

Peer Review File

Manuscript Title: A pan-cancer compendium of chromosomal instability

Reviewer Comments & Author Rebuttals

Reviewer Reports on the Initial Version:

Referees' comments:

Referee #1 (Remarks to the Author):

Dreus and colleagues report on a method for quantification chromosomal instability (CIN) using copy number alterations detected in 7880 tumour samples from 33 cancer types from the TCGA database. According to their analysis, approximately 80% tumours have CIN, although how the threshold is set is questionable and may be an overestimate. They assemble 17 CIN genomic signatures, yet only the top two are characterised in more detail and are assigned to mitotic errors and homologous recombination deficiency. How a specific signature is linked to a given aetiology is unclear and the proposed aetiologies seem largely speculative. Overall, it is not clear what the impact of such a signature database might have on cancer prevention or predicting responses to treatment, or indeed whether there is any use for it at all. In addition, the manuscript is unlikely to be understood by a non-specialist readership, thus it is best suited for a more specialised journal.

Specific points:

1) Could the authors explain the essential difference between copy number signatures, as termed in their 2018 paper (Macintyre et al. Nature Genetics, PMID: 30104763) and CIN signatures computed here? It seems that copy number signatures are just termed CIN for this study, as broadly, similar features were taken into account to assess the signatures. In Macintyre et al. 2018: “to identify copy number signatures, we computed the genome-wide distributions of six fundamental copy number features for each sample: the breakpoint count per 10 Mb, the copy number of the segments, the difference in copy number between adjacent segments, the breakpoint count per chromosome arm, the lengths of oscillating copy number segment chains and the size of segments”. In the present manuscript (p.22, lines 620-629): “we derived genome-wide distributions of five fundamental copy number features. Using all copy number profiles, we computed distributions of segment size, copy number change to the left neighbouring event (called change point), breakpoints per 10 Mb, breakpoints per chromosome arm, and the number of neighbouring segments that oscillated between two copy number states. [...] In a major deviation from Macintyre et al. we did not use the absolute copy number of a segment as a feature as this can artificially split a signature by ploidy status despite representing the same mutational process”;

2) P.4, lines 95-99: “Using the proportion of patients with detectable levels of CIN per tumour type, we assessed the relationship between CIN and cancer-specific five-year survival rates (Fig. 1b). Similar to previous studies⁴, we found that higher rates of CIN are significantly associated with a decreased survival (Beta regression model, $p=0.0035$) and an increased variability in survival (Beta regression model, $p=0.0025$)”.

All the high CIN rate cancers in the upper-right quarter of Figure 1b are associated with good five-

year survival (at the extreme, “Testis” at almost 100% 5-year survival and near 100% detectable CIN). I don’t know about the beta regressions (coefficients are not explicitly shown along with the p-values here, to get a better idea of the odds) but it doesn’t seem to be any relationship between 5-year survival and detectable CIN rates when looking at all the high CIN rate samples together;

3) In their previous description of copy number signatures, the authors found similar component weights for the identified signatures in ovarian cancer in the two independent cohorts (namely, PCAWG-OV and TCGA). How robust are the CIN signatures found in this work, as assessed by detection in other cohorts (e.g. METABRIC has CNA data for breast cancer, MSKCC has CNA data for metastatic colorectal, prostate and skin cancers among others)?;

4) It seems, from the comparison with the ovarian copy number signatures previously described, that the correlation with any of the 17 CIN signatures is low (Extended Figure 5, it looks like the maximal corrected cosine similarity value found is around 0.5). Could this be due to differences in the approaches used to detect CNA events in the two studies? The coverage at genomic positions corresponding to SNPs on the SNP6 array should be, on average, about 40x; whereas the coverage of the shallow WGS (sWGS) conducted in Macintyre et al. 2018 was ~0.1x. Could some events not have been detected by sWGS?

