Peer Review File

Manuscript Title: Deep whole genome ctDNA chronology of treatment resistant prostate cancer.

Reviewer Comments & Author Rebuttals

Reviewer Reports on the Initial Version:

Referees' comments:

Referee #1 (Remarks to the Author):

This manuscript by Herberts et al describes an effort to elucidate the clonal architecture of metastatic tumors using whole-genome sequencing of circulating tumor DNA. The basis for this project comes is a series of plasma specimens from two different cohorts: 1) a subset of patients from the "West Coast Stand Up to Cancer Dream Team" in which the plasma sample was generally drawn at the same time ("synchronous") as the metastasis biopsy; and 2) a series of patients from a British Columbia Cancer Agency clinical trial, without matched tissue, and 1-3 plasma draws per patient.

Having performed deep whole-genome sequencing on these samples (plus matched control DNA), the authors describe multiple analyses of these data. These analyses include somatic mutation and copy-number number calls, which are presented as a function of whole-genome duplication event timing. From these data, the manuscript presents inferences of subclonal phylogenetic architecture for each WGS sample, taking into account multiple samples from the same individual as applicable. As these are all prostate cancer samples acquired from patients receiving androgen-receptor targeted therapies, the paper focuses on AR status as inferred by these data. The manuscript concludes with an analysis of fragmentomics, in which transcriptional activity can be inferred by nucleosome depletion near TSSs.

The major conclusions of this work are significant and reflect the novelty of the specimens analyzed, especially in a clinical context. 1) Although tissue and plasma have been sequenced, extensively, in prior studies, a major goal of those efforts has been to demonstrate the utility of ctDNA sequencing as a surrogate for analyzing solid tissue DNA. While it is well accepted that ctDNA reflects the entire pool of contributions from metastases, that assertion has never been challenged analytically before. The phylogenetic reconstruction approaches used in this paper identify hallmarks of genomic diversity (subclonal tumor populations) that differ between synchronous solid tissue and plasma, validating this model. 2) Whole-genome duplication is rigorously addressed in this manuscript, including how it's status as a confounder of copy number calls and subclonal reconstruction can be addressed. The very high depth WGS performed here allows for clonal "fit" models to by proposed and either accepted or rejected on a per-case basis. From this data, more accurate phylogenies could be conferred. This is a very significant finding. 3) The high depth WGS further enabled mapping AR-binding sites as a function of read depth and establishing a model for calling AR activity based on less nucleosomal-protected ctDNA at these sites. While these analyses are only possible for genome-wide transcription-factor activity and not suitable for gene-level analyses, this is the first time I've seen a continuous range of values inferred for TF binding (rather than calling presence or absence) from ctDNA (see Ulz et al Nat Comms 2019). Although the ability to predict neuroendocrine prostate cancer from adenocarcinoma is an important ability of fragmentomics, this was also performed in the prior study (Ulz et al) from seemingly lower depth WGS.

The major conceptual advances afforded by this work deal with how the authors managed the

confounding nature of whole-genome duplication when fitting each case to a clonal composition of ctDNA and inferring the subclonal architecture via discrete subclone-adjusted copy numbers. This is an extremely challenging task, both conceptually and technically, and the authors are to be commended for their efforts in both walking the reader through it, step-by-step, in a number of supplementary figures (S7-S13) as well as by showing visual depictions of data that are necessary for completing each iterative step. This level of transparency is unusual and much appreciated. However, the challenge with this approach, as well as the custom software written by the authors for performing these tasks, is that these methods are being used on the same samples the were tested against. What is missing is some form of orthogonal validation. For example, many manuscripts describing the utility of new phylogeny software (usually standalone publications) go into depth describing either simulated data or single-cell sequencing data (or both). For example, this approach was used for similar tools, such as PyClone, PhyloWGS and Phylogic. In this current manuscript, the only comparisons shown are benchmarks against ASCAT and Battenberg. This reveals <50% concordance for copy number fitting with ASCAT. While variance is to be expected, ASCAT is a highly cited and well-documented software tool. By contrast, the author's GitHub repository contains code for this manuscript that is neither well-commented or contains instructions for use. As a result, before accepting the authors' findings as truth, there is an additional burden to show an expected result from previously-analyzed data (such as single-cell or simulated data) and account for differences from competing tools. This issue is critical since all of the paper's conclusions rest on their novel analysis methodology.

In addition, there are a number of issues or opportunities for deeper analysis that should be addressed in a revision:

1) In Supplementary Table 1, the sites of "DTB" mCRPC biopsies do not match what is published in the NCI Genomic Data Commons for those samples. Please clarify which listing is erroneous.
2) One of the "DTB" samples listed in Supplementary Table 1 is "DTB-174" but only has a listing for ctDNA. There is no tissue biopsy from that patient?
3) Line 91: the p value for the correlation does not match what is in the figure.
4) Figure 1 color coding in panel D should switch the location of the "Signature" and "Tissue Type" boxes to match their order of appearance in the figure.
5) The mutational signatures presented in Figure 1d are presumed to be from bulk (not deconvolved) samples. What happens to these signatures if decomposed by subclone using a tool such as SigProfilerExtractor with TrackSig? If the decomposed signatures show significant variance, is there a better way to display this data?
6) The data and methodologies in Supplementary Figures 7, 9 and 10 are extensive but important. Showing these methods within the framework of a complete supplement might be better than some figures with legends. This way the authors are not limited for space and can provide much more detail than as limited by text boxes in a figure. Examples can be provided this way too.
7) As mentioned in the previous major comment, the phylogeny estimation approach developed for this work was not benchmarked against other approaches. What happens when other tools, such as Phylogic, PyClone, or PhyloWGS are used with these data? Do the same trees emerge? If not, perhaps the cases with tree structure uncertainty reveal potential sources of confounding mutation or copy number data that highlights the differences between approaches?
8) It would be helpful if samples are labelled in all figures. For example the identity of each case is not clear in Figure 2b. It would be helpful if they are shown in the same order as Supp Fig 9.
9) Why is Supplementary Fig 9 only shown for a subset of patients? The data detail shown is extremely informative.
10) Is the case shown in Supp Fig 10 a specific case?
11) Line 143: the MWU test P value shown does not match the figure.
12) Line 149: Why does WGD occurring multiple independent times in the same tumor indicate "convergent evolution"? Perhaps the fact that the same truncal mutations occur in both lineages within each patient set a genotype that is conducive or selects for WGD events?
13) Lines 164-167 and Figure 2h: Please annotate the figure to support the finding of AR gene alteration diversity within a patient. For example, markings to show which rows belong to the

same patient. In addition, how do you distinguish the timing of a mutation versus amplification for considering AR alterations in the context of a tumor phylogeny?
14) Line 186: "but amenable to reciprocal redetection" -- what does this mean?
15) Line 191: "In one patient" -- which one?
16) Line 194: "early truncal fixation" -- can you assess any truncal genotypes that are more or less associated with subclonal discordance (i.e. more diversity)?
17) Figure 3d -- what happens to discordance when you deconvolve the genotype of ctDNA? Can you resolve a ctDNA "subclone" to be highly concordant to the "bulk" metastasis alteration spectrum in a case that is discordant as currently presented? This would lend weight to the argument that cfDNA is comprised of multiple clones but that the synchronous metastasis is not the only clone. (Is there any clinical data from those cases to support a major metastasis at a site not biopsied?)
18) Line 201: The comment that AR genotypic diversity drives clonal expansion is in direct contraction to data presented in Espritu et al (Cell 2018) in which a subset of localized prostate cancers show clonal diversity. Of course, localized prostate tumors do not harbor AR alterations. Consider an alternative interpretation that tumors exploit alterations to the AR gene for driving progression with underlying genomic instability.
19) Line 212: AR mutations are not shown as highly recurrent in Figure 4.
20) Line 214: What is the data supporting ligand binding domain (LBD) mutation switching?
21) Figure 4b: "1L" means first line but I could not find this defined in the manuscript.
22) It is unclear how clonal dynamics of tumor subpopulations are estimated in Figure 4e. Curve smoothing can be used between samples and each corresponding proportion, but for case AE-035, how do you show the timing for subclone "drop out" at Day 100 when progression does not occur until day ~190? Also, AE-035 is shown in both Fig 4e and Supp Fig 15. Is this intentional?
23) Line 228: It would be helpful to clarify how the occupancy maps were generated at this point.
24) Line 235 - Supp Fig 16, what is "r" used and the statistical test used in 16a and 16b? Given the interval level data, Spearman correlation would be the most likely correlation method.
25) Line 237: Spearman correlation should be used instead of Pearson
26) Lines 237-241: Can you provide this source data?
27) Line 245: the weak correlation between Tissue mRNA abundance with cfDNA nucleosome depletion - is that supposed to be gray in Figure 5d? For 5d and 5e, Spearman rank correlations should be used for percentiles.
28) Line 246: Figure 5f does not show an "anticorrelation". Do you mean no association?
29) Line 258: Use Spearman correlation instead of Pearson, but if the goal is to show enrichment within a subset, you might find Fisher's Exact Test more useful.
30) Supp Fig 18 - Spearman only.
31) Line 268 - The data presented in 5i supports progression with PSA increasing, but there is loss of AR signaling in 5j. How can you rectify this contradiction?
32) Lines 291-294 and 331-333 are in contradiction to each other. The deep data provided in most of this manuscript was possible to due to the deep (and thus expensive) WGS. Targeted cfDNA or lower pass coverage will not be as informative for understanding the composition of tumors.
33) Line 297: If the ancestral tumor lineage is clinically important, does that mean that the tumor composition of ctDNA subclones reflects the clinical aggressivity of the tumor?

Referee #2 (Remarks to the Author):

The authors present a major advance in the field of ctDNA sequencing of prostate cancer samples. The use of Whole Genome Sequencing of ctDNA and concomitant sequencing of tissue samples is key to understand how these 2 sources of tumor derived DNA relate to each other. The manuscript has been well-written and is therefore an interesting paper for the readership of Nature.
My major remarks are:
1) The use of custom and/or unpublished and non-peer reviewed software and algorithms (without compar-isons to contemporary and often-used alternatives) complicates the validity of the presented results. It is not clear whether the tissue biopsies were examined using the same cfDNA

somatic variant calling, CNV and SV-detection pipelines. This could introduce some bias as there are well-suited standards for somatic variant-calling (from tissue) which could outperform the custom cfDNA-trained variant calling pipeline. A comparison of the custom cfDNA-based workflow vs. another validated and contemporary pipeline for tis-sue-samples (e.g., the GATK-workflow) on the tissue-samples could provide clarity of any potential biases of using the cfDNA-based workflow on tissue-samples. Any potential high-quality somatic variants detect-ed using this workflow in tissue and not detected in the matched cfDNA samples (using the cfDNA-based workflow) should also be shown as discrepant.
In addition, the references for the "validation" of this cfDNA-based somatic caller lack validation or com-parison against tissue-based somatic variant callers (line 839 - 840). The references given such as Van-dekerkhove et al. (2021) reference simply refers back to other publications ("Alignment and analysis of DNA-sequencing data were performed utilizing an established pipeline9,13,54,56") and then lists custom and arbitrary criteria again such as "Somatic mutations were required to be supported by a minimum of eight mutant allele reads, with a minimum VAF of 1% in ctDNA and 8% in tissue." Moreover, Annala et al. (2018) does also not perform a comparison against tissue biopsies. The thresholds used in these 3 manuscripts differ, and thus, the cited studies do not support the use of the current thresholds. These technical issues make the manuscript somewhat difficult to understand completely.
2) The conclusion that the tissue based sequencing may "misinform targeted treatment" seems a bit bold based on the data the authors present. To conclude this the authors need to present data on actual dis-crepant targetable genes. The loss or gain of mutation unique to ctDNA is limited to 2 AR variants (fig 3c) in the current manuscript making the basis in my view meager. The difference in genomic landscape may be obvious, to conclude ctDNA is better for making a diagnosis the heterogeneity must be relevant.
3) A major question remains whether indeed clonal architecture is driving the emergence of resistance and the data presented here do not resolve this issue. Clonal reshaping upon treatment has been suggested by fig4e and supplemental figure 15. The data presented seem to rely mainly on the estimation of copy number alterations. These are found indeed to drive the difference between ctDNA and tissue sequencing. The final conclusion whether these findings indeed represent true clonal expansion or adaptation by vari-able amplification of the X-chromosome will need to come from single cell analyses. Therefore, I would suggest to tone down the conclusion or to present more detailed data showing that apart from the AR CNV reshaping there is indeed strong evidence for clonal evolution.

Minor comments:
Figure 1:
♣ 1b: y-axis of timepoint is unclear (1st/2nd/3rd cfDNA), what is the (average) time between peripheral blood draws / cfDNA extraction per patient?
♣ 1d: Legend text does not match y-axis titles.
• E.g., cancer fraction in the legend whilst tumor content (%) is used in figure 1d, total mutation count vs. total variant counts etc.
♣ 1d - 2nd track: Please describe which variant categories underly "total variant count"/ " total mutation count"; is this the summation of somatic SNVs, InDels and MNVs? Could a separation be made to distinguish between each class to determine whether there is a change in SNVs/InDels ratio between cfDNA and tissue capture?
♣ 1d: Please highlight the paired samples (tissue + cfDNA) in 1d, this would ease interpre-tation on the conformality and differences between the tissue and cfDNA profiles per pa-tient. I.e., an additional track denoting (via color or symbol) each of the 13 matched pa-tients and whether the cfDNA was time-matched with the tissue.
♣ 1d - 7th track: >4 and non-integer is not mutually-exclusive. E.g., 4.1 is also >4 and it is unclear whether that will be colored darkred (>4) or grey (non-integer).
♣ 1d - Track 5,6 and 8: Missing y-axis ticks (i.e., 0%, 50%, 100%).
• M&M:
o Please specify whether algorithms/tools (e.g., bowtie2) was performed using default settings or specify the non-default settings.

o Please state how HHR, MMR and CDK12 deficiency status was determined.
o Sequence alignment:
♣ Were lanes (if applicable) merged with respective read-groups during alignment?
♣ The sentences are ordered as if the paired-end reads were trimmed using cutadapt and low-quality bases were removed after Bowtie2-alignment? This would be counterintui-tive and needed to be done prior to alignment to improve mapping accuracy and allow soft-clipping (of remaining artifacts/low quality bases) by Bowtie2. Please check whether this was the case or if the alignment-step should be placed after specifying the trimming M&M.
♣ An "in-house algorithm" was for base-quality trimming, is this code/algorithm publicly available?
♣ Please specify how the single-nucleotide polymorphism (SNP) genotype profiles compari-son between cfDNA and WBC was performed.
♣ Please specify which gene annotations were used during alignment (if applicable).
o Somatic variant calling
♣ Please specify which tools (and settings plus version-numbers) and strategy were current-ly used for germline and somatic variant calling. Currently, the authors only refer to pre-vious literature whilst this encompasses a significant technical element of the manuscript and subsequent results.
• A brief inspection of "mutato" reveals that this mostly consists out of pileups without further extensive modeling employed by most industry-standard alter-natives such as GATK Mutect2 / Strelka (for tissue).
♣ Please specify which ANNOVAR databases and gene-references (Refseq/UCSC etc.) were used to signify Protein-level consequences.
♣ Please specify if all mutational signatures (COSMIC v2) were used for optimal reconstruc-tion or if a restricted fitting was performed.
♣ Please state the version of SigProfilerMatrixGenerator.
o Structural variant calling
♣ Currently, a non-peer reviewed (unpublished) and relatively unknown software package (breakfast) is used for SV-calling instead of common software packages such as GRIDSS2 . manta / delly etc. The resulting SV calls could therefore be suspect as this is also coupled with an arbitrary threshold such as a min. coverage of 5 supporting junction-spanning reads. Could the authors show concordance of their employed method vs. another es-tablished methodology for SV-calling such as GRIDSS2.
♣ Tissue-cfDNA mutation concordance
• Is this based only on SNVs, InDels and MNVs or also on the SVs?
• The authors compared concordance of somatic mutations between time-matched cfDNA/tissue using different settings than when calling final list of so-matic variants per sample. Instead of using the (arbitrary) >1 reads threshold, it would be more relevant to overlap the final list of somatic variants as derived from cfDNA and tissue sequencing and then compare the overlap. The current analysis does not fully take into account the (potential) differences in detecting / calling somatic variants between cfDNA and tissue (with matched normal) se-quencing and somatic variant calling strategies.
♣ Germline variant calling
• Please specify which tools (or in-house software) was used.
• Please specify the GNOMAD version used (and add the respective reference).
• Please specify which ANNOVAR databases and gene-references (Refseq/UCSC etc.) were used to signify Protein-level consequences.
• Please specify the ClinVar database (date of access).

Referee #3 (Remarks to the Author):

Herberts et al analyzed deep whole genome sequencing (WGS) data from plasma cell-free DNA (cfDNA) of patients with metastatic castrate-resistant prostate cancer (mCRPC) and a high circulating tumor DNA (ctDNA) fraction. They created a comprehensive framework for reconstructing subclonal architectures and tumor phylogenies from cfDNA. They found that ctDNA

has high subclonal diversity with often a small contribution from a macrometastatic site of disease. Through serial samples over time, they showed the AR locus to the only site of recurrent alterations, reinforcing its critical importance for resistance to AR signaling inhibitors. Additionally, by using cfDNA fragmentation patterns, the authors were able to predict the transcriptomic landscape of the corresponding biopsied site of disease. Taken together, the authors demonstrated that deep WGS of cfDNA can be a powerful tool for understanding the global evolutionary history, state, and trajectory of mCRPC in individual patients, which may be of significant biological relevance.

This is a comprehensive, relevant, thoughtful, and interesting piece of work from an outstanding research team. To my knowledge, this is the first study to utilize deep, serial WGS of plasma cfDNA to study metastatic prostate cancer (and perhaps in any solid tumor, though I am unfamiliar if similar work has been done in other diseases (e.g. metastatic breast cancer)).

Through analyzing the genomic alterations observed in ctDNA and applying an evolutionary framework to study subclonal diversity, the authors were able to provide novel insights into the diverse evolutionary trajectories of mCRPC. The study would be strengthened by providing understanding on what the biological/translational relevance of these findings are relative to prior studies in this general domain, and how this approach may fit in compared to other profiling techniques. My comments and suggestions are provided below:

Major comments
-While the authors do develop a comprehensive framework for inferring subclonal architecture and phylogeny from ctDNA and apply it extensively in this study, they notably do not provide validation of their method (outside of a comparison of their copy number model to ASCAT and Battenberg). It would be informative to see the performance of their method when applied to samples with a known ground truth regarding its subclonal status. Understandably, this may be challenging to do experimentally, though it should be possible in silico through simulated data (e.g. using MASCoTE, PMID 32879317). Alternatively, explaining the limitations of knowing "ground truth" for this study would be helpful for the reader.
-The authors assert that "the variable contribution to total ctDNA from synchronous metastases challenge the notion that global disease features can be precisely inferred through tissue biopsy" (lines 299-300). However, the authors show that, in their 13 pairs of ctDNA with same-day tissue biopsy, there is a high degree of concordance in the mutations detected between cfDNA and biopsy, with only 9 discordant mutations which were protein truncating or in known hotspots out of 94,675 detected in total. Given potential reasons for false negatives (e.g. variability in tumor content and sequencing depth), it is conceivable that the number of discordant, possibly functionally important mutations may be even lower than 9. Additionally, there was also high concordance in the copy number status of key driver genes between ctDNA and biopsy (lines 182-194). Lastly, in paired patient samples, the authors found that the transcriptional landscape of the biopsy site to be strongly correlated with the magnitude of the transcription start site (TSS) nucleosome depletion, suggesting that the transcriptomics of the biopsied site strongly recapitulates the overall transcriptomic landscape of disease in the patient. Hence, it seems that while a single biopsy may not capture all global disease features, it does capture a very large portion of them, particularly those of likely functional (and clinical) importance. Furthermore, as noted in Figure 3F, biopsy of a metastatic site may sometimes be an earlier indicator than cfDNA of important molecular alterations—in this case, AR gene body and AR enhancer copy gains. The authors could consider softening their claim here, especially since the primary clinically relevant event is AR locus related that is a perpetual source of therapeutic development strategies. More broadly. can the authors comment more on the biological or clinical relevance to distinguishing "precise" disease features vs. those identified from a single tumor biopsy, especially if they ultimately point to AR (albeit with subtleties about the specific molecular events in AR)?
-Through analysis of serial, patient-matched samples, the authors found the AR locus to be the only of recurrent alterations. However, their methods for calling somatic mutations vary for AR vs non-AR regions, with the former requiring variants with ≥8 supporting reads and a variant allele

fraction (VAF) of ≥10%, while for AR gene hotspot mutations, they only required 3 mutant reads, VAF >1%, and ≥15 times background error rate. It would be informative to see if the AR locus still remains the only site of recurrent alterations if the conditions for calling non-AR somatic alterations were relaxed to be comparable to those used for calling AR alterations.
-Regarding the central role of AR locus alterations, how do the authors reconcile this observation with existing and emerging combination therapeutic strategies that leverage other means of inhibition (e.g. AR + PI3K inhibitors)? Similarly, framing the overarching findings from ctDNA WGS relative to prior autopsy/multiregional or serial tissue based studies (e.g. Kumar et al Nature Med 2016, Gundem et al Nature 2015) would be helpful for the reader.
-Regarding Figure 5E, it is surprising that there exists such a strong correlation between cfDNA TSS nucleosome depletion and gene expression for unmatched cfDNA and metastatic site biopsy. I wonder how much of this correlation may be because cancers in general up/downregulate some common transcriptional programs. As a negative control, it would be helpful to see a graph of cfDNA TSS nucleosome depletion of this cohort vs gene expression percentile for different, unrelated carcinomas, e.g. bladder cancer.
-It is noteworthy that the median ctDNA fraction in the study cohort was 48% (range 17-88%), and many of the methods applied here require a high ctDNA fraction. However, as demonstrated previously, the ctDNA fraction in mCRPC can be quite low (e.g. in the abiraterone◊enzalutamide vs enzalutamide◊abiraterone phase 2 trial which comprised part of the patient population here, 43% of the cohort had ctDNA fraction as <2%, and another 36% had 2-30%; PMID 29367197). As such, while the authors contend that "our work advances ctDNA profiling from an emerging tool for detection of selected clinically-actionable gene mutations, towards a modality of genome-scale discovery and deep clinical-biological insight," it is unclear whether deep WGS can be feasibly applied to the study of cancer in clinical states where the ctDNA fraction is low, which represents the majority of mCRPC patients. Can the authors reframe this point?

Minor comments
-Lines 150-152: Can the authors provide citations on methods that behave this way? This would be helpful for the reader
-Figure 1E: Can the authors provide a definition of the subclonal CN diversity score?
-Figure 2C: Can the authors provide a definition of copy segments?
-Figure 3D: For the two patients with a high %ctDNA contribution (patient 183 and 205), did the authors consider reasons why this may be? I am curious whether this is a function of a low overall disease burden (hence a single macrometastatic site of disease contributes a significant amount to the ctDNA subclonal diversity) or whether there is a high degree of homogeneity across metastatic sites of disease in the patient.
-In Figure 3E, the authors quantify the average absolute copy number difference between biopsy and cfDNA. It would be interesting to see the signed difference in copy number—is it that the biopsy site tends to more likely have a copy number gain compared to the cfDNA, which is a composite of many sites of disease in the patient?
-Figure S1D: The y-axis label for the middle row should not have the % sign. It should read "Heterozygous SNP allele fraction."
-Figure S10: Which patient is being used as an example here?
-Lines 244-245: The author state "Tissue mRNA-abundance was weakly correlated with cfDNA nucleosome depletion in the first exon-intron junction." Can the authors explain this statement further since it appears that r=0.98 for this correlation?
-Line 830-831: Do the authors mean 5, rather than 4, non-cancerous cfDNA samples?
-Line 836: What is considered the background error rate in this case?

Author Rebuttals to Initial Comments:

Referees' comments:

Referee #1 (Remarks to the Author):

This manuscript by Herberts et al describes an effort to elucidate the clonal architecture of metastatic tumors using whole-genome sequencing of circulating tumor DNA. The basis for this project comes is a series of plasma specimens from two different cohorts: 1) a subset of patients from the "West Coast Stand Up to Cancer Dream Team" in which the plasma sample was generally drawn at the same time ("synchronous") as the metastasis biopsy; and 2) a series of patients from a British Columbia Cancer Agency clinical trial, without matched tissue, and 1-3 plasma draws per patient.

Having performed deep whole-genome sequencing on these samples (plus matched control DNA), the authors describe multiple analyses of these data. These analyses include somatic mutation and copy-number number calls, which are presented as a function of whole-genome duplication event timing. From these data, the manuscript presents inferences of subclonal phylogenetic architecture for each WGS sample, taking into account multiple samples from the same individual as applicable. As these are all prostate cancer samples acquired from patients receiving androgen-receptor targeted therapies, the paper focuses on AR status as inferred by these data. The manuscript concludes with an analysis of fragmentomics, in which transcriptional activity can be inferred by nucleosome depletion near TSSs.

The major conclusions of this work are significant and reflect the novelty of the specimens analyzed, especially in a clinical context. 1) Although tissue and plasma have been sequenced, extensively, in prior studies, a major goal of those efforts has been to demonstrate the utility of ctDNA sequencing as a surrogate for analyzing solid tissue DNA. While it is well accepted that ctDNA reflects the entire pool of contributions from metastases, that assertion has never been challenged analytically before. The phylogenetic reconstruction approaches used in this paper identify hallmarks of genomic diversity (subclonal tumor populations) that differ between synchronous solid tissue and plasma, validating this model. 2) Whole-genome duplication is rigorously addressed in this manuscript, including how it's status as a confounder of copy number calls and subclonal reconstruction can be addressed. The very high depth WGS performed here allows for clonal "fit" models to be proposed and either accepted or rejected on a per-case basis. From this data, more accurate phylogenies could be conferred. This is a very significant finding. 3) The high depth WGS further enabled mapping AR-binding sites as a function of read depth and establishing a model for calling AR activity based on less nucleosomal-protected ctDNA at these sites. While these analyses are only possible for genome-wide transcription-factor activity and not suitable for gene-level analyses, this is the first time I've

seen a continuous range of values inferred for TF binding (rather than calling presence or absence) from ctDNA (see Ulz et al Nat Comms 2019). Although the ability to predict neuroendocrine prostate cancer from adenocarcinoma is an important ability of fragmentomics, this was also performed in the prior study (Ulz et al) from seemingly lower depth WGS.

The major conceptual advances afforded by this work deal with how the authors managed the confounding nature of whole-genome duplication when fitting each case to a clonal composition of ctDNA and inferring the subclonal architecture via discrete subclone-adjusted copy numbers. This is an extremely challenging task, both conceptually and technically, and the authors are to be commended for their efforts in both walking the reader through it, step-by-step, in a number of supplementary figures (S7-S13) as well as by showing visual depictions of data that are necessary for completing each iterative step. This level of transparency is unusual and much appreciated. However, the challenge with this approach, as well as the custom software written by the authors for performing these tasks, is that these methods are being used on the same samples the were tested against. What is missing is some form of orthogonal validation. For example, many manuscripts describing the utility of new phylogeny software (usually standalone publications) go into depth describing either simulated data or single-cell sequencing data (or both). For example, this approach was used for similar tools, such as PyClone, PhyloWGS and Phylogic. In this current manuscript, the only comparisons shown are benchmarks against ASCAT and Battenberg. This reveals <50% concordance for copy number fitting with ASCAT. While variance is to be expected, ASCAT is a highly cited and well-documented software tool.

By contrast, the author's GitHub repository contains code for this manuscript that is neither well-commented or contains instructions for use.

