which 0.32 mL of PBS was added (total 4.0 mL). This was followed by the second addition of 220 μL proteinaseK and 3.2 mL ACL buffer (containing 1 μg carrier RNA) and vortexed gently for 30 s. The mixture was incubated for 30 min at 60 °C for a better recovery of DNA after protein digestion and lipid solubilization. The ACB buffer (7.2 mL) was added to the mixture, vortexed thoroughly for 30 s, and incubated on ice for 5 min. The lysate was passed through the QIAamp mini spin column during approximately 13 min under a pressure of −900 mbar, as applied by the QIAvac 24 Plus vacuum system. In line with the manufacturer protocol, ACW1 (600 μL) buffer followed by ACW2 buffer (750 μL) and 100% ethanol (750 μL) were passed through the column sequentially, next placed in a 2 mL collection tube, and centrifuged at $20,000\times g$ for 3 min at room temperature. It was incubated at 56 °C for 10 min. Different volumes of warm water (56 °C) were used for elution instead of AVE buffer, as explained above when preparing the library. It was applied in the center of the column, incubated at room temperature for 3 min, and centrifuged at $20,000\times g$ for 1 min. The elution step was repeated 4 times, and nucleic acids were recuperated into separate aliquots. Elution volumes for each step were 85 μL (elute #1), 40 μL (elute #2), 20 μL (elute #3), and 20 μL (elute #4). After gentle pipetting, 2 μL cfDNA aliquots were tested for quality and determining quantity. cfDNA aliquots were stored at −20 °C.

The DNA concentration was measured on a Qubit 3.0 Fluorometer (ThermoFisher Scientific) using a Qubit dsDNA HS assay Kit according to the manufacturer's protocol. To evaluate the quality and precise quantity of cfDNA, 1 μL of sample from each aliquot was run on the Bioanalyzer2100 (Agilent) using the high-sensitivity DNA kit. Based on peaks in the Bioanalyzer electropherogram, cfDNA fragments were categorized into four groups according to their size, including 50 to 250 bp, 250–350 bp, 350–550 bp, 550–5000 bp, and >5000 bp for gDNA contamination. The total quantity of fragments within the 50–550 bp range was determined by summing the quantities across the four eluates, unless otherwise mentioned in the Result section. Their sum was used to report on cfDNA concentrations (ng) per mL of plasma.

4.5. Purification of cfDNA by Removal of Contaminating gDNA

When Bioanalyzer showed evidence of gDNA contamination, a size selection procedure was applied to remove large fragments using AMPure XP magnetic beads (Beckman; Coulter, Brea, CA, USA) in 96-well plates. The first step was carried out using 40 μL of sample and 20 μL of AMPure beads (0.5×), incubated at room temperature for 5 min, and maintained on the magnetic stand (ThermoFisher Scientific) for 3 min. The supernatant was collected in a separate well and mixed thoroughly with 120 μL of magnetic beads (2×). The incubation at room temperature and magnetic stand step was done a second time for 5 min. A 70% ethanol wash (with 200 μL) was performed twice, and the liquid was removed. The magnetic beads were dried in air for 5 min. Then, 40 μL of water was added to the beads and placed on a magnet to collect pure cfDNA in the elute. The gDNA was extracted from the beads and tested on the Bioanalyzer along with the cfDNA to validate the procedure.

4.6. cfDNA Library Preparation and Sequencing

Modifications were brought to the TruSeq Nano DNA low-throughput library prep kit (Illumina; San Diego, CA, USA) protocol, with help from Illumina, to render this kit compatible for cfDNA. The modified protocol was first tested on 30 ng of Multiplex I cfDNA Reference Standard Set (Horizon Discovery, Waterbeach, UK) as a positive control and 30 ng from one of our isolated cfDNA samples. The fragmentation step was skipped for both the cfDNA Reference Standard Set and our cfDNA sample. Repair-end was performed by adding 40 μL of ERP mix to 50 μL of cfDNA sample to convert the natural single-strand ends into blunt ends. Selection library size was done in only one step through adding 2× SPB magnetic beads provided in the kit and removing the supernatant after 6 min on the magnetic stand. Beads were washed two times with 70% ethanol, then diluted with 20 μL of RSB buffer. Adenylate 3′ ends and ligate adapters steps were performed according to the manufacturer's instructions, and IDT for Illumina–TruSeq DNA UD Indexes v2 (Illumina) were used for index oligo ligation. To amplify cfDNA libraries, 8–10 and 13 PCR cycles were compared to find the optimal number of cycles. Library concentrations were tested by qPCR using the sparQ Universal Library Quant Kit (Quantabio; Beverly, MA, USA), as per the manufacturer's protocol.