5) Related to the previous point, are the ovarian cancer copy number signatures from the previous study not cancer-specific signatures then, as they showed some degree of similarity with many different pan-cancer CIN signatures in this study?;

6) The definition of cancer-specific CIN signatures is unclear (p.23, line 680): are these signatures specific to one cancer type? To a small subset? From Figure 2a, it seems that none of the 17 signatures have activity in only one, or very few, TCGA cohorts. In addition, why are the 17 signatures termed “pan-cancer” (e.g. p.4, line 109) when the authors stated in the methods that “therefore the 10 pan-cancer signatures were combined with the 7 cancer-specific signatures to yield a set of 17 signatures of CIN in human cancers” (p.24, lines 709-710)?;

7) Activity estimation is referred to as a heuristic process, there is mention to an activity matrix, on it not necessarily being strictly dependent on the number of CNA events detected, etc. But could the authors clearly state in the main text what they mean precisely by signature “activity”? In p.5, lines 134-135, “we numbered the signatures by their prevalence across the TCGA cohort. Signatures 1 and 2 show pervasive activity patterns”, is it linked to prevalence? In addition, could it be shown that there is no linear relationship between CNA numbers and signature activity in a given cancer type?;

8) Signature 17 seems intriguing: from the heatmap in panel 2b, it seems like there were only a few CNAs called in PCPG samples (>1%, short segment size), the rest of the heatmap column appears in white, so I would assume 0% CNAs detected in every other cohort. However, activity was detected in multiple cancer types, as shown in panel 2a. In relation to point 2), could this signature, or even signature 6 that does not seem to be associated with any biological covariates either (Table 1), be detected in other independent, validation cohorts?;

9) Signature 1 is said to be possibly related to a molecular clock-like mutational process and include events acquired sporadically through the normal ageing process (p.6, lines 181-192). In that sense,

from only association studies as carried out here, it is difficult to understand its real value/confidence in relation to specific cancer-associated processes. I wonder if this same signature would be detectable in normal tissue samples, as a control?

Minor:

1) P. 7, lines 233-234: “signatures 11 and 14 were dependent on signature 1, suggesting they are associated with errors during mitosis”; also p.15, lines 381-382 “For example, samples with signature 1 also carry signatures 11 and 14 more often than expected by chance”. Yet it seems S11 and S14 are mutually exclusive (Extended Data Figure 9), could the authors comment on that?;

2) Figure 2A: the colour code used is confusing, as in the heatmap in panel (a) samples having <25% activity are coded in white; at first look of the heatmap it seems like white spaces correspond to no activity at all, especially when comparing to panel (b) in which 0% CNAs is also white-encoded;

3) Extended Data Figure 5: the colour legend is somewhat misleading, as it does not use the same bounds in panels (b,c) compared to panels (e,f): dark blue seems to be close to -1 in (b,c), but rather close to -0.25 in (e,f);

4) Extended Data Figure 9: with that colour code, it’s almost impossible to see any differences (only clear between white=0% vs the rest of the green-coloured cells of the heatmap);

5) P.5, lines 126-127: “Combined, these data form the basis for a detailed description of the biological processes underlying the CIN signatures”. This should be changed to “putative biological processes underlying the CIN signatures”, as it arises from associations only (as stated in the immediately preceding sentences, “we associated the signature definitions and ...”, “we observed associations with ...”, “we associated differences in ...”);

Referee #2 (Remarks to the Author):

Drews et al. describe a new pan-cancer signature method for chromosomal instability (CIN). They used TCGA data SNP array profiles to expand CIN signatures they previously reported for ovarian cancer. The features they used to define signatures are segment size, copy number changes, breakpoints per 10 Mb, breakpoints per chromosome arm, and the number of switching between two copy number states. They found 17 CIN signatures. For each signature, the authors report:

- 1) The type of features that contribute to the signatures;
- 2) The similarity to ovarian CIN signatures;
- 3) Correlations with gene mutations;
- 4) Present evidence that these associated genes may contribute to DNA repair deficiency.

The authors focus on two signatures CS1 and CS2, that present in nearly all samples with CIN. They report that CS1 is correlated with shortening telomeres and indicate evidence of a negative correlation between telomere activity with TERT and TERC. CS2 was associated with HRD.

The paper is well written and presents results with potential importance, particularly the association of CS signatures with mutations in DNA repair genes that have not been demonstrated before. The authors focused on establishing CS1 and CS2 signature validity with established signatures such as telomere length or HRD signatures (SBS3). Moreover, the authors associate genes such as BRCA1/2 and TERT/TERC with these signatures. However, there are multiple biomarkers for mutations in these genes. It might be worth focusing more on novel gene-signature associations that will demonstrate better the CIN signatures' usefulness.