We admit that the code repository for the initial submission was sparse and not sufficiently documented. For our revised submission, we have added complete code and documentation describing the methodology used in the manuscript (https://github.com/annalam/cfdna-wgs-manuscript-code**), including simulated test data for experimenting with our algorithms.**

As a result, before accepting the authors' findings as truth, there is an additional burden to show an expected result from previously-analyzed data (such as single-cell or simulated data) and account for differences from competing tools. This issue is critical since all of the paper's conclusions rest on their novel analysis methodology.

We appreciate your detailed and thoughtful appraisal of our work. To help address your comments, we have performed extensive additional validation and benchmarking of all aspects of

our initial submission using established, state-of-the-art public software. Summarizing these changes (as it pertains to your specific comments), this includes:

- **Validation of our subclonal reconstruction results using PyClone and PhyloWGS across all patients in our cohort.**
- **Comprehensive model-fitting visualizations for all cfDNA and tissue samples in our cohort (including truncal copy number fitting, SNV CCF clustering and subclonal reconstruction, and deconvolution of per-patient subclonal copy-number landscapes).**
- **Step-by-step methodological breakdowns (with integrated visualizations) of our truncal copy number fitting, subclonal reconstruction process, and inference of whole-genome duplication.**
- ***In silico* validation of our subclonal reconstruction approach via blinded reconstructions of the clonal compositions of 22 simulated patients (2 WGS samples per patient) with known subclone constitutions and genomic profiles.**
- **New visualizations showing a comparison of our truncal copy number model and the Battenberg copy number model for all samples where the two approaches disagreed.**
- **Further exploration of temporal and inter-subclone variance in mutation signature composition using DeconstructSigs, TrackSigFreq, and SigProfileExtractor.**
- **A complete code supplement describing the methodology used in the manuscript (https://github.com/annalam/cfdna-wgs-manuscript-code).**

In total, our new validation results and data/methodology descriptions occupy over 40 pages of additional text, 8 new Supplementary Tables, and 205 pages of supplementary data (see Supplementary Information document). Unprocessed sequencing data is also available to researchers through the EGA. Collectively, we believe that your comments have helped substantially elevate the methodological rigor, reproducibility, and ultimate scientific value of our revised work. Below we offer a point-by-point response to your comments (expanding on the broad analyses summarized above).

In addition, there are a number of issues or opportunities for deeper analysis that should be addressed in a revision:

1) In Supplementary Table 1, the sites of "DTB" mCRPC biopsies do not match what is published in the NCI Genomic Data Commons for those samples. Please clarify which listing is erroneous.

We confirm that the anatomical metastatic biopsy location sites listed in Supplementary Table 1 for the DTB patients are correct. The NCI Genomic Data Commons annotation is erroneous. The relevant study chair(s) have been notified and are working toward correcting the NCI Genomic

Data Commons information. Note that we have now significantly expanded the detail of clinico-genomic annotation contained within Supplementary Table 1.

2) One of the "DTB" samples listed in Supplementary Table 1 is "DTB-174" but only has a listing for ctDNA. There is no tissue biopsy from that patient?

Unfortunately sequencing library construction failed for the matched tissue biopsy of this patient. We decided to still include the cfDNA sequencing data in the manuscript given our focus on genome-wide analyses of cfDNA samples. We have now added the DTB-174 tissue sample to Supplementary Table 1, with a note indicating that its sequencing failed.

3) Line 91: the p value for the correlation does not match what is in the figure.

Thank you for spotting this typo—we have corrected this such that the Results text and Figure are now concordant. We have also updated this statistic (and statistics contingent on these data) to reflect several minor improvements in our truncal copy number fitting and WGS tumor content estimates.

4) Figure 1 color coding in panel D should switch the location of the "Signature" and "Tissue Type" boxes to match their order of appearance in the figure.

We have now implemented this change—thank you.

5) The mutational signatures presented in Figure 1d are presumed to be from bulk (not deconvolved) samples. What happens to these signatures if decomposed by subclone using a tool such as SigProfilerExtractor with TrackSig? If the decomposed signatures show significant variance, is there a better way to display this data?

Thank you for the suggestions! We have revised our manuscript to include a more granular analysis of signature differences in subclonal populations.

Figure 1d displays mutational signatures inferred from bulk (i.e. non-deconvoluted) sequencing data. We designed Figure 1d to provide a broad overview of per-sample genomic information, and

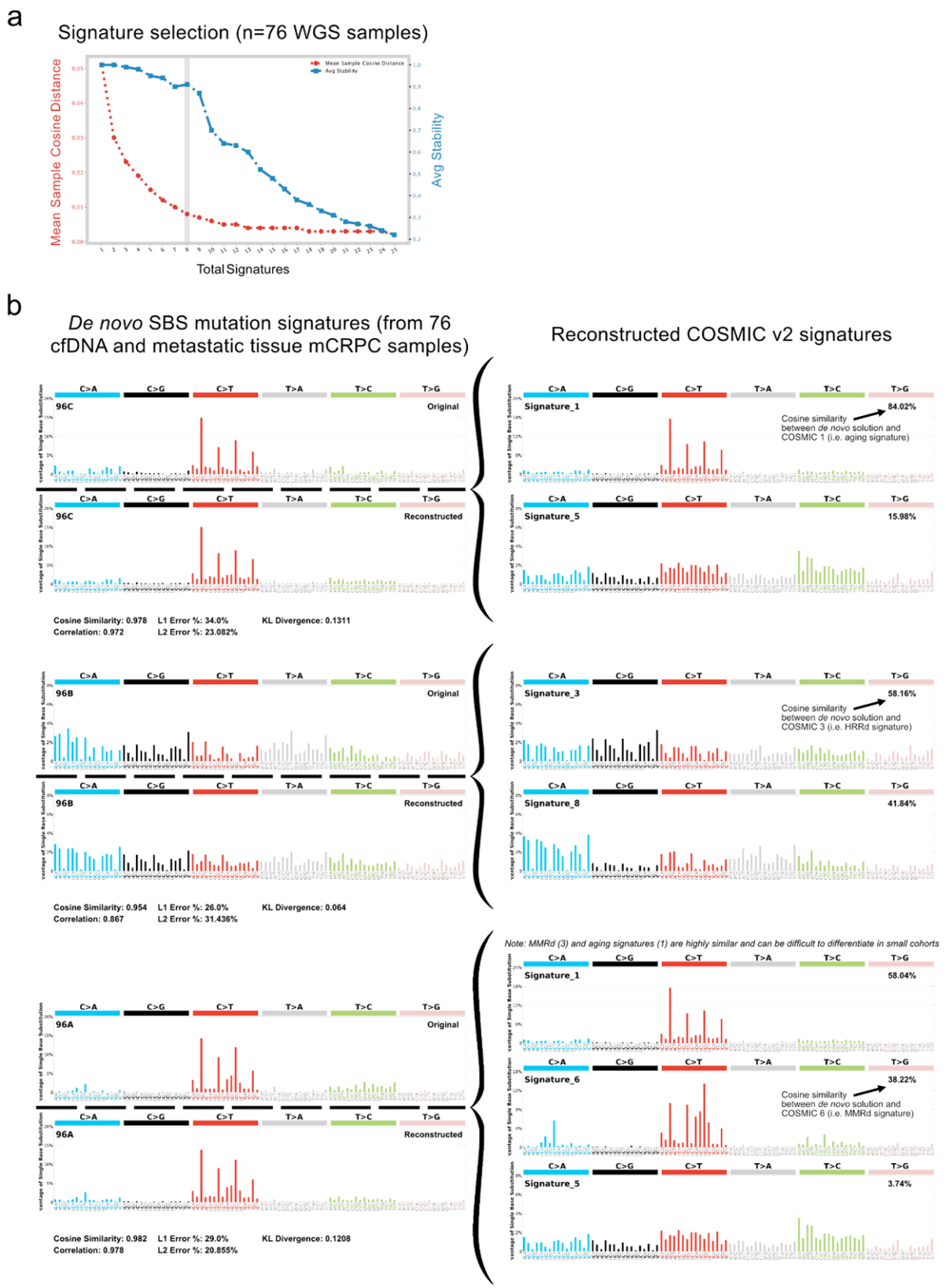
therefore elected to exclude annotation contingent on results explored in detail later in the manuscript (including subclonal reconstruction). We have revised the Figure 1 legend to clarify that the signatures shown are from bulk data (new text is underlined):

"... copy number distribution (as genome fractions), select SNV mutational trinucleotide signature spectra from bulk sequencing data (signatures with weights <0.15 are grouped as 'Other')..."

Subsequently in Figure 2, we explore differences in signature composition between truncal and subclonal populations (including signature differences between pre- and post- truncal whole-genome duplication). Our analyses focus on signatures relevant for the majority of prostate cancers (i.e. the mitotic clock-like 'aging' COSMIC signature 1), as well as DNA damage repair signatures (i.e. *BRCA2* and mismatch-repair deficiency associated signatures) in select samples harboring putatively causal defects.

Reflecting your suggestions, we have performed several new analyses to further characterize possible variance in signature composition across subclonal populations—as well as offer orthogonal validation of both new and original data:

1. **To verify that the COSMIC V2 mutational signatures (which were originally inferred primarily from pan-cancer tissue-based WGS) are appropriate for application in prostate cancer ctDNA, we applied SigProfileExtractor (https://github.com/AlexandrovLab/SigProfilerExtractor) across all bulk cfDNA and metastatic tissue sequencing samples to generate *de novo* mutation signatures. The most prevalent *de novo* signature was characterized by frequent C>T transitions and appeared overall highly similar to the COSMIC 1 aging signature (cosine similarity: 0.84). Signatures associated with homologous recombination repair and mismatch-repair deficiency were also recapitulated (see Figures below). Prior publications (in other cancers but also in mCRPC tissue, i.e. (Quigley et al., 2018) (PMID: 30033370) which includes 13 of the same patients analyzed in our current submission) have implied that the majority of *de novo* signatures inferred from single cancer-type bulk NGS data strongly recapitulate established signatures extracted from larger pan-cancer efforts (Alexandrov et al., 2020). Our analysis using SigProfileExtractor in mCRPC ctDNA and metastatic tissue broadly reaffirms this observation and supports our use of COSMIC signatures to characterize our samples (especially insofar as characterizing the select few dominant mutagenic processes present in prostate cancer which are widely established).**



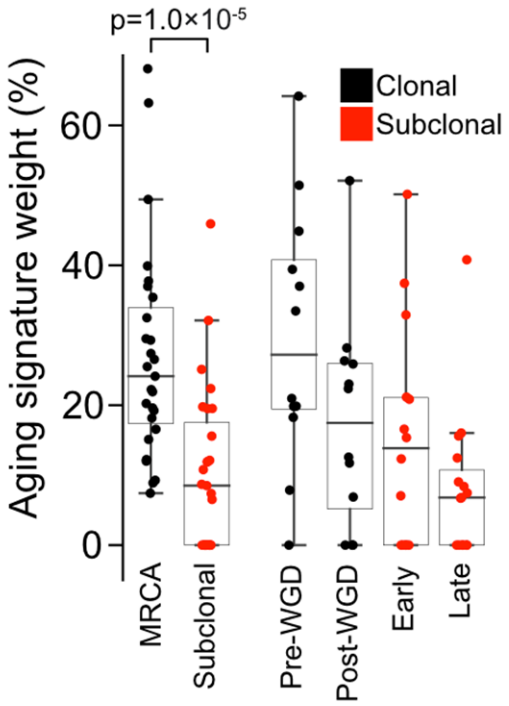
(a) Derivation of optimal number of *de novo* signatures via SigProfileExtractor applied to bulk sequencing data from 76 whole-genome ctDNA and metastatic tissue mCRPC samples (b) The three most dominant *de novo* single-base substitution (SBS) mutation signatures inferred from our cohort. All *de novo* signatures (trinucleotide context plots shown left) broadly recapitulate established COSMIC signatures with known activity in prostate cancer (reconstructions shown right).

2. Next, we applied DeconstructSigs (leveraging matrix-factorization to assign fitted signature fractions of the established COSMIC V2 signatures) to each separate per-patient subclonal population (i.e. including each colored segment of a phylogenetic tree in Figure 2b), rather than all subclones binned together as a single entity (as originally displayed in Figure 2f). This enabled the following new analyses:

    a. We first tested the hypothesis that intra-patient subclones harbor significant variance in fitted signature weights. For each DeconstructSigs COSMIC signature, we first calculated the variance in signature fraction across same-patient subclones and then calculated a weighted average of variances over all patients (i.e. a pooled variance). Weights were equal to the number of same-patient subclones. The median pooled variance across all COSMIC signatures was $\sigma^2 = 9.9\times10^{-4}$ (the mean signature fraction is $\bar{x} = 2.9\times10^{-2}$ across all subclones). Low inter-subclone variance ($\sigma^2 = 5.0\times10^{-3}$) was also observed even after exclusively retaining signatures with nonzero fraction in at least one same-patient subclone (whose prior inclusion would *decrease* overall variance). Finally, inter-subclone variance appeared low for COSMIC signatures 1 (aging: $\sigma^2 = 2.4\times10^{-2}$), 3 (homologous recombination repair deficiency: $\sigma^2 = 3.1\times10^{-2}$), 6 and 21 (mismatch-repair associated: $\sigma^2 = 8.6\times10^{-3}$ and $\sigma^2 = 4.2\times10^{-3}$, respectively), which constitute the main focus of the data originally presented in Figure 2. Collectively, these tests indicate that there is relatively minimal variance in signature fraction between individual same-patient subclones. Manual inspection of Supplementary Table 6 (containing all per-subclone signature fractions) enables intuitive appreciation of this observation.

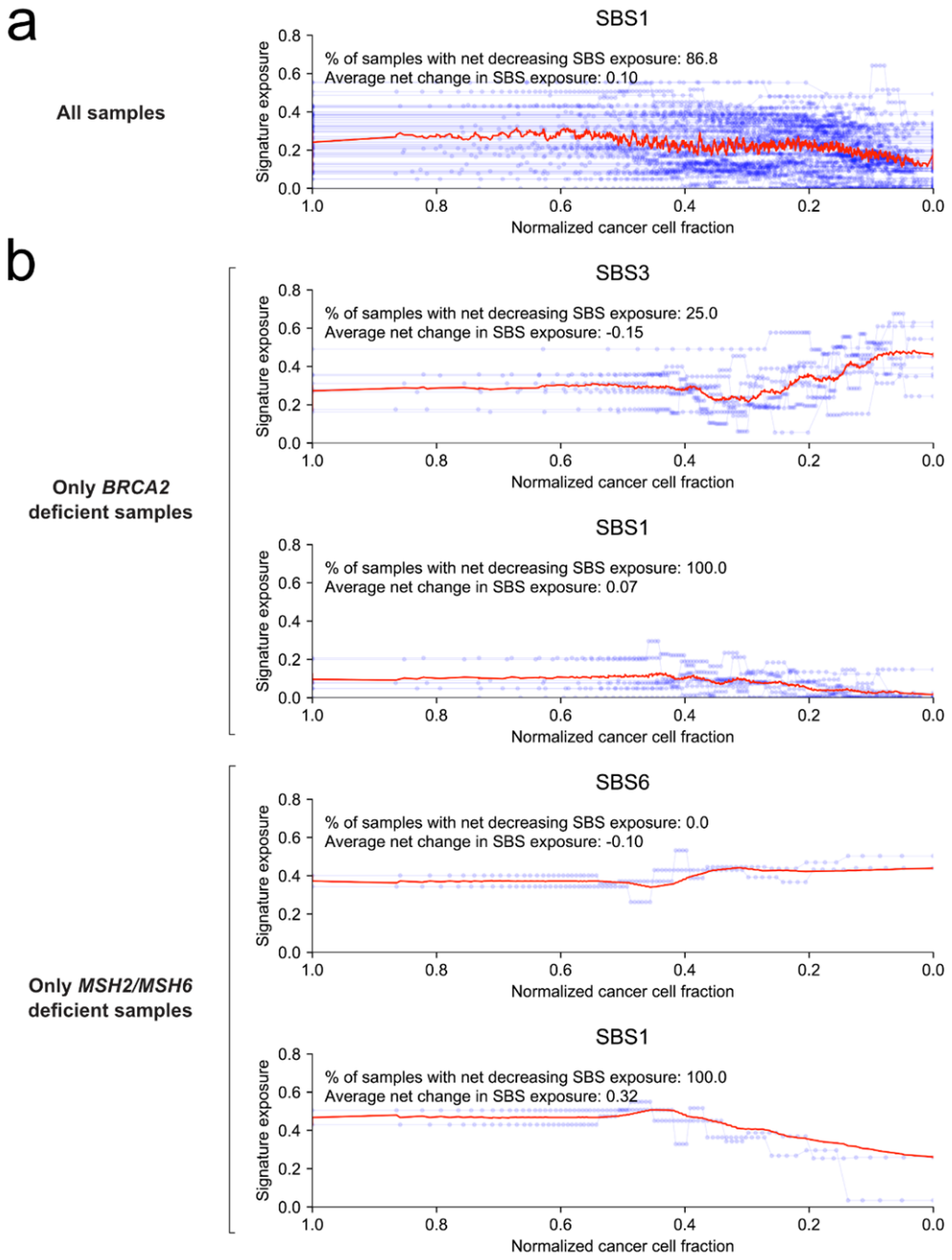
    b. We have revised Figure 2f to show additional biological granularity (also shown below). Instead of dichotomizing mutations into two broad categories (truncal and subclonal) as per our original submission, we further dissected the subclonal mutations by their relative evolutionary timing (inferred from multisample subclonal reconstruction and phylogenetic tree construction shown in Figure 2b). Specifically, we compare the DeconstructSigs COSMIC signature weight composition across truncal mutations (pre-WGD and post-WGD), early-subclonal mutations (non-truncal branch populations excluding leaf nodes), and late-subclonal mutations (leaf nodes) (the latter two being new categories). Although these two new evolutionary epochs (early- and late-arising subclones) show no statistically significant differences in COSMIC signature 1 weights (compatible with the small inter-subclone signature variances shown above), we did observe an overall monotonic decrease in median aging signature weight across the four temporally-ordered evolutionary categories (truncal pre-WGD, truncal post-WGD, early-subclonal, late-subclonal). This latter overall trend was statistically significant, and a new line in the manuscript Results has been added to reflect this observation. We also preserved the original 'subclonal' category (that does not

distinguish between early and late subclones) since not all patients had subclones amenable to further temporal dissection (e.g. if they only harbored a single subclonal cluster).

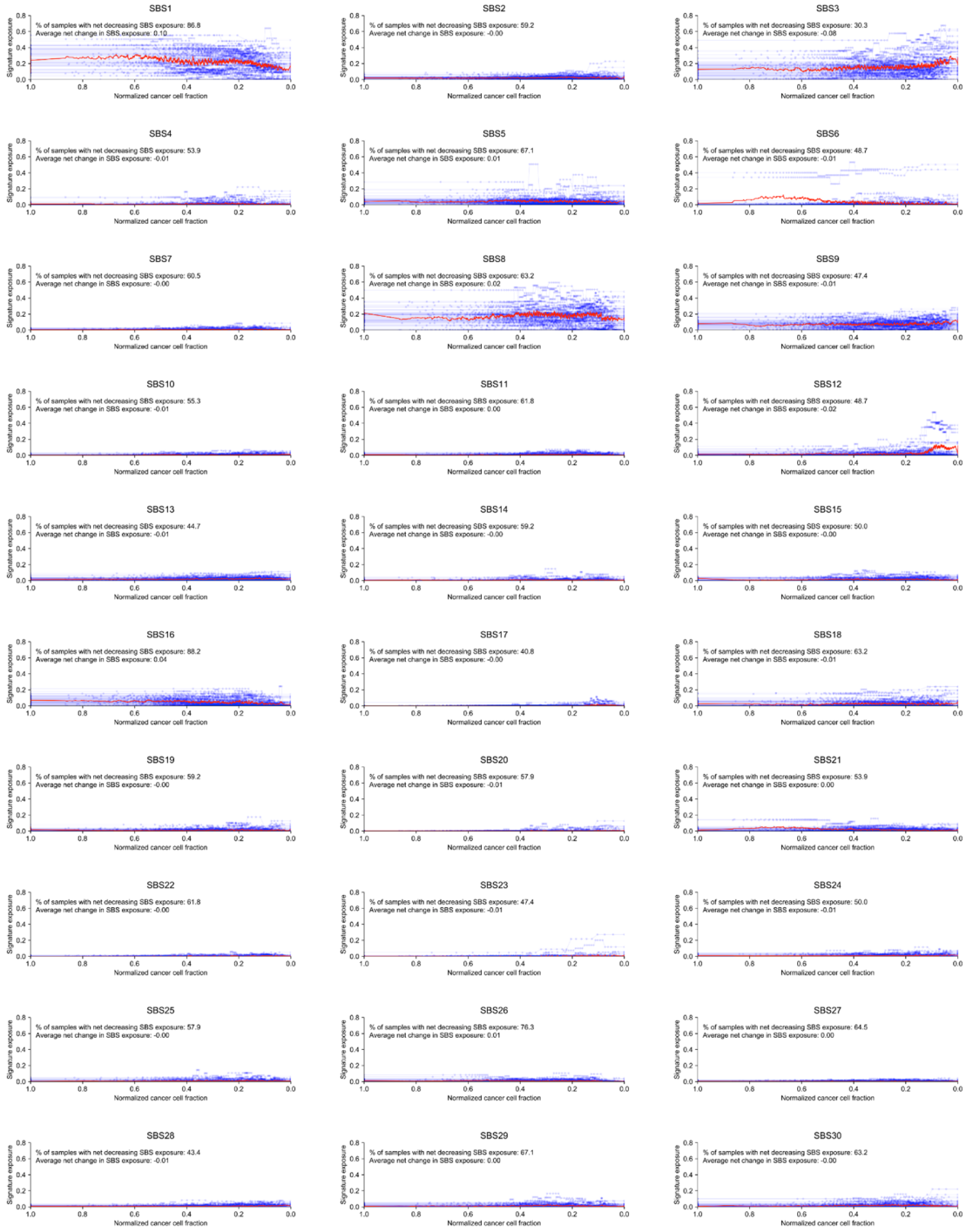


Differences in DeconstructSigs COSMIC v2 aging signature weights for the truncal most-recent common ancestor (MRCA) versus subclonal populations (left), as well as across phylogenetically ordered evolutionary epochs (i.e. truncal pre whole-genome duplication (WGD), truncal post-WGD, early subclonal populations, and late subclonal populations)

3. To validate our results showing mutation signature changes over evolutionary epochs, we ran TrackSigFreq (https://github.com/morrislab/TrackSigFreq) on each individual sample in our cohort (using the COSMIC V2 signature catalog). Note that TrackSigFreq is the updated version of TrackSig. Broadly speaking, TrackSigFreq orders mutations in descending order of their cancer clonal fraction (i.e. VAF corrected for copy-number state and tumor purity), using CCF as a surrogate for evolutionary time. A key feature of TrackSigFreq (and its predecessor TrackSig) is that it does not require prior mutation CCF clustering (excluding this as a potential source of technical error). We used our own allele-specific copy number deconstruction as an input for TrackSigFreq and excluded regions of subclonal copy number heterogeneity.
   In nearly all samples (87%), TrackSigFreq shows a generally consistent decrease of the COSMIC 1 signature over evolutionary pseudotime. Similarly, in all samples with DNA repair defects, the signatures associated with these subtypes uniformly increased over evolutionary pseudotime. This new analysis (shown in our new Supplementary Information document as Supplementary Figure 21-22 and copied below for the Reviewer's convenience) offers orthogonal validation of our original and new data now displayed in Figure 2f.



Supplementary Figure 21. Aggregated TrackSigFreq signature weight trajectories in (**A**) all 76 cfDNA and metastatic tissue WGS samples in our cohort and (**B**) the subset of samples with known DNA repair deficiencies. Only select mutational signatures are displayed (the mitotic 'aging' signature as well as signatures associated with DNA repair deficiencies in samples with putatively causal defects). Blue lines show individual sample trajectories (where dots indicate signature exposure changepoints). To visualize all per-sample trajectories on the same horizontal axis, individual per-sample cancer cell fraction (CCF) values were minmax normalized to fall between [0,1]. Red lines show the rolling average signature trajectory.



Supplementary Figure 22. Aggregated TrackSigFreq signature weight trajectories for all 76 cfDNA and metastatic tissue WGS samples in our cohort shown for all 30 COSMIC v2 signatures.

A limitation of TrackSigFreq is that it operates on individual samples and must therefore assume that mutation CCF is inversely proportional to evolutionary time (i.e. higher CCF corresponds to mutations acquired early, whereas lower CCF corresponds to mutations

acquired more recently). This assumption is not necessarily always true, since recently differentiated subclones can become the dominant population in clonally expanded lesions. For example, if mutation clusters α and β in Sample 1 harbor CCFs of 0.2 and 0.3, but in Sample 2 (from the same patient) harbor CCFs of 0.3 and 0.2, analysis of each sample separately (via TrackSigFreq) would 'time' the α and β mutations oppositely (as β → α and α → β, respectively). However, multi-sample reconstruction would indicate that clones α and β violate the crossing rule and cannot be descendents of one another, rendering both single-sample interpretations incorrect. We observe evidence of this phenomenon from our multi-sample subclonal reconstruction, where some mutation clusters have a low CCF in one sample but a high CCF in another sample (Figure 2; Supplementary Information section "Patient truncal copy number models and subclonal reconstructions").

Therefore, we have elected to use our approach of phylogenetically-aware partitioning of subclonal clusters (e.g. Figure 2f) as our primary data and leverage TrackSigFreq for supplementary orthogonal validation (rather than the other way around).

Detailed methodology accompanying these new results are included in our Supplementary Information document (see "Validation of evolutionary changes in mutational signature activity using TrackSigFreq" subsection, page 38).

6) The data and methodologies in Supplementary Figures 7, 9 and 10 are extensive but important. Showing these methods within the framework of a complete supplement might be better than some figures with legends. This way the authors are not limited for space and can provide much more detail than as limited by text boxes in a figure. Examples can be provided this way too.

Thank you for this useful suggestion. We have now added a complete supplement describing the subclonal reconstruction methodology in considerable detail (with integrated visualizations and step-by-step walkthroughs). The supplement also includes copy number models, subclonal reconstructions, and phylogenetic trees for all samples in the cohort, collectively comprising 246 pages of new material. Although Supplementary Figures 7 - 12 in our original submission focused on phylogenetically or genomically complex samples, most samples have simpler models, as is evident in the new comprehensive Supplementary Information document.

We have also extended the original Supplementary Figure 10 so that it now also shows our analysis of the subclonal WGD event in patient DTB-119 (the original figure only showed our analysis for AE-018). Furthermore, we have simplified and clarified the mathematical analysis that we used for inferring the cancer fraction and subclone CCFs in AE-018 and DTB-119 samples that

were admixtures of populations with differing WGD status. These updated results can be found in section "Detection of subclonal whole genome duplication events" of our new Supplementary Information document.

7) As mentioned in the previous major comment, the phylogeny estimation approach developed for this work was not benchmarked against other approaches. What happens when other tools, such as Phylogic, PyClone, or PhyloWGS are used with these data? Do the same trees emerge? If not, perhaps the cases with tree structure uncertainty reveal potential sources of confounding mutation or copy number data that highlights the differences between approaches?