Library quantification was carried out through both our Bioanalyzer and the Tapestation at the MGC sequencing facility. Libraries were built from one sample with two cfDNA inputs. Pooled libraries were sequenced on Novaseq6000-S4 using 150-bp paired-end sequencing reads (2 × 150 bp; Illumina) at low depth (~10×) and high depth (~170×).

4.7. Sequence Data Analysis

The Genpipe pre-built pipelines were used in this study. Genpipe is a Python-based platform developed at the Canadian Center for Computational Genomics (C3G) that provides various genomic sequencing pipelines, including for WGS [56]. The tumorpair pipeline version 4.6.0 was adapted, and additional steps were added for analyzing cfDNA WGS. Details of data analysis will be further explained below.

For sequence alignment and quality control, adapters were trimmed using Skewer [57], and the end quality and mean quality values were set to 25. Paired-end reads were aligned against the hg38 reference genome using Burrows-Wheelers Aligner (BWA-MEM) version 0.7.17 [58] and realigned by GATK-realigner [59]. Mark duplication was done using Picard version 2.23.3. Recalibration was performed through GATK BaseRecalibrator. The Conpair version 0.2 was used to estimate contamination and sample concordance [60].

4.8. Analysis of CNVs

The CNV calling was performed using IchorCNA [61] for both cfDNA low-depth and deep sequencing. Sequencing data of DNA extracted from the blood buffy coat layer was used as a source of germline DNA.

4.9. Analysis of Somatic Mutations

SNVs were called using four variant callers, including Mutect2 [62], Strelka2 [63], Vardict [64], and Varscan2 [65]. An ensemble vcf file of variations from four tools was developed using Bcbio.variation.recall. Then, GATK VariantAnnotator was used for annotating variants. The five criteria we considered to call true somatic variations were mentioned in the Result section. To further exclude false-positive results, all mutations were manually reviewed using Integrative Genomics Viewer (IGV) [66].

4.10. Downsampling

The Sambamba version V0.8.2 [67] tool was used to downsample the aligned, de-duplicated, recalibrated bam file. Samtools version V1.20 [68] was used to estimate the average depth of sequencing.

4.11. Statistical Analysis

The cfDNA concentration quantified by Qubit or Bioanalyzer and fragment size were compared among healthy controls, patients at the time of RP, disease-free, and mCRPC patients using the Wilcoxon rank sum test. Also, the Wilcoxon test was used for comparison of cfDNA concentration by Gleason Grade (GG) of RP, disease-free and mCRPC cases, pathologic features at RP (intraductal carcinoma, lymphovascular, lymph node or seminal vesicle invasions, pathological staging, and surgical margins), and progression at inclusion or progression on Abiraterone, Enzalutamide, or Docetaxel treatments for mCRPC cases.

For comparing continuous characteristics, Pearson correlation coefficients were used, including cfDNA concentrations versus percentages of fragments collected at first elusion or versus age at diagnosis, PSA at biopsy or inclusion in the study, or average cfDNA fragment size (bp). The mCRPC cases were categorized into low or high cfDNA levels based on the median plasma cfDNA concentration. The association between cfDNA levels and overall survival of mCRPC patients was evaluated by Kaplan-Meier plot. The correlation between low-pass and high-depth sequencing data was assessed by Pearson correlation coefficients. All data analyses were performed using R version 4.4.2 [69].

