

Tracking clonal evolution during treatment in ovarian cancer using cfDNA

Corresponding Author: Professor Sohrab Shah

Any redactions in this file are there to maintain patient confidentiality, the confidentiality of unpublished data, or to remove third-party material.

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Referee #2

(Remarks to the Author)

A. Summary of the key results:

- Baseline shallow single cell WGS was performed at diagnosis on 19 patients with advanced HGSC who were followed over up to 5 years, with plasma cfDNA collected during the time of treatment and/or recurrence.
- Clonal composition was determined at baseline and clone specific variants determined in each patient, to enable patient-specific clonal tracking in plasma cfDNA using targeted capture probes, with structural variants shown to perform at a lower error rate than SNVs.
- Focussing on SVs due to their unique sequence composition is a good approach and HGSC the ideal cancer type to test this.
- Clone abundances were measured over time, with clone specific variants allowing inference of clone specific response to therapy, e.g. in patient 044 initial chemotherapy appeared to eradicate the wild-type ERBB2 clone, while the dominance of the ERBB2 amplified clone at relapse meant the tumour was susceptible to trastuzumab treatment.

B. Originality and significance:

- While cfDNA has previously been used to track the emergence of resistance in HGSC (PMID: 28414925), this approach is unique in combining shallow single cell WGS with targeted capture of clone specific SVs in cfDNA, allowing a more comprehensive view of clone prevalence over time.
- The ability to perform 'real-time' tracking of cancer evolution in patients over time is likely to be clinically useful. Although, a key limitation is not being able to comprehensively characterise the emergence of multiple new alterations/rare clones associated with treatment resistance, which are known to be present in end-stage disease (PMID: 26017449, PMID: 26017449).

C. Data & methodology:

- The patient cohort and samples are appropriate for the aims of the study. The data generated is high quality and sufficient to identify clonal and subclonal variants, and clone-specific complex chromosomal rearrangements, e.g. chromothripsis.
- It would be useful to know the accuracy of the variants called using shallow scWGS and targeted capture of cfDNA compared to those detected using conventional deep bulk tumor WGS of matched patient samples. Please show the concordance & proportion of overlap between high-confidence SNVs, INDELs and SVs called by each approach, as well as a comparison of the VAFs. Are there any high confidence variants detected by one approach were not detected by another?
- The intention of Figure 2b is not clear. The figure and the figure legend don't provide sufficient information to understand which variants are included in this analysis. The Results text states that a mixture of clonal and subclonal variants were selected, so these boxplots should be subdivided into clonal and subclonal variants.
- How was gene amplification status determined and verified, e.g. with chromogenic in situ hybridisation (CISH)? Has ploidy been considered? A copy number of 6 (line 302) may be gain rather than amplification.
- In samples with matched scRNAseq, how well does gene amplification/alteration status correlate with transcript levels in individual clones for genes of interest, i.e. ERBB2, CCNE1, BRCA1, RAB25, etc?

D. Appropriate use of statistics and treatment of uncertainties:

- There is no description of statistical analyses in the Methods. Statistical tests and p values are not defined in Figure legends.
- Please include a discussion on the limitation of not being able to detect new alterations or emerging / initially rare clones that have increased in representation at relapse. It is also plausible that the clone-specific variants or “endogenous barcodes” are lost due to further genomic variation, e.g. deletion of the variant.

E. Conclusions: robustness, validity, reliability

- Amplification of oncogenes is a focus, but homozygous deletions/inactivation of tumour suppressor genes are also common in HGSC. Are any of the key HGSC tumour suppressors lost/inactivated in these samples/clones, e.g. BRCA1/2, NF1, PTEN, RB1, CDK12, etc.? Is there a focus on amplifications because they are easier to detect using shallow WGS? If so, authors should reflect on this limitation. Which gene alterations were characterized and how were certain alterations selected to be described?
- Do all samples have TP53 alterations? This is information that assists with confirmation of HGSC histotype. Are TP53 alterations clonal/truncal in all samples, as shown for case 004 (lines 158-160)?

F. Suggested improvements:

- Lines 70-71, “Serial tumor sampling for genomic profiling from multiple time points is often impractical or contra-indicated”. Agreed, but should briefly explain why this is impractical for readers not familiar with HGSC.
- Line 125, please briefly define the “DLP+ protocol”.
- Line 455, the method of DNA extraction was not included.
- What proportion of patients had known BRCA1/2 or other HR gene alterations at baseline and how many were positive for reversions?

G. References:

- Lines 91-94, please check that appropriate references follow each statement, e.g. shallow whole-genome sequencing was not used in reference 19 (Wang, Y. K. et al. Nat. Genet. 49, 856–865, 2017) nor reference 20 (Funnell, T. et al. PLoS Comput. Biol. 15, e1006799, 2019).
- Lines 309-311, this finding is supported by a previous study that showed CCNE1 amplification in 10% of HGSC long-term survivors (PMID: 36456881).
- Polyclonal disease is mentioned several times, including lines 349-350, “multiple clones were detected in post-recurrence plasma samples, highlighting that chemo-resistance may be polyclonal in many patients”. Polyclonal disease, including multiple resistance mechanisms per patient, has been previously shown in end-stage HGSC (PMID: 36849657).
- Note that a previous study (PMID: 36672479) used shallow WGS of cfDNA in ovarian cancer and one of the major findings was that RAB25 gene amplification was a predictor of patient survival.

H. Clarity and context:

- The sentence on lines 49-50 of the abstract is slightly misleading, as it indicates that in multiple cases there was a therapeutic intervention informed by genomic characterization, but it appears that there was one such example, i.e. patient 044 with the ERBB2 amplification at relapse who then was treated with trastuzumab. Please clarify and amend this statement in the abstract for accuracy.

(Remarks on code availability)

Referee #4

(Remarks to the Author)

This manuscript by Williams and colleagues uses personalized structural variant tracking in circulating tumor DNA to trace clonal evolution through drug treatment. They use a low-noise personalized hybrid capture approach that is most interesting and results in impressive claimed limits of detection. By tracking cfDNA over several years and sampling at multiple time points, the study provides a detailed view of clonal evolution and the development of drug resistance. This is particularly valuable because traditional tissue biopsies often fall short in capturing the full spatial and temporal heterogeneity of tumors. The cohort is well-constructed, linking scWGS with scRNAseq and cfDNA and permitting multi-omic interrogation of tumor evolution at the level of genotype and phenotype in unison, and in particular providing descriptive examples of drug resistance phenotypes. Wright-Fisher modeling is utilised to analyze clonal dynamics in response to treatment.

Despite the strengths of the longitudinal design, the study’s sample size is limited, with only ten patients experiencing recurrence. This restricts its statistical power and limits the generalizability of its findings, making it difficult to identify common genetic patterns associated with drug resistance.

The Wright-Fisher model, though useful, assumes uniform selective pressure across all clones, which may oversimplify the complex interactions and spatial variations within the tumor microenvironment that drive clonal evolution. Additionally, sampling at longer intervals might miss crucial moments when clonal shifts and drug resistance develop, complicating the analysis of clonal dominance in response to treatment.

For a more thorough understanding of drug resistance mechanisms, future research could expand the study to incorporate epigenetic and transcriptional markers alongside genetic ones. Increasing the cohort size and diversity would also enhance the study’s statistical validity and make its findings more broadly applicable.

Major comments

The limit of detection of 10^{-7} is a significant claim. If it is to be reported, it should be determined using spike in data. It is possible that the authors are double counting molecules by taking $1/(\text{raw number of reads})$ when lots of duplicates are present. I'm not sure that a LOD of 10^{-7} tallies with the approximate numbers of structural variants (~100 clonal SVs) and duplex depth (2-3k) – spike in data would help confirm this.

I disagree that estimates of clone fractions should not be done in a manner that corrects for copy number. Given that structural variants specifically are likely to be associated with copy number alterations this might be likely to cause greater error than methods using SNVs. Given that there is a ~linear relationship between the number of supporting reads in ctDNA and mutation multiplicity, a cancer cell fraction estimate of 20% could easily change to 40% (or more). Whilst it is true that when comparing across samples copy number would not affect the overall VAF to the extent that it changes by orders of magnitude, for estimates of cancer cell fraction (that control for ctDNA fraction) it might still be expected to make a big difference. Previous work has shown a linear relationship between that work that does not correct for copy number systematically underestimates the cancer cell fraction.

In the Fisher-Wright model, am I correct that the authors assume that any deviation from neutral evolution would be driven by treatment? Could the authors elaborate on the potential bottlenecks that might arise from (debulking) surgery and the metastatic transition that might be happening concomitantly and would not be associated with treatment.

Minor comments

Did the degree of tissue sampling used to construct the patient specific panels affect the assessment of the pattern of relapse as monoclonal vs polyclonal? Also - did effective coverage influence this?

The error rates in 2D are not reflective of the state of the art – could the y axis on this plot be extended to include error rates in keeping with best in class SNV based approaches?

(Remarks on code availability)

Referee #5

(Remarks to the Author)

I thoroughly enjoyed reading this fantastic paper. The integration of high quality scWGS data and analyses with longitudinal monitoring of evolutionary trajectories using ctDNA in patients is incredibly exciting to me and I believe of great interest to many people from multiple disciplines. Harnessing this approach, Williams et al are able to track evolving clonal frequencies through real time and view dynamics in the context of treatment received and patterns of recurrence. They are also able to track (pre-identified) SVs through tumor evolution which is a current challenge for the field but will be critical for monitoring the consequences of ongoing chromosomal instability in tumors. For a paper packed with novel approaches and insight, I found the analyses straightforward to follow, the statistical analyses sound and the figures beautiful. I do, however, have the following minor points of clarification that I believe would aid understanding:

Minor comments

- To assign mutations and SVs to clones the authors test the distribution of the subset of cells positive for a mutation across clones in the tree. In the examples given in Fig 1 this is straightforward but how is this done for more ambiguous cases? Would it be possible to add detail to the assigning SVs to clones section of the methods?

- What is the impact of site particularly on the phylogenetic inference? It looks like clonal structure correlates very strongly with site. How much would the authors expect this to impact the patterns of clonal evolution that they observe? For example, are responses to treatment and therefore selective pressures likely to differ between sites?

- The authors report clone specific complex SVs in Figure 3. Presumably these were detected in scWGS pseudobulks and the characteristic breakpoints sequenced in cfDNA. Please could you provide more details in the methods of how the patterns of SVs were predicted in pseudobulks?

- There are a couple of typos in y-axis labels e.g. Figs 2c, 5a.

(Remarks on code availability)

The code provided by the authors with the exception of the code for Articul (manuscript in prep) were available on github and I was able to run the Mondrian-scwgs nextflow pipeline with relative ease.

Version 1:

Reviewer comments:

Referee #2

(Remarks to the Author)

Thank you for the comprehensive responses and detailed clarifications. No further comments.

(Remarks on code availability)

Referee #4

(Remarks to the Author)

This revised manuscript from Williams and colleagues is substantially methodologically improved compared to the initial draft. I also note they have included further samples to address previous concerns about novelty.

The corrections for copy number in the calculations related to clone frequency seem appropriate, and I am pleased that they have removed the claim about the claimed limit of detection of their approach being 10^{-7} .

Supp Fig 5 still contains an error – they underestimate the LOD of SNV-based approach - the true limit of detection is $\sim 10^{-7}$, not 10^{-5} (e.g. see Bae et al, Nat Genet, 2023).

I recommend acceptance pending this minor revision.

(Remarks on code availability)

Open Access This Peer Review File is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

In cases where reviewers are anonymous, credit should be given to 'Anonymous Referee' and the source.