Thank you for the suggestion. We have now compared the subclones identified by our analysis against those identified by PyClone (https://github.com/Roth-Lab/pyclone) and PhyloWGS (https://github.com/morrislab/phylowgs). The analysis is described in the new Supplementary Information section "Validation of subclonal reconstruction using PyClone and PhyloWGS". Side-by-side comparisons of the detected mutation clusters (i.e. subclones) are shown in the new Supplementary Information section "PyClone and PhyloWGS subclonal reconstruction validation data". In brief, PyClone and PhyloWGS confirmed the mutation clusters identified by our analysis, but sometimes explained a cloud of mutations as arising from two overlapping clusters, rather than the single cluster that our analysis deemed sufficient. We believe that these differences were caused by two primary factors:

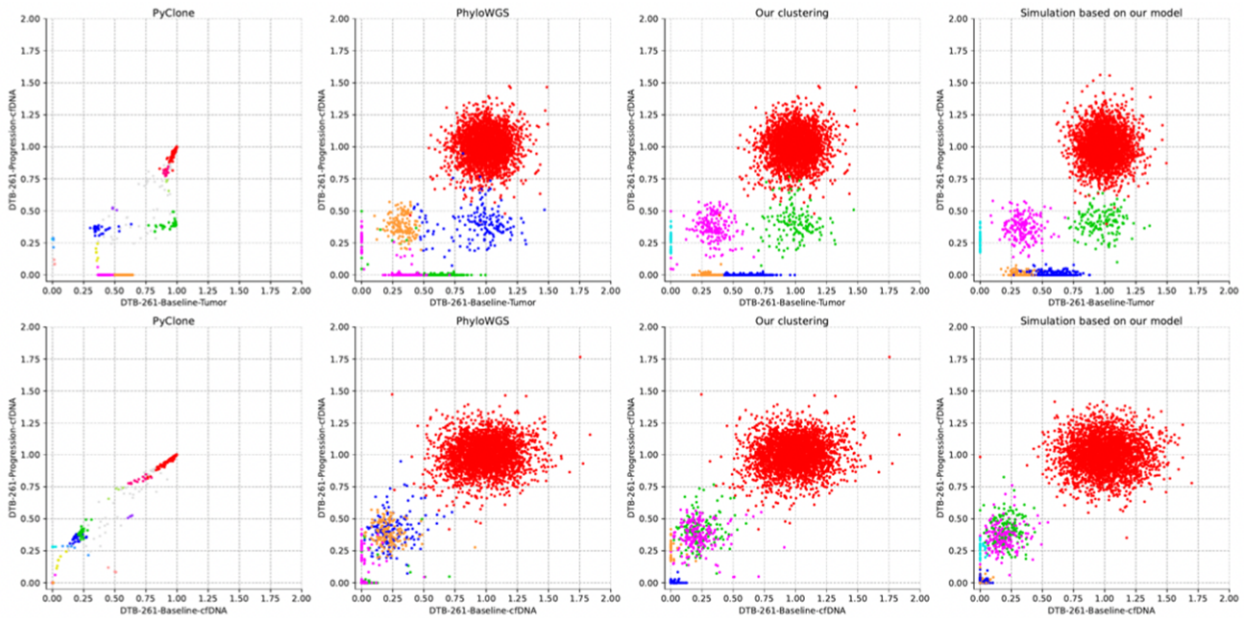
1) **Improper modeling of the cluster CCF variances (i.e. how wide the mutation clusters are expected to be along each dimension) by PyClone and PhyloWGS. Their documentation requests that somatic mutation allele fractions are given as the number of mutation-supporting reads and total overlapping reads. But in paired-end sequencing the individual reads are not all independent observations of the allele, since paired mate reads from the same original DNA template display the same allele. Therefore if one models the variance of the mutation clusters using a binomial distribution Bin(n, p) where n = total number of overlapping reads, the model will underestimate the variance in CCF space. This increases the likelihood that model-based clustering will model an observed mutation cluster as arising from multiple overlapping clusters. In our own mutation clustering analysis, we avoided this issue by using DNA fragment counts (rather than read counts).**

2) **Sporadic somatic mutations with CCF values falling on the diagonal between the truncal mutation cluster and the origin (CCF = 0). We found that many of these outlier mutations localized to genomic sites with challenging DNA sequences, resulting in alignment artifacts and inaccurate allele fraction estimation. But it has also been shown in literature that mutation clusters generally have an associated "tail" of mutations with increasing density towards the origin. These "tails" are thought to arise from neutral accumulation of mutations during the initial stages of clonal expansion (Caravagna et al., 2020). In other**

words, when a cell acquires a proliferative advantage and creates a new clonal expansion, the new mutations that accumulate during the earliest subsequent cell divisions will be present in a significant fraction of cells in the new population. These mutations "tails" (whether genuine biological events or technical artifacts) can theoretically result in the detection of spurious clusters, although in practice these mutations are often sparse.

To provide further support for our subclonal reconstructions, we augmented our side-by-side comparisons with PyClone and PhyloWGS with a fourth panel where we took the copy number and subclonal reconstruction models of each patient (i.e. subclones and their CCF in each sample), and used a generative physical simulation to investigate how the mutation CCF scatter plot was expected to look for such a patient. We sampled mutant allele fractions (MAF) from a binomial distribution to model the stochastic allele sampling that occurs when a random subset of circulating DNA fragments are analyzed, accounting for the copy number state and sequencing depth of each individual mutation. We then converted the MAFs into CCF values using our standard procedure. This analysis was able to recapitulate the observed mutation CCF scatter plots with high fidelity, demonstrating that our subclonal reconstructions provide a parsimonious explanation of the observed data. This analysis is described in Supplementary Information section "Validation of mutation CCF clusters using a generative simulation" (page 32).

The figure below shows an example side-to-side comparison of PyClone, PhyloWGS, and our mutation clustering, together with the generative simulation (based on our model) in the fourth column. A high resolution version of this figure can be found on page 227 of the Supplementary Information document.

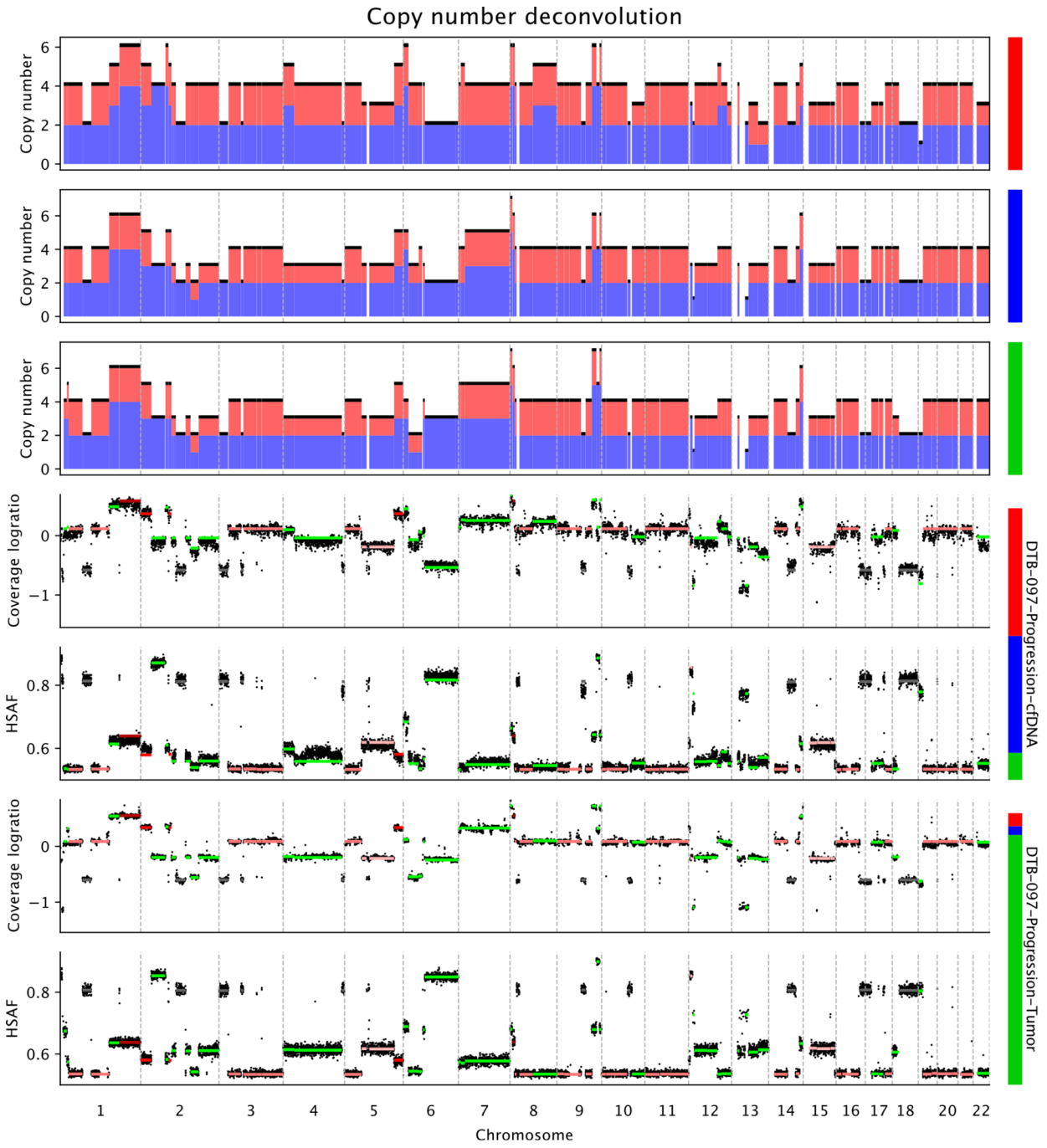


8) It would be helpful if samples are labelled in all figures. For example the identity of each case is not clear in Figure 2b. It would be helpful if they are shown in the same order as Supp Fig 9.

We have now added patient identifiers to Figure 1d, Figure 2b, and Figure 2h. In the revised Figure 2b, the patients are now ordered based on their samples' average cancer fraction, since this makes it visually clear that low cancer fractions are associated with diminished capability for detecting subclones. Our new Supplementary Information document shows truncal copy number models and subclonal reconstructions for all patients, ordered by patient ID for ease of access.

9) Why is Supplementary Fig 9 only shown for a subset of patients? The data detail shown is extremely informative.

We are glad that you recognize the value in these deconvolution visuals; we also find them fascinating. Only six patients were shown in our original Supplementary Figure 9 but we now provide copy number deconvolutions for all evaluable patients. These are presented in a new visual format, which includes allele-specific copy numbers for each subclone, and expected-vs-observed coverage logratio and HSAF values for each sample (example shown below). They can be found on pages 42 to 216 of Supplementary Information. Our revised Supplementary Information document now also includes a new section "Validation of subclonal reconstruction using a blinded simulation" where we validate our subclonal CN deconvolution methodology using simulated ground truth samples containing multiple cancer cell subpopulations.



10) Is the case shown in Supp Fig 10 a specific case?

The original Supplementary Figure 10 provided a representative example of one patient (AE-018) with a subclonal WGD. In total there were five patients in our cohort where we determined that a WGD event had occurred in a branch of the phylogenetic tree (i.e. was not truncal). These patients were AE-018, AE-030, AE-180, DTB-119 and DTB-216. For patients AE-018 and DTB-119, we found individual samples composed of a mixture of WGD and non-WGD populations. For patient DTB-216, we detected two independent WGD events in the phylogenetic tree. Some DTB-216 samples

represented a mixture of two distinct WGD populations, but no samples carried ancestral diploid populations. For patients AE-030 and AE-180, all individual samples were composed entirely of either WGD or non-WGD cells, with no mixtures. As such, patients AE-030, AE-180, and DTB-216 did not require any specialized analysis, and our standard process of truncal model fitting was sufficient to establish sample cancer fractions and subclone CCFs.

We have now added a new Supplementary Figure to illustrate the analysis of DTB-119 (which required an analysis similar to the one carried out for patient AE-018). We also simplified and streamlined the mathematical steps used in the analysis of patient AE-018 (see page Supplementary Information section "Detection of subclonal whole genome duplication events" page 20), and used the new streamlined analysis also for patient DTB-119.

11) Line 143: the MWU test P value shown does not match the figure.

Thank you for spotting this typo—this has now been corrected.

12) Line 149: Why does WGD occurring multiple independent times in the same tumor indicate "convergent evolution"? Perhaps the fact that the same truncal mutations occur in both lineages within each patient set a genotype that is conducive or selects for WGD events?

It is possible that the shared reservoir of truncal mutations may have predisposed both subpopulations to acquire WGD (and indeed prior work has implicated specific gene defects or broad genotypic features as being strongly associated with acquiring WGD; e.g. (Bielski et al., 2018) **(PMID: 30013179),** (López et al., 2020) **(PMID: 32139907)). Nevertheless, we agree that it is challenging to conclude whether this is due to convergent evolution or is just a chance occurrence. Therefore we have removed this clause from the sentence.**

13) Lines 164-167 and Figure 2h: Please annotate the figure to support the finding of AR gene alteration diversity within a patient. For example, markings to show which rows belong to the same patient.

We have added patient labels to Figure 2h (samples are grouped by patient; cfDNA and tissue samples are distinguished via heatmap annotation). For the Reviewer's interest, the time-matched cfDNA and tissue samples from patient DTB-119 shown in Figure 2h offer a representative example of intrapatient *AR* genotype diversity: the tissue sample from DTB-119 harbors a focal *AR* enhancer gain that is absent from his cfDNA, and conversely DTB-119's cfDNA harbors two *AR*

ligand binding domain mutations that are absent from his tissue. Collectively, these differences in *AR* amplicon structure and mutation presence between same-day tissue and cfDNA imply the presence of ≥2 metastatic populations with different *AR* configurations. Figure 3e (showing the average absolute copy-number difference between same-day metastatic tissue and cfDNA) helps quantify this particular phenomenon across all patients providing both cfDNA and metastatic tissue.

In addition, how do you distinguish the timing of a mutation versus amplification for considering AR alterations in the context of a tumor phylogeny?

Allosomal (i.e. sex chromosome) alterations were excluded from our subclonal reconstruction models (and therefore phylogenies as well) because *AR* amplifications can involve a very high number of *AR* copies and have significant effects on mutation allele fractions and overall plasma *AR* copy number even when only present in a small fraction of cancer cells. For this reason, we do not make any explicit claims about the *AR* mutation status of individual subclones, or how many copies of the mutant allele are present in each subclone. In some cases, when a population sweep occurs in parallel to a dramatic shift in *AR* genotype, it is possible to qualitatively assume that the emergent AR genotype is related to the newly dominant population. Nevertheless, we have now updated Figure 4d to show the average number of mutant allele copies per cfDNA-contributing cancer cell in each sample. This number can be unambiguously calculated based on the *AR* coverage logratio, sample cancer fraction, and mutation allele fraction. But it is not possible to assign the *AR* mutations to specific subclonal populations.

The Figure 4d legend was also updated to reflect the changes:

"Estimated number of ligand binding domain mutant AR copies in each sample is indicated using colored bar segments. The number of wildtype AR copies is shown in black."

14) Line 186: "but amenable to reciprocal redetection" -- what does this mean?

Differences in tumor purity and sequencing coverage (both overall coverage, coverage differences as a consequence of intersample copy-number heterogeneity, and stochastic variability) can impact mutation detection characteristics. In comparing mutation overlap between cfDNA and matched tissue, we therefore used a statistical approach to help control for these confounders (reducing the likelihood of false negative comparisons).

In our revised manuscript we have now provided additional granularity within our Methods (new text shown with underline):

"To compare mutation presence/absence across same-day metastatic tissue and cfDNA samples, we defined mutations as concordant if independently detected in one sample and had ≥1 supporting mutant reads in the other sample. Mutations were defined as discordant if independently detected in one sample and harbored zero supporting reads in the other sample. However, variability in tumor content and sequencing depth can impact mutation detection sensitivity and result in false negative comparisons. To mitigate against potential false negatives, we only analyzed variants amenable to reciprocal redetection. Specifically, mutations independently called in sample A were considered not amenable for detection in a paired sample B if the expected allele fraction $MAF_A * (F_B / F_A)$ was below 10%, where F_B and F_A were tumor fractions of the B and A samples, respectively. In addition, if the probability of observing 8 or more mutant reads in sample B (based on a binomial distribution of observed read depth in B and expected allele fraction from sample A) was <90%, the mutation was considered not amenable for detection *(Annala et al., 2021).***"**

For the reviewer's interest, we have now conducted an alternative version of this analysis using only independently detected variants (as per Reviewer #2's suggestion: "*The authors compared concordance of somatic mutations…*").

15) Line 191: "In one patient" -- which one?

This references patient DTB-119, who shows a genome-wide lower CN in the cfDNA sample in Figure 3d. We have modified the Results sentence to state: "*In patient DTB-119, the dominant cfDNA population was diploid in contrast to a hexaploid tissue population.*"

16) Line 194: "early truncal fixation" -- can you assess any truncal genotypes that are more or less associated with subclonal discordance (i.e. more diversity)?

Thank you for the suggestion. We were also interested to see whether specific truncal genotypes are differentially associated with metrics of subclonal discordance and complexity. To address your question, we have performed several additional exploratory analyses, focusing on alterations found in a large fraction of samples. We tested whether deleterious *TP53* mutations, *BRCA2* defects (i.e. associated with homologous recombination repair deficiency phenotype), and ETS fusions were differentially associated with the following variables:

1. **Number of per-sample subclones (inferred from multisample subclonal reconstruction)**
2. **Per-sample subclonal copy number diversity score.**
3. **Per-sample whole-genome duplication status.**

	Clonal complexity feature		
	Number of subclones	Subclonal CN diversity score	WGD status
TP53 mutation (mutant vs. WT)	p=0.13 (mean # of subclones: 2.6 vs 3.1); MWU	p=0.027 (mean score: 0.12 vs. 0.08); MWU	p=0.1; FET
ETS fusion (present vs. absent)	p=0.43 (mean# of subclones: 2.6 vs. 2.7); MWU	p=0.05 (mean score: 0.12 vs. 0.10); MWU	p=0.002; FET (ETS fusions enriched in WGD tumors)
BRCA2 defect (present vs. absent)	p=0.5 (mean # of subclones: 2.6 vs. 2.7); MWU	p=0.35 (mean score: 0.12 vs. 0.11); MWU	p=0.28; FET

MWU = Mann-Whitney U test; FET = Fisher's Exact Test

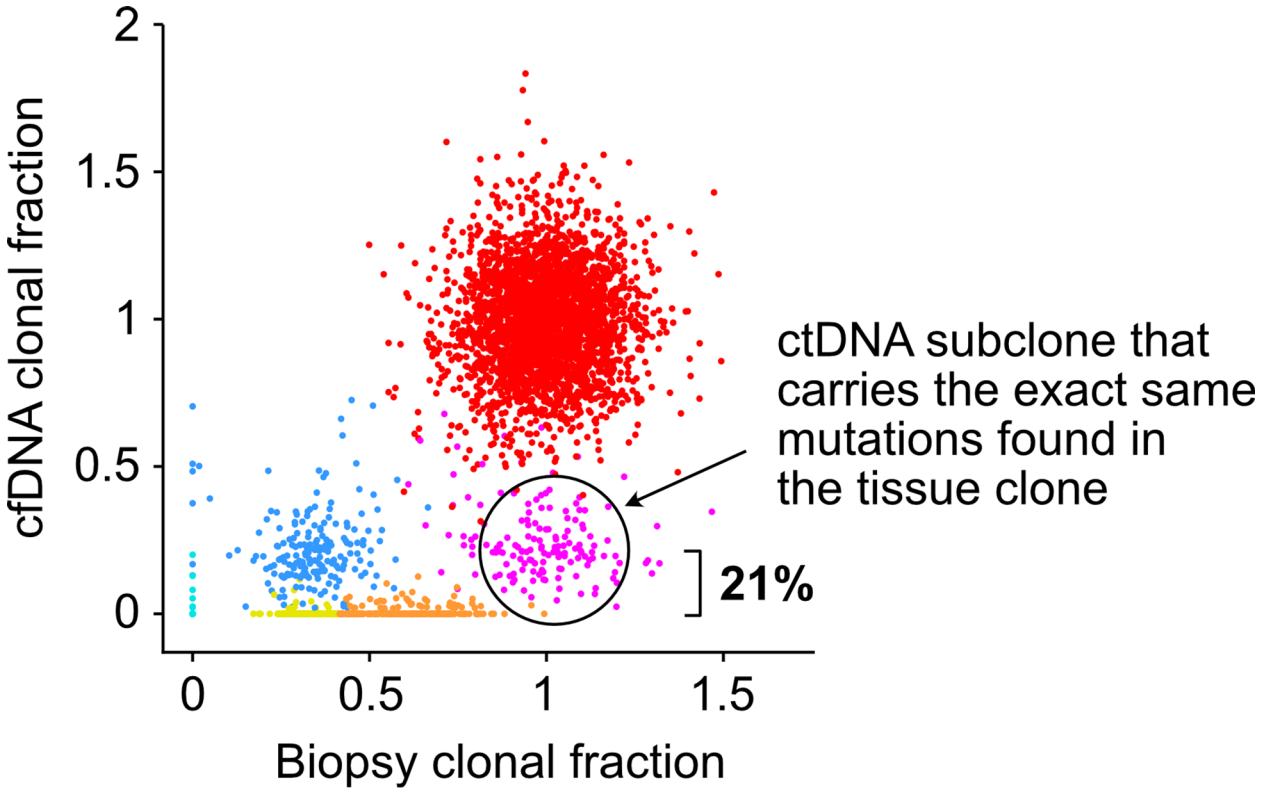
Two hypothesis tests reached statistical significance (prior to correction for multiple comparisons). No other clinically or biologically relevant genotypes (e.g. truncal *SPOP* MATH-domain mutation, *FOXA1* forkhead domain mutations; each of which occurred in fewer than 2 patients) were sufficiently recurrent in our cohort to merit testing.

It is important to acknowledge that although we have analyzed nearly 80 samples through deep WGS, the total number of patients is relatively small for subgroup comparisons (n=33 prostate cancer patients). Patients are also pre-selected based on clinical presentation and the requirement for all samples to have reasonably high tumor fraction. Finally—when coupled with the per-patient heterogeneity in technical limits of detection for subclonal reconstruction (which strongly affects features like tree topology and number of subclones, as shown in our revised Figure 2b)—we feel our cohort is not ideally suited to systematically interrogate these hypotheses. Therefore, we have omitted these new analyses from the revised manuscript to avoid overinterpretation of these largely negative results.

17) Figure 3d -- what happens to discordance when you deconvolve the genotype of ctDNA? Can you resolve a ctDNA "subclone" to be highly concordant to the "bulk" metastasis alteration spectrum in a case that is discordant as currently presented? This would lend weight to the argument that cfDNA is

comprised of multiple clones but that the synchronous metastasis is not the only clone. (Is there any clinical data from those cases to support a major metastasis at a site not biopsied?)

The somatic mutation CCF analysis shown in Figure 3A (and described in Supplementary Information section "Estimation of metastatic tissue contribution to ctDNA": page 19) allows us to accurately quantify the relative fraction of plasma ctDNA contributed by the biopsied metastasis populations. If hypothetically the synchronous metastasis represented the only clone present within the patient's entire metastatic ecosystem, then the matched ctDNA would appear identical to the biopsied lesion (i.e. tissue contribution to ctDNA would be 100%). By virtue of our observation that individual metastases typically contribute only a minority of total ctDNA (average: 19%), this analysis already demonstrates that ctDNA must be composed of multiple clones, and that the biopsied metastasis is not the only clone present in ctDNA. Our analysis also directly shows that one ctDNA subclone carries a somatic mutation profile that is identical with the characteristic mutation profile of the biopsied metastasis. See example figure below.



Performing a similar analysis for copy number profiles is more challenging, because the measurements are less accurate, and there is a significant probability that two populations independently acquire copy number alterations to the same broad chromosomal regions (whereas the probability of identical somatic mutations arising independently in two populations is very small).

To address these challenges, for our initial submission we developed a computational approach that performs copy number deconvolution in a phylogeny-aware manner, utilizing information about the number of detected subclones and their CCFs across all same-patient samples. The algorithm finds a copy number profile for each subclone such that when the subclone copy number profiles are mixed together at the known sample-specific CCF ratios, the theoretic mixed copy number profile matches with the observed coverage logratios and heterozygous SNP allele fractions in all individual bulk (i.e. non-deconvoluted) samples. This algorithm is now extensively described in the revised Supplementary Information document section "Reconstruction of subclone specific copy number profiles". We have also now performed additional validation of this algorithm using simulated *in silico* ground truth genomes. These validation results are shown in the section "Validation of subclonal reconstruction using a blinded simulation" on pages 32-35 of the Supplementary Information document. All subclonal copy number deconvolutions inferred by our approach are also now shown in Supplementary Information section "Patient truncal copy number models and subclonal reconstructions", featuring several new types of visualization (as described on Supplementary Information page 42).

In most patients the cancer population present in the biopsied metastasis contributed only a small fraction of total plasma ctDNA (e.g. DTB-097: 11%, DTB-102: 7%), making it difficult to observe faint plasma copy number changes unique to the metastasis (if any exist).

Nonetheless, we can provide a few convincing examples where the dominant subclone in the biopsied metastasis had an observable effect on the plasma cfDNA copy number profile:

1) **If you look at the phylogenetic tree and CN deconvolution of patient DTB-175 in our Supplementary Information (pages 172-174), you can see that the Baseline-Tumor sample contains the blue population and its green and pink subpopulations. Although the patient's Progression-cfDNA sample was collected at a different clinical timepoint, the same blue population and its subpopulations were detected at ~50% CCF in the cfDNA sample. In the CN deconvolution plot, it is clear that the copy number changes characteristic to the blue population and its subpopulations (such as 13q deletion, 6p gain, 15q gain, and two small deletions in chr1) are seen to perturb the coverage logratios and heterozygous SNP allele fractions of those regions in the amounts expected from the blue population CCF in that cfDNA sample.**

2) **If you study the phylogenetic tree and CN deconvolution of patient DTB-261 in our Supplementary Information (pages 209-211), you can see that the Baseline-Tumor sample contains the green and pink populations, which are descendants of the blue population. The blue, green and pink populations all harbor a chr20 gain that was not present in the red ancestor population. As shown in the CN deconvolution (page 211), the coverage logratio and HSAF values in the Baseline-Tumor sample suggest that the chr20 gain is**

clonal in that sample (i.e. present in all cancer cells in that sample). However, in the two plasma cfDNA samples we can see that the chr20 gain is only faintly visible, and the magnitude of the change (in both coverage logratios and HSAF) is mathematically consistent with the comparatively low cancer cell fractions (CCF) of the green and blue populations in those cfDNA samples. A similar phenomenon is seen in 18q, where the blue and green populations have a copy number state of 2+0, and the ancestral red population has a copy number state of 1+0, and we can see that in the plasma cfDNA samples the magnitude of the coverage logratio and HSAF signals is consistent with the CCFs of the green and blue populations.

We were very interested in exploring potential clinical variables that might have influenced the percent contribution of single biopsied metastases to bulk ctDNA (e.g. size of the biopsied lesion from imaging data, metrics of global metastatic burden, organ-specific distribution of metastases, etc.). Unfortunately we were not able to obtain sufficiently detailed clinical annotation to explore these hypotheses. Patients in our cohort that provided matched metastatic tissue and cfDNA were accrued under the SU2C/WCDT clinical study protocol (NCT02432001) which formally completed in 2020, without continued funding for retrospective clinical data collection (and most of these patients are now deceased).

18) Line 201: The comment that AR genotypic diversity drives clonal expansion is in direct contraction to data presented in Espritu et al (Cell 2018) in which a subset of localized prostate cancers show clonal diversity. Of course, localized prostate tumors do not harbor AR alterations. Consider an alternative interpretation that tumors exploit alterations to the AR gene for driving progression with underlying genomic instability.

We agree that the sentence on line 201 was not precisely worded. Our intended message was that the greater degree of subclonal genotypic diversity observed at the *AR* locus suggests that somatic alterations within the *AR* locus in the mCRPC setting are more prone to give rise to clonal expansions than alterations in other genomic regions. Importantly, the critical qualifier (missing from our original phrasing) is that this is only relevant for the specific therapeutic context in which our samples are analyzed (i.e. androgen suppression via ADT plus prior potent AR-axis inhibition in some patients). *AR* alterations (in this highly androgen-suppressive environment) are more likely to provide a differential fitness advantage relative to other genomic alterations, and are therefore key candidates for clonal selection. We did not mean to imply that *AR* genotypic diversity drives clonal expansions *overall* (e.g. including in localized ADT-naive disease, where selective pressures are likely to be significantly different compared to mCRPC).