References

1.   Berenguer, C.V.; Pereira, F.; Camara, J.S.; Pereira, J.A.M. Underlying Features of Prostate Cancer-Statistics, Risk Factors, and Emerging Methods for Its Diagnosis. *Curr. Oncol.* **2023**, *30*, 2300–2321. [CrossRef] [PubMed]
2.   Correas, J.M.; Halpern, E.J.; Barr, R.G.; Ghai, S.; Walz, J.; Bodard, S.; Dariane, C.; de la Rosette, J. Advanced ultrasound in the diagnosis of prostate cancer. *World J. Urol.* **2021**, *39*, 661–676. [CrossRef]
3.   Eldred-Evans, D.; Ahmed, H.U. Population-Based Prostate Cancer Screening With Magnetic Resonance Imaging or Ultrasonography-The IP1-PROSTAGRAM Study-Reply. *JAMA Oncol.* **2021**, *7*, 1575–1576. [CrossRef]
4.   Tikkinen, K.A.O.; Dahm, P.; Lytvyn, L.; Heen, A.F.; Vernooij, R.W.M.; Siemieniuk, R.A.C.; Wheeler, R.; Vaughan, B.; Fobuzi, A.C.; Blanker, M.H.; et al. Prostate cancer screening with prostate-specific antigen (PSA) test: A clinical practice guideline. *BMJ* **2018**, *362*, k3581. [CrossRef] [PubMed]
5.   Thompson, I.M.; Ankerst, D.P.; Chi, C.; Goodman, P.J.; Tangen, C.M.; Lucia, M.S.; Feng, Z.; Parnes, H.L.; Coltman, C.A., Jr. Assessing prostate cancer risk: Results from the Prostate Cancer Prevention Trial. *J. Natl. Cancer Inst.* **2006**, *98*, 529–534. [CrossRef] [PubMed]
6.   Nakanishi, K.; Tanaka, J.; Nakaya, Y.; Maeda, N.; Sakamoto, A.; Nakayama, A.; Satomura, H.; Sakai, M.; Konishi, K.; Yamamoto, Y.; et al. Whole-body MRI: Detecting bone metastases from prostate cancer. *Jpn. J. Radiol.* **2022**, *40*, 229–244. [CrossRef]
7.   Li, R.; Ravizzini, G.C.; Gorin, M.A.; Maurer, T.; Eiber, M.; Cooperberg, M.R.; Alemozzaffar, M.; Tollefson, M.K.; Delacroix, S.E.; Chapin, B.F. The use of PET/CT in prostate cancer. *Prostate Cancer Prostatic Dis.* **2018**, *21*, 4–21. [CrossRef]
8.   Bjurlin, M.A.; Carter, H.B.; Schellhammer, P.; Cookson, M.S.; Gomella, L.G.; Troyer, D.; Wheeler, T.M.; Schlossberg, S.; Penson, D.F.; Taneja, S.S. Optimization of initial prostate biopsy in clinical practice: Sampling, labeling and specimen processing. *J. Urol.* **2013**, *189*, 2039–2046. [CrossRef]
9.   Trujillo, B.; Wu, A.; Wetterskog, D.; Attard, G. Blood-based liquid biopsies for prostate cancer: Clinical opportunities and challenges. *Br. J. Cancer* **2022**, *127*, 1394–1402. [CrossRef]
10. Crocetto, F.; Russo, G.; Di Zazzo, E.; Pisapia, P.; Mirto, B.F.; Palmieri, A.; Pepe, F.; Bellevicine, C.; Russo, A.; La Civita, E.; et al. Liquid Biopsy in Prostate Cancer Management-Current Challenges and Future Perspectives. *Cancers* **2022**, *14*, 3272. [CrossRef]
11. Kim, C.J.; Dong, L.; Amend, S.R.; Cho, Y.K.; Pienta, K.J. The role of liquid biopsies in prostate cancer management. *Lab Chip* **2021**, *21*, 3263–3288. [CrossRef] [PubMed]
12. Garofoli, M.; Maiorano, B.A.; Bruno, G.; Giordano, G.; Falagario, U.G.; Necchi, A.; Carrieri, G.; Landriscina, M.; Conteduca, V. Circulating Tumor DNA: A New Research Frontier in Urological Oncology from Localized to Metastatic Disease. *Eur. Urol. Oncol.* **2024**, *8*, 805–817. [CrossRef] [PubMed]
13. 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. [CrossRef]
14. Wyatt, A.W.; Annala, M.; Aggarwal, R.; Beja, K.; Feng, F.; Youngren, J.; Foye, A.; Lloyd, P.; Nykter, M.; Beer, T.M.; et al. Concordance of Circulating Tumor DNA and Matched Metastatic Tissue Biopsy in Prostate Cancer. *J. Natl. Cancer Inst.* **2017**, *109*, djx118. [CrossRef]
15. De Laere, B.; van Dam, P.J.; Whitington, T.; Mayrhofer, M.; Diaz, E.H.; Van den Eynden, G.; Vandebroek, J.; Del-Favero, J.; Van Laere, S.; Dirix, L.; et al. Comprehensive Profiling of the Androgen Receptor in Liquid Biopsies from Castration-resistant Prostate Cancer Reveals Novel Intra-AR Structural Variation and Splice Variant Expression Patterns. *Eur. Urol.* **2017**, *72*, 192–200. [CrossRef]
16. Herberts, C.; Annala, M.; Sipola, J.; Ng, S.W.S.; Chen, X.E.; Nurminen, A.; Korhonen, O.V.; Munzur, A.D.; Beja, K.; Schonlau, E.; et al. Deep whole-genome ctDNA chronology of treatment-resistant prostate cancer. *Nature* **2022**, *608*, 199–208. [CrossRef]
17. Lallous, N.; Volik, S.V.; Awrey, S.; Leblanc, E.; Tse, R.; Murillo, J.; Singh, K.; Azad, A.A.; Wyatt, A.W.; LeBihan, S.; et al. Functional analysis of androgen receptor mutations that confer anti-androgen resistance identified in circulating cell-free DNA from prostate cancer patients. *Genome Biol.* **2016**, *17*, 10. [CrossRef] [PubMed]
18. Wyatt, A.W.; Azad, A.A.; Volik, S.V.; Annala, M.; Beja, K.; McConeghy, B.; Haegert, A.; Warner, E.W.; Mo, F.; Brahmbhatt, S.; et al. Genomic Alterations in Cell-Free DNA and Enzalutamide Resistance in Castration-Resistant Prostate Cancer. *JAMA Oncol.* **2016**, *2*, 1598–1606. [CrossRef]
19. Vandekerkhove, G.; Struss, W.J.; Annala, M.; Kallio, H.M.L.; Khalaf, D.; Warner, E.W.; Herberts, C.; Ritch, E.; Beja, K.; Loktionova, Y.; et al. Circulating Tumor DNA Abundance and Potential Utility in De Novo Metastatic Prostate Cancer. *Eur. Urol.* **2019**, *75*, 667–675. [CrossRef]
20. Herberts, C.; Murtha, A.J.; Fu, S.; Wang, G.; Schonlau, E.; Xue, H.; Lin, D.; Gleave, A.; Yip, S.; Angeles, A.; et al. Activating AKT1 and PIK3CA Mutations in Metastatic Castration-Resistant Prostate Cancer. *Eur. Urol.* **2020**, *78*, 834–844. [CrossRef]
21. Del Re, M.; Conteduca, V.; Crucitta, S.; Gurioli, G.; Casadei, C.; Restante, G.; Schepisi, G.; Lolli, C.; Cucchiara, F.; Danesi, R.; et al. Androgen receptor gain in circulating free DNA and splicing variant 7 in exosomes predict clinical outcome in CRPC patients treated with abiraterone and enzalutamide. *Prostate Cancer Prostatic Dis.* **2021**, *24*, 524–531. [CrossRef] [PubMed]