The images or other third party material in this Peer Review File are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>

Reply to reviewer reports on Nature submission of the manuscript:

“Tracking clonal evolution of drug resistance in ovarian cancer patients by exploiting structural variants in cfDNA”

We thank the reviewers for their enthusiastic and thoughtful feedback on our paper, please find below a point by point response. Major changes include:

- Inclusion of an additional 8 patients with longitudinal tracking as well as additional samples from patients included in the original manuscript to address concerns of sample size. This brings the total cohort size to 153 cfDNA plasma samples from 24 patients (18 with longitudinal samples).
- A more complete accounting of important genomic alterations found in each tumor and clone, including tumor suppressors
- New quantitative analysis of the limit of detection, demonstrating when structural variant approaches provide quantitative advantages over SNV-based methods
- To address concerns around identification of clones and emergence of new clones we have i) generated deep WGS of plasma samples at baseline and recurrence from 6 patients ii) generated scWGS data from 2 recurrence patient samples and one interval debulking sample iii) adjusted the text to appropriately caveat any results.
- Modified the clone frequencies calculations to correct for copy number
- Overall, our conclusions are reinforced by the new data and the responses to reviewer comments. We include a point-by-point response below.

Referee #2:.....	1
Referee #4:.....	15
Referee #5:.....	24

We note that all line numbers refer to lines in the tracked changes version of the manuscript file.

Referee #2:

A. Summary of the key results:

- Baseline shallow single cell WGS was performed at diagnosis on 19 patients with advanced HGSC who were followed over up to 5 years, with plasma cfDNA collected during the time of treatment and/or recurrence.
- Clonal composition was determined at baseline and clone specific variants determined in each patient, to enable patient-specific clonal tracking in plasma cfDNA using targeted capture probes, with structural variants shown to perform at a lower error rate than SNVs.
- Focussing on SVs due to their unique sequence composition is a good approach and HGSC the ideal cancer type to test this.
- Clone abundances were measured over time, with clone specific variants allowing inference of clone specific response to therapy, e.g. in patient 044 initial chemotherapy appeared to eradicate the wild-type ERBB2 clone, while the dominance of the ERBB2 amplified clone at relapse meant the tumour was susceptible to trastuzumab treatment.

B. Originality and significance:

- While cfDNA has previously been used to track the emergence of resistance in HGSC (PMID: 28414925), this approach is unique in combining shallow single cell WGS with targeted capture of clone specific SVs in cfDNA, allowing a more comprehensive view of clone prevalence over time.

- The ability to perform 'real-time' tracking of cancer evolution in patients over time is likely to be clinically useful. Although, a key limitation is not being able to comprehensively characterise the emergence of multiple new alterations/rare clones associated with treatment resistance, which are known to be present in end-stage disease (PMID: 26017449, PMID: 26017449).

C. Data & methodology:

- The patient cohort and samples are appropriate for the aims of the study. The data generated is high quality and sufficient to identify clonal and subclonal variants, and clone-specific complex chromosomal rearrangements, e.g. chromothripsis.

- It would be useful to know the accuracy of the variants called using shallow scWGS and targeted capture of cfDNA compared to those detected using conventional deep bulk tumor WGS of matched patient samples. Please show the concordance & proportion of overlap between high-confidence SNVs, INDELs and SVs called by each approach, as well as a comparison of the VAFs. Are there any high confidence variants detected by one approach were not detected by another?

This is an important point raised by the reviewer. There are of course non-overlapping variants. Exact matching of DNA pools is not possible here since by definition the libraries constructed by single cell approaches are comprised of different DNA fragments from bulk extractions. We have published the degree of overlap from site-matched DLP+ and bulk approaches in several papers (e.g. Laks *et al.* and Funnell *et al.*) and provide a similar analysis here. Importantly, we do not claim to have complete coverage of all variants in a tumor - this is not possible and we discuss limitations in the Discussion.

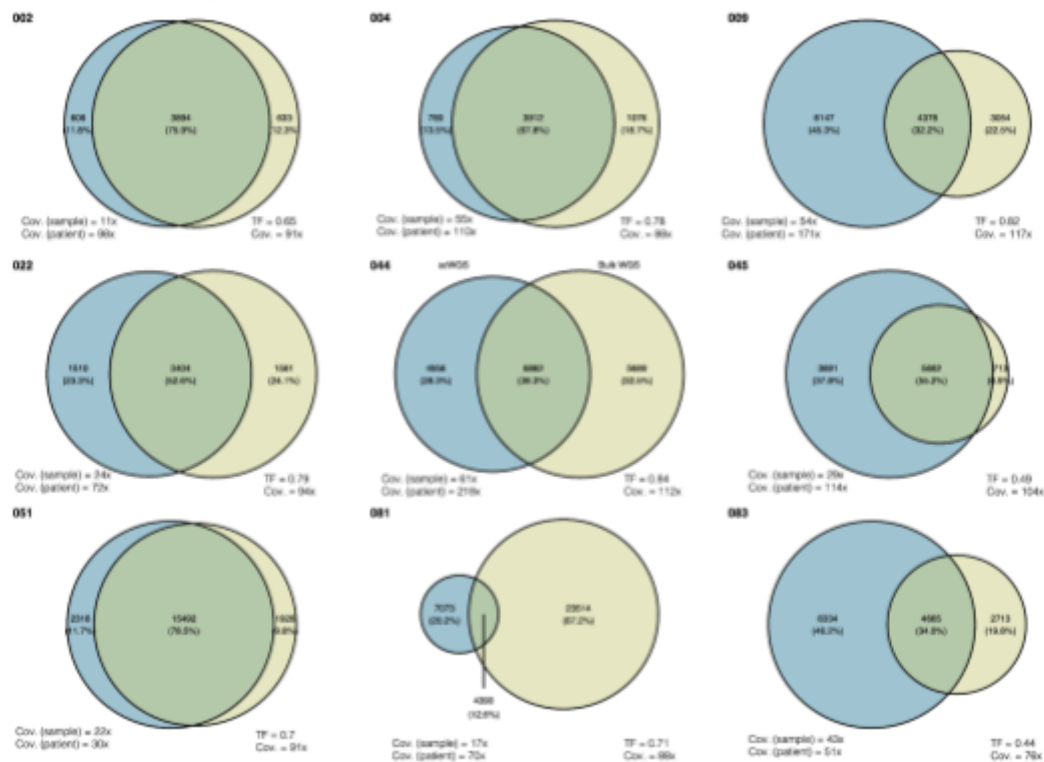
Figures showing the analysis requested are provided on the next page. **Figure R1** show venn diagrams of the overlap of mutations called in scWGS and bulk WGS for SNVs and SVs. This was done for 9 patients where data was generated from anatomically similar sites. We see there is a large amount of overlap (average 50% for SNVs and 58% for SVs) but also a large number of assay specific variants. There are a number of confounders which make interpreting this comparison challenging:

1. Site-to-site heterogeneity, although the comparison was done on tissue from similar anatomical sites, the specimens were different.
2. Differences in sequencing coverage
3. Differences in purity
4. Differences in variant calling approaches. For the scWGS data we pool sequencing data from all cells across all sites, call SNVs using mutect, genotype SNVs in each cell and then remove artifacts specific to scWGS using ArtiCull (Satas et al. Biorxiv 2025). Although this comparison only assess mutation overlap in the sites with matched bulk WGS, our mutation calling approach means that information from other sites is used.

Filtering for higher VAF mutations can mitigate some of these confounders. For example, filtering for VAF > 0.1 increases the fraction of overlapping SNVs from 50% to 70%. This suggests that a major component of assay-specific differences stems from intra-tumor heterogeneity, likely due to differences in sampling site and variations in tumor purity/sequencing coverage that affect the detection of lower frequency variants.

Figure R2 shows the correlation between VAF values for SNVs. We see that for all 9 patients the VAFs are significantly correlated (Pearson correlation coefficient range 0.6-0.85, all p-values significant).

SNV overlap between scWGS and bulk WGS



SV overlap between scWGS and bulk WGS

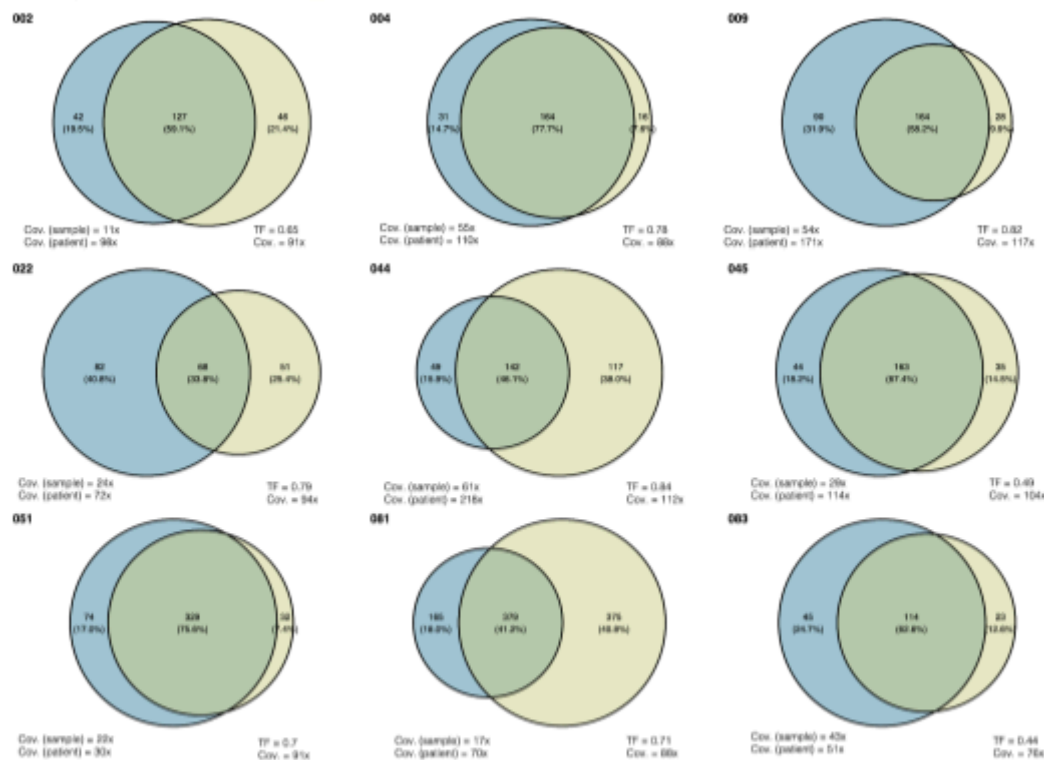


Figure R1 Venn diagrams showing the overlap in SNVs and SVs called between scWGS and bulk WGS. Only data from the same anatomical site were included. Each venn diagram is annotated with the number and percentage called for each group. Below each venn diagram text labels indicate the total coverage of tumor cells per sample and per patient (combining all sites) on the left and the tumor fraction (TF) and coverage of the bulk WGS on the right hand side