The alternative hypothesis is that the *AR* locus is intrinsically subject to more genomic instability (per cell division) than other genomic regions. However, the lack of increased subclonal genotypic diversity at the *AR* locus in studies of localized prostate cancer supports the former theory.

Compared to mCRPC, tumor dynamics in pre-treatment localized prostate cancer (e.g. (Espiritu et al., 2018)) are not influenced by the strong selective pressure of prior or active androgen/AR-axis suppression.

We have revised this sentence to read:

"These data implicate AR locus alterations as a critical substrate driving clonal expansions in the context of AR targeted therapy."

Relatedly (primarily in response to Reviewer #3), we have added a new Discussion paragraph addressing the role of therapeutic context in shaping emergent resistance mechanisms, as well as the role of *AR* alterations in mCRPC more broadly:

"Clinical suppression of AR-signaling is essential for mCRPC management—all drugs are administered on a backbone of continuous androgen-deprivation, and emerging standard-of-care and investigational agents (e.g. ^{177}Lu–PSMA-617, AR protein degraders, AR N-terminal domain inhibitors) offer renewed strategies for exploiting dependence on AR-signaling. This constant selective pressure underscores the importance of understanding nuanced variation in AR genomic resistance mechanisms and underlying (sub)clonal diversity. More broadly, genomic mediators of resistance to clinically-active targeted therapies are likely coupled to the specific pathways under selective pressure, implying that ctDNA can nominate distinct mechanisms of acquired resistance in other cancers and therapeutic contexts (e.g. PI3K-pathway or PARP inhibition in prostate and breast cancer)."

19) Line 212: AR mutations are not shown as highly recurrent in Figure 4.

Thank you for highlighting this shortcoming in our visualizations. Changes in the *AR* mutation status between serial plasma cfDNA collections were detected in 5 / 21 patients analyzed. We have updated Figure 4d to quantitatively illustrate the average number of *AR* mutant allele copies per cancer cell in the serial plasma cfDNA samples. This makes *AR* mutation changes in patients 124, 035, 092, 099, and 261 apparent.

Of these five patients, three showed a statistically significant increase in the allele fraction of an AR LBD mutation (patients AE-035, AE-092, and DTB-261). We have updated Figure 4b to highlight that AR stood out in an analysis of the frequency of acquired protein altering mutations. The statistical methodology underlying this approach is described in our new Supplementary Information document.

20) Line 214: What is the data supporting ligand binding domain (LBD) mutation switching?

Temporal changes in *AR* mutation status after therapy were identified in 4 patients. In our revised version of Figure 4d, we show the *AR* LBD mutation status in all samples, and the average number of mutant allele copies in the cancer cells present in each sample.

21) Figure 4b: "1L" means first line but I could not find this defined in the manuscript.

We agree that the label was confusingly worded. Indeed, the label should have more correctly read *"Genomic changes after one line of AR targeted therapy in serial cfDNA biopsies"*. We have also clarified the figure legend to state:

"Aggregate of acquired genomic changes in ctDNA after treatment with AR signaling inhibitors."

22) It is unclear how clonal dynamics of tumor subpopulations are estimated in Figure 4e. Curve smoothing can be used between samples and each corresponding proportion, but for case AE-035, how do you show the timing for subclone "drop out" at Day 100 when progression does not occur until day ~190? Also, AE-035 is shown in both Fig 4e and Supp Fig 15. Is this intentional?

Showing AE-015 in both figures was not intentional. In our revised submission we have removed the previous Supplementary Figure 15 and we are showing these clonal dynamics plots for all evaluable patients (including the patients in Figure 4) in the Supplementary Information document (pages 42-217).

Briefly, subclonal reconstruction was used to infer the subclonal composition of each discrete cfDNA collection timepoint, with subclones visualized as fractions of synchronous PSA concentration. The subclonal composition at all intervening timepoints was interpolated using the following framework, as now described in our revised Supplementary Information (page 17):

"Monotone cubic interpolation using Steffen's method was applied to PSA measurements obtained at 4-week intervals to visualize tumor burden over time. This interpolation method guarantees that the interpolation curve passes through all data points, and never switches derivative sign (i.e. overshoots) between data points. Population CCFs were estimated at plasma cfDNA collection

timepoints, and interpolated between timepoints by fitting an exponential function $f(t) = ae^{bt}$ to the boundary conditions $f(t_1) = F_1$ and $f(t_2) = F_2$, simulating the exponential nature of cell population growth. If a new subclone emerged or an old subclone disappeared between two cfDNA timepoints, we visualized the subclone as having crossed the detectability threshold in the exact middle of the two timepoints. If N nested subclones emerged during time interval T between two cfDNA timepoints, the time interval was divided evenly such that each nested subclone emerged after T / N time. The combination of dense PSA and sparse subclone data from plasma cfDNA allowed us to simultaneously visualize treatment-driven changes in tumor burden (i.e. treatment responses), and subclonal dynamics. To indicate phylogenetic relationships between subclones already present in the first available cfDNA timepoint, a qualitative rendering of clonal expansion starting from the MRCA of all cancer cells was drawn on the left-hand side of the figure, without a quantitative time axis. No time series data was available from the pre-mCRPC stage of disease."

We have also modified the Figure 4 legend to better articulate what these plots describe:

"Clonal population dynamics (inferred from subclonal reconstruction) across treatment timelines (horizontal axis). Vertical axis represents monotonically interpolated prostate-specific antigen (PSA) measurements at 4-week intervals, indicating tumor burden and quantifying therapy response and progression. Subclone fractions are interpolated between plasma cfDNA timepoints using exponential functions (see Supplementary Information section "Visualization of tumor burden and subclone fractions over time")."

23) Line 228: It would be helpful to clarify how the occupancy maps were generated at this point.

We have revised the sentence to read: "*We used orthogonal methods (i.e. windowed protection score and relative read-depth depletion) to generate per-sample nucleosome occupancy maps from cfDNA and contrasted results with whole-transcriptome sequencing of same-day metastatic biopsy.***"**

24) Line 235 - Supp Fig 16, what is "r" used and the statistical test used in 16a and 16b? Given the interval level data, Spearman correlation would be the most likely correlation method.

For both Supplementary Figures, we performed a Spearman rank-order correlation ('r' therefore denotes Spearman's rho) for each individual patient (line) and reported the mean Spearman's rho (and a combined p-value using Fisher's Method). We have now annotated these details on the Figures directly. Note that these two Figures are now included as part of Extended Data 3.

25) Line 237: Spearman correlation should be used instead of Pearson

Thank you—we have replaced this statistic with the Spearman correlation.

26) Lines 237-241: Can you provide this source data?

The following sources data were used for our nucleosome cfDNA profiling analyses:

1. **RNA-Seq data from 101 mCRPC metastatic tissue samples from (Quigley et al., 2018) (PMID: 30033370)**
2. **A list of 3804 housekeeping genes from (Eisenberg and Levanon, 2013) (PMID: 23810203) identified from an expression analysis of Human BodyMap 2.0 Project data.**
3. **A list of 350 genes highly expressed in hematopoietic lineages derived from (Ulz et al., 2016) (PMID: 27571261).**
4. **3224 AR binding sites (ARBS) identified from prior chromatin immunoprecipitation followed by ChIP-sequencing of 13 primary prostate cancer tissue samples, (Pomerantz et al., 2015) (PMID: 26457646).**

These data are all publicly accessible and are currently referenced and described in our Methods (Data Availability subsection). We have now included three new Supplementary Tables containing the raw source data for #2, #3, and #4 (Supplementary Tables 8, 9, and 10, respectively). The RNA-Seq data described in #1 is available for unrestricted download here (https://quigleylab.ucsf.edu/data), and this link is now placed in our revised Data Availability Methods subsection.

27) Line 245: the weak correlation between Tissue mRNA abundance with cfDNA nucleosome depletion - is that supposed to be gray in Figure 5d?

In Figure 5d the gray lines show the correlation between gene TSS nucleosome depletion and tissue mRNA expression, while the black lines show the correlation between the first exon-intron junction and tissue mRNA expression. The original sentence on line 245 stated that the correlation between the first exon-intron junction and tissue mRNA expression is weak, referring to the weaker magnitude of the effect.

In the revised manuscript we have improved the visual clarity of panels 5d and 5e by using colors that are easier to distinguish, and we have improved the manuscript wording (see comment #28 below).

For 5d and 5e, Spearman rank correlations should be used for percentiles.

Thank you for the suggestion, we have updated these panels to show Spearman rank correlations in the revised manuscript.

28) Line 246: Figure 5f does not show an "anticorrelation". Do you mean no association?

We intended 'anticorrelation' to mean an inverse or negative correlation. Nonetheless we have rephrased this paragraph to make the message more clear:

"Tissue mRNA-abundance was also strongly correlated with cfDNA nucleosome depletion at the first exon-intron junction (Zhu et al., 2021) (Figure 5d), although the magnitude of effect (relative to the TSS) was weaker and exhibited an association with first exon length."

29) Line 258: Use Spearman correlation instead of Pearson, but if the goal is to show enrichment within a subset, you might find Fisher's Exact Test more useful.

We have replaced these values with Spearman correlation statistics.

30) Supp Fig 18 - Spearman only.

We now only report Spearman correlations for these plots (and have changed the Results text accordingly).

31) Line 268 - The data presented in 5i supports progression with PSA increasing, but there is loss of AR signaling in 5j. How can you rectify this contradiction?

Serum PSA and AR-pathway activity can be relatively uncoupled in advanced prostate cancer. This is supported clinically by the observation that histologically-confirmed neuroendocrine prostate cancer (NEPC) (which is typically AR-low or independent) can sometimes have high serum PSA. For example, (Aggarwal et al., 2018) (PMID: 29985747) showed no significant difference in serum PSA between mCRPC patients with histological evidence of small-cell NEPC (from metastatic biopsy) versus those without. (Labrecque et al., 2019) (PMID: 31361600) provide preclinical evidence of per-patient transcriptomic heterogeneity among well-characterized NEPC-associated and AR-regulated genes (including co-expression in some cases), supporting the idea that NEPC and AR-driven adenocarcinoma exist along a spectrum (rather than as binary and mutually-exclusive disease states). Therefore, one plausible explanation for the data shown in Figure 5j is that the clone rising to dominance at timepoint #3 is both AR-low and PSA-secreting. However, patient AE-180's PSA across all three timepoints is relatively low compared to typical treatment-naive mCRPC (Khalaf et al., 2019; Oudard et al., 2017) despite having widespread metastatic disease, implying that PSA is only loosely informing on his disease burden. Indeed, his serum LDH (associated with liver metastases in prostate cancer and less AR-driven disease) changed from 0.99 to 1.5x the upper limit of normal while serum ALP (associated with bone metastases and AR-driven prostate cancer) decreased from 2.15 to 0.84, together suggestive of more drastic clinical changes than implied by PSA alone. We have indicated this via inclusion of a new in-set line-plot in Figure 5j showing LDH and ALP dynamics (with accompanying text in the Legend): "*In-set line plot shows temporal dynamics of lactate dehydrogenase (LDH; associated with liver metastases) and alkaline phosphatase (ALP) concentrations per upper-limits of normal (ULN), indicating changes in disease composition.*"

32) Lines 291-294 and 331-333 are in contradiction to each other. The deep data provided in most of this manuscript was possible due to the deep (and thus expensive) WGS. Targeted cfDNA or lower pass coverage will not be as informative for understanding the composition of tumors.

Thank you for pointing this out—to remedy this contradiction we have reworded Discussion lines 331-333 to better clarify our key messages:

"*Targeted cfDNA assays incorporating these biologically informative chromatin landmarks may circumvent the need for expensive deep WGS if TSS/TFBS profiling is a primary goal, with the caveat that capacity for comprehensive dissection of other genomic information will be reduced.*"

33) Line 297: If the ancestral tumor lineage is clinically important, does that mean that the tumor composition of ctDNA subclones reflects the clinical aggressivity of the tumor?

Although it is challenging to delineate the precise clinical significance of specific subclones, the populations arising to clonal dominance at progression on Abiraterone/Enzalutamide (Figure 4e) most likely represent populations that have evolved resistance to therapy (i.e. harbored a fitness advantage relative to competing clones and experienced positive selection). At progression, these treatment-resistant clones comprise a patient's active/contemporaneous disease, and therefore strongly influence their subsequent disease trajectory and clinical outcomes.

Interestingly, in several patients we observed that it was the truncal lineage that ultimately superseded its daughter lineages at progression (e.g. patient AE-035 at progression). This implies that these truncal lineages can sometimes be associated with clinical aggressivity. In this scenario, it is not necessarily the truncal clone *per se* (as it existed prior to the evolution of its daughter lineages), but rather clones that descended from the truncal clone but were too minor to be detected in subclonal reconstruction. Taking AE-035 as an example, the red 'truncal' population at second progression in reality is an amalgamation of many smaller subclones below our detection capacity (and may constitute both clones arising from neutral evolution and clones shaped by selection forces). Sequencing bulk ctDNA to higher depths may enable detection of these smaller subclones (Turajlic et al., 2019) (N.B. for the Reviewer's interest, we have provided a detailed discussion of this concept as part of our response to Reviewer #3's comment: "*It is noteworthy that the median ctDNA fraction in the study cohort…*").

Future work could explore how broader features of ctDNA clonal composition (e.g. number of subclones greater than some minimum subclonal fraction, variance in the clonal fractions among quantified subclones, mutation/ploidy/copy-number diversity *between* subclones) affect various metrics of disease aggressiveness (e.g. response rates to specific therapies and overall prognosis).

Referee #2 (Remarks to the Author):

The authors present a major advance in the field of ctDNA sequencing of prostate cancer samples. The use of Whole Genome Sequencing of ctDNA and concomitant sequencing of tissue samples is key to understand how these 2 sources of tumor derived DNA relate to each other. The manuscript has been well-written and is therefore an interesting paper for the readership of Nature.

Thank you for your positive feedback and helpful suggestions (addressed below).

My major remarks are:

1) The use of custom and/or unpublished and non-peer reviewed software and algorithms (without comparisons to contemporary and often-used alternatives) complicates the validity of the presented results. This could introduce some bias as there are well-suited standards for somatic variant-calling (from tissue) which could outperform the custom cfDNA-trained variant calling pipeline. A comparison of the custom cfDNA-based workflow vs. another validated and contemporary pipeline for tis-sue-samples (e.g., the GATK-workflow) on the tissue-samples could provide clarity of any potential biases of using the cfDNA-based workflow on tissue-samples.

We have now performed comprehensive benchmarking and orthogonal validation of all aspects of our original submission using established, state-of-the-art public software. Briefly (as pertaining to your specific comments), this includes:

- **Validation of SNVs/indels using the GATK Mutect2 pipeline.**
- **Validation of structural variants using the GRIDSS2 pipeline.**
- **Validation of copy-number calling using the Battenberg and ASCAT algorithms.**

We have also performed additional validation with other tools, focusing on subclonal reconstruction and mutational signature data (using PyClone, PhyloWGS, TrackSigFreq, and SigProfileExtractor—and described in other Reviewer responses). All validation methodology and results are contained within our new 246 page Supplementary Information file (including 41 pages of dedicated methodology text), and we have significantly expanded our online code repository (including simulated test data for experimenting with our algorithms). Overall, we feel that addressing your and the other Reviewer comments has substantially elevated the rigor and validity of our work.

It is not clear whether the tissue biopsies were examined using the same cfDNA somatic variant calling, CNV and SV-detection pipelines.

The tissue and cfDNA samples were analyzed using the same methods and thresholds. We have made this more more clear in the revised Methods:

"Somatic mutations (base substitutions and indels) were called in cfDNA and tumor samples by …"

"Somatic structural rearrangements were identified in cfDNA and tumor samples using split-read methodology …"

In copy number analysis we used a different reference coverage profile for cfDNA and tissue samples, since processes of DNA fragmentation and degradation are different for these two sources of DNA. This is now articulated in our new Supplementary Information document (subsection: "*Copy number model fitting and cancer fraction estimation*", page 3):

"A set of 5 whole genome sequenced ctDNA-negative cfDNA samples and 34 leukocyte DNA samples were used to generate two reference coverage profiles for copy number analysis (one for cfDNA analysis and one for tissue DNA analysis)."

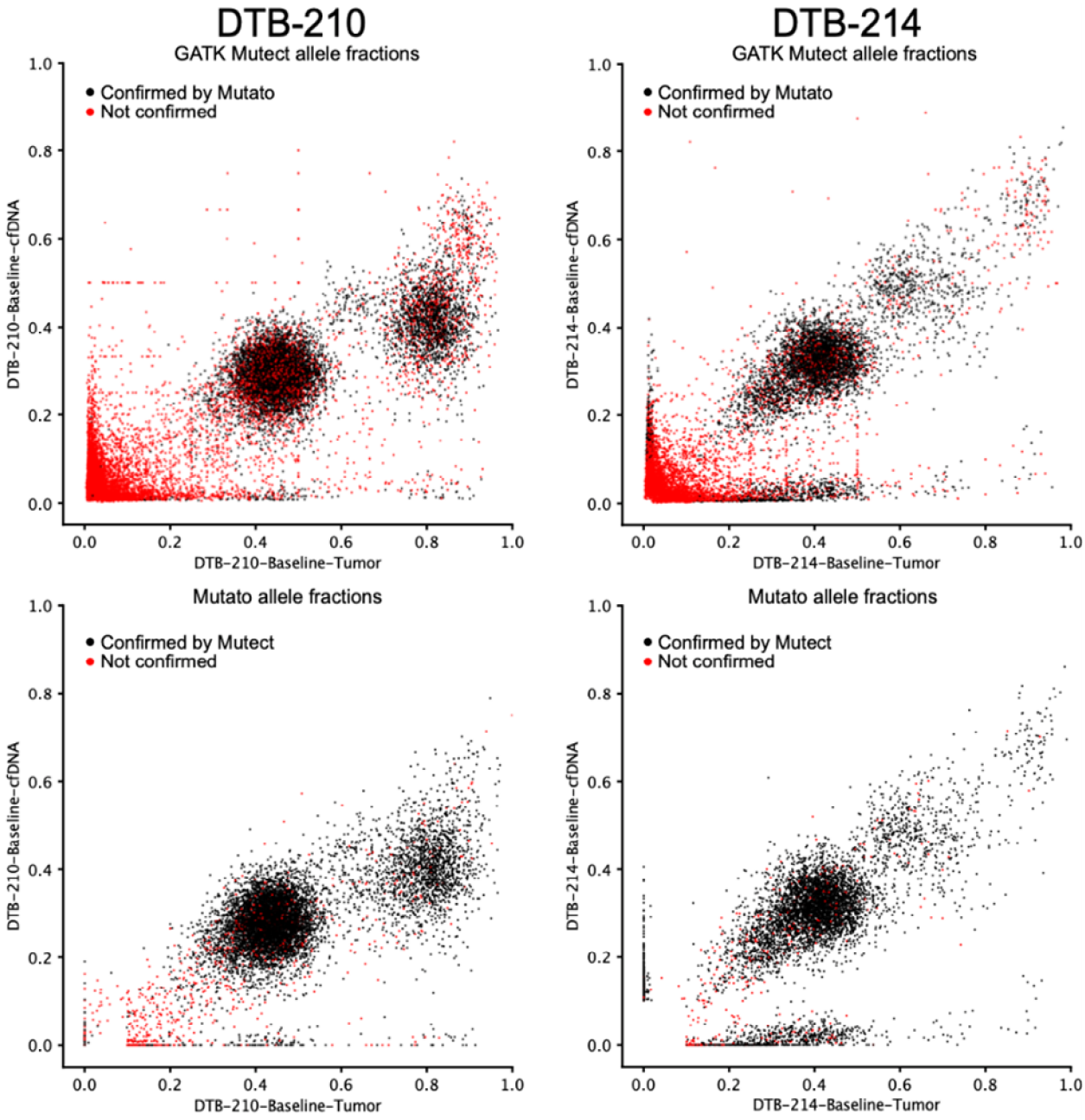
Any potential high-quality somatic variants detected using this workflow in tissue and not detected in the matched cfDNA samples (using the cfDNA-based workflow) should also be shown as discrepant. In addition, the references for the "validation" of this cfDNA-based somatic caller lack validation or comparison against tissue-based somatic variant callers (line 839 - 840). The references given such as Vandekerkhove et al. (2021) reference simply refers back to other publications ("Alignment and analysis of DNA-sequencing data were performed utilizing an established pipeline9,13,54,56") and then lists custom and arbitrary criteria again such as "Somatic mutations were required to be supported by a minimum of eight mutant allele reads, with a minimum VAF of 1% in ctDNA and 8% in tissue." Moreover, Annala et al. (2018) does also not perform a comparison against tissue biopsies. The thresholds used in these 3 manuscripts differ, and thus, the cited studies do not support the use of the current thresholds. These technical issues make the manuscript somewhat difficult to understand completely.

We have now carried out additional validation of the mutation calling pipeline by analyzing multiple tissue and plasma cfDNA samples using the established GATK Mutect2 pipeline. A head-to-head comparison of the somatic mutation calls made by the two analysis pipelines revealed that GATK Mutect2 confirmed 92.2% of the mutations detected by our analysis. The GATK Mutect2 pipeline identified on average 2.06-fold more mutations than our pipeline, primarily at low allele fractions. In a known cancer-negative plasma cfDNA sample subjected to 202x depth WGS, GATK Mutect2 identified 15,125 somatic mutations, compared to 36 somatic mutations identified by our analysis. The somatic mutations that GATK Mutect2 detected in the negative control sample had low allele fractions, and were consistent with the low allele fraction mutations that only GATK Mutect2 had detected in the cancer samples. Concordance between the two tools was extremely high for somatic mutations detected with an allele fraction ≥ 10%.

The somatic mutation calling thresholds utilized in our manuscript were calibrated to yield a low rate of false positives, based on analysis of a cancer-negative plasma cfDNA sample sequenced to a similar 202x depth. Our Methods section in our original manuscript draft did not make this sufficiently clear, so we have clarified this in the revised manuscript: *"Our mutation calling*

thresholds were calibrated to achieve a false discovery rate of 1.0%, based on analysis of a cancer-negative control plasma cfDNA sample sequenced to 202x WGS depth."

The figure below shows two patients where we compared Mutato and GATK Mutect2 results (using default GATK Mutect2 parameters). The mutations (dots) are colored according to whether the mutation was also independently called by the other pipeline. For more detail on the GATK Mutect2 validation and an extended version of the figure, please refer to Supplementary Information section "Validation of somatic mutation calls using GATK Mutect2".



We have additionally validated our structural variant calls through using the established GRIDSS2 software (described separately below). Importantly, both the GATK2 Mutect2 (for SNVs / indels) and GRIDSS2 pipelines (for structural variants) achieved similar concordance with our own tools' results across both tissue and cfDNA samples (i.e. there were no systemic performance differences between the two analytes). These comparisons suggest that utilizing the identical pipeline for both tissue and cfDNA samples is not a major source of bias.

We have also rewritten several parts of our main Methods to provide richer methodological detail about our SNV/indel and structural variant calling approach, as well as to direct readers to our comprehensive validation experiments (located in our Supplementary Information document). Additionally, specific analyses within our Supplementary Information document (including validation) are now appropriately highlighted through the manuscript Results and Methods.

2) The conclusion that the tissue based sequencing may "misinform targeted treatment" seems a bit bold based on the data the authors present. To conclude this the authors need to present data on actual dis-crepant targetable genes. The loss or gain of mutation unique to ctDNA is limited to 2 AR variants (fig 3c) in the current manuscript making the basis in my view meager. The difference in genomic landscape may be obvious, to conclude ctDNA is better for making a diagnosis the heterogeneity must be relevant.

We agree that our original phrasing was bold—we have now toned down this Discussion paragraph (copied below):

ORIGINAL SENTENCE:

"From a clinical perspective, reliance on tissue alone may misinform choice of targeted therapy by failing to identify primary resistance mechanisms present in other metastases, or by misclassifying site-specific mutations as truncal features."

REVISION:

"Nevertheless, reliance on a single tissue biopsy may provide an incomplete view of polyclonal treatment resistance mechanisms (e.g. AR alterations, BRCA2 reversion mutations), misclassify alterations as truncal features, or underestimate disease subclonal complexity."

3) A major question remains whether indeed clonal architecture is driving the emergence of resistance and the data presented here do not resolve this issue. Clonal reshaping upon treatment has been suggested by fig4e and supplemental figure 15. The data presented seem to rely mainly on

the estimation of copy number alterations. These are found indeed to drive the difference between ctDNA and tissue sequencing. The final conclusion whether these findings indeed represent true clonal expansion or adaptation by vari-able amplification of the X-chromosome will need to come from single cell analyses. Therefore, I would suggest to tone down the conclusion or to present more detailed data showing that apart from the AR CNV reshaping there is indeed strong evidence for clonal evolution.

Figure 4e shows the *AR*-region copy number structure (i.e. breakpoint locations and relative number of copies of each segment) at sequential cfDNA-collection timepoints. There are three possible mechanisms by which the *AR*-neighborhood could become copy amplified over time (as observed from bulk ctDNA sequencing):

1. **Adaptive model: The majority of ctDNA-shedding cells independently acquire *AR*-region copy gains where each cell's breakpoints are unique (and are randomly distributed). There is no appreciable differential selection of any given cancer cell (predicated on its specific *AR* genotype).**

2. **Synchronous adaptive model: The majority of ctDNA-shedding cells independently acquire *AR*-region copy gains where each cell's breakpoints are *identical*.**

3. **Clonal expansion model: At treatment initiation, the patient's body contains billions of cancer cells, some of which carry randomly acquired mutations that provide resistance against the therapy (e.g. an *AR* gain). These cells expand during treatment (i.e. are positively selected) while other cancer cells are extinguished. The expanded resistant population(s) become detectable subclones, each harboring a unique *AR*-region structure with different magnitudes of average *AR* amplification—and that the proportions of these clones (relative to one another) change over time (and shift in favor of the clones with higher number of *AR* copies).**

The adaptive model would mean that all *AR*-region copy-number profiles in ctDNA (which are aggregates of the *AR* profiles of individual ctDNA-shedding cells) would appear as smooth, bell-like curves (without appreciable breakpoints) with its local maxima centered on the AR gene/enhancer. However, for all individual patients in our cohort, the observed *AR*-amplicon is always composed of discrete copy number segments (with distinct and a finite number of breakpoints), therefore ruling out this model. The synchronous adaptive model would require all tumor cells to independently acquire the exact same patterns of copy-gains (*and* losses in some cases) to achieve the overall *AR*-amplicon structures shown in Figure 4e. The probability of all tumor cells independently acquiring copy number defects with the exact same breakpoints is infinitesimally small, therefore ruling out this model. Consequently, the clonal expansion model represents the most parsimonious explanation for these AR-copy number data. Note that this

conclusion can be arrived at *based on the features of these copy number plots alone and without any other data*.

Notwithstanding, our original and revised manuscript offer numerous other pieces of evidence in support of dynamic clonal population shifts over time—whose validity for informing on evolution is supported by precedence from prior research. Briefly:

1. **We exploited the thousands of mutations detected in each of our samples as lineage markers to accurately quantify the number of somatic populations present and their proportions relative to one another (described extensively in our 246 page Supplementary Information document). This general approach (i.e. multisample subclonal reconstruction) is an established technique for population composition inference from bulk sequencing data and has been used extensively in tissue-based studies (alongside a significant methodological pedigree) (Caravagna et al., 2020; Dentro et al.; Deshwar et al., 2015; Gundem et al., 2015; Liu et al., 2020; Roth et al., 2014; Salcedo et al., 2020; Sun et al., 2017; Tarabichi et al., 2021; Turajlic et al., 2018); (Caravagna et al., 2020; Deshwar et al., 2015; Gundem et al., 2015; Tarabichi et al., 2021). A schematic of our subclonal reconstruction approach is provided in Figure 2a, and the specific technical advancements offered by our approach (relative to prior work) are detailed in our Supplementary Information.**

   The distinct cancer populations identified from subclonal reconstruction are shown as the components of the stacked barplot in Figure 2b. For patients providing serial ctDNA samples (e.g. patients AE-030, AE-046), these barplots illustrate that these somatic populations change in relative prevalence over time, suggesting clonal restructuring.