22. Chen, E.; Cario, C.L.; Leong, L.; Lopez, K.; Marquez, C.P.; Li, P.S.; Oropeza, E.; Tenggara, I.; Cowan, J.; Simko, J.P.; et al. Cell-Free DNA Detection of Tumor Mutations in Heterogeneous, Localized Prostate Cancer Via Targeted, Multiregion Sequencing. *JCO Precis. Oncol.* **2021**, *5*, 710–725. [CrossRef] [PubMed]

23. Fonseca, N.M.; Maurice-Dror, C.; Herberts, C.; Tu, W.; Fan, W.; Murtha, A.J.; Kollmannsberger, C.; Kwan, E.M.; Parekh, K.; Schonlau, E.; et al. Prediction of plasma ctDNA fraction and prognostic implications of liquid biopsy in advanced prostate cancer. *Nat. Commun.* **2024**, *15*, 1828. [CrossRef]

24. Xia, S.; Kohli, M.; Du, M.; Dittmar, R.L.; Lee, A.; Nandy, D.; Yuan, T.; Guo, Y.; Wang, Y.; Tschannen, M.R.; et al. Plasma genetic and genomic abnormalities predict treatment response and clinical outcome in advanced prostate cancer. *Oncotarget* **2015**, *6*, 16411–16421. [CrossRef] [PubMed]