Major revisions:

1) The survival analysis was done on tumor type. Although, the regression analysis found a significant association, it seems, as the author mentioned, that a small subset of tumors with a low proportion of CIN drive this association. Tumor types with high CIN, as the author said, have variable OS. I suggest showing the association of survival with CIN within tumor type to convince the reader that CIN is truly associated with OS.

2) While the association of CS2 with BRCA1/2 is convincing, 70% of the samples identified as CS2 lack any event. None of the previous studies showed a high fraction of unexplained HRD cases (when methylation was performed). This may be concerning. I would like the authors to compare CS2 to other copy number based HRD signatures such as HRD scores and HRDetect.

3) For a subset of ovarian and breast TCGA samples, WGS is included in PCAWG. Hence, HRDetect can be calculated. HRD scores (PMID: 26957554) can be calculated from SNP arrays.

4) Recent WGS studies have produced a very detailed description of somatic copy number alterations at higher resolution than SNP arrays. Can the authors compare the presence of CSs to the PCAWG copy number alterations profiles that were recently published and briefly mentioned in the text?

5) Since the biological interpretation of CIN signatures is of paramount importance, how will these CSs be identified based on WGS? The ability to detect these is tricky since the number of segments and their sizes will be different.

6) The methodology for the signature derivation is not fully clear. Please clarify the method, maybe by using an illustration or a flow chart.

Minor questions:

1) The authors have removed cancer-specific signatures that have a cosine similarity over 0.74 with any Pan-cancer signature. Can they explain why this cut-off?

2) "This suggests that telomere shortening is indeed a key factor contributing to the mutational process represented by CIN signature 1." Does the author mean to say telomere shortening is a consequence of the mutational process?

3) In lines 158-161, the authors show that CS1 is associated with shortening telomere length. In Barthel et al. cited by the authors, shortening telomere length was already anti-correlated with amplification of TERC and TERT. So, results on CS1 negatively correlated with telomere gene activity in lines 163-172 is expected. Can the authors test if the association of CS1 with TERC and TERT is independent of telomere length?

4) CS3 is associated with the CDK12-associated signature in ovarian cancer. However, the authors did not report an association of CS3 with CDK12 mutations. Can the authors present the CS3 levels in CDK12 mutated vs. non-mutated?

Referee #3 (Remarks to the Author):

Drews et al. present an attempt to quantify the pan-cancer landscape of copy number instability using a mutational signature approach. They apply their previously published (Macintyre et al.) computational method (originally applied to the TCGA ovarian cancers) to the rest of the tumor types in the TCGA dataset, and claim that the sources of chromosomal instability (CIN) are mitotic errors and impaired HR. Although the idea that CNV signatures can be informative for understanding CIN is promising, the approach described is not sufficiently novel or informative, and the findings lack solid supporting evidence. Here are some major issues:

1. Too many of authors' conclusions are based on correlation coefficients. As the authors are undoubtedly aware, when there are many data points, even weak relationships can become statistically significant. But if the effect size is small, the relationship is often not meaningful. For instance, the paper claims that there is a correlation between 5-years survival and CIN. But it can be plainly seen from Fig 1b that this correlation is driven by a small number of outliers, and I am surprised that the authors are claiming based on this figure that CIN has association with survival. Cancers of different tumor types have vastly different outcomes not because of CIN but more likely because of several other factors, e.g., testis and liver cancers have similar values for the median number of CINs and proportions of patients with CIN, but the five-year survival of the patients with these tumor types are on the two ends of the spectrum. Similarly, the conclusion that the majority of CINs arise due to errors during mitosis or impaired homologous recombination also is based on the co-occurrence of signatures with gene alterations or other signatures. But co-occurrence of minor CIN signatures with CS1 and C2 given that these two are present in 97.4% and 71.4% of samples is not conclusive in determining their source. The telomere shortening is discussed in depth, but the effect is prevalent mostly in the top 5% of samples with CS1, making it difficult to claim that CS1 is directly connected with telomere shortening.

2. The study fails to describe how the present paper is different from and adds to other papers on the same topic, especially given that the SNP data are very old and have been analyzed previously,

e.g., Zack et al (Nat Gen 2013), which analyzed a large subset (~5000 samples) of the data analyzed in the present paper.