2. **In Figure 4e (left) we project the clonal fractions of each population from subclonal reconstruction onto the patient's PSA concentration over time. This plot enables joint interpretation of patients' disease burden and clinical endpoints (i.e. PSA response and progression), as well as their metastatic clonal population dynamics (captured by serial ctDNA). Four of the five patients showcased in Figure 4e show evidence for clonal remodeling insofar as positive selective sweeps of population(s) (color) that harbored a comparatively lower clonal fraction at baseline. In all five patients, the average AR copy number increases over sequential cfDNA collections (often in the context of a different amplicon structure), implying that selection for *AR* resistance alterations is driving these clonal expansions. Clonal expansion plots for all eligible patients are shown in the Supplementary Information document (pages 42-216).**

3. **In Figure 4b we compare same-patient serial ctDNA collected before and after potent AR-targeted therapy (abiraterone or enzalutamide). Here we demonstrate that the only highly**

recurrent temporal changes in copy-number (as highlighted by the reviewer in context of Figure 4e) but also protein-altering mutations and structural variants occur at the *AR* locus. Interpreted via the clonal expansion model, these data imply positive selection of specific *AR*-augmented populations that harbor comparatively low baseline clonal fractions (i.e. were not dominant at the earliest cfDNA timepoint). The emergence of pathogenic resistance mutations in the *AR* ligand-binding domain at treatment progression is better visualized in our improved version of Figure 4d (showing temporal increase in average *AR* mutant-allele copy number in several patients).

Importantly, our use of deep whole-genome sequencing in searching for acquired genomic changes at treatment progression avoids the selection bias ensconced in serial targeted- (focusing only on known cancer genes) or exome-sequencing (focusing only on coding regions), which have dominated prior efforts in this space—including our own.

For the Reviewer's interest, we have significantly expanded our Discussion to better delineate the potential role the *AR* plays in contributing to treatment resistance in mCRPC. Although our data shows strong links between clonal expansion, clinical resistance to AR-targeted drugs, as well as progressive temporal augmentation of established AR resistance alterations, we agree with the Reviewer that these data are ultimately associative. Determining the precise causality (in an *in vivo* setting) of specific populations in driving treatment resistance will be extremely challenging, and will ultimately need to come from synthesizing evidence across 1) analyses of large numbers of patients in standardized clinical trial cohorts, 2) prospective clinical interventional studies using *AR*-genotype for patient stratification, and 3) patient-derived preclinical model experimentation (e.g. organoid and patient-derived xenografts).

Emerging hybrid approaches to subclonal reconstruction combine bulk reconstruction techniques (analogous to what we have performed here) with single-cell tissue analysis (Malikic et al., 2019; Tarabichi et al., 2021). Integrating these techniques (to some extent) provides reciprocal validation and offsets the limitations of each technique used as standalone. Nevertheless, supplementing cfDNA-based subclonal reconstruction with tissue single-cell data is rarely realistic due to the substantial pragmatic and clinical constraints of repeat tissue biopsy (including multi-region sampling) in live metastatic patients, as well as complications associated with spatial sampling bias (detailed in our Discussion). In the future, analysis of circulating tumor cells (CTCs) may theoretically overcome some of these tissue-specific challenges and offer evidence complementary to ctDNA-based subclonal reconstruction. Unfortunately CTC analysis is currently associated with technical limitations, including variable cell-selection specificity (e.g. choice of enrichment strategy), extremely low CTC count per mL of whole-blood (i.e. 1-10/mL) in patients with metastatic disease, and a high degree of noise—collectively prohibiting their current integration into ctDNA-based subclonal reconstruction paradigms (Alix-Panabières and Pantel, 2021).

Overall, we posit that interrogating active clonal evolution in living patients is presently the most achievable using serial ctDNA. Here we have provided a comprehensive framework for achieving ctDNA-based clonal monitoring, and have exploited this methodology (together with other reconstruction-agnostic pieces of evidence) to reveal clonal selection and evolution in men with metastatic prostate cancer.

Minor comments:

Figure 1:

1b: y-axis of timepoint is unclear (1st/2nd/3rd cfDNA), what is the (average) time between peripheral blood draws / cfDNA extraction per patient?

The y-axis refers to the ordinal timing of sequential cfDNA collections. For the samples prefixed with 'AE', the 1st cfDNA sample is always collected at baseline mCRPC diagnosis prior to systemic mCRPC-therapy initiation. We have reworded the Figure 1 legend to make this more clear:

"The cfDNA samples are shown in three rows to indicate the ordinal timing of sequential cfDNA collections."

We have also added a new Supplementary Table 2 showing the exact amount of time between sequential blood draws for each patient providing >1 cfDNA sample (average time between sequential blood draws is 194 days across the entire cohort).

1d: Legend text does not match y-axis titles.

• E.g., cancer fraction in the legend whilst tumor content (%) is used in figure 1d, total mutation count vs. total variant counts etc.

Thank you. We have revised Figure 1d to use the same nomenclature as its legend.

1d - 2nd track: Please describe which variant categories underly "total variant count"/ " total mutation count"; is this the summation of somatic SNVs, InDels and MNVs? Could a separation be made to distinguish between each class to determine whether there is a change in SNVs/InDels ratio between cfDNA and tissue capture?

Total mutation count includes all somatic SNVs and indels (structural rearrangements are tabulated separately). Adjacent somatic mutations were considered separate, and were not merged into multi-nucleotide variants. Since we use a logarithmic y-axis to accommodate the range of mutation count in our cohort, it is not possible to visually depict the SNV:indel ratio within Figure 1d. The median SNV:indel ratio across all samples was 8.98 (IQR: 7.47-10.61). We have included a new Supplementary Table 1 containing all information plotted in Figure 1d, enabling convenient manual inspection of variant distribution characteristics between sample subsets.

For the reviewer's interest, we elected to exclude the category of MNV due to two potential analytical challenges:

1. **Challenge of determining appropriate criteria for when to collapse adjacent variants into a single MNV. What is an acceptable upper limit of disagreement in mutant read count and overall read depth (i.e. accommodating for noise) between adjacent variants for them to be annotated as a single MNV?**

2. **Relatedly—in the context of sequencing multiple samples from the same patient—even if adjacent variants have identical mutant-read support and VAF in one sample (i.e. strictest possible criteria classifying MNVs), it is possible that in another sample the two variants have starkly different clonal fractions (indicating that they are biologically independent SNVs that occurred as separate somatic events). Erroneously classifying such variants as a single MNV could pose problems for multi-sample subclonal reconstruction and measurements of inter-sample concordance.**

1d: Please highlight the paired samples (tissue + cfDNA) in 1d, this would ease interpretation on the conformality and differences between the tissue and cfDNA profiles per pa-tient. I.e., an additional track denoting (via color or symbol) each of the 13 matched patients and whether the cfDNA was time-matched with the tissue.

Thank you for the suggestion! We agree that being able to visually compare genomic features between same-patient samples is helpful for interpreting these data. We have therefore generated a new Extended Data Figure 1 showing the identical data as per Figure 1d, except now same-patient samples are grouped together. In both the revised Figure 1d oncoprint (ordered by total mutation count) and its new Supplementary version (grouped by patient), we have added patient and sample collection timepoint labels such that same-patient samples can be easily identified.

We have elected to preserve our original Figure 1d since we believe this is visually cleaner (easier to interpret) and better suited for providing a broad cohort-wide overview of our data.

1d - 7th track: >4 and non-integer is not mutually-exclusive. E.g., 4.1 is also >4 and it is unclear whether that will be colored darkred (>4) or grey (non-integer).

It is true that these categories (0, 1, 2, 3, 4, >4, and non-integer) are not mutually exclusive. However, discriminating between integer and non-integer CN values for CN > 4 is challenging because the coverage logratio values corresponding to those CN levels become progressively less separated, whereas coverage noise is multiplicative (and therefore constant in logratio space). In other words, segments with high absolute CN are more likely to be artificially classified as integer (since for this summary plot we allow for a small constant CN buffer around integer values to accommodate sequencing noise and uncertainty in model fitting).

To remedy this apparent contradiction, we have revised the Figure 1d legend to better define these categories: "*Note that the 'non-integer copy number' category only includes non-integer values < 4*".

1d - Track 5,6 and 8: Missing y-axis ticks (i.e., 0%, 50%, 100%).

We have added these missing y-ticks.

• M&M:

o Please specify whether algorithms/tools (e.g., bowtie2) was performed using default settings or specify the non-default settings.

We have now specified the exact command-line arguments used for Bowtie2, cutadapt, and samblaster in the Methods. These details are also captured within our online code supplement (https://github.com/annalam/cfdna-wgs-manuscript-code/).

o Please state how HHR, MMR and CDK12 deficiency status was determined.

We have added the following new paragraph to our Supplementary Information methods:

"Homologous recombination repair defects (HRRd), mismatch-repair defects (MMRd), and CDK12-defects were originally determined from prior targeted exon sequencing data (enabling selection of samples with these subtypes for subsequent WGS). Specifically, we selected samples harboring evidence for putatively deleterious defects in BRCA2 (HRRd), MSH2/MSH6 (MMRd), or CDK12. Mutation pathogenicity was assessed on the basis of predicted protein-level consequences: truncating variants—constituting frameshift insertions/deletions, splice site mutations (mutations within ±2bp from an exon/intron junction), and nonsense (stopgain) mutations—as well as deep deletions (predicated on gene logratio and lack of HSAF deviation) were presumed deleterious. Germline missense mutations classified as 'pathogenic' or 'likely pathogenic' in the ClinVar database were also considered deleterious. Missense variants disrupting the CDK12 kinase domain are linked to protein inactivation and were additionally classified as deleterious if predicted to be at least 'probably damaging' by Polyphen2. Samples with putatively deleterious defects in these genes that were included in our WGS cohort had strong genome-wide evidence for associated phenotypes (including established COSMIC v2.0 mutational signatures for HRRd (signature 3) and MMRd (signatures 6, 15, and 21), as well as the CDK12-associated tandem duplicator phenotype (Wu et al., 2018))."

o Sequence alignment:

Were lanes (if applicable) merged with respective read-groups during alignment?

The sentences are ordered as if the paired-end reads were trimmed using cutadapt and low-quality bases were removed after Bowtie2-alignment? This would be counterintuitive and needed to be done prior to alignment to improve mapping accuracy and allow soft-clipping (of remaining artifacts/low quality bases) by Bowtie2. Please check whether this was the case or if the alignment-step should be placed after specifying the trimming M&M.

We apologize for the confusing wording. The paired-end reads were trimmed using cutadapt and low-quality bases were masked prior to Bowtie2 alignment. We have reworded these sentences in our revised Methods for better clarity (and have included some additional methodological details to reflect your other helpful suggestions).

These procedural steps are also made clear in our expanded code repository (https://github.com/annalam/cfdna-wgs-manuscript-code). For the Reviewer's interest, we have copied the relevant piece of code below that pertains to the question above:

```
# ALIGN READS USING BOWTIE2, MARKING DUPLICATES USING SAMBLASTER
```

```
fasta interleave sample_1.fq.gz sample_2.fq.gz | cutadapt --interleaved -f fastq -m 20 -a
AGATCGGAAGAGC -A AGATCGGAAGAGC - | fasta trim by quality - 30 | bowtie2 -p20 -X 1000 --
score-min L,0,-0.6 --ignore-quals -x hg38 --interleaved - | samblaster | samtools view -u -
| samtools sort -@ 8 -m 4G -o sample.bam
```

An “in-house algorithm” was for base-quality trimming, is this code/algorithm publicly available?

Yes, the code is publicly available. We used the “fasta trim by quality” command found in the Seqkit software package: https://github.com/annalam/seqkit

We have now added this detail to the Methods section.

Please specify how the single-nucleotide polymorphism (SNP) genotype profiles comparison between cfDNA and WBC was performed.

We have added the following description to Supplementary Information section “Validation of sample identities via SNP profile comparison” (page 41):

“To validate sample identities, we determined high confidence genotypes for known human SNPs (GNOMAD database version 3) in each sample. A SNP genotype was considered high confidence if a sample had 30x or higher sequencing depth at the locus, and had an allele fraction between 30 - 70% (heterozygous) or within the ranges 0 - 1% or 99 - 100% (homozygous). Two samples were considered to have a matching SNP profile if at least 99% of high confidence genotypes were in agreement between the samples. SNPs that did not have a high confidence genotype in both compared samples were omitted from analysis.”

The analysis was applied not just between cfDNA and WBC samples, but between all same-patient samples of any kind.

Please specify which gene annotations were used during alignment (if applicable).

Whole genome DNA sequencing reads were aligned against the hg38 human reference genome. No gene annotations were used during the read alignment step, although gene annotations were

later used in predicting the protein-level effects of somatic mutations and chromosomal rearrangements. We now describe the gene annotations used in the relevant Methods sections:

"Protein-level consequences of variants were predicted using ANNOVAR (version 20191024) using ANNOVAR RefGene annotations downloaded on 18th August 2020 (Wang et al. 2010)."

The gene annotations used for somatic structural rearrangement analysis are now also mentioned in our revised Methods paragraph:

"Somatic structural rearrangements were identified in cfDNA and tumor samples using split-read methodology implemented in the Breakfast software version 0.5 (github.com/annalam/breakfast) with the `--max-frag-len=1000 --anchor-len=30` options. A minimum of 5 unique junction-spanning reads were required. We discarded rearrangements with ≥2 supporting reads in the patient-matched leukocyte sample or in any samples from other patients (since rearrangement breakpoint positions are not recurrent between patients). Rearrangement breakpoints were annotated with adjacent Ensembl 84 genes. Rearrangements with breakpoints adjacent to a list of 177 known human cancer genes were collected into a spreadsheet and investigated by aligning the junction sequence using UCSC BLAT."

o Somatic variant calling

Please specify which tools (and settings plus version-numbers) and strategy were current-ly used for germline and somatic variant calling. Currently, the authors only refer to previous literature whilst this encompasses a significant technical element of the manuscript and subsequent results.

We have revised our Methods to specify all software version numbers and settings used for germline and somatic variant calling. The exact mutation calling thresholds and criteria for all variant calling analyses are described in Methods (as well as in our significantly expanded and commented code repository). In the revised Supplementary Information, we have carried out a validation of our somatic mutation calls using the GATK Mutect2 pipeline (see Supplementary Information section "Validation of somatic mutation calls using GATK Mutect2"—as well as our summary of these analyses in response to one of your previous comments).

• A brief inspection of "mutato" reveals that this mostly consists out of pileups without further extensive modeling employed by most industry-standard alter-natives such as GATK Mutect2 / Strelka (for tissue).

Yes, Mutato itself only generates a list of mutation candidates (using very simple and relaxed thresholds), and counts the number of reads supporting those mutations in each sample in a cohort. In other words, it generates a matrix of read level evidence for each mutation candidate, where rows are mutation candidates, and columns are samples.

This read count matrix is then post-processed using the thresholds specified in Methods, to generate the final list of somatic mutations and germline variants. Our mutation analysis pipeline has been designed to utilize a large number of known negative control samples to empirically estimate the real background error rate at each genomic position, rather than utilizing the mathematical error modeling that is normally required if a tumor sample is compared against a single matched normal sample. In our experience, this data-driven approach provides superior results over mathematical approximation of background error.

We have now made the post-processing scripts available in the Github repository for this manuscript.

Please specify which ANNOVAR databases and gene-references (Refseq/UCSC etc.) were used to signify Protein-level consequences.

We used version 20191024 of the ANNOVAR software and its RefGene annotations database, which was downloaded on 18th August 2020. We have added this information to the Methods section.

Please specify if all mutational signatures (COSMIC v2) were used for optimal reconstruc-tion or if a restricted fitting was performed.

All single base-substitution signatures (SBS) in the COSMIC v2 catalog were used for signature inference in all analyses (i.e. no signatures were excluded prior to signature fitting). We used default parameters for DeconstructSigs (which automatically disregards fitted signature weights below 6%). In Figure 1d, only mutation signatures with ≥15% weight are visualized (the remaining are grouped into "other") to focus interpretation on well-supported mutational processes (i.e. disregarding signatures with low weights that are more likely to arise due to factorization error). We have better articulated these details in our revised Methods.

Note that we have added additional mutation signature analyses relevant to Figure 2 (in response to Reviewer #1's comments).

Please state the version of SigProfilerMatrixGenerator.

We used SigProfileMatrixGenerator v1.2 and have revised our Methods to include this detail.

o Structural variant calling

Currently, a non-peer reviewed (unpublished) and relatively unknown software package (breakfast) is used for SV-calling instead of common software packages such as GRIDSS2 . manta / delly etc. The resulting SV calls could therefore be suspect as this is also coupled with an arbitrary threshold such as a min. coverage of 5 supporting junction-spanning reads. Could the authors show concordance of their employed method vs. another established methodology for SV-calling such as GRIDSS2.

To validate the somatic structural variants (SVs) called by Breakfast, as per your suggestion we ran the established SV-calling software suite GRIDSS2 (version 2.10.0; https://github.com/PapenfussLab/gridss) on the plasma cfDNA and metastatic tissue samples of 9 patients in our cohort (9 plasma cfDNA, 2 metastatic tissue, and 8 leukocyte DNA samples), as well as one ctDNA-negative plasma sample (to estimate rate of false-positives).

A head-to-head comparison of the somatic SV calls made by Breakfast and GRIDSS2 (using default parameters) revealed that GRIDSS2 recapitulated 62.4% (1705 of 2729) of the SVs originally identified by our analysis—with similar levels of redetection across all samples (median 63.6%), as well as all classes of SVs as determined by Breakfast (i.e. deletions, inversions, duplications, and translocations). GRIDSS2 identified on average 3.4-fold (range: 1.6-6.6) more SVs than our pipeline (among tumor-positive samples), and identified 273 SVs in the ctDNA-negative control sample (whereas Breakfast identified 9). There were no major differences in software concordance between tissue and cfDNA (median 54.1% vs. 63.6%, p=0.27).

Reassuringly, this degree of concordance appears similar to what is typically achieved in the context of contemporary consensus-SV calling strategies (where the intersection of calls from two or more SV-calling tools are taken as the truth set) (Cameron et al., 2019; Gong et al., 2021). Importantly, the concordance of 62.4% between Breakfast and GRIDSS2 was achieved without any optimization of GRIDSS2 on our data (i.e. only using default GRIDSS2 parameters and following all recommended best practices)—therefore representing a plausible lower limit of technical concordance. We posit that concordance can likely be improved by optimization of GRIDSS2.

Detailed methodology is supplied within our Supplementary Information document (and copied below for the Reviewer's convenience):

"To mitigate against unexpected tool behavior and maximize compatibility with the GRIDSS2 pipeline, we realigned the adapter-trimmed sequencing data to the hg38 reference genome with BWA-mem (version 0.7.15-r1140) using default parameters, and performed duplicate marking with Picard (version 2.18.0). GRIDSS2 was run using default parameters on sample-patient tumor-

normal sample pairs (i.e., joint calling mode as per recommended practices). GRIDSS2 output files were then post-processed using the somatic filtering tool GRIPSS (version 2.0) using default parameters and the default 'Panel of Normals' blacklist. Only SVs labeled as PASS in the FILTER column of the resulting post-GRIPSS VCF files were retained. The only custom filter applied was that we required all somatic SVs to have at least 15 supporting split-reads and be >10bp length (this latter criteria was also applied to Breakfast).

DNA sequence homology between rearrangement breakpoints can make it impossible to localize the chromosomal breakpoints at a single base accuracy. Differences in SV calling methodologies mean that identical true-positive SVs may be reported with slightly different breakpoint coordinates. Therefore to evaluate concordance between Breakfast and GRIDSS2, we performed a fuzzy intersect between same-sample SVs identified by both approaches. Specifically, two SVs were considered matching if both their breakpoint start and end coordinates were within ±6bp of each other (i.e. `abs(SVb_start - SVg_start) ≤ 6)` and `abs(SVb_end - SVg_end) ≤ 6`; where SVb and SVg denote individual SVs as called by Breakfast and GRIDSS2). The reciprocal comparison was also made (i.e. comparing start to end and vice versa) to accommodate potential differences in genome annotation conventions between the two tools."

Finally, we have made a minor technical improvement to our Breakfast SV-calling algorithm (better identification and elimination of duplicate split-reads), reducing our rate of false-positives by ~1-2%. This has been propagated through all analyses.

Tissue-cfDNA mutation concordance

• Is this based only on SNVs, InDels and MNVs or also on the SVs?

We evaluated tissue-cfDNA mutation concordance using SNVs and indels. We did not include MNVs as a category (as explained above), but genuine MNVs (i.e. occurring as a single biological event) would register as two or more separate SNVs (and would be both included in concordance assessment).

• The authors compared concordance of somatic mutations between time-matched cfDNA/tissue using different settings than when calling final list of somatic variants per sample. Instead of using the (arbitrary) >1 reads threshold, it would be more relevant to overlap the final list of somatic variants as derived from cfDNA and tissue sequencing and then compare the overlap. The current analysis does not fully take into account the (potential) differences in detecting / calling somatic variants between cfDNA and tissue (with matched normal) se-quencing and somatic variant calling strategies.

Thank you for the suggestion! There are at least two possible approaches to mutation concordance analysis:

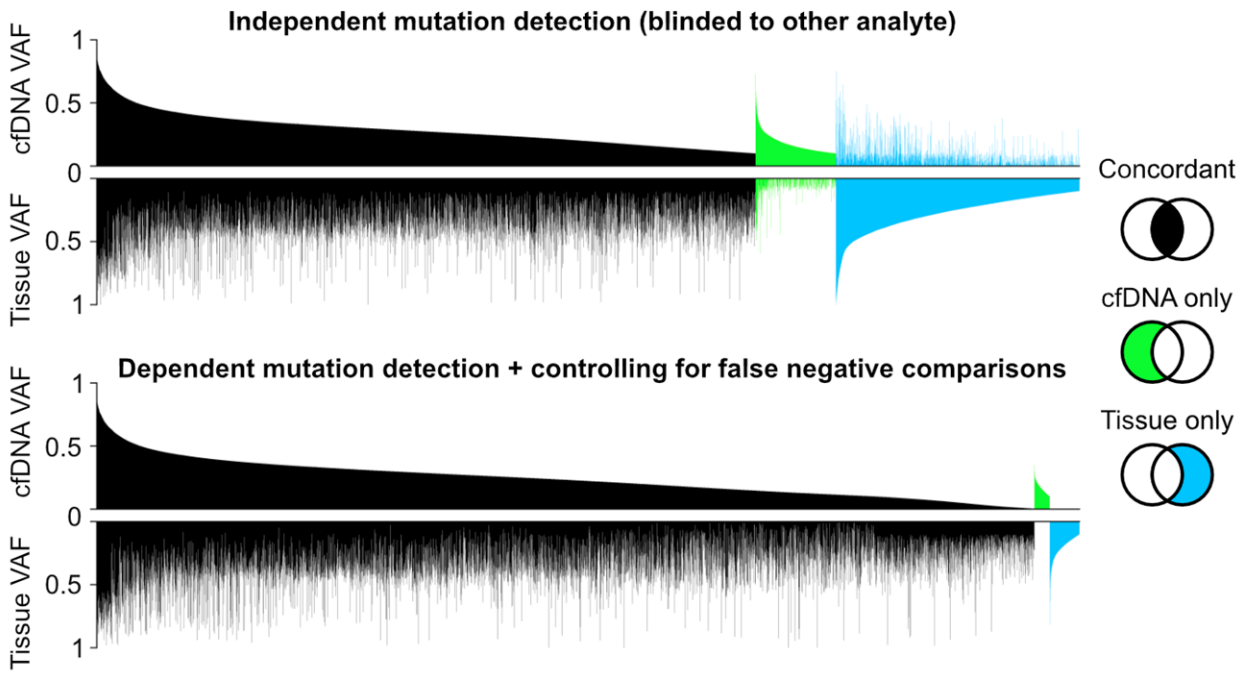
1. **Using identical independent variant calling parameters on both cfDNA and tissue (i.e. requiring ≥8 supporting mutant reads and ≥10% VAF in our data), perform variant calling on same-patient cfDNA and tissue and check overlap of independently detected mutations (as suggested by the reviewer).**

2. **Using identical independent variant calling parameters on both cfDNA and tissue (as above), perform variant calling on same-patient cfDNA and tissue, and among the *union of independently detected same-patient variants*, check for any evidence of mutation presence in the other sample (cfDNA or tissue) where it was not independently identified (our original approach).**

In our view, these two approaches address slightly different questions (which are both important). Approach #1 evaluates concordance from the perspective of assay operating characteristics within a typical clinical genotyping workflow (i.e. if you had only sequenced cfDNA or tissue and were blinded to the other results). Approach #2 is better poised to evaluate true biological concordance (i.e. evidence of shared clonal populations), which was the original objective of our analysis. Note that Approach #2 is analogous to methodology used in the context of minimal-residual disease in other cancers (e.g. bladder, breast, colorectal), whereby *a priori* knowledge of mutation presence (from tissue genotyping) is used to improve limits of detection in cfDNA (i.e. rescuing variants that may otherwise be indiscriminable from background noise) (Coombes et al., 2019; Magbanua et al., 2021; Powles et al., 2021).

Our mutation-based subclonal reconstruction analysis also incorporated all somatic mutations, even if they were only independently detected in a single sample. This is crucial because mutations that only appear in some samples and cell populations provide key information about subclonal dynamics, and therefore it is important to quantify their MAF values even in samples where the mutation was not independently detected.

As per your suggestion, we have performed a new analysis using Approach #1 (see Figure below; top row). 67% of all mutations (n= 62,647 / 93,453) were independently detected in both cfDNA (where n=93,453 is the union of all independently detected mutations across same-patient cfDNA and tissue). 8% (n=7647) and 25% (n=23,159) of mutations were detected in cfDNA but not tissue (and vice versa: tissue but not cfDNA). The reason why many 'discordant' (green and cyan) mutations have nonzero VAFs in both compartments is because our criteria for independent mutation detection also utilizes several additional heuristics/thresholds beyond mutation VAF (e.g. these variants may have harbored low average mapping quality scores, an insufficient number of supporting mutant reads, or the mutant bases may have localized to read ends).



Given that the focus of Figure 3 are biological/populational differences between matched cfDNA and tissue, we have elected to preserve our original Figure 3c.

We agree that it is challenging to comprehensively control for potential differences in somatic variant detection characteristics in tissue versus cfDNA. Nevertheless, in our original analysis we used a statistical framework (described in our Methods) to help mitigate against false negative comparisons between tissue and cfDNA, addressing two considerable sources of differential detection sensitivity (which affect all cross-sample comparisons regardless of analyte type):

1. **Differences in sequencing coverage (including global sequencing depth (i.e. median target depths), as well as copy-number heterogeneity and stochastic variability).**
2. **Differences in per-sample tumor purity.**

Importantly, our new benchmarking of SNVs/Indels called by Mutato against those called by the GATK Mutect2 pipeline (which was designed/tested using tumor tissue) indicate a similar degree of technical concordance regardless of analyte type (cfDNA vs. tissue). This supports the idea that potential differences in intrinsic sample properties between cfDNA and tissue likely have a relatively minimal impact on variant calling performance in our data.

Germline variant calling

• Please specify which tools (or in-house software) was used.