25. Eagle, S.H.C.; Robertson, J.; Bastedo, D.P.; Liu, K.; Nash, J.H.E. Evaluation of five commercial DNA extraction kits using Salmonella as a model for implementation of rapid Nanopore sequencing in routine diagnostic laboratories. *Access Microbiol.* **2023**, *5*, 000468.v3. [CrossRef]

26. Annala, M.; Vandekerkhove, G.; Khalaf, D.; Taavitsainen, S.; Beja, K.; Warner, E.W.; Sunderland, K.; Kollmannsberger, C.; Eigl, B.J.; Finch, D.; et al. Circulating Tumor DNA Genomics Correlate with Resistance to Abiraterone and Enzalutamide in Prostate Cancer. *Cancer Discov.* **2018**, *8*, 444–457. [CrossRef]

27. Chapman, L.; Ledet, E.M.; Barata, P.C.; Cotogno, P.; Manogue, C.; Moses, M.; Christensen, B.R.; Steinwald, P.; Ranasinghe, L.; Layton, J.L.; et al. TP53 Gain-of-Function Mutations in Circulating Tumor DNA in Men With Metastatic Castration-Resistant Prostate Cancer. *Clin. Genitourin. Cancer* **2020**, *18*, 148–154. [CrossRef]

28. Ye, S.; Wang, H.; He, K.; Peng, M.; Wang, Y.; Li, Y.; Jiang, S.; Li, J.; Yi, L.; Cui, R. Clinical Characterization of Mismatch Repair Gene-Deficient Metastatic Castration-Resistant Prostate Cancer. *Front. Oncol.* **2020**, *10*, 533282. [CrossRef]

29. Fan, L.; Fei, X.; Zhu, Y.; Pan, J.; Sha, J.; Chi, C.; Gong, Y.; Du, X.; Zhou, L.; Dong, B.; et al. Comparative Analysis of Genomic Alterations across Castration Sensitive and Castration Resistant Prostate Cancer via Circulating Tumor DNA Sequencing. *J. Urol.* **2021**, *205*, 461–469. [CrossRef]

30. Chen, E.; Cario, C.L.; Leong, L.; Lopez, K.; Marquez, C.P.; Chu, C.; Li, P.S.; Oropeza, E.; Tenggara, I.; Cowan, J.; et al. Cell-free DNA concentration and fragment size as a biomarker for prostate cancer. *Sci. Rep.* **2021**, *11*, 5040. [CrossRef]

31. Kohli, M.; Tan, W.; Zheng, T.; Wang, A.; Montesinos, C.; Wong, C.; Du, P.; Jia, S.; Yadav, S.; Horvath, L.G.; et al. Clinical and genomic insights into circulating tumor DNA-based alterations across the spectrum of metastatic hormone-sensitive and castrate-resistant prostate cancer. *EBioMedicine* **2020**, *54*, 102728. [CrossRef] [PubMed]

32. Fettke, H.; Kwan, E.M.; Docanto, M.M.; Bukczynska, P.; Ng, N.; Graham, L.K.; Mahon, K.; Hauser, C.; Tan, W.; Wang, X.H.; et al. Combined Cell-free DNA and RNA Profiling of the Androgen Receptor: Clinical Utility of a Novel Multianalyte Liquid Biopsy Assay for Metastatic Prostate Cancer. *Eur. Urol.* **2020**, *78*, 173–180. [CrossRef] [PubMed]

33. Reimers, M.A.; Yip, S.M.; Zhang, L.; Cieslik, M.; Dhawan, M.; Montgomery, B.; Wyatt, A.W.; Chi, K.N.; Small, E.J.; Chinnaiyan, A.M.; et al. Clinical Outcomes in Cyclin-dependent Kinase 12 Mutant Advanced Prostate Cancer. *Eur. Urol.* **2020**, *77*, 333–341. [CrossRef] [PubMed]

34. Prakash, K.; Aggarwal, S.; Bhardwaj, S.; Ramakrishna, G.; Pandey, C.K. Serial perioperative cell-free DNA levels in donors and recipients undergoing living donor liver transplantation. *Acta Anaesthesiol. Scand.* **2017**, *61*, 1084–1094. [CrossRef]