3. For impaired HR signature, the length of copy number segments (4.9-30.8 Mb) seems to be much larger than what is reported by the SV signatures (1-10 kb) (Nik Zainal et al. Nature 2016), and it is unexpected for HR deficiency to be as prevalent as reported in this study. Authors claim that nine distinct signatures are related to homologous recombination deficiency, but they don't provide any mechanistic explanation. It is puzzling why such a large variety of patterns was not identified by studies carrying out SV analysis in whole genomes. We do not see why the authors did not extend their analysis to WGS and demonstrate that their findings are reproducible and consistent with the literature.

4. To determine whether CIN is present in a sample or not, authors use a somewhat arbitrary 20 CNA threshold which may inflate the fraction of samples with actual genomic instability. Again, the same samples have been analyzed before (not just SNP arrays but also WGS), but there is no comparison or examination of robustness of the conclusions on these parameters.

5. Possible overfitting and false positive assignments of signatures should be investigated further. Authors only tested the necessity of applying a minimum threshold on signature activity for two of their signatures. However, the two signatures they tested are those that associate with whole chromosome or chromosome arm copy number changes, which are easier to detect. Signatures that have high probabilities of generating smaller size CNAs and/or that have distributions that overlap with other signatures are likely to suffer more from false-positive assignments. In their supplementary table for sample specific assignments (SM_7_Estimated_CNAs_per_signatu), there are nonzero assignments with very low values. For example 88% of nonzero exposures for Signature 1, 82% of nonzero exposures for Signature 7, 80% of nonzero exposures for Signature 6 are less than 5. A signature of 43 dimensions cannot be assigned accurately with such low values of exposures. The modelling with Gaussian/Poisson distributions can also aggravate this issue by increasing the overlap of probability distributions across different signatures. If the signature assignments are inflated due to overfitting, the co-occurrence based association of various signatures with HRD would lead to misleading conclusions.

Drews et al; Nature_2020-07-14327

Point-to-point response to reviewer comments

We thank the reviewers for their constructive comments. This revision includes major improvements and additional results that have completely transformed the manuscript, including:

1. **Validating the robustness of our approach** by comparing samples from the same patients across five sequencing platforms and settings. We also defined signature-specific noise-thresholds, which makes the detection of low-activity signatures more robust and avoids overfitting.
[Manuscript: page 4 and Ext Data Fig 5 and 6; Methods: pages 15-21]
2. **Transparently identifying aetiologies** by combining associations with mutated driver genes, copy number patterns and multiple orthogonal data sets. We are now much more explicit about the many sources of evidence that went into each proposed aetiology, including a star rating. We present all evidence in a web resource for readers to explore at <https://markowetz.cruk.cam.ac.uk/cincompendium> (user: cnsigs2; password: souptomatobasil)
[Manuscript: pages 4-7 and Ext Data Fig 7,8,9; Methods: pages 22-33]
3. **Highlighting the clinical impact** of our approach by (a) using our signatures to identify novel drug targets and predict drug responses and (b) developing a signature-based classifier for predicting platinum-sensitivity in ovarian cancer.
[Manuscript: pages 7-10 and Ext Data Fig 10, Methods pages 34-37]
4. **Substantial extension of the data and resources** we integrate, including the PCAWG and ICGC cohorts, additional independent annotations and covariates, and measures of different types of CIN.
[Overview in Figure 1]

On the following pages we respond to all major reviewer comments in detail. We cite reviewer comments in full and highlight their key points in bold, followed by our responses in blue boxes.

General notes:

- Because the manuscript has been completely rewritten, several comments no longer apply to this resubmission.
- To keep the naming scheme simple, signatures are again ordered by prevalence across the TCGA cohort, but with the new signature-specific thresholds this ordering has changed compared to the initial submission. We renamed them to CX1-17 (rather than CS1-17) to make this difference explicit.

Referee #1

i) Drews and colleagues report on a method for quantification chromosomal instability (CIN) using copy number alterations detected in 7880 tumour samples from 33 cancer types from the TCGA database.

According to their analysis, approximately 80% tumours have CIN, although **how the threshold is set is questionable and may be an overestimate.**

The estimate of 80% is in line with previous estimates using different methodologies. The major advance is that our approach is more nuanced than previous approaches by building on our previously presented ovarian copy number signatures (Macintyre et al., Nat Gen, 2018). Details of the computation are in **Methods, page 6**, section “Computing detectable chromosomal instability (dCIN)”.