Germline variant calling was carried out by post-processing the pileup matrix generated by Mutato. This is now specified in our revised Methods:
"We used Mutato version 0.7 (https://github.com/annalam/mutato) with arguments --alt-reads=5 --alt-frac=0.05 to generate a table of candidate single nucleotide and indel variants, and their supporting read counts in all samples in the cohort. This table was post-processed using in-house scripts to identify somatic mutations and germline variants."

The source code can be found in the manuscript's code repository.

• Please specify the GNOMAD version used (and add the respective reference).

We used GNOMAD version 3.0 and have specified this detail in our revised Methods:
"Germline variants with population allele frequency 0.5% or higher in the GNOMAD v3.0 database were discarded."

• Please specify which ANNOVAR databases and gene-references (Refseq/UCSC etc.) were used to signify Protein-level consequences.

We used ANNOVAR (version 20191024) with RefGene annotations downloaded on 18th August 2020. We have specified these details in our revised Methods:

"Protein-level consequences of variants were predicted using ANNOVAR (version 20191024) using ANNOVAR RefGene annotations downloaded on 18th August 2020 (Wang et al. 2010)."

• Please specify the ClinVar database (date of access).

Our methods now specify that we used a snapshot of the ClinVar database dated 6th June 2020.

Referee #3 (Remarks to the Author):

Herberts et al analyzed deep whole genome sequencing (WGS) data from plasma cell-free DNA (cfDNA) of patients with metastatic castrate-resistant prostate cancer (mCRPC) and a high circulating tumor DNA (ctDNA) fraction. They created a comprehensive framework for reconstructing subclonal architectures and tumor phylogenies from cfDNA. They found that ctDNA has high subclonal diversity with often a small contribution from a macrometastatic site of disease. Through serial samples over time, they showed the AR locus to the only site of recurrent alterations, reinforcing its critical importance for resistance to AR signaling inhibitors. Additionally, by using cfDNA fragmentation patterns, the authors were able to predict the transcriptomic landscape of the corresponding biopsied site of disease. Taken together, the authors demonstrated that deep WGS of cfDNA can be a powerful tool for understanding the global evolutionary history, state, and trajectory of mCRPC in individual patients,

which may be of significant biological relevance.

This is a comprehensive, relevant, thoughtful, and interesting piece of work from an outstanding research team. To my knowledge, this is the first study to utilize deep, serial WGS of plasma cfDNA to study metastatic prostate cancer (and perhaps in any solid tumor, though I am unfamiliar if similar work has been done in other diseases (e.g. metastatic breast cancer)).

Thank you for the positive feedback!

Through analyzing the genomic alterations observed in ctDNA and applying an evolutionary framework to study subclonal diversity, the authors were able to provide novel insights into the diverse evolutionary trajectories of mCRPC. The study would be strengthened by providing understanding on what the biological/translational relevance of these findings are relative to prior studies in this general domain, and how this approach may fit in compared to other profiling techniques. My comments and suggestions are provided below:

Major comments:

-While the authors do develop a comprehensive framework for inferring subclonal architecture and phylogeny from ctDNA and apply it extensively in this study, they notably do not provide validation of their method (outside of a comparison of their copy number model to ASCAT and Battenberg). It would be informative to see the performance of their method when applied to samples with a known ground truth regarding its subclonal status. Understandably, this may be challenging to do experimentally, though it should be possible in silico through simulated data (e.g. using MASCoTE,

PMID 32879317). Alternatively, explaining the limitations of knowing "ground truth" for this study would be helpful for the reader.

We agree that subclonal reconstruction can be a challenging problem, and validation of identified solutions (models) is key. Our initial manuscript relied on detailed visualizations that illustrate how the raw sequencing coverages, heterozygous SNP allele fractions, and somatic mutation allele fractions (i.e. easily understood, direct measurements) match with the expected values from the proposed subclonal reconstruction. However, our original supplementary figures only provided these visualizations for a handful of samples. In the revised manuscript, we provide a far more complete validation by providing the following:

1. **Truncal copy number model and subclonal reconstruction visualizations for all cfDNA and tissue samples in the cohort.**
2. **Subclonal reconstructions for 22 simulated patients (2 samples each) with known computer-generated subclone constitutions and genomic profiles (investigators were blinded to the ground truth during analysis).**
3. **Physically-based generative simulations demonstrating that our subclonal reconstruction models recapitulate the observed mutation-level data, proving that the inferred models provide a parsimonious explanation of the samples**
4. **Comparisons between our models and:**
    a. **Cancer fraction estimates and CN profiles by ASCAT and Battenberg**
    b. **Subclone compositions by PyClone and PhyloWGS**
5. **A new visualization showing our analysis of DTB-119, a second patient with a sample that contained subclones with a differing WGD status**

These new validation analyses are part of the new 246 page Supplementary Information document.

We have also extended the Discussion to articulate some of the complexities in benchmarking/validating subclonal reconstruction:

"Our study reveals that plasma ctDNA-based subclonal reconstruction is associated with a unique set of challenges and opportunities. Heightened subclonal diversity invalidates many simplifying assumptions on which contemporary subclonal reconstruction methods rely, and regionally varying sequencing depth due to cfDNA nucleosome footprints may pose a challenge for ultra-sensitive copy number analysis. Conversely, serial ctDNA sampling also presents new opportunities, as the increased data dimensionality allows resolution of previously ambiguous scenarios such as distinguishing between a WGD and a 50% CCF subclone (Supplementary Information). Importantly, the fractal nature of clonal evolution means that the level of detail in

reconstructions from bulk sequencing can vary based on sampling characteristics (e.g. sequencing depth, breadth, number of samples).”

-The authors assert that “the variable contribution to total ctDNA from synchronous metastases challenge the notion that global disease features can be precisely inferred through tissue biopsy” (lines 299-300). However, the authors show that, in their 13 pairs of ctDNA with same-day tissue biopsy, there is a high degree of concordance in the mutations detected between cfDNA and biopsy, with only 9 discordant mutations which were protein truncating or in known hotspots out of 94,675 detected in total. Given potential reasons for false negatives (e.g. variability in tumor content and sequencing depth), it is conceivable that the number of discordant, possibly functionally important mutations may be even lower than 9. Additionally, there was also high concordance in the copy number status of key driver genes between ctDNA and biopsy (lines 182-194). Lastly, in paired patient samples, the authors found that the transcriptional landscape of the biopsy site to be strongly correlated with the magnitude of the transcription start site (TSS) nucleosome depletion, suggesting that the transcriptomics of the biopsied site strongly recapitulates the overall transcriptomic landscape of disease in the patient. Hence, it seems that while a single biopsy may not capture all global disease features, it does capture a very large portion of them, particularly those of likely functional (and clinical) importance. Furthermore, as noted in Figure 3F, biopsy of a metastatic site may sometimes be an earlier indicator than cfDNA of important molecular alterations—in this case, AR gene body and AR enhancer copy gains. The authors could consider softening their claim here, especially since the primary clinically relevant event is AR locus related that is a perpetual source of therapeutic development strategies.

Thank you for your detailed points. Our original intention was to emphasize that ctDNA analysis appears to inform a more complete snapshot of global disease genomic features compared to single metastatic biopsy. We recognize that most alterations are shared between subclones within a patient, and our revised discussion now makes this point more clear. However, while the clinical relevance of variables such as subclonal complexity, differential WGD status, and distinct AR genotypes are not yet fully explored, we believe that it is highly plausible that they hold biological relevance.

Our revised discussion paragraph now reads:

“In almost all patients in our cohort, ctDNA was composed of multiple subclones with distinct genomic features, and each biopsied metastasis contributed only a small fraction of total ctDNA (Figure 3a; 3d). Most prostate cancer genomic alterations of established functional importance (e.g. in TP53, PTEN) are early truncal events, and were consequently shared between subclones and between ctDNA and tissue—this is corroborated by the relative homogeneity in driver gene alterations between metastases in published mCRPC autopsy studies. Nevertheless, reliance on a single tissue biopsy may provide an incomplete view of polyclonal treatment resistance

mechanisms (e.g. AR alterations, BRCA2 reversion mutations), misclassify alterations as truncal features, or underestimate disease subclonal complexity."

Regarding the note from the reviewer "*Given potential reasons for false negatives (e.g. variability in tumor content and sequencing depth)…*". We agree that differences in tumor purity and sequencing coverage (both overall coverage, coverage differences as a consequence of intersample copy-number heterogeneity, and stochastic variability) can impact mutation detection characteristics. However, in comparing mutation overlap between cfDNA and matched tissue, we used a statistical approach to help control for these confounders (reducing the likelihood of false negative comparisons). Please see our reply to comment #14 from Reviewer 1 for a more detailed methodological breakdown (also included in the manuscript Methods).

More broadly, can the authors comment more on the biological or clinical relevance to distinguishing "precise" disease features vs. those identified from a single tumor biopsy, especially if they ultimately point to AR (albeit with subtleties about the specific molecular events in AR)?

Our data from the 33 patients in this study suggests that somatic alterations to the *AR* gene remain the major individual genomic mechanism through which resistance arises against contemporary AR pathway inhibitors such as abiraterone and enzalutamide. Parallel positive selection of resistant clones within multiple metastatic foci creates disproportionally high degrees of subclonal heterogeneity within these key genomic loci (such as *AR*). Awareness of the full repertoire of resistance mutations within each patient (which is not necessarily achievable via profiling single tissue foci) may affect therapeutic decision making, especially as new therapies targeting the AR pathway continue to emerge (e.g. AR disruptors, AR NTD inhibitors).

The different classes of *AR* alteration (e.g. pathogenic LBD hotspot mutations, LBD-truncating structural variants, gene and/or enhancer copy amplification)—together with the considerable heterogeneity within each of these classes and potential for combinatoriality—likely equate to complex associations with treatment outcomes that are currently incompletely understood. There are many common clinical observations for which we do not have a compelling genomic/mechanistic explanation, but where high-resolution ctDNA profiling promises to help clarify. For example, interrogating precise genomic features of metastatic disease may help us understand per-patient heterogeneity in response/progression to different classes of standard-of-care AR-targeted therapies. Beyond cataloging resistance mechanisms to specific targeted therapies, ctDNA offers a global lens through which to view metastatic disease, capturing variables like subclonal complexity which single tissue foci will typically underestimate. Importantly, we believe that our study promotes ctDNA analysis from a predominantly clinical-oriented tool for cancer driver genotyping, to a more nuanced approach that can help researchers understand fundamental properties of metastatic biology.

We have better addressed these considerations throughout our revised Discussion.

-Through analysis of serial, patient-matched samples, the authors found the AR locus to be the only of recurrent alterations. However, their methods for calling somatic mutations vary for AR vs non-AR regions, with the former requiring variants with ≥8 supporting reads and a variant allele fraction (VAF) of ≥10%, while for AR gene hotspot mutations, they only required 3 mutant reads, VAF >1%, and ≥15 times background error rate. It would be informative to see if the AR locus still remains the only site of recurrent alterations if the conditions for calling non-AR somatic alterations were relaxed to be comparable to those used for calling AR alterations.

We only used relaxed calling criteria for reporting the *AR* mutations in Figure 4d and Supplementary Table 7. Importantly, these criteria were not used in the genome-wide analyses shown in Figure 4a-b. We have clarified this in the Methods:

"For Figure 4d and Supplementary Table 7 showing AR mutations, we required only 3 mutant reads with an allele fraction VAF >1% and ≥15 times the background error rate. These relaxed thresholds were only used for visualization, not in calculations of mutation frequencies and temporal changes within the cohort."

Somatic rearrangements and copy number alterations were always called using uniform criteria throughout the genome.

-Regarding the central role of AR locus alterations, how do the authors reconcile this observation with existing and emerging combination therapeutic strategies that leverage other means of inhibition (e.g. AR + PI3K inhibitors)?

All patients in our study had previously received treatment with one or more drugs inhibiting AR axis signaling—including a backbone of androgen deprivation therapy (ADT; as per standard of care) plus in some patients subsequent potent AR pathway inhibitors for mCRPC. The consecutive cfDNA samples analyzed in Figure 4 were collected before and after monotherapy treatment with AR signaling inhibitors abiraterone or enzalutamide. For this reason, we did not intend to make any claims about the role of *AR* locus alterations in driving clonal expansions outside this specific (albeit very common) therapeutic context. We anticipate that if the second active drug in these combination therapies is sufficiently effective, a combination of acquired resistance-driving alterations will emerge, or a different pathway will be perturbed to circumvent both drugs. The

exact genomic outcomes should prove specific to each drug combination. We agree that these are very important points that must be made clear, and have therefore reworded elements of our Discussion (reiterating that our results are in the context of AR-targeted monotherapy) and have added the following new Discussion text in our revised manuscript:

"Clinical suppression of AR-signaling is essential for mCRPC management—all drugs are administered on a backbone of continuous androgen-deprivation, and emerging standard-of-care and investigational agents (e.g. ^{177}Lu–PSMA-617, AR protein degraders, AR N-terminal domain inhibitors) offer renewed strategies for exploiting dependence on AR-signaling. This constant selective pressure underscores the importance of understanding nuanced variation in AR genomic resistance mechanisms and underlying (sub)clonal diversity. More broadly, genomic mediators of resistance to clinically-active targeted therapies are likely coupled to the specific pathways under selective pressure, implying that ctDNA can nominate distinct mechanisms of acquired resistance in other cancers and therapeutic contexts (e.g. PI3K-pathway or PARP inhibition in prostate and breast cancer)."

Similarly, framing the overarching findings from ctDNA WGS relative to prior autopsy/multiregional or serial tissue based studies (e.g. Kumar et al Nature Med 2016, Gundem et al Nature 2015) would be helpful for the reader.

(Kumar et al., 2016) and (Gundem et al., 2015) exploit multi-lesion metastatic tissue profiling to characterize intrapatient spatial heterogeneity in mCRPC tissues collected during rapid autopsy. These articles broadly harbor the complementary conclusion that key prostate cancer driver genotypes (including deleterious *TP53* and *PTEN* defects, 8p loss / 8q gain, *TMPRSS2-ERG* fusions) are concordant between same-patient metastatic lesions (implying they are truncal to all mCRPC populations). However, Kumar et al. 2016 and Gundem et al. 2015 differ in their conclusions regarding intrapatient AR genotype heterogeneity (with Kumar et al. 2016 advocating for relatively limited AR heterogeneity, while Gundem et al. 2015 offers evidence to the contrary and further suggests that AR-axis alterations are relatively late/subclonal events).

Although the primary objective of our Figure 3 is to describe the genomic relationship between synchronous cfDNA and metastatic tissue, our data also carry implications about intralesion heterogeneity that are compatible with several of the conclusions proffered by Kumar et al. 2016 and Gundem et al. 2015. Our observation that cfDNA and tissue harbor a large reservoir of shared mutations (Figure 3c)—coupled with the highly similar copy number of established prostate cancer driver genes (*TP53*, *PTEN*; Figure 3b)—suggest relative genomic driver homogeneity between lesions. Therefore, the majority of interlesion genomic diversity is likely attributable to relatively unknown drivers and/or passenger events that rise to clonal prevalence due to founder effects associated with metastatic seeding or invasion into new anatomic niches. The notable exception to this was the *AR* locus, which displayed pervasive heterogeneity in copy-number amplicon

structure between cfDNA and synchronous matched tissue (similarly to what Gundem et al. 2015 show between individual same-patient metastatic lesions in their Figure 4c. This contrasts with the conclusion offered by Kumar et al. 2016 (suggesting that *AR* genomic aberrations are generally concordant within individuals). However, this difference is at least partially attributable to Kumar et al. 2016 treating *AR*-genotype as a noncontinuous variable (i.e. discrete categories of gain versus amplification as determined by array comparative genomic hybridization, as well as mutation presence versus absence from whole-exome sequencing). Nevertheless, Kumar et al. 2016 show that several patients do contain discordant *AR*-genotype status between same-patient metastatic lesions.

Critically, our work is distinguished from Kumar et al. 2016 and Gundem et al. 2015 in we profile live mCRPC patients receiving active systemic therapy in an earlier disease context (predominantly first- and second-line mCRPC), contrasted with the rapid autopsy profiling of Kumar et al. 2016 and Gundem et al. 2015. Samples collected at rapid autopsy likely reflect cancer populations after a period of time unconstrained by systemic anti-cancer drugs (since patients are typically switched to palliative/comfort-care prior to death). Differences in fitness landscape therefore complicate the interpretation of rapid autopsy studies for earlier clinically-controllable disease.

Importantly (in contrast to Kumar et al. 2016 and Gundem et al. 2015), we also focus on treatment-driven genome-wide evolution occurring across time in patients receiving potent AR-inhibitor therapies. At time of resubmission there are no other studies to deploy ultra-deep serial ctDNA WGS in living patients as they respond to and progress on standard-of-care therapy. A key conclusion of our serial profiling is that intra-lesion heterogeneity is ultimately dynamic and highly reflective of current therapeutic context. Understanding how and why populations expand and contract over time is essential for optimizing future treatment strategies. These insights are not possible to glean from interrogation of a single timepoint nor via rapid autopsy profiling.

We have included several new lines in our Discussion to address these concepts (and place our work in context of Kumar et al. 2016 and Gundem et al. 2015).

-Regarding Figure 5E, it is surprising that there exists such a strong correlation between cfDNA TSS nucleosome depletion and gene expression for unmatched cfDNA and metastatic site biopsy. I wonder how much of this correlation may be because cancers in general up/downregulate some common transcriptional programs. As a negative control, it would be helpful to see a graph of cfDNA TSS nucleosome depletion of this cohort vs gene expression percentile for different, unrelated carcinomas, e.g. bladder cancer.

The result in Figure 5e represents a broad analysis of ~20,000 gene transcription start sites (TSS). Therefore even though housekeeping and hematopoiesis-associated genes are excluded, the overall transcriptomic profile will be broadly similar between mCRPC patients, and will overlap with that of other epithelial phenotypes, especially those with similar lineage maintenance programs. The current language in our results notes that we are reporting a relationship with "*global transcriptomic patterns in metastatic lesions*". In our revised Figure 5e, we have changed the inset legend "*tumor-specific genes*" to "*non-housekeeping and non-hematopoietic genes*" to make clear that we are not claiming a patient- or lesion- specific relationship at this point (*c.f.* later in the result sub-section where we interrogate nucleosome occupancy at AR transcription factor binding sites in the context of per-sample genomic features).

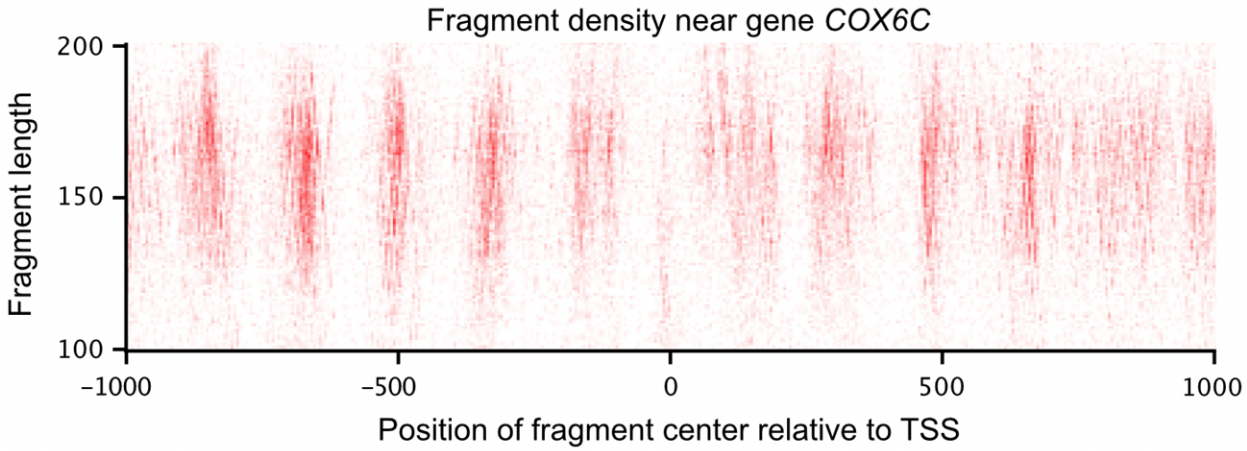
A core element of novelty of our cfDNA fragmentomic data is highlighted by Figure 5h, where we exploit nucleosome depletion to measure AR transcription factor binding activity as a continuous variable, and connect this phenotype information to AR-genotype with per-sample resolution. The considerable range in AR transcription-factor binding activity between patients—as well as absence of activity in patients with histologically-confirmed neuroendocrine disease—nominates cfDNA fragmentomic profiling as a strategy for evaluating mCRPC androgen-signaling reliance. Translationally, this may be useful for therapeutic stratification (e.g. choosing between AR pathway inhibitors or broad chemotherapies), as well as for early real-time detection of treatment resistance via neuroendocrine transdifferentiation (i.e. transition to AR-indifferent disease).

Nevertheless, an essential goal for future research will be to optimize techniques for inferring transcriptional activity from TSS cfDNA fragmentomic features. Excitingly, a variety of methodologies are being actively developed, variably utilizing read-depth depletion, windowed protection scores, fragment-end positions, fragment-size diversity, fragment-end motifs, and other information sources. Our original manuscript focused on read-depth depletion and the windowed protection score for quantifying cfDNA TSS fragmentomic features (as well as broader expression-linked nucleosome depletion across entire gene-bodies).

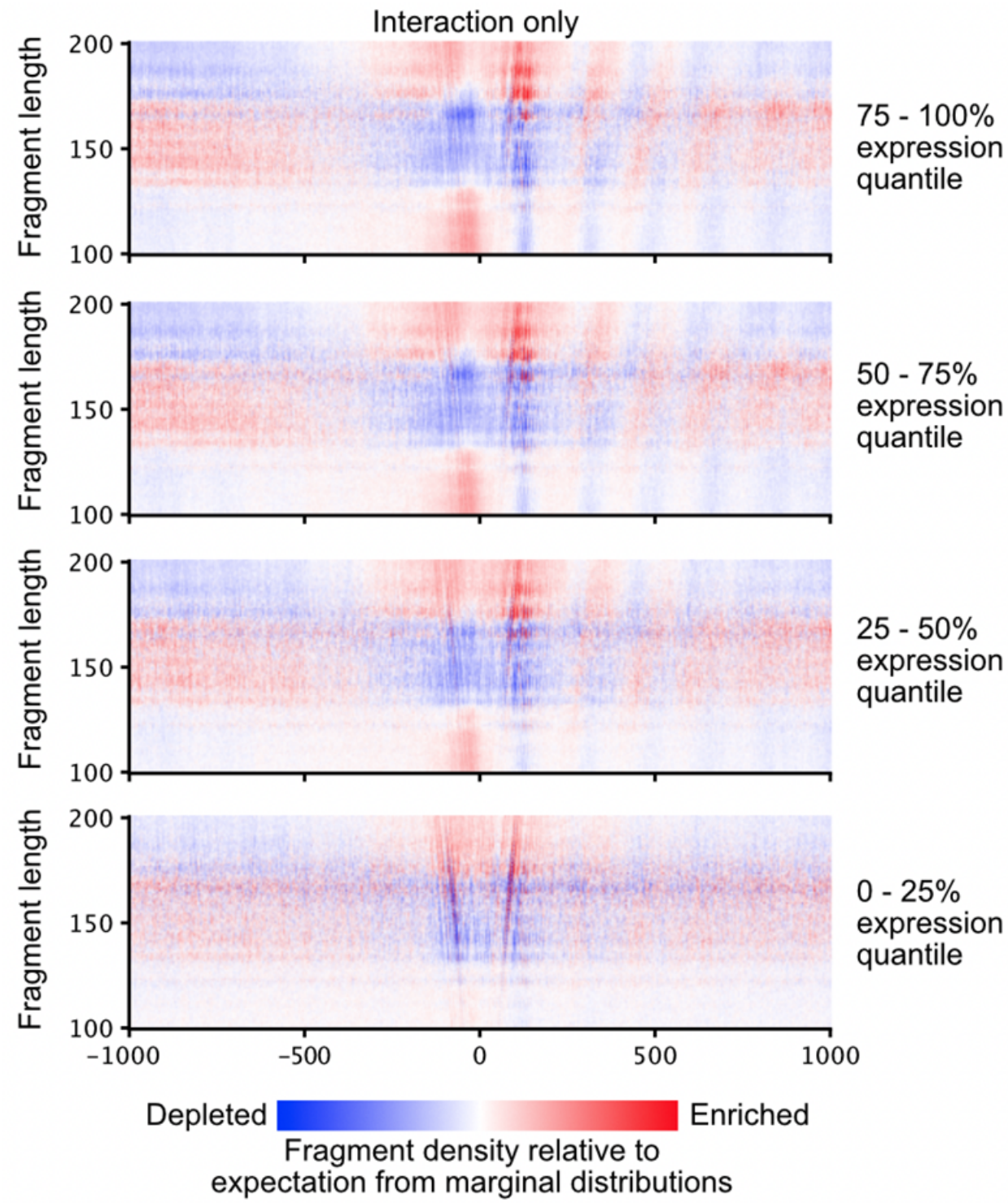
To this end, we have been investigating a new potential strategy that leverages joint analysis of cfDNA fragment position and size characteristics. Specifically, we quantified the distribution of cfDNA fragment lengths as a function of positional coordinate, around the TSS of one or more genes. Importantly, this new matrix representation (coined the 'nucleogram') captures all fragment-length and fragment-position information present in a given sample, at single-base resolution. Prior methodologies in this space have only attempted to exploit one (but not both) of these variables (Cristiano et al., 2019; Esfahani et al., 2022; Snyder et al., 2016).

Shown below is the fragment-density nucleogram for *COX6C* (ranked within the top 10% expressed genes in mCRPC metastatic tissue RNA-Seq), aggregated across 63 mCRPC cfDNA samples in our cohort. The periodic high-density streaks demarcate regions of high nucleosome

occupancy (i.e. greater overall fragment density). Fascinatingly, at the TSS there is both an overall dearth of fragments (i.e. suggestive of nucleosome depletion) in addition to a pronounced decrease in modal fragment size.



Expanding upon this observation, we can leverage these nucleograms to reveal complex interactions between cfDNA fragment length and genomic position (relative to the TSS or other positions). Visualized below is the interaction between cfDNA fragment position and fragment length, where we can see an enrichment of short (~100bp) cfDNA fragments at the TSS, and an enrichment of larger (>167bp) cfDNA fragments at the adjacent -1 and +1 nucleosomes (but with decaying enrichment extending away from the TSS). These new nucleogram plots are shown as part of our new Extended Data Figure 4 (and copied below for your convenience).