35. Du, M.; Tian, Y.; Tan, W.; Wang, L.; Wang, L.; Kilari, D.; Huang, C.C.; Wang, L.; Kohli, M. Plasma cell-free DNA-based predictors of response to abiraterone acetate/prednisone and prognostic factors in metastatic castration-resistant prostate cancer. *Prostate Cancer Prostatic Dis.* **2020**, *23*, 705–713. [CrossRef]

36. Liu, H.; Gao, Y.; Vafaei, S.; Gu, X.; Zhong, X. The Prognostic Value of Plasma Cell-Free DNA Concentration in the Prostate Cancer: A Systematic Review and Meta-Analysis. *Front. Oncol.* **2021**, *11*, 599602. [CrossRef]

37. Mouliere, F.; Robert, B.; Arnau Peyrotte, E.; Del Rio, M.; Ychou, M.; Molina, F.; Gongora, C.; Thierry, A.R. High fragmentation characterizes tumour-derived circulating DNA. *PLoS ONE* **2011**, *6*, e23418. [CrossRef]

38. Knutson, T.P.; Luo, B.; Kobilka, A.; Lyman, J.; Guo, S.; Munro, S.A.; Li, Y.; Heer, R.; Gaughan, L.; Morris, M.J.; et al. AR alterations inform circulating tumor DNA detection in metastatic castration resistant prostate cancer patients. *Nat. Commun.* **2024**, *15*, 10648. [CrossRef]

39. Sonpavde, G.; Agarwal, N.; Pond, G.R.; Nagy, R.J.; Nussenzveig, R.H.; Hahn, A.W.; Sartor, O.; Gourdin, T.S.; Nandagopal, L.; Ledet, E.M.; et al. Circulating tumor DNA alterations in patients with metastatic castration-resistant prostate cancer. *Cancer* **2019**, *125*, 1459–1469. [CrossRef]

40. Azad, A.A.; Volik, S.V.; Wyatt, A.W.; Haegert, A.; Le Bihan, S.; Bell, R.H.; Anderson, S.A.; McConeghy, B.; Shukin, R.; Bazov, J.; et al. Androgen Receptor Gene Aberrations in Circulating Cell-Free DNA: Biomarkers of Therapeutic Resistance in Castration-Resistant Prostate Cancer. *Clin. Cancer Res.* **2015**, *21*, 2315–2324. [CrossRef]

41. Dang, H.X.; Chauhan, P.S.; Ellis, H.; Feng, W.; Harris, P.K.; Smith, G.; Qiao, M.; Dienstbach, K.; Beck, R.; Atkocius, A.; et al. Cell-free DNA alterations in the AR enhancer and locus predict resistance to AR-directed therapy in patients with metastatic prostate cancer. *JCO Precis. Oncol.* **2020**, *4*, 680–713. [CrossRef] [PubMed]

42. Tukachinsky, H.; Madison, R.W.; Chung, J.H.; Gjoerup, O.V.; Severson, E.A.; Dennis, L.; Fendler, B.J.; Morley, S.; Zhong, L.; Graf, R.P.; et al. Genomic Analysis of Circulating Tumor DNA in 3,334 Patients with Advanced Prostate Cancer Identifies Targetable BRCA Alterations and AR Resistance Mechanisms. *Clin. Cancer Res.* **2021**, *27*, 3094–3105. [CrossRef]

43. Sumiyoshi, T.; Mizuno, K.; Yamasaki, T.; Miyazaki, Y.; Makino, Y.; Okasho, K.; Li, X.; Utsunomiya, N.; Goto, T.; Kobayashi, T.; et al. Clinical utility of androgen receptor gene aberrations in circulating cell-free DNA as a biomarker for treatment of castration-resistant prostate cancer. *Sci. Rep.* **2019**, *9*, 4030. [CrossRef] [PubMed]

44. Romanel, A.; Gasi Tandefelt, D.; Conteduca, V.; Jayaram, A.; Casiraghi, N.; Wetterskog, D.; Salvi, S.; Amadori, D.; Zafeiriou, Z.; Rescigno, P.; et al. Plasma AR and abiraterone-resistant prostate cancer. *Sci. Transl. Med.* **2015**, *7*, 312re10. [CrossRef]