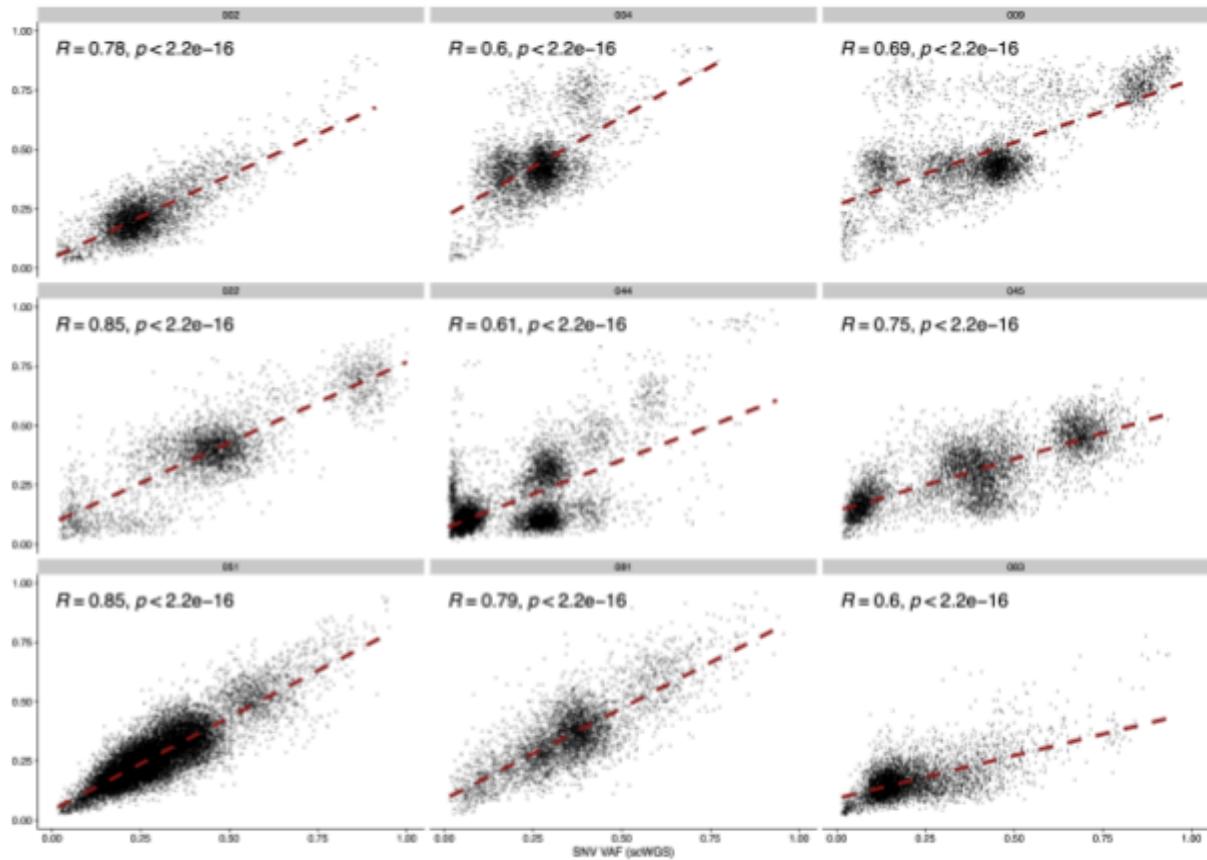


Figure R2 Correlation of VAFs from scWGS and bulk WGS for SNVs for 9 patients. Dashed red line shows the best fit linear regression line. Each plot is annotated with pearson correlation coefficient (R) and associated p-value in the top left. Only SNVs detected with both methods are shown.

As a final piece of analysis, we felt that the key point the reviewer is raising is whether the scWGS data is sufficiently robust to infer base pair level mutations with high accuracy, which is fundamental to all the analysis in the paper. We reasoned that the best way to confirm this would be to take mutations that are truncal with high confidence and then assess the recovery of these mutations in the bulk WGS and cfDNA samples at baseline where tumor burden is high. Clonality can be determined with high confidence in scWGS data and focussing on truncal mutations mitigates issues related to intra-tumor heterogeneity and tumor cell purity. This shows that more than 75% of high confidence SNVs and SVs are detected in bulk WGS and more than 90% of these are found in pre-op cfDNA (**Figure R3**).

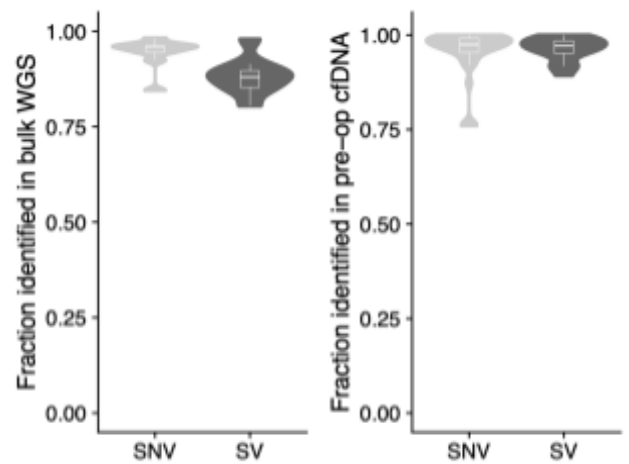


Figure R3 Fraction of high confidence truncal SNVs and SVs also called in bulk WGS (left) and identified in pre-op cfDNA (right). Plots show distributions over patients.

Satas, Gryte, et al. "Inferring active mutational processes in cancer using single cell sequencing and evolutionary constraints." *bioRxiv* (2025): 2025-02.

Laks, Emma, et al. "Clonal decomposition and DNA replication states defined by scaled single-cell genome sequencing." *Cell* 179.5 (2019): 1207-1221.

Funnell, Tyler, et al. "Single-cell genomic variation induced by mutational processes in cancer." *Nature* 612.7938 (2022): 106-115.

- The intention of Figure 2b is not clear. The figure and the figure legend don't provide sufficient information to understand which variants are included in this analysis. The Results text states that a mixture of clonal and subclonal variants were selected, so these boxplots should be subdivided into clonal and subclonal variants.

While a mixture of clonal and subclonal variants were included in each patient's panel, the main purpose of Figure 2 is to provide technical validation of our novel approach using structural variants for cfDNA detection. As this represents one of the first applications of SVs for liquid biopsy, we focused on clonal/truncal genomic variants to ensure our assessment was not confounded by intra-tumor heterogeneity. This design allowed us to establish the analytical robustness of SV detection and directly compare it to conventional single nucleotide variant analysis. We have attempted to make this clearer by modifying the axis and the figure legends. The axis for Fig2b is now "VAF at baseline (truncal mutations)". We note that Figure 3f includes boxplots split by clonal/subclonal status.

- How was gene amplification status determined and verified, e.g. with chromogenic in situ hybridisation (CISH)? Has ploidy been considered? A copy number of 6 (line 302) may be gain rather than amplification.

We have refined our amplification classification, we now distinguish the following categories:

- Amplification: $\geq 2x$ ploidy:
- High-level amplification: > 10 copies

Under this classification, the *CCNE1* amplification the reviewer refers to is classified as an amplification but not a high level amplification. We have reserved the word gain to refer to larger chromosome scale alterations rather than focal events.

To confirm amplification status for the *CCNE1* cases we have now generated fluorescent in situ hybridization (FISH) data for the 2 patients presented in Main Figure 4 (107 and 045). We stained cells with DAPI and included probes for the centromere on chromosome 19 as well as for *CCNE1*. Example cells exhibiting the amplification are now included in main Figure 4 which we show below in **Figure R4** for the reviewers benefit. All FISH images will also be included as supplementary data.

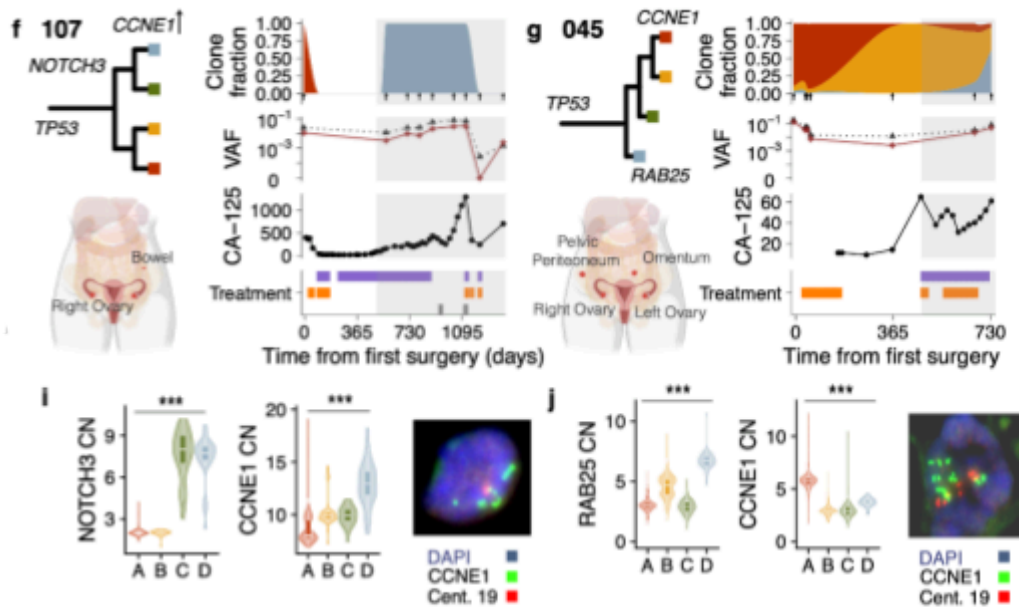


Figure R4 Panels f-j) from main Figure 4. Clone trajectory plots alongside treatment for 2 CCNE1 amplified patients. i) and j) show FISH data confirming the amplification status

We also note that patient 044 who received T-DXd was eligible for the trial based on IHC2+ of Her2, this has now been noted in the manuscript (line 561):

“Subsequently, following a second disease recurrence at day 730, the patient was enrolled onto a clinical trial based on HER2 IHC2+ (DESTINY-PanTumor02(Meric-Bernstam et al. 2024)) and treated with trastuzumab deruxtecan (T-DXd), an antibody-drug conjugate that targets HER2 (ERBB2).”

In addition we have added the following sentence to the introduction referencing previous publications using the DLP+ technology that have validated the ability to identify a range of cell and clone specific alterations including amplifications (line 258):

“scWGS data was generated using the DLP+ platform which enables the identification of whole chromosome and chromosome-arm aneuploidies, high-level amplifications and complex genome rearrangements in single cells at 0.5Mb resolution(Williams et al. 2024; Laks et al. 2019; Funnell et al. 2022; Salehi et al. 2021)”

- In samples with matched scRNAseq, how well does gene amplification/alteration status correlate with transcript levels in individual clones for genes of interest, i.e. ERBB2, CCNE1, BRCA1, RAB25, etc?

We were unable to generate scRNAseq for 044 (ERBB2) and clone D in 045 (RAB25). Nevertheless, we show below some examples that are discussed in the text (**Figure R5**).

- Increased copy number of NOTCH3 in clone C and D results in increased expression in those clones. This data is included in Main Figure 5.

- For CCNE1, increased copy number results in increased expression in 045 but not 107. As CCNE1 is a cell-cycle dependent gene with expression naturally fluctuating across cell cycle phases, direct mapping between copy number and expression is inherently challenging and difficult to interpret
- For MYC in 022 we do not see any relationship between copy number and expression
- 083 has a clone specific chromothripsis on chromosome 8 where some regions are highly amplified, MYC is not in an amplified region but we show below increased copy number of EIF3E which results in increased expression

These results are consistent with our understanding that genes vary in sensitivity to copy number changes, with transcriptional adaptive mechanisms (including epigenetic modifications and downstream regulation) modulating gene dosage effects, as we previously explored using TreeAlign method (Shi et al). Notably, regions like 8q containing MYC also harbor super-enhancer elements that may influence expression beyond simple copy number changes. Finally, we note that while not the central focus of this study, we have published methods for analysis of gene-dosage effects on clone-specific transcription using scWGS and scRNA (Shi et al).

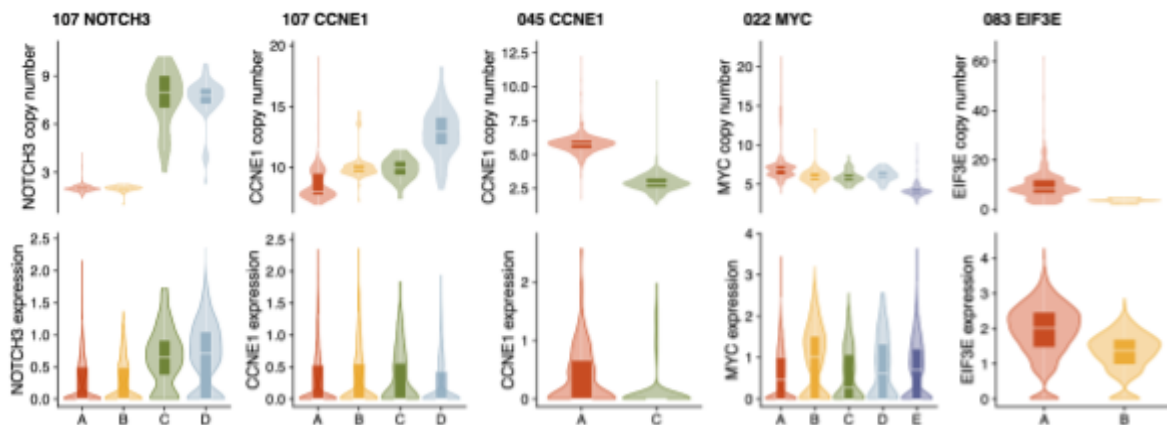


Figure R5 scWGS based copy number distributions (top) and expression distributions (bottom) for 5 genes.