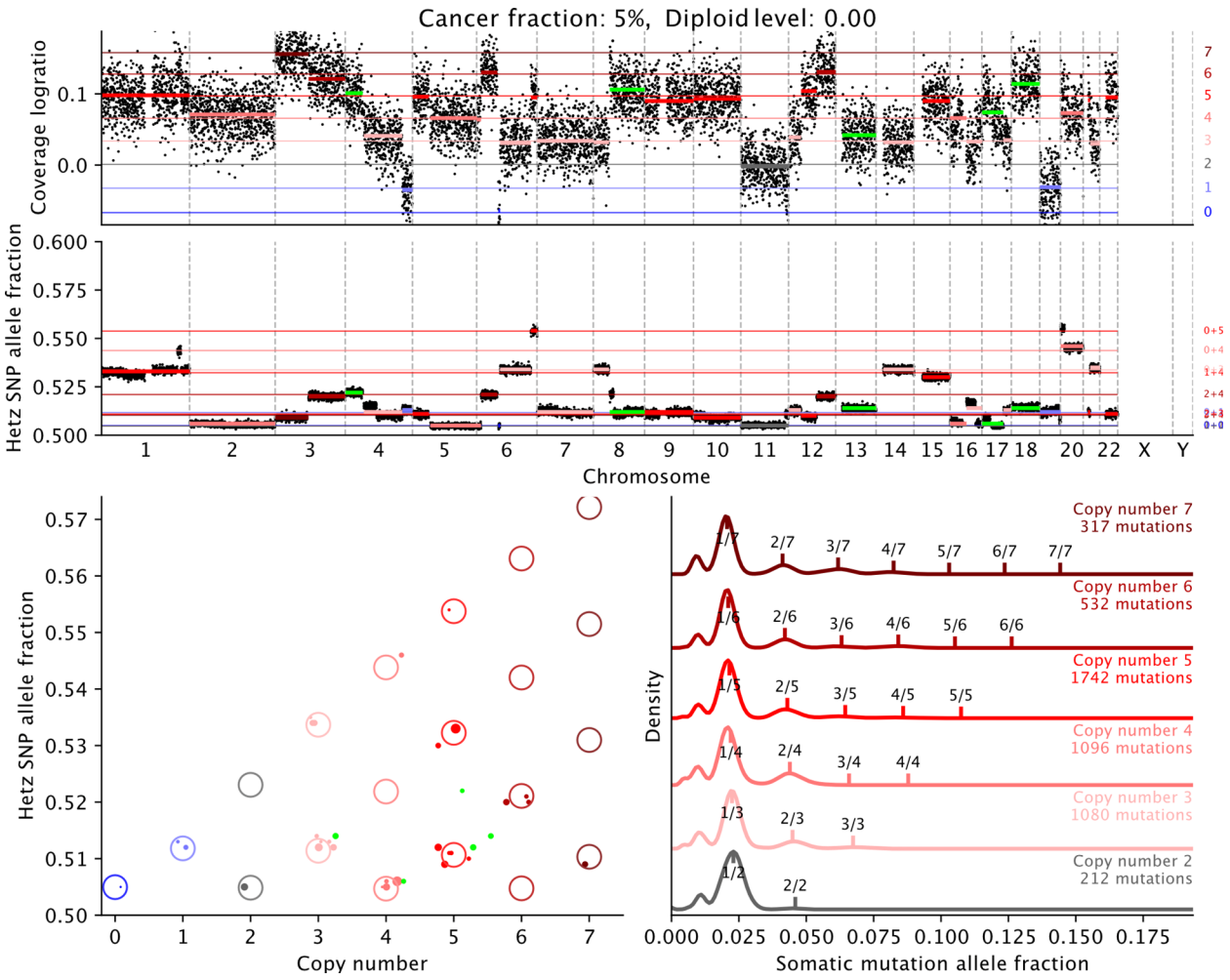


We anticipate this novel integration of complementary fragmentomic features (fragment-length and spatial density) will be important for future methodological developments focused on quantifying gene expression from cfDNA TSS features—especially with the goal of achieving sample- and gene-level resolution. Importantly, the high depth (~200x) of our cfDNA WGS bolsters our capacity to resolve per-sample TSS trends (using mCRPC metastatic tissue as a benchmark), both here and in our original Figure 5 data. This is more challenging using low-pass WGS (lpWGS) (~0.5-10x depth), where quantification of biological TSS features often necessitates aggregating

multiple samples together. To help overcome this limitation for other researchers, we have made our sequencing data publicly available through the European Genome-Phenome Archive. Together with the rich genomic annotation included in our Supplementary Tables and new nucleogram data/methods, we anticipate our data will be an unparalleled resource for future exploratory analyses of cfDNA fragmetomics. We expect that emerging tools designed for existing lpWGS datasets (e.g. GRIFFIN—medarXiv preprint https://doi.org/10.1101/2021.08.31.21262867) will be able to extract additional biological insight using our deep cfDNA WGS samples, where the high depth will be advantageous for both biological discovery and technical/algorithmic development (e.g. characterization of tool sensitivity/specificity via coverage downsampling).

-It is noteworthy that the median ctDNA fraction in the study cohort was 48% (range 17-88%), and many of the methods applied here require a high ctDNA fraction. However, as demonstrated previously, the ctDNA fraction in mCRPC can be quite low (e.g. in the abiraterone-enzalutamide vs enzalutamide-abiraterone phase 2 trial which comprised part of the patient population here, 43% of the cohort had ctDNA fraction as <2%, and another 36% had 2-30%; PMID 29367197). As such, while the authors contend that "our work advances ctDNA profiling from an emerging tool for detection of selected clinically-actionable gene mutations, towards a modality of genome-scale discovery and deep clinical-biological insight," it is unclear whether deep WGS can be feasibly applied to the study of cancer in clinical states where the ctDNA fraction is low, which represents the majority of mCRPC patients. Can the authors reframe this point?

ctDNA-fraction (i.e. tumor purity) and sequencing depth are two critical determinants of the resolution of information that can be obtained via cfDNA WGS (and indeed any sequencing assay) (Herberts and Wyatt, 2021). For the detection of discrete tumor-specific features (i.e. somatic mutations, methylation markers, etc.) and analyses contingent on these data, low ctDNA-fraction can be overcome by increasing sequencing depth and methods such as duplex consensus error correction (Schmitt et al., 2012). Therefore, boosting sequencing depth can improve the lower ctDNA-fraction limit of detection for subclonal reconstruction (where accurate mutation allele frequencies are critical inputs). For example, we expect it is already technically feasible to perform ~5000x depth WGS on plasma cfDNA to enable genome-wide analysis of samples with ctDNA fractions as low as 5% (in contrast to the median ctDNA fraction of ~45% in our cohort). To illustrate this, we generated an *in silico* simulated sample with 5% tumor content whole-genome sequenced at a depth of ~5000x (with realistic amounts of noise for coverage logratios and heterozygous SNP allele fractions). From this example, we can see that the cancer fraction and truncal copy numbers can be solved (figure attached below).



At present, the only prohibitive factor (with respect to application of subclonal reconstruction across a broader range of ctDNA-fraction) is sequencing cost. Given the continual decreases in sequencing cost, we posit that the applicability of ctDNA whole-genome sequencing analysis in patient populations will continue to increase.

Nevertheless, a limiting factor for subclonal reconstruction will be the minimum ctDNA fraction required to infer allele-specific copy number landscape. Although increased depth will reduce overall noise and improve copy-number resolution, we anticipate there will be a fundamental lower limit of ctDNA-fraction below which whole genome analysis cannot reliably be performed from an conventional clinically-pragmatic blood sample (in contrast to collecting large quantities of blood or exploiting plasmapheresis). This is because sequencing depth is fundamentally limited by the number of fragmented haploid cfDNA genomes present in a blood sample. Secondly, there may be sources of noise beyond pure mathematical sampling noise that cannot be overcome with additional sequencing depth. Another limitation concerns the detection of biological features that are inseparably admixed with normal cfDNA (e.g. nucleosome-related TSS coverage depletion), and the cfDNA contribution of benign tumors or other expanded cell populations (e.g. clonal hematopoiesis) that carry somatic mutations which cannot be distinguished from cancer

mutations. We expect that these factors will significantly limit the applicability of whole genome cfDNA analysis at ctDNA fractions < 1%.

To address these considerations, we have added a line to the Discussion highlighting that the current technical characteristics of utilized approaches theoretically enable biological deep resolution on samples with lower ctDNA fraction (aided by increased sequencing depth)—i.e. these analyses can be made applicable to a much broader group of patients:

"We anticipate that the diminishing cost of WGS will enable higher sequencing depths (>1000×) in future studies, unlocking similar genome-wide clonal analysis in plasma samples with ctDNA fractions as low as 5%"

Minor comments:

-Lines 150-152: Can the authors provide citations on methods that behave this way? This would be helpful for the reader

Subclonal whole genome duplication events pose a problem both for software packages designed for inferring a single sample's copy number profile (e.g. ASCAT, Battenberg, HATCHET), and software packages designed for inferring subclone fractions and phylogenetic trees (e.g. PyClone, PhyloWGS).

Established software packages for sample copy number profile estimation suffer from the following problems:

- **Battenberg is designed to detect genomic regions that have a different copy number state in two or more clones present in a sample. However, Battenberg makes an explicit assumption that among all subclones present in a sample, there are only two distinct copy number states, and those copy number states differ by a maximum of one copy of one allele. Quoting from Nik-Zainal et al. (2012) where the Battenberg method was first described:**
  ***"To estimate τ, the fraction of tumor cells with the specific subclonal copy number change, we further assume that the subclonal copy number change resulted in a gain or a loss of exactly one copy of one allele"*. If a sample is an admixture of a WGD and non-WGD clone, most genomic regions will not be compatible with this starting assumption, and will be incorrectly analyzed.**

- **The HATCHET software can only detect WGDs that are clonal and involve a single duplication of the genome. Higher degrees of genome duplication and subclonal WGDs are not currently supported. Quoting from Zaccaria et al. (2020) where the HATCHet method was described:**
  "Third, HATCHet's modeling of WGD could be further generalized. While recent pan-cancer studies show that the current assumptions used in HATCHet (namely that a WGD occurs at most once as a clonal event and that additional clonal CNAs also occur) are reasonable for most tumors, HATCHet's model could be extended to allow for multiple WGDs (e.g. hexaploid or higher ploidy), subclonal WGDs, or WGDs occurring without any other clonal CNAs."
- **ASCAT assumes that each sample is composed of only a single cancer cell population (clone), and does not support subclonal CN heterogeneity.**

Established software packages for subclone fraction and phylogenetic tree inference suffer from a different set of problems:

- **The PyClone input format requires an integer copy number state to be provided for each genomic region with a mutation. If a sample contains two or more subclones with a different copy number state within some genomic region, any mutations from those regions cannot be included in the input file. If a sample is an admixture of a WGD and non-WGD clone, all genomic regions will have a subclonally heterogeneous copy number state and will thus be invalid for PyClone analysis.**
- **The PhyloWGS input format is designed to support mutations in genomic regions with sublonal CN heterogeneity, but since there are limitations in the software packages designed for detecting a sample's CN profile (listed above), the input data for carrying out the analysis is not available.**

-Figure 1E: Can the authors provide a definition of the subclonal CN diversity score?

We have added additional text to our expanded Supplementary Information defining the subclonal copy-number diversity score:

"Subclonal copy-number heterogeneity was quantified as the difference in segment absolute copy number and its nearest integer value (i.e. `abs(round(CN) - CN)`***), averaged over all same-sample segments produced from truncal copy number fitting."***

Note that a similar technique for measuring subclonal heterogeneity is also explored in detail in van Dijk et al. 2021, Nature Communications (PMID: 34045449) (which we have cited).

-Figure 2C: Can the authors provide a definition of copy segments?

The y-axis label "copy segments" was referring to the number of genomic segments harboring a copy number state that differs from their flanking regions. To increase clarity, we have relabeled the y-axis as "Copy number segments". Somatic copy number alterations that affect only a part of a chromosome give rise to segments with distinct copy number states, and these segments can be identified computationally (the methodology for identifying segments is extensively described in our new Supplementary Information document; see Supplementary Figure 1 and associated text (pages 3-6)). If a cancer cell's genome has undergone many somatic copy number alteration events throughout its evolution, the number of chromosomal segments with locally distinct copy number states will be higher. In Figure 2C, this is evident in the samples harboring a *CDK12* double mutant genotype, which is known to cause a severe increase in the rate of small tandem duplication events.

-Figure 3D: For the two patients with a high %ctDNA contribution (patient 183 and 205), did the authors consider reasons why this may be? I am curious whether this is a function of a low overall disease burden (hence a single macrometastatic site of disease contributes a significant amount to the ctDNA subclonal diversity) or whether there is a high degree of homogeneity across metastatic sites of disease in the patient.

We were also very interested in exploring potential clinical variables that might have influenced the percent contribution of single biopsied metastases to bulk ctDNA. We agree with your speculation that single lesion biopsies in the context of low-volume metastatic disease may contribute a larger fraction to bulk ctDNA (in contrast to patients with high disease burden). Similarly, the size of the biopsied lesion may also influence its relative ctDNA contribution.

Patients DTB-183 and DTB-205 had the lowest overall plasma ctDNA fractions (29% and 28%, respectively) among all of the patients that had a matched tissue available. They were also the only two patients whose tissue biopsies were obtained from a lymph node metastasis. This loosely supports the hypothesis that these two patients had an overall low tumor burden, and their biopsied metastases were shedding the majority of ctDNA. However, since this association is largely anecdotal, we have chosen to omit it from the revised manuscript.

Unfortunately we were not able to obtain clinical imaging data for further exploration of these hypotheses. Patients in our cohort that provided matched metastatic tissue and cfDNA were accrued under the SU2C/WCDT clinical study protocol (NCT02432001) which formally completed in

2020, without continued funding for retrospective clinical data collection (and most of these patients are now deceased).

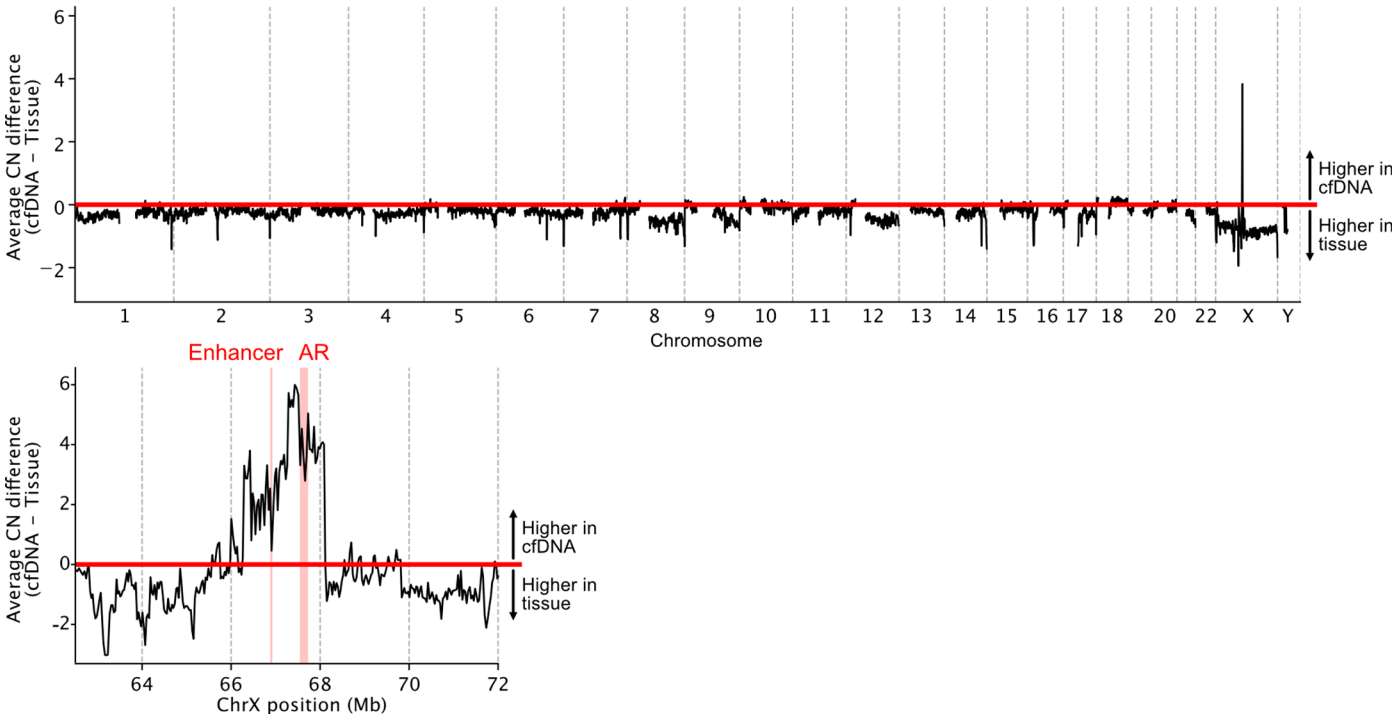
Nevertheless, we felt it was important to bring these interesting clinical questions to the reader's attention via the following Discussion text:

"Future work will need to understand how metastatic site (e.g. bone versus soft tissue) and size relative to total tumor burden influences contribution to ctDNA, and whether emerging functional imaging (e.g. PSMA PET-CT, [^{18}F]FDG-PET) can help select lesions likely to reflect clinically-dominant disease (Abbosh et al., 2017; Unterrainer et al., 2020)."

-In Figure 3E, the authors quantify the average absolute copy number difference between biopsy and cfDNA. It would be interesting to see the signed difference in copy number—is it that the biopsy site tends to more likely have a copy number gain compared to the cfDNA, which is a composite of many sites of disease in the patient?

We have replotted a version of Figure 3e showing the signed copy number (CN) difference between cfDNA and tissue (genome-wide plot shown above; AR-locus shown below). The genome-wide CN profile is generally less than zero due to the large difference in ploidy between the tissue and cfDNA of patient DTB-119 (whose tissue sample was hexaploid in contrast to diploid cfDNA). Most of the sharp downward spikes in the genome-wide plot we see at the chromosome ends are likely due to poorly aligned sequences in telomeres and longer fragment lengths in tissue samples.

Overall we feel that 13 cfDNA-tumor sample pairs is not a sufficient number to robustly interrogate signed CN differences. Since individual genomic loci show cfDNA-tumor differences in only a few samples, the null hypothesis that the CN alterations have no underlying preference for cfDNA or tumor cannot be rejected. Therefore we have decided not to add this new analysis to our revised manuscript.



-Figure S1D: The y-axis label for the middle row should not have the % sign. It should read "Heterozygous SNP allele fraction."

Thank you for pointing this out—we have corrected this typo.

-Figure S10: Which patient is being used as an example here?

It is patient AE-018. We have now communicated this in the new Supplementary Information, where the original S10 was moved:

"Below we demonstrate an example of how we detected a subclonal WGD event in the plasma cfDNA samples of patient AE-018, through a process of eliminating simpler hypotheses."

-Lines 244-245: The author state "Tissue mRNA-abundance was weakly correlated with cfDNA nucleosome depletion in the first exon-intron junction." Can the authors explain this statement further since it appears that r=0.98 for this correlation?

In the manuscript we intended to point out that the slope of this line (i.e. cfDNA nucleosome depletion at first intron-exon junction) was shallower relative to the TSS nucleosome depletion response curve (also plotted on Figure 5d in gray). The Pearson correlation coefficient associated with both curves was similarly high. We apologize for the incorrect wording and have revised this paragraph to read:

"Tissue mRNA-abundance was also strongly correlated with cfDNA nucleosome depletion at the first exon-intron junction (Zhu et al., 2021) (Figure 5d), although the magnitude of effect (relative to the TSS) was weaker and exhibited an association with first exon length."

The overarching objective of this analysis (Figure 5d; as well as Figure 5f) was to indicate that the relationships between cfDNA nucleosome depletion and tissue mRNA expression are potentially complex insofar as varying between specific chromatin features (e.g., TSS, TF binding sites, 1st exon-intron junction, TTS, etc.).

-Line 830-831: Do the authors mean 5, rather than 4, non-cancerous cfDNA samples?

Yes, the correct number is 5 non-cancerous cfDNA samples, and we have now corrected the text. Thank you for spotting the error.

-Line 836: What is considered the background error rate in this case?

We defined the background error rate of a mutation as the total number of mutant allele reads, divided by the total number of any allele reads, across a set of control samples including leukocyte and ctDNA-negative cfDNA samples. We have made the definition of this term more clear in the revised Methods:

"To remove germline SNPs, common sequencing or alignment errors, and mutations of clonal hematopoietic origin, we required the mutant allele fraction to be at least 50× higher than the background error rate (i.e. the average fraction of the mutant allele across all germline control samples, including 39 leukocyte samples and 5 non-cancerous cfDNA samples)"

References

Abbosh, C., Birkbak, N.J., Wilson, G.A., Jamal-Hanjani, M., Constantin, T., Salari, R., Le Quesne, J., Moore, D.A., Veeriah, S., Rosenthal, R., et al. (2017). Phylogenetic ctDNA analysis depicts early-stage lung cancer evolution. Nature *545*, 446–451.

Aggarwal, R., Huang, J., Alumkal, J.J., Zhang, L., Feng, F.Y., Thomas, G.V., Weinstein, A.S., Friedl, V., Zhang, C., Witte, O.N., et al. (2018). Clinical and Genomic Characterization of Treatment-Emergent Small-Cell Neuroendocrine Prostate Cancer: A Multi-institutional Prospective Study. J. Clin. Oncol. *36*, 2492–2503.

Alexandrov, L.B., PCAWG Mutational Signatures Working Group, Kim, J., Haradhvala, N.J., Huang, M.N., Ng, A.W.T., Wu, Y., Boot, A., Covington, K.R., Gordenin, D.A., et al. (2020). The repertoire of mutational signatures in human cancer. Nature *578*, 94–101.

Alix-Panabières, C., and Pantel, K. (2021). Liquid Biopsy: From Discovery to Clinical Application. Cancer Discovery *11*, 858–873.

Annala, M., Taavitsainen, S., Khalaf, D.J., Vandekerkhove, G., Beja, K., Sipola, J., Warner, E.W., Herberts, C., Wong, A., Fu, S., et al. (2021). Evolution of castration-resistant prostate cancer in ctDNA during sequential androgen receptor pathway inhibition. Clin. Cancer Res.

Bielski, C.M., Zehir, A., Penson, A.V., Donoghue, M.T.A., Chatila, W., Armenia, J., Chang, M.T., Schram, A.M., Jonsson, P., Bandlamudi, C., et al. (2018). Genome doubling shapes the evolution and prognosis of advanced cancers. Nat. Genet. *50*, 1189–1195.

Cameron, D.L., Di Stefano, L., and Papenfuss, A.T. (2019). Comprehensive evaluation and characterisation of short read general-purpose structural variant calling software. Nat. Commun. *10*, 3240.

Caravagna, G., Heide, T., Williams, M.J., Zapata, L., Nichol, D., Chkhaidze, K., Cross, W., Cresswell, G.D., Werner, B., Acar, A., et al. (2020). Subclonal reconstruction of tumors by using machine learning and population genetics. Nat. Genet. *52*, 898–907.

Coombes, R.C., Page, K., Salari, R., Hastings, R.K., Armstrong, A., Ahmed, S., Ali, S., Cleator, S., Kenny, L., Stebbing, J., et al. (2019). Personalized Detection of Circulating Tumor DNA Antedates Breast Cancer Metastatic Recurrence. Clin. Cancer Res. *25*, 4255–4263.

Cristiano, S., Leal, A., Phallen, J., Fiksel, J., Adleff, V., Bruhm, D.C., Jensen, S.Ø., Medina, J.E., Hruban, C., White, J.R., et al. (2019). Genome-wide cell-free DNA fragmentation in patients with cancer. Nature *570*, 385–389.

Dentro, S.C., Leshchiner, I., Haase, K., Tarabichi, M., Wintersinger, J., Deshwar, A.G., Yu, K., Rubanova, Y., Macintyre, G., Demeulemeester, J., et al. Characterizing genetic intra-tumor heterogeneity across 2,658 human cancer genomes.

Deshwar, A.G., Vembu, S., Yung, C.K., Jang, G.H., Stein, L., and Morris, Q. (2015). PhyloWGS: reconstructing subclonal composition and evolution from whole-genome sequencing of tumors. Genome Biol. *16*, 35.

Eisenberg, E., and Levanon, E.Y. (2013). Human housekeeping genes, revisited. Trends in Genetics

29, 569–574.

Esfahani, M.S., Hamilton, E.G., Mehrmohamadi, M., Nabet, B.Y., Alig, S.K., King, D.A., Steen, C.B., Macaulay, C.W., Schultz, A., Nesselbush, M.C., et al. (2022). Inferring gene expression from cell-free DNA fragmentation profiles. Nat. Biotechnol. *40*, 585–597.

Espiritu, S.M.G., Liu, L.Y., Rubanova, Y., Bhandari, V., Holgersen, E.M., Szyca, L.M., Fox, N.S., Chua, M.L.K., Yamaguchi, T.N., Heisler, L.E., et al. (2018). The Evolutionary Landscape of Localized Prostate Cancers Drives Clinical Aggression. Cell *173*, 1003–1013.e15.

Gong, T., Hayes, V.M., and Chan, E.K.F. (2021). Detection of somatic structural variants from short-read next-generation sequencing data. Brief. Bioinform. *22*.

Gundem, G., Van Loo, P., Kremeyer, B., Alexandrov, L.B., Tubio, J.M.C., Papaemmanuil, E., Brewer, D.S., Kallio, H.M.L., Högnäs, G., Annala, M., et al. (2015). The evolutionary history of lethal metastatic prostate cancer. Nature *520*, 353–357.

Herberts, C., and Wyatt, A.W. (2021). Technical and biological constraints on ctDNA-based genotyping. Trends Cancer Res. *7*, 995–1009.

Khalaf, D.J., Annala, M., Taavitsainen, S., Finch, D.L., Oja, C., Vergidis, J., Zulfiqar, M., Sunderland, K., Azad, A.A., Kollmannsberger, C.K., et al. (2019). Optimal sequencing of enzalutamide and abiraterone acetate plus prednisone in metastatic castration-resistant prostate cancer: a multicentre, randomised, open-label, phase 2, crossover trial. The Lancet Oncology *20*, 1730–1739.

Kumar, A., Coleman, I., Morrissey, C., Zhang, X., True, L.D., Gulati, R., Etzioni, R., Bolouri, H., Montgomery, B., White, T., et al. (2016). Substantial interindividual and limited intraindividual genomic diversity among tumors from men with metastatic prostate cancer. Nat. Med. *22*, 369–378.

Labrecque, M.P., Coleman, I.M., Brown, L.G., True, L.D., Kollath, L., Lakely, B., Nguyen, H.M., Yang, Y.C., da Costa, R.M.G., Kaipainen, A., et al. (2019). Molecular profiling stratifies diverse phenotypes of treatment-refractory metastatic castration-resistant prostate cancer. J. Clin. Invest. *129*, 4492–4505.

Liu, L.Y., Bhandari, V., Salcedo, A., Espiritu, S.M.G., Morris, Q.D., Kislinger, T., and Boutros, P.C. (2020). Quantifying the influence of mutation detection on tumour subclonal reconstruction. Nat. Commun. *11*, 6247.

López, S., Lim, E.L., Horswell, S., Haase, K., Huebner, A., Dietzen, M., Mourikis, T.P., Watkins, T.B.K., Rowan, A., Dewhurst, S.M., et al. (2020). Interplay between whole-genome doubling and the accumulation of deleterious alterations in cancer evolution. Nat. Genet. *52*, 283–293.

Magbanua, M.J.M., Swigart, L.B., Wu, H.-T., Hirst, G.L., Yau, C., Wolf, D.M., Tin, A., Salari, R., Shchegrova, S., Pawar, H., et al. (2021). Circulating tumor DNA in neoadjuvant-treated breast cancer reflects response and survival. Annals of Oncology *32*, 229–239.

Malikic, S., Jahn, K., Kuipers, J., Sahinalp, S.C., and Beerenwinkel, N. (2019). Integrative inference of subclonal tumour evolution from single-cell and bulk sequencing data. Nat. Commun. *10*, 2750.

Oudard, S., Fizazi, K., Sengeløv, L., Daugaard, G., Saad, F., Hansen, S., Hjälm-Eriksson, M., Jassem, J., Thiery-Vuillemin, A., Caffo, O., et al. (2017). Cabazitaxel Versus Docetaxel As First-Line Therapy for Patients With Metastatic Castration-Resistant Prostate Cancer: A Randomized Phase III Trial—

FIRSTANA. Journal of Clinical Oncology *35*, 3189–3197.

Pomerantz, M.M., Li, F., Takeda, D.Y., Lenci, R., Chonkar, A., Chabot, M., Cejas, P., Vazquez, F., Cook, J., Shivdasani, R.A., et al. (2015). The androgen receptor cistrome is extensively reprogrammed in human prostate tumorigenesis. Nat. Genet. *47*, 1346–1351.

