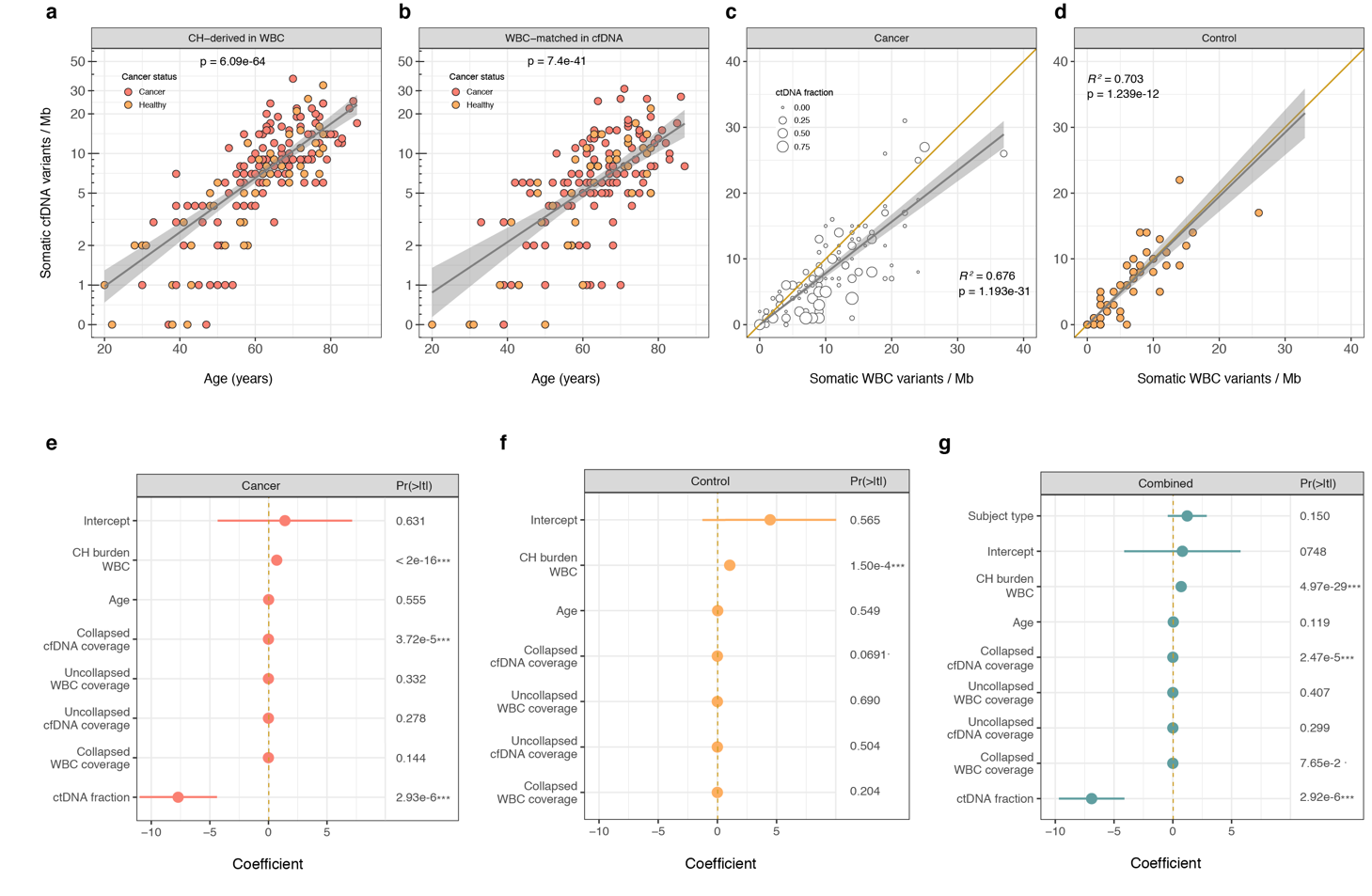
The main new finding as highlighted within the abstract appears to be that the majority of cfDNA mutations are most likely derived from leukocytes due to clonal hematopoiesis (CH) and that CH is a "pervasive" biological phenomenon. While confirming prior findings is of significant value, this observation is not a new one especially because CH is very common with age, “trending towards inevitability” when considering somatic variants in circulating leukocytes (Zink et al 2017 Blood). Seeing that circulating leukocytes are the dominant source of cfDNA as demonstrated by several studies, finding CH-derived variants masquerading in cfDNA seems hardly surprising. More specifically, as cited by the authors (Ref #26), a prior study of 259 healthy adults (more than their 124+47 combined) previously found that the majority of these subjects those had ≥1 nonsynonymous mutation (>75% for those older than 50) in the plasma cfDNA, and that most of these somatic variants were present in matched blood leukocytes. That same study also showed that many mutations were passengers and not classic CHIP mutations. The authors themselves state twice that their findings are "consistent with" Ref 26 thus acknowledging that this result it is not novel.