Shi, Hongyu, et al. "Allele-specific transcriptional effects of subclonal copy number alterations enable genotype-phenotype mapping in cancer cells." *Nature Communications* 15.1 (2024): 2482.

D. Appropriate use of statistics and treatment of uncertainties:

- There is no description of statistical analyses in the Methods. Statistical tests and p values are not defined in Figure legends.

We have added statistical tests to all relevant figures and the following sentence to the methods:

"For between-group comparisons, we used two-sided t-tests. To assess correlations we used Pearson correlation. All statistical analysis was done using R."

- Please include a discussion on the limitation of not being able to detect new alterations or emerging / initially rare clones that have increased in representation at relapse. It is also plausible that the clone-specific variants or “endogenous barcodes” are lost due to further genomic variation, e.g. deletion of the variant.

We agree that this is an important point and have clarified this further in the discussion (line 1031):

“The inability to comprehensively identify all clones present at diagnosis, and the emergence of novel clones post tissue sampling may limit our ability to fully resolve and track clonal populations. Due to ongoing structural mutational processes such as chromosomal instability, breakage fusion bridge and ongoing genome doubling, copy number alterations and SVs will accrue over time, and could also result in deletion of breakpoints/mutations. Interestingly, additional scWGS from recurrence samples and WGS of the cfDNA samples points to relatively minor acquisition of genomically divergent clonal expansions. Nevertheless, further study as to the granularity of evolutionary change post diagnosis, for example in the neoadjuvant setting where pre- and post- chemotherapy tissue samples are resected, would further inform the degree to which significantly divergent clones emerge on therapy.”

In addition, to address this important point directly, we have generated whole genome deep sequencing of plasma cfDNA from pre and post-recurrence samples as well as scWGS from 3 patients where we were able to collect tissue and perform scWGS from multiple timepoints.

scWGS from multiple timepoints

For 2 patients (026 and 139) we were able to collect tissue from debulking surgeries in the post-recurrence setting. For patient 026 this was 4.7 years after their initial diagnosis and for 139 2.6 years. From 1 patient (046) we generated scWGS from tissue collected during an interval debulking procedure as well as from the initial diagnostic laparoscopic biopsy. Due to the challenges inherent with these types of specimens (tissue samples are typically small and are fragile due to treatment exposure) the recovery of cells was limited (range 5-20 cells from the second time point). Despite the limited number of cells they were of sufficient quality to assess similarity to clones/cells identified from the first surgery. We found the following:

- **OV-026** Cells recovered from the recurrence sample were most similar to clone B. This is consistent with our cfDNA measurements where SVs associated with clone B were the only ones detected at the final timepoint. However, cells from the recurrence were all whole genome doubled. Interestingly, differently from all other clones, clone B in the primary sample contained a small population of WGD cells (n=4), suggesting the precursor to these recurrent WGD cells may have already been present at the time of diagnosis
- **OV-139** All cells recovered from the recurrence sample were most similar to clone B. Again this is consistent with our cfDNA measurements which confirmed the presence of clone B at recurrence. Although these cells were highly similar to cells from clone B, they did contain some novel events such as gains on chromosomes 7q and 12p.
- **OV-046** Cells from the interval debulking were most similar to clones B and E. Compared to cfDNA measurements from the same timepoint, clone B was detected but clone E was not. However clone E was detectable in subsequent timepoints following the conclusion of adjuvant chemotherapy.

Trajectories for these patients are now included in Main Figure 4 and the scWGS data from the recurrence is shown in a new supplementary Figure 10 and provided below as **Figure R6**. Relevant panels from Main Figure 4 are also included below as **Figure R7**.

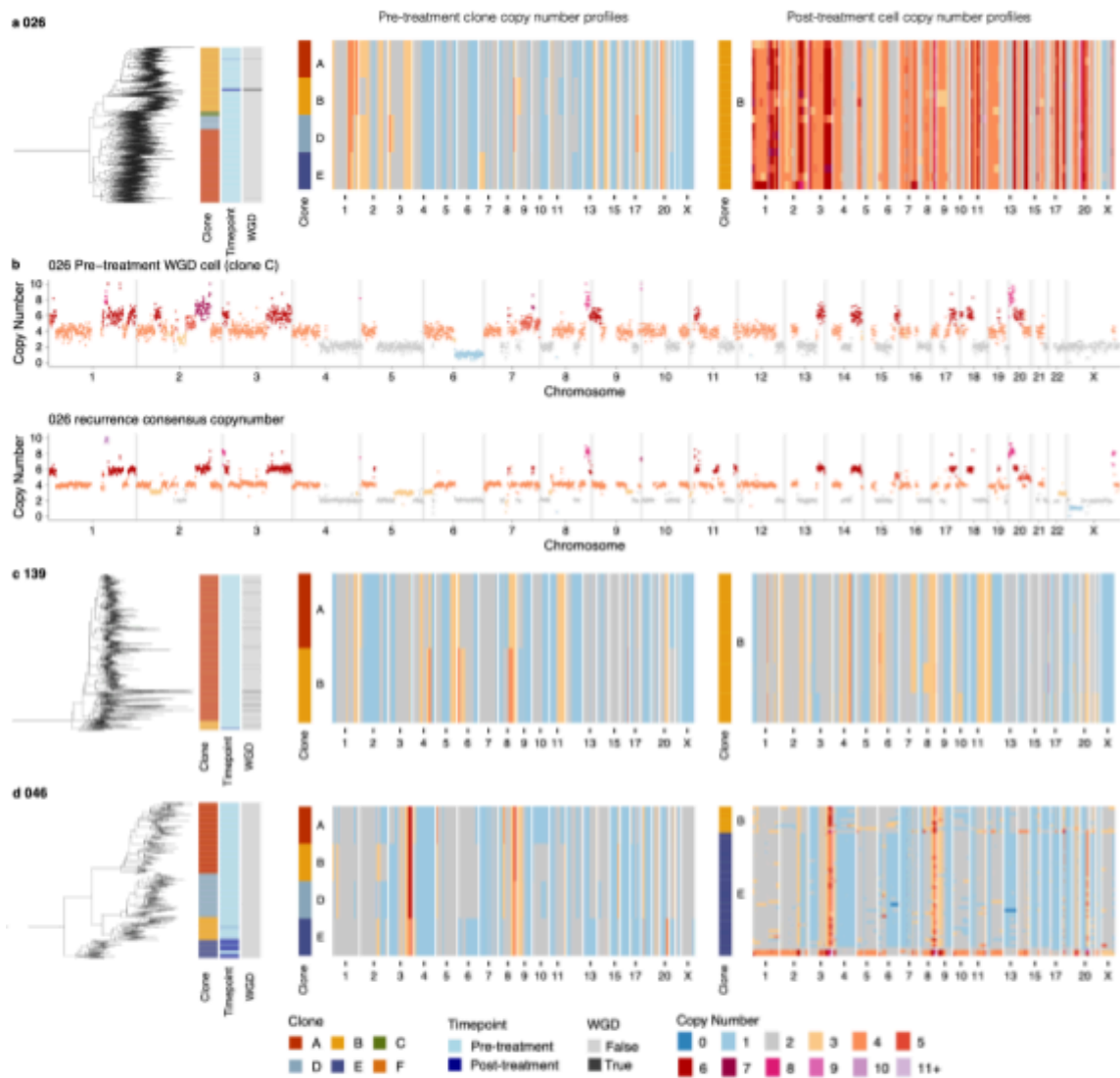


Figure R6 Comparison of pre-and post treatment scWGS data for 3 patients: a) 026 c) 139 and d) 046. For each row we show the phylogenetic tree of all cells annotated by clone, timepoint (pre or post treatment) and cell WGD state. Then the consensus total copy number heatmap for each clone based on pre-treatment cells and the total copy number heatmap for all post-treatment cells. b) Shows a pre-treatment WGD cell from clone C and the consensus copy number profile of all post-treatment cells.

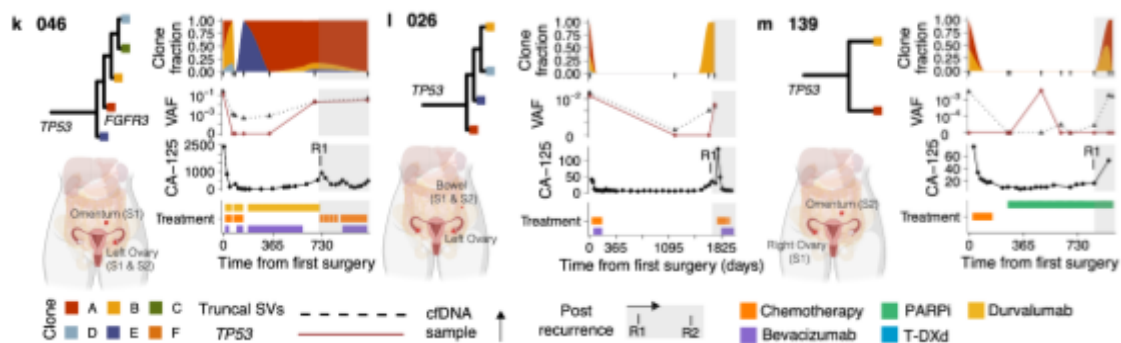


Figure R7 Clonal evolution trajectories in patients 046, 026 and 139 (panels k-m from Main Figure 4).

Plasma cfDNA WGS

We have also generated WGS from 2 plasma samples (1 pre-op and 1 post recurrence) from 6 patients. We initially generated this data in order to confirm our inference of the dominant clone at each timepoint was correct but this can also be used to identify novel copy number alterations in the recurrence sample. Two patients had a recurrence sample with sufficiently high tumor fraction to analyse the data in this way (045 and 022). For 045 we did not identify any regions of the genome that strongly deviated from the predictions based on cfDNA and scWGS derived copy number profiles. For OV-022, there were losses on chromosome 3p, 5q and 9p that were not present in the clone A consensus copy number profile, the only clone inferred to be present at the recurrence timepoint. However, upon closer inspection of clone A we identified a minor subset of cells that harboured these alterations. Therefore, the recurrence clone is likely derived from a subclone within A. This data is shown below in **Figure R8** (panels derived from Supplementary Figures 9 and 6).