45. Kohli, M.; Li, J.; Du, M.; Hillman, D.W.; Dehm, S.M.; Tan, W.; Carlson, R.; Campion, M.B.; Wang, L.; Wang, L.; et al. Prognostic association of plasma cell-free DNA-based androgen receptor amplification and circulating tumor cells in pre-chemotherapy metastatic castration-resistant prostate cancer patients. *Prostate Cancer Prostatic Dis.* **2018**, *21*, 411–418. [CrossRef]

46. Kang, M.; Cho, E.; Jang, J.; Lee, J.; Jeon, Y.; Jeong, B.C.; Seo, S.I.; Jeon, S.S.; Lee, H.M.; Choi, H.Y.; et al. Genomic analysis of Korean patients with advanced prostate cancer by use of a comprehensive next-generation sequencing panel and low-coverage, whole-genome sequencing. *Investig. Clin. Urol.* **2019**, *60*, 227–234. [CrossRef]

47. Ritch, E.; Fu, S.Y.F.; Herberts, C.; Wang, G.; Warner, E.W.; Schonlau, E.; Taavitsainen, S.; Murtha, A.J.; Vandekerkhove, G.; Beja, K.; et al. Identification of Hypermutation and Defective Mismatch Repair in ctDNA from Metastatic Prostate Cancer. *Clin. Cancer Res.* **2020**, *26*, 1114–1125. [CrossRef] [PubMed]

48. Chi, K.N.; Barnicle, A.; Sibilla, C.; Lai, Z.; Corcoran, C.; Barrett, J.C.; Adelman, C.A.; Qiu, P.; Easter, A.; Dearden, S.; et al. Detection of BRCA1, BRCA2, and ATM Alterations in Matched Tumor Tissue and Circulating Tumor DNA in Patients with Prostate Cancer Screened in PROfound. *Clin. Cancer Res.* **2023**, *29*, 81–91. [CrossRef]

49. Chung, S.; Tamura, K.; Furihata, M.; Uemura, M.; Daigo, Y.; Nasu, Y.; Miki, T.; Shuin, T.; Fujioka, T.; Nakamura, Y.; et al. Overexpression of the potential kinase serine/ threonine/tyrosine kinase 1 (STYK 1) in castration-resistant prostate cancer. *Cancer Sci.* **2009**, *100*, 2109–2114. [CrossRef]

50. Wang, Z.; Qu, L.; Deng, B.; Sun, X.; Wu, S.; Liao, J.; Fan, J.; Peng, Z. STYK1 promotes epithelial-mesenchymal transition and tumor metastasis in human hepatocellular carcinoma through MEK/ERK and PI3K/AKT signaling. *Sci. Rep.* **2016**, *6*, 33205. [CrossRef]

51. Ma, Z.; Liu, D.; Li, W.; Di, S.; Zhang, Z.; Zhang, J.; Xu, L.; Guo, K.; Zhu, Y.; Han, J.; et al. STYK1 promotes tumor growth and metastasis by reducing SPINT2/HAI-2 expression in non-small cell lung cancer. *Cell Death Dis.* **2019**, *10*, 435. [CrossRef] [PubMed]

52. Valenca, I.; Ferreira, A.R.; Correia, M.; Kuhl, S.; van Roermund, C.; Waterham, H.R.; Maximo, V.; Islinger, M.; Ribeiro, D. Prostate Cancer Proliferation Is Affected by the Subcellular Localization of MCT2 and Accompanied by Significant Peroxisomal Alterations. *Cancers* **2020**, *12*, 3152. [CrossRef]

53. Costa, E.T.; Barnabe, G.F.; Li, M.; Dias, A.A.; Machado, T.R.; Asprino, P.F.; Cavalher, F.P.; Ferreira, E.N.; Del Mar Inda, M.; Nagai, M.H.; et al. Intratumoral heterogeneity of ADAM23 promotes tumor growth and metastasis through LGI4 and nitric oxide signals. *Oncogene* **2015**, *34*, 1270–1279. [CrossRef] [PubMed]

54. Solorzano, S.R.; Imaz-Rosshandler, I.; Camacho-Arroyo, I.; Garcia-Tobilla, P.; Morales-Montor, G.; Salazar, P.; Arena-Ortiz, M.L.; Rodriguez-Dorantes, M. GABA promotes gastrin-releasing peptide secretion in NE/NE-like cells: Contribution to prostate cancer progression. *Sci. Rep.* **2018**, *8*, 10272. [CrossRef]