Powles, T., Assaf, Z.J., Davarpanah, N., Banchereau, R., Szabados, B.E., Yuen, K.C., Grivas, P., Hussain, M., Oudard, S., Gschwend, J.E., et al. (2021). ctDNA guiding adjuvant immunotherapy in urothelial carcinoma. Nature *595*, 432–437.

Quigley, D.A., Dang, H.X., Zhao, S.G., Lloyd, P., Aggarwal, R., Alumkal, J.J., Foye, A., Kothari, V., Perry, M.D., Bailey, A.M., et al. (2018). Genomic Hallmarks and Structural Variation in Metastatic Prostate Cancer. Cell *175*, 889.

Roth, A., Khattra, J., Yap, D., Wan, A., Laks, E., Biele, J., Ha, G., Aparicio, S., Bouchard-Côté, A., and Shah, S.P. (2014). PyClone: statistical inference of clonal population structure in cancer. Nat. Methods *11*, 396–398.

Salcedo, A., Tarabichi, M., Espiritu, S.M.G., Deshwar, A.G., David, M., Wilson, N.M., Dentro, S., Wintersinger, J.A., Liu, L.Y., Ko, M., et al. (2020). A community effort to create standards for evaluating tumor subclonal reconstruction. Nat. Biotechnol. *38*, 97–107.

Schmitt, M.W., Kennedy, S.R., Salk, J.J., Fox, E.J., Hiatt, J.B., and Loeb, L.A. (2012). Detection of ultra-rare mutations by next-generation sequencing. Proc. Natl. Acad. Sci. U. S. A. *109*, 14508–14513.

Snyder, M.W., Kircher, M., Hill, A.J., Daza, R.M., and Shendure, J. (2016). Cell-free DNA Comprises an In Vivo Nucleosome Footprint that Informs Its Tissues-Of-Origin. Cell *164*, 57–68.

Sun, R., Hu, Z., Sottoriva, A., Graham, T.A., Harpak, A., Ma, Z., Fischer, J.M., Shibata, D., and Curtis, C. (2017). Between-region genetic divergence reflects the mode and tempo of tumor evolution. Nat. Genet. *49*, 1015–1024.

Tarabichi, M., Salcedo, A., Deshwar, A.G., Ni Leathlobhair, M., Wintersinger, J., Wedge, D.C., Van Loo, P., Morris, Q.D., and Boutros, P.C. (2021). A practical guide to cancer subclonal reconstruction from DNA sequencing. Nat. Methods *18*, 144–155.

Turajlic, S., Xu, H., Litchfield, K., Rowan, A., Chambers, T., Lopez, J.I., Nicol, D., O'Brien, T., Larkin, J., Horswell, S., et al. (2018). Tracking Cancer Evolution Reveals Constrained Routes to Metastases: TRACERx Renal. Cell *173*, 581–594.e12.

Turajlic, S., Sottoriva, A., Graham, T., and Swanton, C. (2019). Resolving genetic heterogeneity in cancer. Nat. Rev. Genet. *20*, 404–416.

Ulz, P., Thallinger, G.G., Auer, M., Graf, R., Kashofer, K., Jahn, S.W., Abete, L., Pristauz, G., Petru, E., Geigl, J.B., et al. (2016). Inferring expressed genes by whole-genome sequencing of plasma DNA. Nat. Genet. *48*, 1273–1278.

Unterrainer, M., Eze, C., Ilhan, H., Marschner, S., Roengvoraphoj, O., Schmidt-Hegemann, N.S., Walter, F., Kunz, W.G., af Rosenschöld, P.M., Jeraj, R., et al. (2020). Recent advances of PET imaging in clinical radiation oncology. Radiation Oncology *15*.

Wu, Y.-M., Cieślik, M., Lonigro, R.J., Vats, P., Reimers, M.A., Cao, X., Ning, Y., Wang, L., Kunju, L.P.,

de Sarkar, N., et al. (2018). Inactivation of CDK12 Delineates a Distinct Immunogenic Class of Advanced Prostate Cancer. Cell *173*, 1770–1782.e14.

Zhu, G., Guo, Y.A., Ho, D., Poon, P., Poh, Z.W., Wong, P.M., Gan, A., Chang, M.M., Kleftogiannis, D., Lau, Y.T., et al. (2021). Tissue-specific cell-free DNA degradation quantifies circulating tumor DNA burden. Nat. Commun. *12*, 1–11.

Reviewer Reports on the First Revision:

Referees' comments:

Referee #1 (Remarks to the Author):

The revised manuscript Herberts et al comprehensively addresses numerous requests for clarification, correction and validation during the previous round of review. Although this revision had a combined length of manuscript, extended data, rebuttal letter, and supplementary information exceeding the number of pages of Dostoevsky's Crime and Punishment, such a lengthy revision was neither a "crime" nor a "punishment" to this reviewer, who thoroughly appreciated the attention to detail the authors made when drafting their revision. Of particular note, I was not the only reviewer who expressed concern regarding the robustness or performance of the in-house algorithms used by the authors. A substantial portion of the supplementary data, which I have reviewed, contains these extensive validation studies. I commend the authors on their efforts to compare their methodology transparently with previously published approaches.

I have the following minor concerns and questions:

1) In Figure 5, panel h, the ranking and separation of samples (rows) is not clear in the mCRPC portion. What is being compared between groups at p=.003? Or is this the gene/enh copy number correlation? It is (still) neither obvious nor intuitive how these are grouped/compared.

2) In Extended data 2a, it took me some time to understand why the tissue and cfDNA had similar copy numbers despite amplicon structure. In the legend, consider being more explicit with what the data shows -- for example, the broad vs. focal amplifications, and that you are considering a gene body copy number of approximately 3 vs. approximately 2 to be similar.

3) Further explanation of the validation steps described in the Supplement would be helpful. For example, "Validation" is shown against ASCAT and Battenberg (Supp Fig 14) but is disclaimed as a comparison between unsupervised algorithms. In the case of discordance, is this because manual adjustment was performed using the authors' own algorithms and these tools cannot be adjusted for WGD? Can the authors explain reasons for discordance with GRIDSS2 (3.4-fold more SVs with GRIDSS2 in tumor-positive samples)? Reasons for discordance are appropriately explained for PyClone and PhyloWGS.

4) In the supplementary data (page 32) a Supplementary Table number is missing ("X").

5) I really enjoyed reviewing the individual case reconstructions in the Supplement. Please pay close attention to some of the annotations or edits as they overlap with the figures in some panels. For example, "DTB-175 mutation clustering" has text overlapping with the colored bars, and "Patient DTB-205" has text overlapping with the allele density curves.

Referee #2 (Remarks to the Author):

The answers of the authors to the comments/question raised have been very extensive and adequate. I would like to thank the authors for providing such a detailed rebuttal. Upon the second reading of the manuscript I am now convinced that the methodologic rigor that should be present in such a paper has been met. Moreover, the impact of this paper is impressive. It opens up a methodology to study the impact of treatment on the cancer evolution using a simple blood draw and therefore is able to transform the way we treat our patients. The main problem in medical

oncology to date is the emergence of resistance to treatment and understanding the dynamics of this process in real life is crucial in making the next steps. This paper represents a next step into that direction. I strongly recommend Nature to publish this paper.

Referee #3 (Remarks to the Author):

Overall, I greatly appreciate the thoughtful and detailed consideration of my comments and feedback (along with those of the other reviewers), the team should be commended for such extensive revisions. The main findings supported by this study (as I understand it) include deep sequencing/analysis to reveal complex subclonal genomic architecture of CRPC, differences between tissue biopsies and ctDNA, and fragmentomics to infer transcriptional features, with associated methods that may be broadly useful for the cancer genomics community. These findings are then placed into context and what they might mean for subsequent clinical consideration of AR directed therapies (as well as other approaches). I have a few additional minor comments and feedback for certain aspects, summarized below:

Method Validation:
The authors make a commendable and comprehensive effort at validating the results of their subclonal and phylogeny reconstruction work which does provide overall reasonable reassurance regarding the general validity of their methodology. There are a couple of specific points I would like to raise:

1. Under "Validation of subclonal reconstruction using a blinded simulation" (page 32), the authors note that they generated 22 patient cases with 2 samples each, with each sample carrying 2-3 cancer cell subpopulations and note excellent concordance between their inferred models and the simulated data. However, in the cfDNA samples, the authors noted cases with 4-5 subpopulations (AE-030 and AE-098). It may provide additional reassurance if the authors are able to demonstrate that their methods produce comparably high degrees of concordance on simulated data of 4-5 subpopulations.
2. To the authors' point regarding regionally varying sequencing depth of cfDNA due to nucleosome footprints, it would appear that this effect is relatively modest (median change in logratio of approximately -0.1 for the 90th percentile of expressed genes based on Figure 5g) and likely only significantly affects focal copy number alterations (as averaging of gene body coverage with surrounding flanking regions decreases this effect). The authors may want to note these points in their Discussion excerpt above.

ctDNA conclusions:
The potential advantages of ctDNA that the authors noted here are noteworthy, though (1) the point regarding the detection of polyclonal treatment resistance mechanisms is not necessarily fully supported by their data and (2) the clinical implications of misclassifying alterations as truncal events or underestimating subclonal complexity are not clear. The authors may wish to consider these points further while recognizing the unexplored potential of comprehensive ctDNA profiling.

TSS analyses:
I would like to thank the authors for the clarification. It appears from Figure 5e that cfDNA TSS nucleosome depletion is able to provide a lineage and context-specific recapitulation of the transcriptomics of metastases (e.g. in this manuscript, cfDNA from a group of mCRPCs is able to recapitulate the metastases' transcriptomics from a separate cohort of mCRPCs), presumably as cancers from the same lineage and context exhibit many similarities in the transcriptionally active programs. Additionally, for AR, the authors are able to demonstrate in one patient the utility of cfDNA footprinting in identifying changes in global transcriptomics.
1. In the Discussion, the authors note that "cfDNA footprinting at gene TSSs can recapitulate metastatic tissue mRNA expression…" (lines 351-352). It may be worthwhile to clarify that this

occurs largely on a lineage and context-specific level as the authors do not systemically analyze detailed patient-level concordance.

2. The authors also note that "we show that cfDNA sequencing enables parallel epigenomic and comprehensive genomic profiling without the need for additional assays (e.g. ATAC-seq, methyl-seq)" (lines 354-355). Without systematic comparisons between these currently complementary techniques, the authors may wish to reconsider this point.

Additional minor comments:
1. Figure 5d legend: the legend within Figure 5d appears mislabeled, with "First exon-intron junction r = 0.89" when it should be 0.98 based on the initial submission?

2. On page 32 of the Supplementary Information, the authors refer to Supplementary Table X. I suspect this is a mistake as X should denote a number (e.g. S1, S2, etc).

Author Rebuttals to First Revision:

Response to reviewers

Referee #1 (Remarks to the Author):

The revised manuscript Herberts et al comprehensively addresses numerous requests for clarification, correction and validation during the previous round of review. Although this revision had a combined length of manuscript, extended data, rebuttal letter, and supplementary information exceeding the number of pages of Dostoevsky's Crime and Punishment, such a lengthy revision was neither a "crime" nor a "punishment" to this reviewer, who thoroughly appreciated the attention to detail the authors made when drafting their revision. Of particular note, I was not the only reviewer who expressed concern regarding the robustness or performance of the in-house algorithms used by the authors. A substantial portion of the supplementary data, which I have reviewed, contains these extensive validation studies. I commend the authors on their efforts to compare their methodology transparently with previously published approaches.

Thank you for the positive feedback!

I have the following minor concerns and questions:

1) In Figure 5, panel h, the ranking and separation of samples (rows) is not clear in the mCRPC portion. What is being compared between groups at p=.003? Or is this the gene/enh copy number correlation? It is (still) neither obvious nor intuitive how these are grouped/compared.

The cfDNA samples (rows) are ordered by the magnitude of read-depth depletion (of the central AR binding site (ARBS) position relative to flanking regions). The annotated p-value reflected a Mann-Whitney U test of AR copy number in a top quantile of samples (ranked by ARBS depletion) versus a bottom quantile—however we have now elected to remove this statistical annotation on the Figure since the quantile dichotomization is somewhat arbitrary (and the effect of higher AR gene/enhancer copy number in the samples with highest ARBS depletion is already visually clear). This relationship is also more precisely quantified in Extended Data Figure 3g.

We have clarified the ordering of samples (rows) in the revised Manuscript Figure 5 caption:

"Per-sample (row) cfDNA read-depth depletion at 3224 AR binding sites (ARBS) ordered by magnitude of nucleosome depletion."

2) In Extended data 2a, it took me some time to understand why the tissue and cfDNA had similar copy numbers despite amplicon structure. In the legend, consider being more explicit with what the data shows -- for example, the broad vs. focal amplifications, and that you are considering a gene body copy number of approximately 3 vs. approximately 2 to be similar.

This figure legend now includes the following additional text: "*In this patient, the tissue sample shows a focal amplification of the AR enhancer region and a broad copy gain of the entire locus, while the cfDNA sample shows a more complex pattern of apparent nested copy gains. Despite these contrasting structures, the overall AR enhancer and gene body copy numbers are relatively similar (considering that the range of AR gene body absolute copy number across all samples was 1 - 91).*"

3) Further explanation of the validation steps described in the Supplement would be helpful. For example, "Validation" is shown against ASCAT and Battenberg (Supp Fig 14) but is disclaimed as a comparison between unsupervised algorithms. In the case of discordance, is this because manual adjustment was performed using the authors' own algorithms and these tools cannot be adjusted for WGD?

Our own algorithm for truncal copy number model fitting proposes up to three different models with different degrees of WGD, with a built-in preference for lower ploidy models unless a higher ploidy model explains the data significantly better. Since all three algorithms (ours, ASCAT, and Battenberg) only perform single-sample analysis, a human armed with knowledge garnered from other same-patient samples can improve the model selection. In this study, our primary goal was to produce maximally accurate models for our WGS samples, and therefore we curated all of the automatically generated models, and selected one of the alternative WGD state models for 11 samples, and in the case of 2 samples (AE-018 and DTB-119) we rejected the automated result and carried out a more detailed manual inference (the samples described in Supplementary Methods section "Detection of subclonal whole genome duplication events").

The models reported by ASCAT and Battenberg were not manually curated. We compared their automated result against our curated results to demonstrate that for the majority of samples (primarily the ones with relatively simple models), the results were concordant. For all cases where our curated model and the Battenberg model differed, we show the competing models in our Supplementary section "Battenberg copy number models", for full transparency. For brevity we do not show the ASCAT models, since Battenberg is effectively a more refined version of ASCAT.

Since our own models were manually curated, we felt it was important to clarify in the Methods section that we were not carrying out a comparison between two unsupervised algorithms.

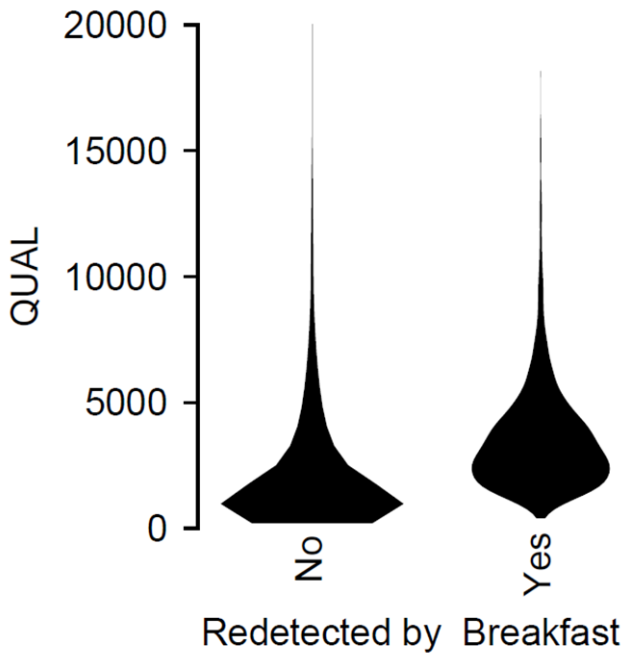
We have now clarified the section in the Supplementary Information / Supplementary Methods document:

"This comparison is shown for validation purposes, and is not intended as a comparison between unsupervised algorithms, since a few of the models reported by our algorithm were manually adjusted based on information obtained from same-patient samples (particularly with respect to whole genome duplication)."

Can the authors explain reasons for discordance with GRIDSS2 (3.4-fold more SVs with GRIDSS2 in tumor-positive samples)? Reasons for discordance are appropriately explained for PyClone and PhyloWGS.

We interrogated the features of structural variants (SVs) called by GRIDSS2. Broadly speaking, GRIDSS2 SVs that failed to be recapitulated by Breakfast expectedly had a lower QUAL (i.e. quality) score (versus SVs successfully redetected by Breakfast). QUAL is the sum of individual quality metrics for each evidence supporting variable

utilized by GRIDSS2 (e.g. counts of breakpoint-supporting assembly contigs, total number of split/soft clipped/indel-containing reads, number of discordant read pairs, etc.).



Further investigation revealed that 50.1% of the somatic rearrangements found by GRIDSS2 were not unique to a single patient (i.e. GRIDSS2 had detected them in two or more patients). Because it is extremely unlikely that tumors from two patients would actually carry somatic rearrangements with identical breakpoint coordinates, we conclude that at least half of the GRIDSS2 detected rearrangements are false positives. Since the GRIDSS2 validation was only carried out on samples from 8 patients, this percentage is likely an underestimate, and the actual false positive rate is higher. GRIDSS2-assigned QUAL scores were much lower for the rearrangements that were not unique to a single patient (median 1301 vs 1902, p = 5.85e-41, MWU), lending further support to them being false positives.

Since GRIDSS2 does not output a complete list of supporting reads for the rearrangements it detects, we cannot provide a detailed analysis of potential reasons for the GRIDSS2 false positives.

4) In the supplementary data (page 32) a Supplementary Table number is missing ("X").

Thank you for spotting this. We have now corrected it to reference Supplementary Table 5.

5) I really enjoyed reviewing the individual case reconstructions in the Supplement. Please pay close attention to some of the annotations or edits as they overlap with the figures in some panels. For example, "DTB-175 mutation clustering" has text overlapping with the colored bars, and "Patient DTB-205" has text overlapping with the allele density curves.

We have now scanned through the supplement and adjusted overlapping annotations.

Referee #2 (Remarks to the Author):

The answers of the authors to the comments/question raised have been very extensive and adequate. I would like to thank the authors for providing such a detailed rebuttal. Upon the second reading of the manuscript I am now convinced that the methodologic rigor that should be present in such a paper has been met. Moreover, the impact of this paper is impressive. It opens up a methodology to study the impact of treatment on the cancer evolution using a simple blood draw and therefore is able to transform the way we treat our patients. The main problem in medical oncology to date is the emergence of resistance to treatment and understanding the dynamics of this process in real life is crucial in making the next steps. This paper represents a next step into that direction. I strongly recommend Nature to publish this paper.

Thank you for the feedback and positive recommendation.

Referee #3 (Remarks to the Author):

Overall, I greatly appreciate the thoughtful and detailed consideration of my comments and feedback (along with those of the other reviewers), the team should be commended for such extensive revisions. The main findings supported by this study (as I understand it) include deep sequencing/analysis to reveal complex subclonal genomic architecture of CRPC, differences between tissue biopsies and ctDNA, and fragmentomics to infer transcriptional features, with associated methods that may be broadly useful for the cancer genomics community. These findings are then placed into context and what they might mean for subsequent clinical consideration of AR directed therapies (as well as other approaches). I have a few additional minor comments and feedback for certain aspects, summarized below:

Method Validation:

The authors make a commendable and comprehensive effort at validating the results of their subclonal and phylogeny reconstruction work which does provide overall reasonable reassurance regarding the general validity of their methodology. There are a couple of specific points I would like to raise:

1. Under “Validation of subclonal reconstruction using a blinded simulation” (page 32), the authors note that they generated 22 patient cases with 2 samples each, with each sample carrying 2-3 cancer cell subpopulations and note excellent concordance between their inferred models and the simulated data. However, in the cfDNA samples, the authors noted cases with 4-5 subpopulations (AE-030 and AE-098). It may provide additional reassurance if the authors are able to demonstrate that their methods produce comparably high degrees of concordance on simulated data of 4-5 subpopulations.

We have now carried out additional testing where we generated 4 simulated patient cases, each with 3 samples and 5 cancer cell subpopulations arranged in a randomly generated phylogenetic tree. Blinded analysis of these simulated patient cases identified a correct truncal copy number model for all 12 samples, and a correct phylogenetic tree for all 4 patients. Identified subclone fractions were highly accurate (average absolute error 1.3 percentage units). Copy number deconvolution analyses were also accurate despite the higher number of subpopulations, with an average 99.1% of the genome assigned with the correct copy number state in each clone.

2. To the authors’ point regarding regionally varying sequencing depth of cfDNA due to nucleosome footprints, it would appear that this effect is relatively modest (median

change in logratio of approximately -0.1 for the 90th percentile of expressed genes based on Figure 5g) and likely only significantly affects focal copy number alterations (as averaging of gene body coverage with surrounding flanking regions decreases this effect). The authors may want to note these points in their Discussion excerpt above.

It is true that the bias in gene body coverage logratio between the first and last expression deciles is approximately -0.1. However, in samples with low ctDNA fractions, this can correspond to CNA of many copies. Furthermore, the bias may be much larger than |0.1| for very highly/lowly expressed individual genes (our analysis of bulk RNA only captures the average effect of ~1000s of genes binned together by their median expression across many different mCRPC tissue samples). Importantly, nucleosome-mediated coverage depletion is *always* present in cfDNA regardless of ctDNA fraction, but the complement of genes/regions affected may differ as a function of ctDNA fraction, as well as tumor-specific transcriptional program and copy number features.

While surrounding regions could be used to reduce the amount of bias by calculating the gene CN based on the gene body + flanks, this would result in an incorrect CN measurement if a CNA only affected the gene body, or if a CNA affected one or both gene flanks without affecting the gene itself.

ctDNA conclusions:

The potential advantages of ctDNA that the authors noted here are noteworthy, though (1) the point regarding the detection of polyclonal treatment resistance mechanisms is not necessarily fully supported by their data and (2) the clinical implications of misclassifying alterations as truncal events or underestimating subclonal complexity are not clear. The authors may wish to consider these points further while recognizing the unexplored potential of comprehensive ctDNA profiling.

Thank you for raising these additional considerations:

1. **The frequent observation of compound AR resistance alterations in individual cfDNA samples (e.g. Figure 2h) implies some degree of polyclonal selection. Specifically, the often discordant intra-sample clonal prevalences of these compound alterations suggests the presence of ≥2 metastatic populations with distinct AR genotypes. Evidence for potential polyclonal resistance is made more explicit in the context of analysing serial ctDNA samples (e.g. Figure 4), where we see examples of co-occurring AR resistance alterations emerging at clinical progression on AR targeted therapies.**

   Nevertheless, we agree that the differential fitness landscape of AR alterations (including possible combinatorial configurations) and their precise clinical relevance is not yet fully established. Therefore we have therefore added an additional reference to the relevant Discussion paragraph (Quigley et al. 2017; Cancer Discovery), illustrating an example where ctDNA sequencing provides an ostensibly more complete picture of a patient's polyclonal *BRCA2* resistance alterations (i.e. an established example of polyclonal resistance to PARP inhibitors) than synchronous metastatic tissue biopsy.

2. **For the Reviewer's interest, one theoretical implication of misclassifying site-specific mutations as truncal features could arise in the context of tissue-informed minimal residual disease (MRD) profiling. Recently acquired**

mutations that have undergone clonal expansion (due to the founder effect in context of metastatic seeding) may have a similar clonal fraction to genuinely truncal mutations (present in all metastatic cancer populations). Tracking these site-specific expanded mutations in ctDNA over time would fail to inform on temporal changes in global disease burden (and instead only inform on the dynamics of the individual lesion genotyped and all its clonally-related descendent populations). Realistically, risk of misclassification (in selecting mutations appropriate for MRD-tracking) can be largely mitigated via 1) selection of *multiple* clonal (i.e. CCF ~ 1) mutations from tissue genotyping, increasing the probability of selecting genuinely globally truncal alterations (in most patients), and 2) integration of *a priori* knowledge of which gene/mutations are typically truncal events in a given cancer type (this is typically achieved through tissue genotyping with targeted panels that exclusively capture the exons of established cancer genes).

Nonetheless, as a parenthetical, we have now included a specific example (motivated by our observations of discordant WGD across matched tissue versus cfDNA) where analysis of a single metastatic biopsy may provide a potentially misleading view of truncal genomic features (new Discussion text is underlined):

"Nevertheless, reliance on a single tissue biopsy may provide an incomplete view of polyclonal treatment resistance mechanisms (e.g. AR alterations, BRCA2 reversion mutations), misclassify alterations as truncal features (e.g. WGD), or underestimate disease subclonal complexity (Quigley et al. 2017; Cancer Discovery)."

Finally, subclonal complexity and intratumoral heterogeneity are emerging predictors of poor disease prognosis. We believe that accurate quantification of global disease clonal complexity (via ctDNA) promises to improve understanding of its precise prognostic relevance and clarify potential real-world usages for patient management. Note that the previous Discussion paragraph explicitly references some of the potential biological and clinical implications of subclonal diversity/heterogeneity:

"Subclonal diversity may worsen disease prognosis by increasing opportunities for selection of resistant phenotypes (Black and McGranahan 2021; Turajlic et al. 2018). Our results suggest that deep ctDNA WGS can be deployed in large cohorts of patients with high ctDNA fractions to determine the clinical relevance of metastatic tumor clonal composition and understand how to exploit Darwinian evolutionary principles to improve cancer control."

TSS analyses:

I would like to thank the authors for the clarification. It appears from Figure 5e that cfDNA TSS nucleosome depletion is able to provide a lineage and context-specific recapitulation of the transcriptomics of metastases (e.g. in this manuscript, cfDNA from a group of mCRPCs is able to recapitulate the metastases' transcriptomics from a separate cohort of mCRPCs), presumably as cancers from the same lineage and context exhibit many similarities in the transcriptionally active programs. Additionally, for AR, the authors are able to demonstrate in one patient the utility of cfDNA footprinting in identifying changes in global transcriptomics.

1. In the Discussion, the authors note that "cfDNA footprinting at gene TSSs can recapitulate metastatic tissue mRNA expression…" (lines 351-352). It may be

worthwhile to clarify that this occurs largely on a lineage and context-specific level as the authors do not systemically analyze detailed patient-level concordance.

To be clear that this observation was generally at a more gross level than individual genes or patients, we have added two qualifiers to this sentence: "*cfDNA footprinting at gene TSSs can recapitulate broad metastatic tissue mRNA expression patterns, suggesting that cfDNA may also enable RNA subtyping in cancers where clinical relevance is established*"

2. The authors also note that "we show that cfDNA sequencing enables parallel epigenomic and comprehensive genomic profiling without the need for additional assays (e.g. ATAC-seq, methyl-seq)" (lines 354-355). Without systematic comparisons between these currently complementary techniques, the authors may wish to reconsider this point.

We agree with your point, and have revised the sentence to:

***"Ultimately, we show that cfDNA sequencing enables parallel epigenomic and comprehensive genomic profiling, potentially simplifying workflows and conserving often limited DNA yields."* (note the removal of the parenthetical example).**

Additional minor comments:

1. Figure 5d legend: the legend within Figure 5d appears mislabeled, with "First exon-intron junction r = 0.89" when it should be 0.98 based on the initial submission?

The original value of 0.98 was a typo in our initial submission that has been subsequently corrected. R=0.89 (annotated on the Figure) is indeed correct.

2. On page 32 of the Supplementary Information, the authors refer to Supplementary Table X. I suspect this is a mistake as X should denote a number (e.g. S1, S2, etc).

Thank you. We have now changed it to the correct table (Supplementary Table 5).