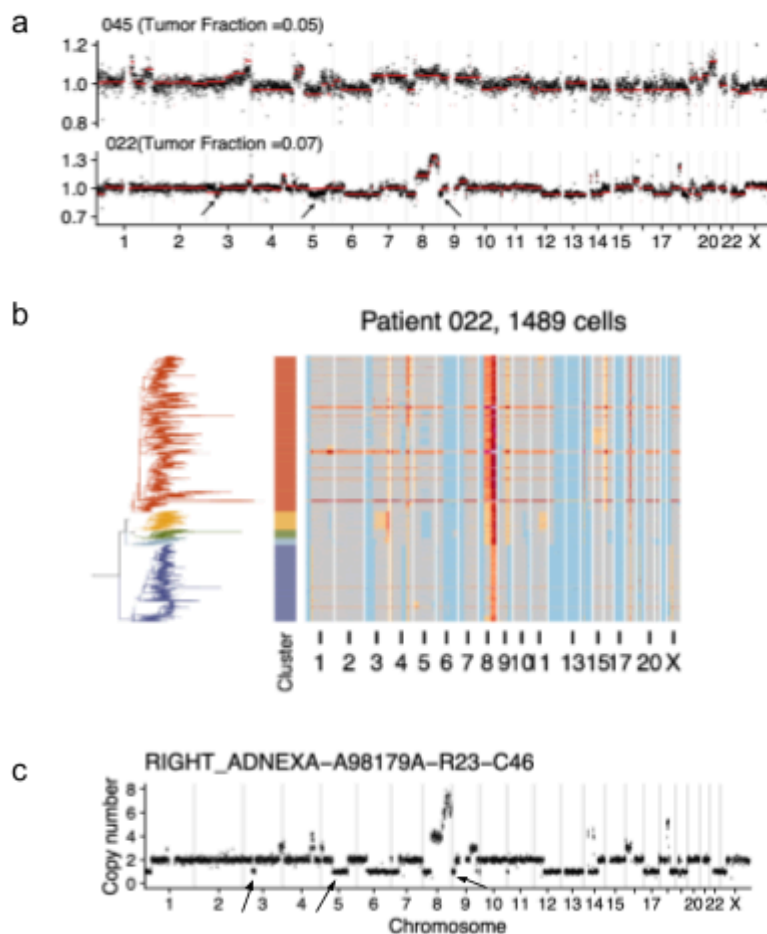


Figure R8 a) Read depth profiles from recurrence cfDNA WGS for 2 patients. Black points are data, red points are predictions based on scWGS and cfDNA targeted assay. 022 is annotated with regions that deviate from predictions. b) Copy number heatmap for 022 c) Single cell from 022 that has the losses on chromosomes 3, 5 and 9 highlighted in a)

Taken together, this additional data re-inforces that pre-existing diversity at the time of diagnosis is selected in relapsed disease. However we recognize that we do not have sufficient data to conclude this for all patients. Furthermore there are alterations that we could not profile accurately such as point

mutations and other non genetic mechanisms may emerge through the course of treatment. This is elaborated at various points in the discussion:

Line 1063:

“Finally, we expect that drug resistance in HGSOc and other cancers to be multi-factorial, with both transcriptional and epigenetic plasticity operating in tandem with pre-adapted and genomically encoded phenotypic states.”

Line 1033:

“Due to ongoing structural mutational processes such as chromosomal instability, breakage fusion bridge and ongoing genome doubling²², copy number alterations and SVs will accrue over time and could also result in deletion of breakpoints/mutations”

Line 1035:

“Interestingly, scWGS and cfDNA from post-recurrence samples revealed only modest genomic divergence between baseline clones and their descendants at the recurrence timepoints. This limited evolution may reflect the disparity between tumor development timescales, spanning decades, versus the relatively brief period between diagnosis and recurrence.”

E. Conclusions: robustness, validity, reliability

- Amplification of oncogenes is a focus, but homozygous deletions/inactivation of tumour suppressor genes are also common in HGSC. Are any of the key HGSC tumour suppressors lost/inactivated in these samples/clones, e.g. BRCA1/2, NF1, PTEN, RB1, CDK12, etc.? Is there a focus on amplifications because they are easier to detect using shallow WGS? If so, authors should reflect on this limitation. Which gene alterations were characterized and how were certain alterations selected to be described?

We have added an oncoplot showing tumor suppressor gene alterations across patients and clones. Our analysis reveals:

1. Tumor suppressors are predominantly truncal. Most apparent clone-specific absences occur in regions with no sequencing coverage in small clones, suggesting false negatives (annotated on the plot).
2. Two notable exceptions exist: a truncating NF1 mutation in patient 009 (present only in clones B,C,D,F, with F being the sole clone post-recurrence) and a somatic *BRCA1* mutation absent in clone E of patient 129. Interestingly, patient 129 also carries inactivating *BRCA2* alterations across all clones (homozygous deletion in 4 clones, germline truncating mutation in the fifth).
3. We tracked VAF trajectories for 5 somatic tumor suppressor mutations (**Figure R9** below and new Supplementary Figure 8). Three mutations persist post-recurrence, consistent with their truncal status or clone specificity (NF1 in clone F). For patients 129 and 139, BRCA1/BRCA2 mutations were not detectable post-recurrence, though tumor fractions in these samples approach our assay's single-SNV detection limit (10^{-3}).
4. Tracking a single mutation inherently provides lower detection sensitivity than our approach of aggregating multiple truncal SVs, as the signal from a single genomic locus is more susceptible to sampling variation and technical noise at low tumor fractions. These plots therefore show similar properties to the TP53 timeseries in our main figures.

Our approach for characterising gene level alterations combined multiple data sources. We used scWGS, bulk WGS and MSK IMPACT for small mutations and indels, these were then genotyped in

individual cells. Amplifications and deletions were called from clone level copy number profiles at 10kb resolution. We therefore are limited to homozygous deletions that are greater than 10kb. This is now mentioned in the methods.

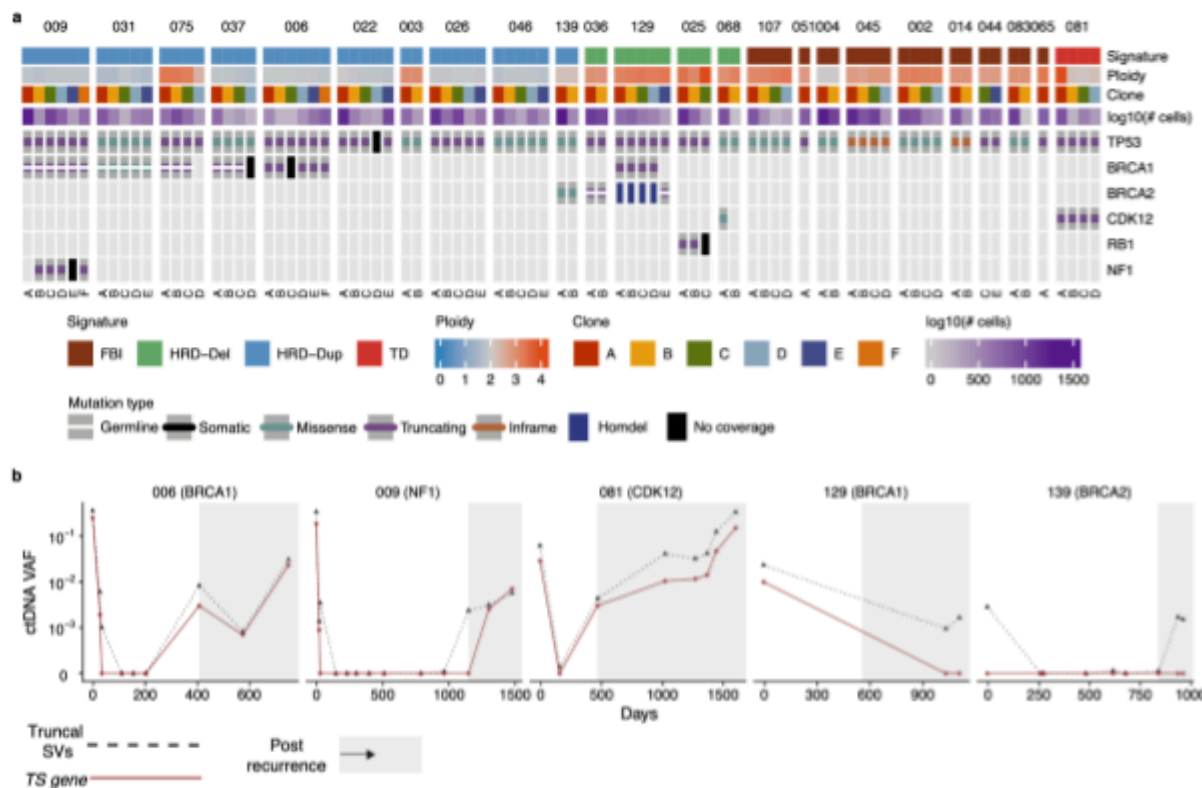


Figure R9 a) Oncoplot showing presence/absence of tumor suppressor mutations across patients and clones. Filled squares indicate instances where at least one sequencing read in a clone shows evidence for the mutation. Absence of reads supporting a mutations is inconclusive in some clones when the number of cells is low, these are annotated with black squares. Patients are ordered by mutational signature (HRD-Dup, HRD-Del, FBI, TD). **b)** VAF of somatic mutations across time, shown are non-TP53 mutations from patients with time series.

- Do all samples have TP53 alterations? This is information that assists with confirmation of HGSC histotype. Are TP53 alterations clonal/truncal in all samples, as shown for case 004 (lines158-160)?

All patients were confirmed HGSOC at diagnosis by specialist gyn pathologists and had TP53 mutation status confirmed through FDA-approved MSK IMPACT clinical panel based sequencing. In the DLP+ data, all samples have TP53 and these are truncal. See oncoplot above (**Figure R9**). The only clone with no read level evidence of the *TP53* mutation is clone D from patient 022 which had no reads covering the position due to the limited cumulative coverage in this small clone.

F. Suggested improvements:

- Lines 70-71, “Serial tumor sampling for genomic profiling from multiple time points is often impractical or contra-indicated”. Agreed, but should briefly explain why this is impractical for readers not familiar with HGSC.

Yes, the paper has been edited to reflect this (line 89):

“Serial tumor sampling for genomic profiling from multiple time points is often impractical or contra-indicated in patients with HGSOE; in the recurrent setting, patients most often develop carcinosarcoma and ascites which is challenging to biopsy.”

- Line 125, please briefly define the “DLP+ protocol”.

We have added the following sentence (line 258):

“scWGS data was generated using DLP+, a high-throughput, tagmentation based shallow single-cell whole genome sequencing approach, yielding approximately 0.05X coverage per cell. This enables identification of whole chromosome and chromosome-arm aneuploidies, segmental alterations, high-level amplifications and complex genome rearrangements in single cells at 0.5Mb resolution(Williams et al. 2024; Laks et al. 2019; Funnell et al. 2022; Salehi et al. 2021; McPherson et al. 2024).”

- Line 455, the method of DNA extraction was not included.

This has been added along with a citation to the MSK-ACCESS protocol which describes the library preparation in detail.

- What proportion of patients had known BRCA1/2 or other HR gene alterations at baseline and how many were positive for reversions?

As can be seen from **Figure R9**, 8 patients had BRCA1/2 (no patients with other HR related genes). Of these only 1 had a reversion mutation, patient 009 which is shown in Main Figure 4.

G. References:

- Lines 91-94, please check that appropriate references follow each statement, e.g. shallow whole-genome sequencing was not used in reference 19 (Wang, Y. K. et al. Nat. Genet. 49, 856–865, 2017) nor reference 20 (Funnell, T. et al. PLoS Comput. Biol. 15, e1006799, 2019).

We have removed these references.

- Lines 309-311, this finding is supported by a previous study that showed CCNE1 amplification in 10% of HGSC long-term survivors (PMID: 36456881).

We thank the reviewer for reminding us of this interesting paper and have added a citation and the following sentence (line 802):

“Interestingly, CCNE1 amplification has been shown to be enriched in long term HGSOE survivors(Garsed et al. 2022).”

- Polyclonal disease is mentioned several times, including lines 349-350, “multiple clones were detected in post-recurrence plasma samples, highlighting that chemo-resistance may be polyclonal in many patients”. Polyclonal disease, including multiple resistance mechanisms per patient, has been previously shown in end-stage HGSC (PMID: 36849657).

We have added a citation to this study

- Note that a previous study (PMID: 36672479) used shallow WGS of cfDNA in ovarian cancer and one of the major findings was that RAB25 gene amplification was a predictor of patient survival.

We have added a citation to this study

H. Clarity and context:

- The sentence on lines 49-50 of the abstract is slightly misleading, as it indicates that in multiple cases there was a therapeutic intervention informed by genomic characterization, but it appears that there was one such example, i.e. patient 044 with the ERBB2 amplification at relapse who then was treated with trastuzumab. Please clarify and amend this statement in the abstract for accuracy.

This has been adjusted to the following:

“In one notable case, clone-specific ERBB2 amplifications, impacted selection of a secondary line of therapy with positive patient outcome.”

Referee #4:

This manuscript by Williams and colleagues uses personalized structural variant tracking in circulating tumor DNA to trace clonal evolution through drug treatment. They use a low-noise personalized hybrid capture approach that is most interesting and results in impressive claimed limits of detection. By tracking cfDNA over several years and sampling at multiple time points, the study provides a detailed view of clonal evolution and the development of drug resistance. This is particularly valuable because traditional tissue biopsies often fall short in capturing the full spatial and temporal heterogeneity of tumors. The cohort is well-constructed, linking scWGS with scRNAseq and cfDNA and permitting multi-omic interrogation of tumor evolution at the level of genotype and phenotype in unison, and in particular providing descriptive examples of drug resistance phenotypes. Wright-Fisher modeling is utilised to analyze clonal dynamics in response to treatment.