55. Chen, J.Y.; Chang, C.F.; Huang, S.P.; Huang, C.Y.; Yu, C.C.; Lin, V.C.; Geng, J.H.; Li, C.Y.; Lu, T.L.; Bao, B.Y. Integrated analysis identifies GABRB3 as a biomarker in prostate cancer. *BMC Med. Genom.* **2024**, *17*, 41. [CrossRef]

56. Bourgey, M.; Dali, R.; Eveleigh, R.; Chen, K.C.; Letourneau, L.; Fillon, J.; Michaud, M.; Caron, M.; Sandoval, J.; Lefebvre, F.; et al. GenPipes: An open-source framework for distributed and scalable genomic analyses. *Gigascience* **2019**, *8*, giz037. [CrossRef]

57. Jiang, H.; Lei, R.; Ding, S.W.; Zhu, S. Skewer: A fast and accurate adapter trimmer for next-generation sequencing paired-end reads. *BMC Bioinform.* **2014**, *15*, 182. [CrossRef]

58. Li, H.; Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **2009**, *25*, 1754–1760. [CrossRef]

59. McKenna, A.; Hanna, M.; Banks, E.; Sivachenko, A.; Cibulskis, K.; Kernytsky, A.; Garimella, K.; Altshuler, D.; Gabriel, S.; Daly, M.; et al. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **2010**, *20*, 1297–1303. [CrossRef]

60. Bergmann, E.A.; Chen, B.J.; Arora, K.; Vacic, V.; Zody, M.C. Conpair: Concordance and contamination estimator for matched tumor-normal pairs. *Bioinformatics* **2016**, *32*, 3196–3198. [CrossRef]

61. Adalsteinsson, V.A.; Ha, G.; Freeman, S.S.; Choudhury, A.D.; Stover, D.G.; Parsons, H.A.; Gydush, G.; Reed, S.C.; 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. [CrossRef] [PubMed]

62. Cibulskis, K.; Lawrence, M.S.; Carter, S.L.; Sivachenko, A.; Jaffe, D.; Sougnez, C.; Gabriel, S.; Meyerson, M.; Lander, E.S.; Getz, G. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat. Biotechnol.* **2013**, *31*, 213–219. [CrossRef] [PubMed]

63. Kim, S.; Scheffler, K.; Halpern, A.L.; Bekritsky, M.A.; Noh, E.; Kallberg, M.; Chen, X.; Kim, Y.; Beyter, D.; Krusche, P.; et al. Strelka2: Fast and accurate calling of germline and somatic variants. *Nat. Methods* **2018**, *15*, 591–594. [CrossRef] [PubMed]

64. Lai, Z.; Markovets, A.; Ahdesmaki, M.; Chapman, B.; Hofmann, O.; McEwen, R.; Johnson, J.; Dougherty, B.; Barrett, J.C.; Dry, J.R. VarDict: A novel and versatile variant caller for next-generation sequencing in cancer research. *Nucleic Acids Res.* **2016**, *44*, e108. [CrossRef]

65. Koboldt, D.C.; Zhang, Q.; Larson, D.E.; Shen, D.; McLellan, M.D.; Lin, L.; Miller, C.A.; Mardis, E.R.; Ding, L.; Wilson, R.K. VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res.* **2012**, *22*, 568–576. [CrossRef]

66. Robinson, J.T.; Thorvaldsdottir, H.; Winckler, W.; Guttman, M.; Lander, E.S.; Getz, G.; Mesirov, J.P. Integrative genomics viewer. *Nat. Biotechnol.* **2011**, *29*, 24–26. [CrossRef]

67. Tarasov, A.; Vilella, A.J.; Cuppen, E.; Nijman, I.J.; Prins, P. Sambamba: Fast processing of NGS alignment formats. *Bioinformatics* **2015**, *31*, 2032–2034. [CrossRef]

68. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R.; Genome Project Data Processing, S. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **2009**, *25*, 2078–2079. [CrossRef]

69. R Core Team. *R: A Language and Environment for Statistical Computing*; R Core Team: Vienna, Austria, 2022.