Authors: We thank the Reviewer for the most pertinent of questions. The Reviewer is correct that clonal hematopoiesis (CH) has been described, and that non-synonymous mutations have been identified in circulating cell-free DNA (cfDNA) of >75% of cancer patients. Our study, however, provides several novel aspects to the characterization of CH in healthy individuals and cancer patients, given the unprecedented approaches employed in this study, namely the depth of sequencing attained in the cfDNA and white blood cell (WBC) samples analyzed coupled with the use of unique molecular identifiers (UMIs), the large ‘genomic footprint’ assessed and the bespoke hierarchical Bayesian model to reduce errors in sequencing data. This novel approach allowed us to go above and beyond the observations made in Liu *et al.* (PMID: 30475948). Here, we demonstrate that:

1. Although CH correlates with age, it can be detected in the cfDNA of a substantial proportion of young patients and is almost always found in patients with advanced cancers, an aspect not investigated in Liu *et al.* given that in that study, the focus was only on healthy individuals.
2. In addition, the average depth of WBC sequencing achieved in Liu *et al.* (i.e. ~406X) does not allow for a fair comparison with our study, where both cfDNA and WBC samples were sequenced with the same approach at comparable depths (i.e. deduplicated uncollapsed target sequencing depth >60,000X).
3. Our results demonstrate objectively and based on direct evidence, the importance of matched WBC sequencing as overall, >50% of the mutations identified in cfDNA of cancer patients originate from CH.
4. We provide direct evidence to demonstrate that cfDNA sequencing without taking into account the results of WBC sequencing can be misleading; this is perhaps best exemplified by mutations affecting `cancer genes` (e.g. *TP53*) which can be present in tumor-derived cfDNA and also be part of CH.
5. Finally, we provide evidence that the assessment of the repertoire of somatic mutations in cfDNA post-therapy (in the context of disease monitoring) can be confounded by CH-associated mutations, and that these post-therapy alterations preferentially affect specific genes.

In support of the above statements and barring any association with smoking history as we show later in this response at point #11 of this Reviewer’s comments, we demonstrate here that the frequency of CH in cfDNA is a function not only of age (ref) but several confounding factors which could only be investigated in the current study on account of the technological innovations. Using a zero-inflated Poisson regression model, **Response to Reviewers Figure XXa** and **XXb** show that the CH-derived mutational burden is better associated with age when measured from WBC than cfDNA (p=6.09e-64 versus p=7.4e-41, respectively). As shown in **Response to Reviewers Figure XXc** and **XXd**, this is due, at least in cancer patients, to the fact that there is on average more CH-derived mutations in WBC than in cfDNA. In univariate analysis the fraction of tumor derived DNA in plasma i.e. ctDNA fraction explains this difference, whereby the ctDNA fraction is, as expected, significantly negatively associated with the burden of WBC-matched variants in cfDNA (two-sided Wald test, p=1.81e-3). In a multivariate model including several additional analytical conditions e.g. mean target collapsed depth in WBC, mean target collapsed depth in cfDNA, etc. and restricting to cancer patients, the ctDNA fraction and the mean collapsed target depth in cfDNA in addition to the burden of CH-derived mutation in WBC were statistically associated with the number of CH-derived mutations in cfDNA (see **Response to Reviewers Figure XXe**). A comparable model restricted to healthy controls who do not have tumor derived DNA in plasma confirmed that the mean target collapsed depth in cfDNA was the only additional variable with a significant positive association (see **Response to Reviewers Figure XXf**). In a combined model, however, including both cancer and healthy controls where the ctDNA fraction of the healthy controls was entered as 0, the mean target collapsed depth in cfDNA and WBC as well as the ctDNA fraction and the CH-derived burden in WBC were significantly associated with the number of WBC-matched variants in cfDNA (see **Response to Reviewer Figure XXg)**.



**:**

In addition, our study is novel as it challenges one of the current paradigms in the field of cfDNA analysis. Cohen *et al.* (PMID: 29348365) stated “there must be a limit on the number of bases queried in the test because the more bases queried, the more likely that artifactual mutations would be identified, reducing the signal-to-noise ratio”. Here, we demonstrate and validate with orthogonal methods that this statement is not necessarily correct. If sequencing is performed with UMIs and a robust error correction method is applied to the sequencing, a relatively large genomic footprint (~1Mb) can be employed for the detection of tumor-derived mutations, but this can only be accurately achieved at present if WBC sequencing is performed concurrently.

Given the technological advancements presented in our study and the shifts to the current paradigms in cfDNA sequencing and data analysis our manuscript provides, we would contend that although some elements of our manuscript are confirmatory, collectively the results are novel and would constitute an important contribution to both the deployment of cfDNA sequencing as a tool for early diagnosis and disease monitoring, as well as to the characterization of CH in WBC and cfDNA samples.