Despite the strengths of the longitudinal design, the study's sample size is limited, with only ten patients experiencing recurrence. This restricts its statistical power and limits the generalizability of its findings, making it difficult to identify common genetic patterns associated with drug resistance.

We have now added an additional 60 cfDNA samples to the study (new total = 153 cfDNA samples), these include 17 additional cfDNA samples from patients included in the first study and 43 cfDNA samples from 5 new patients. This brings the total number of patients experiencing a recurrence to 18. We agree that we are underpowered to detect recurrent genetic alterations, which would require significantly larger sample size, this is something we are excited to pursue in follow-up studies. Nonetheless, we would highlight that with this modest cohort we identify some general patterns of resistance and present some clinically valuable case studies. To summarize these findings:

- Clonal diversity decreases post recurrence, suggesting therapy produces a clonal pruning effect. In many cases, we observe the eradication of clones through the course of treatment.
- We observe expansion of clones carrying oncogenic amplifications in multiple patients
- Our unique approach combining clone-specific SV detection with longitudinal cfDNA sampling allowed us to study patient response in unprecedented detail, revealing the effectiveness of anti-HER2 therapy in one patient, the non-deterministic nature of CCNE1 amplifications in two patients, and BRCA reversion following PARPi therapy in another patient. These case studies provide valuable insights into the clonal dynamics of HGSOc during treatment and demonstrate the clinical utility of our methodology for monitoring treatment response and resistance mechanisms in real time.

We note that in the discussion we highlight that larger sample sizes are needed to discover recurrent genetic alterations driving resistance (line 1016).

"While our study is underpowered to identify recurrent genetic features of drug resistant clones, we do observe some plausibly important features that will require additional study."

The Wright-Fisher model, though useful, assumes uniform selective pressure across all clones, which may oversimplify the complex interactions and spatial variations within the tumor microenvironment that drive clonal evolution. Additionally, sampling at longer intervals might miss crucial moments when clonal shifts and drug resistance develop, complicating the analysis of clonal dominance in response to treatment.

We agree with these points and have highlighted them in the discussion (line 988):

“We recognize that our simulation framework neglects any spatial component whereby some clones may reside in a privileged site with different immunological properties, drug localization propensities, or participate in metastatic dissemination processes that could contribute to clonal frequency shifts.”

Line 1048:

“We note that the granularity of temporal sampling in this study limits the ability to accurately time emergence of drug-resistant clones. Some clone trajectories coincided with therapy modulation. We recognize this is confounded by the timing of sampling at relapse and consequently makes it challenging to establish a causal link between selective sweeps of clones and change of therapeutic selective pressure. In future studies, more granular cfDNA sampling at regular intervals would address this interpretation challenge, but our data is nevertheless indicative of clone-specific selective sweeps on switch of therapy.”

Despite these limitations, our Wright-Fisher modeling framework provides a valuable quantitative benchmark that captures key aspects of clonal dynamics and establishes a foundation for future models incorporating spatial heterogeneity and finer temporal resolution.

For a more thorough understanding of drug resistance mechanisms, future research could expand the study to incorporate epigenetic and transcriptional markers alongside genetic ones. Increasing the cohort size and diversity would also enhance the study’s statistical validity and make its findings more broadly applicable.

We certainly agree these are critical next steps in this work. While out of scope for this contribution, we articulate in the discussion that larger scale studies that use the approach outlined here will be required to fully impact clinical decision making and inform mechanistic studies of treatment failure.

Relevant points in the discussion:

Line 1016:

“While our study is underpowered to identify recurrent genetic features of drug resistant clones, we do observe some plausibly important features that will require additional study.”

Line 1013:

“However, larger scale studies across spatio-temporal measurements may be needed to fully establish these properties.”

Line 1062:

“Finally, we expect that drug resistance in HGSOV and other cancers to be multi-factorial, with both transcriptional and epigenetic plasticity operating in tandem with pre-adapted and genomically encoded phenotypic states. We contend that the framework we establish here is poised to quantify the proportion of drug resistance that is explained by therapeutic selective pressure and can inform future evolution-informed adaptive clinical trials. “

Major comments

The limit of detection of 10^{-7} is a significant claim. If it is to be reported, it should be determined using spike in data. It is possible that the authors are double counting molecules by taking $1/(\text{raw number of reads})$ when lots of duplicates are present. I'm not sure that a LOD of 10^{-7} tallies with the approximate numbers of structural variants (~100 clonal SVs) and duplex depth (2-3k) – spike in data would help confirm this.

This is a good point and has motivated us to clarify our claims on the limit of detection. We now provide a more quantitative discussion on the factors influencing our ability to robustly detect and quantify tumor DNA from plasma using structural variants, which we will summarize below. We first outline and clarify some terminology which was unclear or incorrectly articulated in our first submission before showing some new analysis that has been added to the paper.

We will define the limit of detection (LOD) as the smallest tumor fraction that can be reliably detected, this can be defined as when the theoretical limit of detection (TLOD) is greater than the error rate of the assay. The TLOD is the limit of detection in the absence of any sequence error and depends on the number of genomic targets and the sequencing coverage. A framework for estimating this value is outlined in Zviran et al and given by (equation (4) in Zviran et al.):

$$M = N(1 - (1 - TF)^{cov}) \quad (1)$$

Where M is the number of detected mutations, N is the total number of mutations in the panel, TF is the tumor fraction and cov is the depth of coverage. As in Zviran *et al.*, when $M < 1$, we interpret it as a detection probability. We show this detection probability as a function of tumor fraction, coverage and the number of mutations in **Figure R10a** (this is reproducing Figure 1e from Zviran et al).

We can then define the TLOD as the lowest TF with 99% probability of detection. So TLOD is given by (setting $M = 0.99$ and rearranging for TF):

$$TLOD = 1 - \left(1 - \frac{0.99}{N}\right)^{\frac{1}{cov}} \quad (2)$$

This is shown as a function of N (number of mutations in the panel) in **Figure R10b**, where we have also annotated the error rates for SNVs and SVs in duplex and uncorrected sequencing.

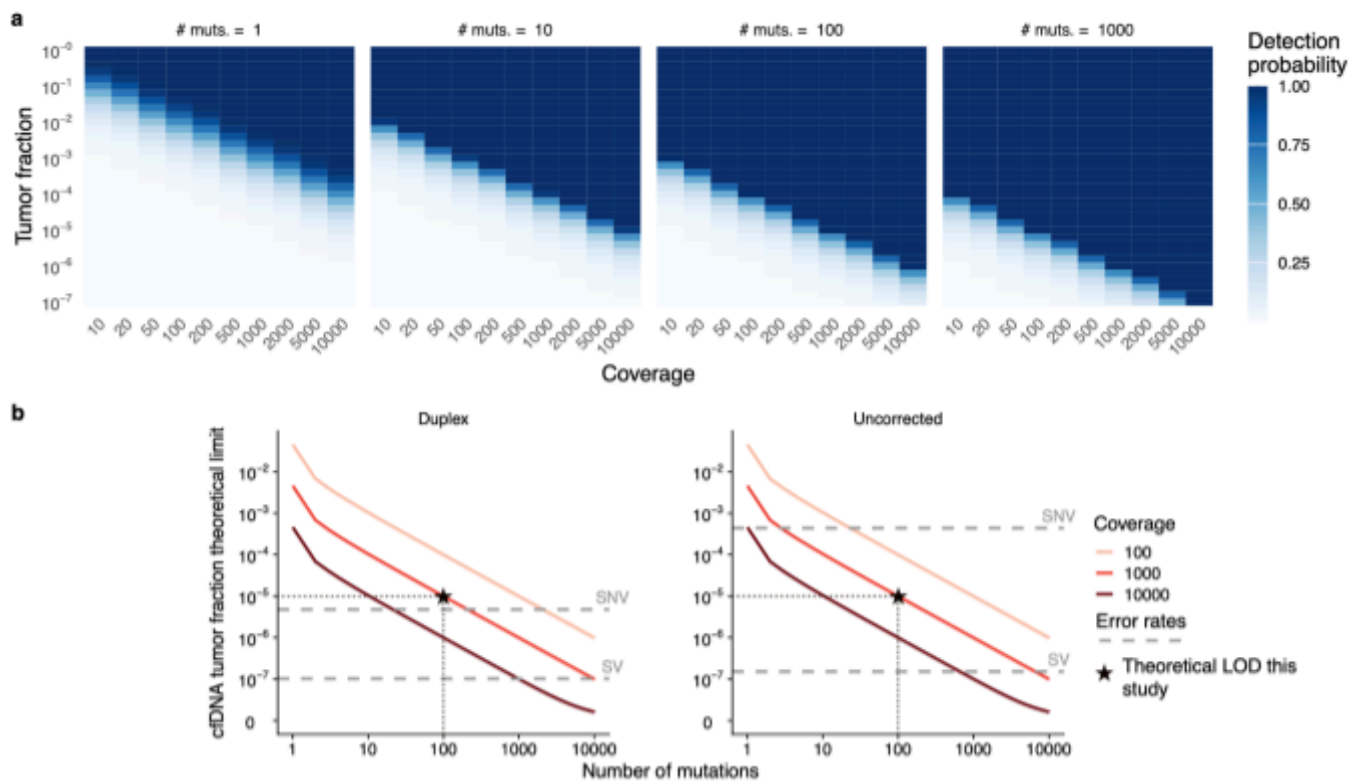


Figure R10 a) Detection probability as a function of cfDNA tumor fraction, number of mutations assayed and coverage using formula provided by Zviran et al. b) Theoretical limit of detection (probability of detection > 0.99) as a function of the number of mutations (x-axis) and Coverage (colors). Error rates of SNVs and SVs estimated in duplex and uncorrected sequencing are shown on the left and right respectively.

We note that the number of structural variants (SVs) detected in a typical high-grade serous ovarian cancer (HGSOC) tumor is lower than the number of single nucleotide variants (SNVs). However many SVs in HGSOC have high copy number, notably those implicated in a high level amplification meaning the effective number of mutations is increased (the above theoretical considerations assume a single copy of a mutation per cell).

We have now incorporated the above analysis into the paper. An overview of the above approach to estimate the LOD is included in the methods, **Figure R10** is included as **Supplementary Figure 5** and the following text has been added to the results section (line 342):

“Detection sensitivity in cfDNA assays depends on genome coverage and number of variants in the panel²⁸ (Methods). With typical parameters (1,000X coverage, 100 mutations), the theoretical limit is 1×10^{-5} . SV error rates (1×10^{-7} for both duplex and uncorrected) are substantially lower than this theoretical limit, while average duplex-corrected SNV error rates (6.7×10^{-6}) are of similar magnitude, creating uncertainty at this detection threshold (Supplementary Figure 5). Uncorrected SNV sequencing produces much higher error rates (4×10^{-4}). The improved signal-to-noise ratio of SVs enables confident detection of tumor DNA even from single events without requiring error correction, though advanced SNV assays with lower error rates are now emerging²⁹. Furthermore SVs are frequently associated with high-level amplifications which can further enhance sensitivity despite being less numerous than SNVs”

Our reported error rates of 10^{-7} for SVs reflect aggregated measurements across all patients in the entire dataset. We also note that pooled probe sets are applied to multiple samples (between 5-10). Therefore with a duplex depth of 1000X, 100 SVs applied to 10 samples and read counts aggregated across 20 patients gives total read counts $\sim 10^7$. We clarified this calculation in the methods and ensured we are not counting duplicate reads in the uncorrected sequencing, the error correction for generating duplex and simplex consensus reads naturally takes into account the read duplicates (line 1273).