A comparison to previous estimates shows that 80% is unlikely to be an overestimate. Pan-cancer and cancer-type specific aneuploidy estimates range between 60% and 95% depending on the method (karyograms: Cimini et al., Biochim Biophys Acta, 2008; genomics case studies: Lens and Medema, Nat Rev Cancer, 2019; computational method: Taylor et al., Cancer Cell, 2018; FISH: Cisyk et al., Neoplasia, 2015; reviewed in Ben-David and Amon, Nat Rev Gen, 2020).

For example, one study based on the Mitelman database found that up to 90% of solid cancers experience aneuploidy in the form of whole-chromosome CNAs (Weaver and Cleveland, Curr Opin Cell Biol, 2006). This result has been highly cited (directly and indirectly) and used as a motivation for studying CIN in human cancers (e.g. Taylor et al., Cancer Cell, 2018; Cimini et al., Biochim Biophys Acta, 2008; Holland and Cleveland, Nat Rev Mol Cell Biol, 2009; Lens and Medema, Nat Rev Cancer, 2019; Ben-David and Amon, Nat Rev Cancer, 2020; Bakhoum and Cantley, Cell, 2018; Duijf et al., Int J Cancer, 2013). Two other studies found between 68 and 88% of samples with aneuploidy depending on whether they included acrocentric chromosomes (Taylor et al., Cancer Cell, 2018; Duijf et al., Int J Cancer, 2013).

We compared our dCIN threshold of 20 CNAs to two competing approaches: First, on a cancer type level to the original study of Weaver and Cleveland and found a striking overlap (**Extended Data Fig. 1c**), apart from two cancer types (AML and UVM) which we discuss in our Methods (**page 7**). Second, on a sample level by comparing the aneuploidy score (Taylor et al., Cancer Cell, 2018), SCNA score and its derivatives whole chromosome scores, arm scores and focal scores (Davoli et al., Science, 2017) and found again a striking correlation (**Extended Data Fig. 1d-e**).

Our dCIN threshold might overlook quiet genomes with mostly whole-chromosome CNAs (e.g. AML, see **Methods page 7**) but we have strong evidence that it is a pragmatic and robust threshold for identifying samples which most likely experience an identifiable type of CIN.

ii) They assemble 17 CIN genomic signatures, yet **only the top two are characterized in more detail** and are assigned to mitotic errors and homologous recombination deficiency.

We have characterised all 17 signatures with descriptions appearing in the main text of the revised manuscript. In this resubmission, we further strengthened the framework for identifying aetiologies and substantially improved its presentation to more explicitly show how much

evidence went into each aetiology. Results are summarised in **Figure 2** with additional details in **Extended Data Figs. 7-9; Methods, pages 22-34; and Supplementary Materials 4.**

iii) How a specific signature is linked to a given **aetiology is unclear** and the proposed aetiologies seem largely **speculative**.

We have improved the presentation of our approach.

Manuscript page 4 (beginning of section “Putative causes underlying each signature”) describes the data integration approach in detail. **Extended Data Fig. 7** visually summarises the process and shows the evidence supporting CX3 in detail. Similar summaries for the other signatures are provided in **Supplementary Materials 4.**

In short, we used driver gene associations as markers for putative pathways involved in the aetiologies and assumed other pathway deregulation for samples where no driver gene was mutated. In many cases, the signature pattern was already suggestive of a mechanism (e.g. whole chromosome missegregation). Once a cause was proposed, we sought additional supporting data, including data from two additional patient cohorts and their clinical metadata, to further refine the aetiologies.

To navigate the evidence supporting each signature, we are developing a web resource, with a prototype available at <https://markowetz.cruk.cam.ac.uk/cincompendium> (user: cnsigs2; password: souptomatobasil)

iv) Overall, it is not clear what the impact of such a signature database might have on **cancer prevention or predicting responses to treatment**, or indeed whether there is any use for it at all.

In this resubmission, we have substantially extended the scope of the manuscript and now show how the signatures and their aetiologies lead to multiple clinically important insights.

We show that copy number signatures predict response to 44 drugs targeting 40 genes (**Fig. 3b**). We also identified 49 novel targets that currently do not have clinically available inhibitors (**Fig. 3c**).