“To compute error rates across sequencing types (duplex, simplex, raw uncorrected) and mutation types (structural variants and single nucleotide variants) we applied the patient specific probe set to multiple samples from off-target patients. We then summed the counts of reference supporting reads and variant supporting reads for off-target variants across all patients and defined the error rate as variants supporting reads divided by total number of reads.”

Zviran, Asaf, et al. "Genome-wide cell-free DNA mutational integration enables ultra-sensitive cancer monitoring." *Nature medicine* 26.7 (2020): 1114-1124. Ref #28

I disagree that estimates of clone fractions should not be done in a manner that corrects for copy number. Given that structural variants specifically are likely to be associated with copy number alterations this might be likely to cause greater error than methods using SNVs. Given that there is a ~linear relationship between the number of supporting reads in ctDNA and mutation multiplicity, a cancer cell fraction estimate of 20% could easily change to 40% (or more). Whilst it is true that when comparing across samples copy number would not affect the overall VAF to the extent that it changes by orders of magnitude, for estimates of cancer cell fraction (that control for ctDNA fraction) it might still be expected to make a big difference. Previous work has shown a linear relationship between that work that does not correct for copy number systematically underestimates the cancer cell fraction.

We agree with the reviewer that we overlooked these possible biases in our first submission and appreciate the comment. To address this, we developed an approach that leverages accurate cancer cell fraction (CCF) from single cell whole genome sequencing data. Using single cell copy number phylogenies, we first identify the most-recent common ancestor (MRCA) of cells containing a structural variant and then calculate the tree derived CCF by dividing the number of descendents of the MRCA by the total number of cells. We then calculate the VAF of each SV from the pseudobulk BAM file. These two values give us a way to infer a correction factor which is used to convert a given VAF measurement to a CCF value. This is then applied to VAFs measured from cfDNA. **Figure R11** below outlines this approach.

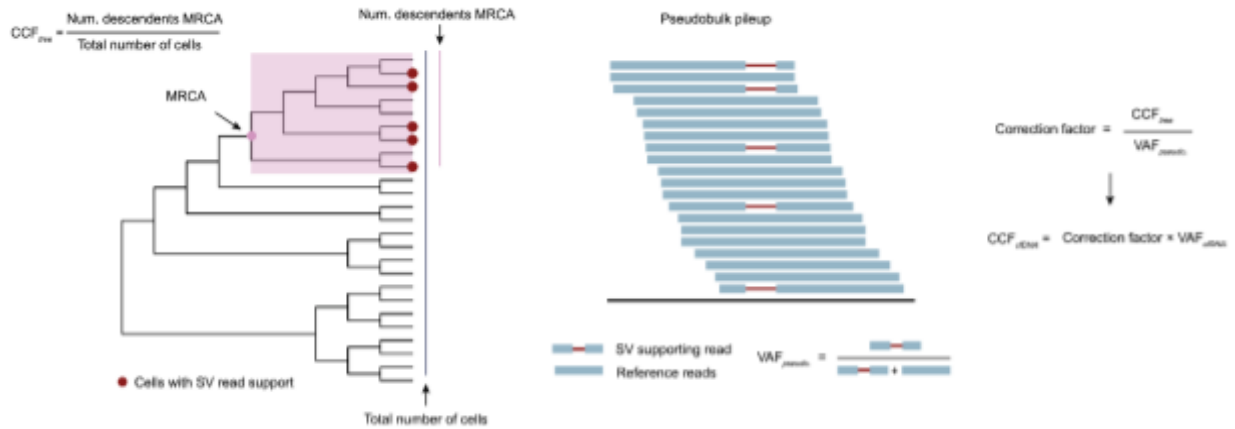


Figure R11 Overview of approach for correcting clone frequency estimates using CCF estimates derived from scWGS data.

Figure R12 compares clone trajectories with and without this correction. Trajectories remained broadly similar, likely because averaging multiple variants with different copy numbers mitigates large biases, and a large number of timepoints show single-clone dominance in cfDNA. Notable differences were observed in patients 083 and 037. In 083, both clones have a number of clone specific high level amplifications with associated SVs, therefore the correction methodology likely makes an outsized difference here.

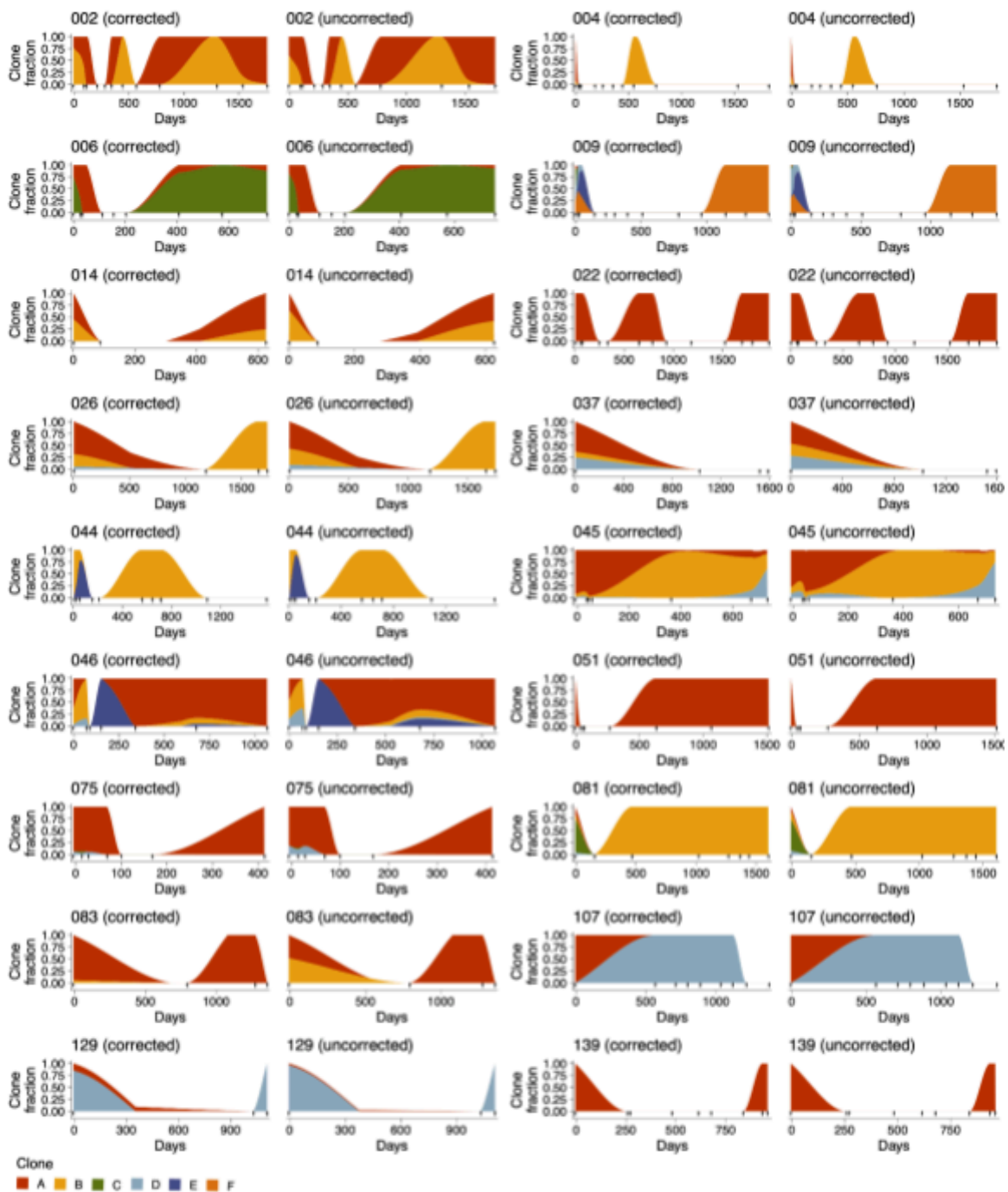
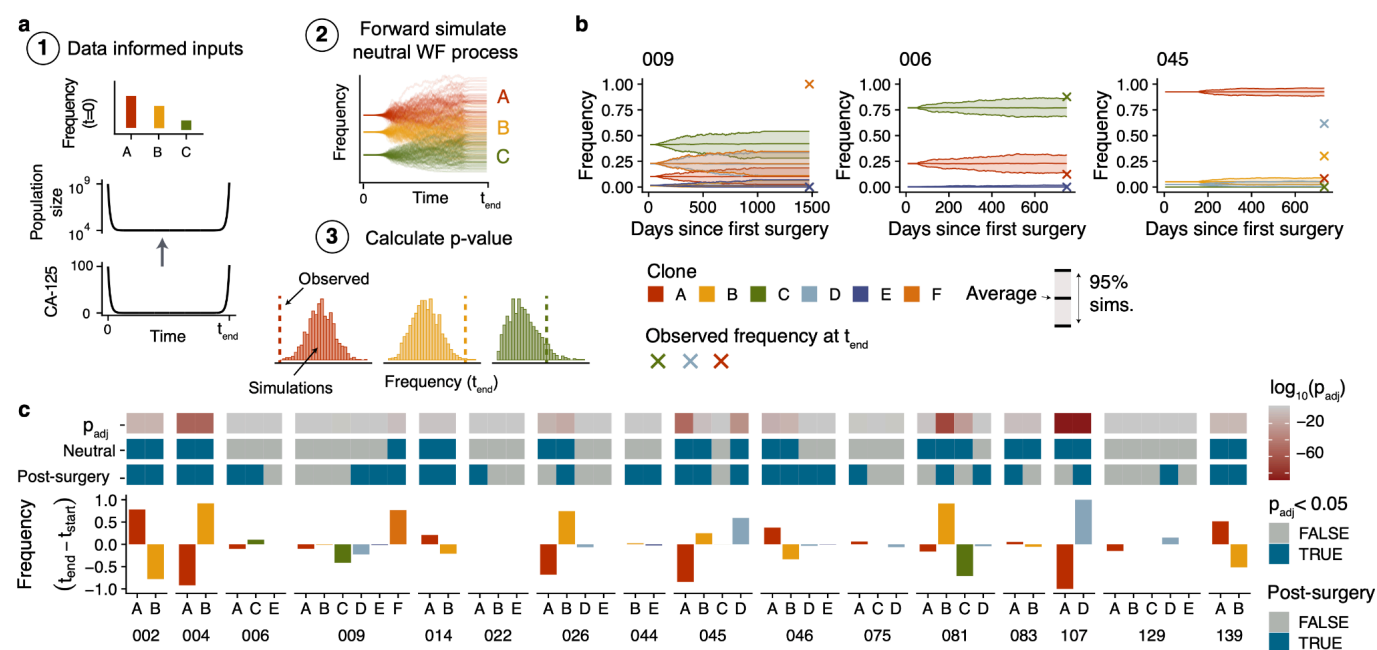


Figure R12 Clone trajectories with (left) and without (right) correction for CCF estimates for all patients.

In the Fisher-Wright model, am I correct that the authors assume that any deviation from neutral evolution would be driven by treatment? Could the authors elaborate on the potential bottlenecks that might arise from (debulking) surgery and the metastatic transition that might be happening concomitantly and would not be associated with treatment.

Yes, the reviewer is correct that our Wright-Fisher model attributes deviations from neutral evolution primarily to treatment-induced selective pressures. We agree with the reviewer and have added text to the discussion (see earlier reply on page 15) to address these important limitations.

In addition, we have added a track to **Main Figure 6c** (see below) to add additional context, we now show whether each clone was detectable at any timepoint post-surgery. This visualization demonstrates that changes in clone frequency cannot always be explained by complete surgical excision of clones, suggesting that treatment-related selection and other biological factors also contribute significantly to the observed dynamics. This is particularly notable in patients 009, 045 and 046 where the majority of clones are detectable in post-surgery cfDNA samples but we observe large changes in clone frequencies over time.



Minor comments

Did the degree of tissue sampling used to construct the patient specific panels affect the assessment of the pattern of relapse as monoclonal vs polyclonal? Also - did effective coverage influence this?

This is a good point and we agree that our inability to comprehensively characterise all clones will likely impact clonality assessment in some cases. We generated multi-site scWGS wherever possible and 13/18 patients have at least 2 sites sequenced. Our results showing that there is a reduction in diversity and the number of clones identified between baseline and recurrence should not be overly influenced by tissue sampling so we have emphasized these results over clonality assessment. We have added text to caveat these findings in appropriate places throughout the results section and discussion:

Line 891:

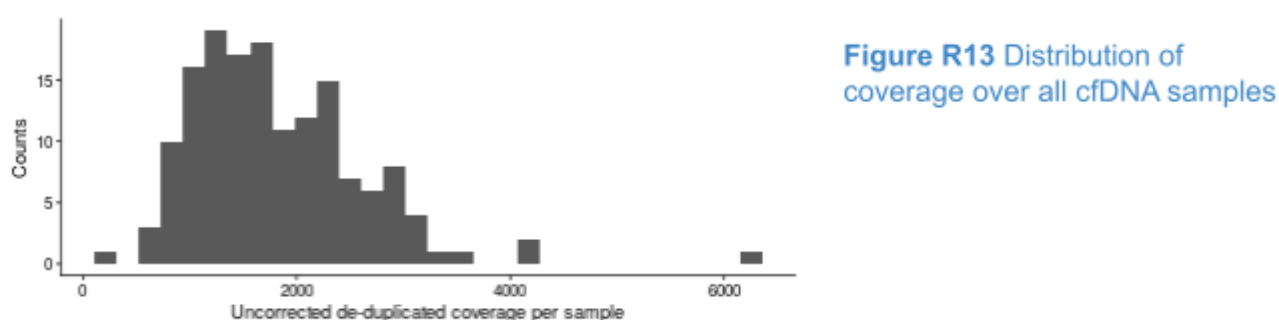
"In 6/18 patients (002, 006, 014, 045, 046, 139; see Fig. 4 & Supplementary Figure 7), multiple clones were detected in post-recurrence plasma samples, highlighting that chemo-resistance is polyclonal in many patients (Burdett et al. 2023), although limitations in comprehensively identifying all clones present at diagnosis may influence these observations."

Line 1030:

“The inability to comprehensively identify all clones present at diagnosis, and the emergence of novel clones post tissue sampling may limit our ability to fully resolve and track clonal populations.”

We also note the additional scWGS data from recurrence tissue samples and WGS of recurrence plasma samples in select patients provides further data to interpret the clonal dynamics (see page 8).

Coverage should not overly influence these results. cfDNA coverage was fairly consistent across samples. Raw de-duplicated coverage (used for VAF estimation of SVs) was >500X for all but one sample, with a mean value of 1819X, see **Figure R13** below that shows the distribution across samples. This data is included in Supplementary Data Table 1.



For scWGS, the pseudobulk coverage of all tumor cells was sufficient to identify subclonal SVs in all samples. In 2 samples we did not identify subclonal SVs, but this was due to clonal homogeneity rather than insufficient coverage. Only one of these samples (051) was part of the subset of patients we followed longitudinally. This data is in Supplementary Figure 2.

The error rates in 2D are not reflective of the state of the art – could the y axis on this plot be extended to include error rates in keeping with best in class SNV based approaches?

We appreciate the suggestion, however we feel it is important to the reader to report the error rates measured from the experimental data in this study. We believe this provides the most accurate representation of the comparative error profiles between SVs and SNVs in our specific assay system.

However, we do acknowledge the reviewer's valid point that more advanced SNV detection methods can achieve lower error rates than those represented in our figure. To address this, we have added the following sentence (line 410) to the manuscript where we cite the NanoSeq method (error rates approaching 10^{-9}). We are not aware of this approach being applied to cfDNA samples as it likely requires substantial modification to be used in cfDNA assays.

“The improved signal-to-noise ratio of SVs enables confident detection of tumor DNA even from single events without requiring error correction, though advanced SNV assays with lower error rates are now emerging([Abascal et al. 2021](#))”

Abascal, Federico, et al. "Somatic mutation landscapes at single-molecule resolution." *Nature* 593.7859 (2021): 405-410.

Referee #5:

I thoroughly enjoyed reading this fantastic paper. The integration of high quality scWGS data and analyses with longitudinal monitoring of evolutionary trajectories using ctDNA in patients is incredibly exciting to me and I believe of great interest to many people from multiple disciplines. Harnessing this approach, Williams et al are able to track evolving clonal frequencies through real time and view dynamics in the context of treatment received and patterns of recurrence. They are also able to track (pre-identified) SVs through tumor evolution which is a current challenge for the field but will be critical for monitoring the consequences of ongoing chromosomal instability in tumors. For a paper packed with novel approaches and insight, I found the analyses straightforward to follow, the statistical analyses sound and the figures beautiful. I do, however, have the following minor points of clarification that I believe would aid understanding:

We thank the reviewer for their enthusiastic feedback and appreciation of our integrative approach to tracking tumor evolution.

Minor comments

- To assign mutations and SVs to clones the authors test the distribution of the subset of cells positive for a mutation across clones in the tree. In the examples given in Fig 1 this is straightforward but how is this done for more ambiguous cases? Would it be possible to add detail to the assigning SVs to clones section of the methods?

We agree that assigning SVs to clones can be challenging in some cases, especially in cases with skewed clone size distributions where small clones are susceptible to false negatives, and in cases with relative clonal homogeneity. To address this, we integrated high-resolution 10kb clone-level copy number profiles with SV locations, requiring clones lacking an SV to show no supporting reads and no associated copy number transitions. In the first submission, we relied solely on computational assignment of SVs to clones and copy number; however, manual review revealed limitations in this approach due to subtle differences in copy number breakpoint locations between clones. We have now manually reviewed all 4,000+ SVs to ensure accurate assignment. **Figure R14** illustrates this refinement process, showing an SV from patient 081 initially classified as clone A-specific but correctly reassigned as truncal after identifying the associated copy number profile in all four clones. This manual validation improved clone trajectory reconstructions compared to our initial submission, notably revealing the absence of clone A from post-recurrence cfDNA samples in patient 081.

The following is included in the methods to explain our approach.

“Assigning SVs to clones was done using the matrix of read counts per SV per cell, the cell to clone label mapping, and the clone level 10kb copy number profiles. First, we summed the SV supporting reads across clones giving an SV by clone matrix. Any SVs with non-zero read counts were assumed to be present in the clone. In addition, when SVs could be mapped to copy number changepoints identified at 10kb resolution, we checked whether there existed other clones that had the same copy number changepoint but lacked read level support for the SV. In these cases, the SV was also assumed to be present in that clone. This was to circumvent cases where the total number of cells was too low to confidently assume the absence of a particular SV. Finally we manually

reviewed all SVs and copy number profiles to produce a high confidence set of clone specific SVs. Manual review involved inspecting the location of all SVs relative to copy number profiles and ensuring absence of SV read support did not coincide with a proximal copy number change.”

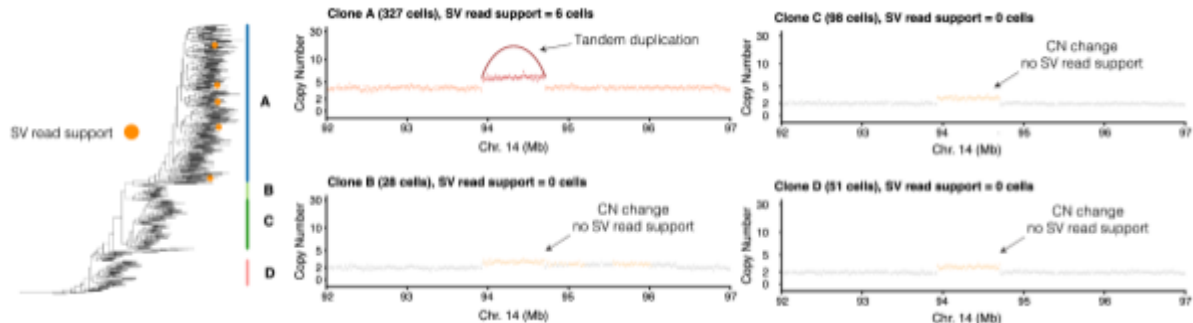


Figure R14 Phylogenetic tree for 081 annotated with clones on the right hand side, cells with read support for a 800kb duplication are highlighted in orange. On the right hand side, copy number profiles for all 4 clones at the locus of the duplication. All 4 clones show a CN change consistent with the SV but only clone A has read support. Following manual review this SV was reassigned from clone A to be clonal.

- What is the impact of site particularly on the phylogenetic inference? It looks like clonal structure correlates very strongly with site. How much would the authors expect this to impact the patterns of clonal evolution that they observe? For example, are responses to treatment and therefore selective pressures likely to differ between sites?

This is an excellent point and something also raised by reviewer #4. We have added text that discusses these limitations throughout the text, see reply to reviewer #4 on page 22. In addition we note that the additional scWGS from recurrence specimens and WGS of plasma specimens provides additional context here, discussed in reply to reviewer #2 page 8 and added to the revised manuscript in Figure 4 and Supplementary Figures 9 and 10.

- The authors report clone specific complex SVs in Figure 3. Presumably these were detected in scWGS pseudobulks and the characteristic breakpoints sequenced in cfDNA. Please could you provide more details in the methods of how the patterns of SVs were predicted in pseudobulks?

Yes this is correct, we have added the following section to the methods:

Identifying complex rearrangement processes

To identify complex structural events we manually reviewed copy number profiles overlaid with SVs for all patients and clones. We then assigned events as follows:

- i) Chromothripsis: high density of rearrangement breakpoints and oscillating copy number profiles (eg 004, 083)
- ii) breakage fusion bridge cycles: Foldback inversion breakpoints coincident with copy number changepoints and staircase like copy number profiles (eg 045, 051)
- iii) Pyrgo: stacked tandem duplications resulting in tower-like copy number profiles (eg 081)
- iv) Complex inter-chromosomal rearrangements: High diversity of copy number states with translocations between two or more chromosomes (eg 002)

v) ecDNA: focal high level amplification with large copy number diversity between cells (none detected)

In addition, oncogenic amplifications (whether they were part of an identifiable complex rearrangement process or not) were defined as high level amplification (>10 copies), amplifications (>2xPloidy).

- There are a couple of typos in y-axis labels e.g. Figs 2c, 5a.
These have been corrected. "Specific" in 2c and "clone" in 5a.

Referee #5 (Remarks on code availability):

The code provided by the authors with the exception of the code for Articull (manuscript in prep) were available on github and I was able to run the Mondrian-scwgs nextflow pipeline with relative ease.

Articull is now available as a preprint and its associated code is now available on GitHub. We have included these citations in the manuscript.

Referee #2 (Remarks to the Author):

Thank you for the comprehensive responses and detailed clarifications. No further comments.

Referee #4 (Remarks to the Author):

This revised manuscript from Williams and colleagues is substantially methodologically improved compared to the initial draft. I also note they have included further samples to address previous concerns about novelty.

The corrections for copy number in the calculations related to clone frequency seem appropriate, and I am pleased that they have removed the claim about the claimed limit of detection of their approach being 10^{-7} .

Supp Fig 5 still contains an error – they underestimate the LOD of SNV-based approach - the true limit of detection is $\sim 10^{-7}$, not 10^{-5} (e.g. see Bae et al, Nat Genet, 2023).

I recommend acceptance pending this minor revision.

We have added a reference to Bae et al/ (ref 20) in the following sentence:

“The improved signal-to-noise ratio of SVs enables confident detection of tumor DNA even from single events without requiring error correction, though advanced SNV assays with lower error rates are now emerging^{12,20}.”

We have elected to keep Supp Fig 5 as is, as the error measurements we report are measured directly from our data, we do not claim to be reporting error rates of the current best in class most sensitive assays. To further clarify we have added the following to the legend:

“Error rates of SNVs and SVs estimated in duplex and uncorrected sequencing from our data are shown on the left and right respectively.”

We feel that using error rates measured from our data is the best way to compare SV detection vs SNV detection given that one of the novel contributions of our paper is highlighting the utility of SVs for ctDNA detection, which to our knowledge has not been explored in depth using the most advanced error corrected sequencing technologies.