We then focussed on the prediction of platinum-sensitivity in ovarian cancer patients (**Fig. 4**). We devised a clinical classifier that is able to distinguish significant overall survival separation within BRCA1 germline mutant carriers and across several cohorts of patients routinely treated with platinum-based therapies (**Fig. 4d, Extended Data Fig. 10, Supplementary Figs. 17-20**). Furthermore, this classifier had comparable performance and in some cases superior performance to state-of-the-art predictors across ovarian, oesophageal and breast cancer cohorts (**Extended Data Fig. 10c-d**).

v) In addition, the manuscript is **unlikely to be understood by a non-specialist readership**, thus it is best suited for a more specialised journal.

We tested the manuscript on a wide range of specialist and non-specialist readers.

Specific points:

1) Could the authors **explain the essential difference** between copy number signatures, as termed in their 2018 paper (Macintyre et al. Nature Genetics, PMID: 30104763) and CIN signatures computed here? It seems that copy number signatures are just termed CIN for this study, as broadly, similar features were taken into account to assess the signatures.

In Macintyre et al. 2018: “to identify copy number signatures, we computed the genome-wide distributions of six fundamental copy number features for each sample: the breakpoint count per 10 Mb, the copy number of the segments, the difference in copy number between adjacent segments, the breakpoint count per chromosome arm, the lengths of oscillating copy number segment chains and the size of segments”.

In the present manuscript (p.22, lines 620-629): “we derived genome-wide distributions of five fundamental copy number features. Using all copy number profiles, we computed distributions of segment size, copy number change to the left neighbouring event (called change point), breakpoints per 10 Mb, breakpoints per chromosome arm, and the number of neighbouring segments that oscillated between two copy number states. [...] In a major deviation from Macintyre et al. we did not use the absolute copy number of a segment as a feature as this can artificially split a signature by ploidy status despite representing the same mutational process”;

The pan-cancer copy number signatures we describe here are a substantial extension of the ovarian signatures we have described previously. The essential difference is that our pan-cancer results are no longer confounded by ploidy status as the ovarian results were.

We list all technical changes in **Methods, pages 10-12**, section “Implementation changes to Macintyre et al. (Nature Genetics, 2018)”.

2) P.4, lines 95-99: “Using the proportion of patients with detectable levels of CIN per tumour type, we assessed the relationship between CIN and cancer-specific five-year survival rates (Fig. 1b). Similar to previous studies⁴, we found that higher rates of CIN are significantly associated with a decreased survival (Beta regression model, $p=0.0035$) and an increased variability in survival (Beta regression model, $p=0.0025$)”.

All the high CIN rates cancers in the upper-right quarter of Figure 1b are associated with good five-year survival (at the extreme, “Testis” at almost 100% 5-year survival and near 100% detectable CIN).

I don’t know about the beta regressions (coefficients are not explicitly shown along with the p-values here, to get a better idea of the odds) but **it doesn’t seem to be any relationship between 5-year survival and detectable CIN rates** when looking at all the high CIN rate samples together;

This survival analysis has been removed from the resubmission.

3) In their previous description of copy number signatures, the authors found similar component weights for the identified signatures in ovarian cancer in the two independent cohorts (namely, PCAWG-OV and TCGA). **How robust are the CIN signatures found in this work**, as assessed by detection in other cohorts (e.g METABRIC has CNA data for breast cancer, MSKCC has CNA data for metastatic colorectal, prostate and skin cancers among others)?

Different cohorts often use different technologies and thus the robustness of our approach will largely depend on how well it translates from SNP 6.0 to other genomic technologies. To test this, we used 478 samples profiled both by TCGA (SNP 6.0 and whole exome sequencing, WES) and PCAWG (whole genome sequencing, WGS).

We compared signature definitions and activities across five platforms: SNP6 without matched normal, WGS downsampled to SNP 6.0 positions, WGS downsampled to shallow WGS, on-target WES, and off-target WES.

This analysis accurately replicated the pan-cancer signatures (cosine similarity threshold 0.95) and found significant similarities in signatures activities across all platforms (**Extended Data Fig. 6**). All details in **Methods, pages 17-21**.

4) It seems, from the comparison with the ovarian copy number signatures previously described, that the correlation with any of the 17 CIN signatures is low (Extended Figure 5, it looks like the maximal corrected cosine similarity value found is around 0.5). **Could this be due to differences in the approaches used to detect CNA events in the two studies?**

The correlation to the previous ovarian signatures are in fact very strong. See **Methods, page 33**, “Comparison to ovarian copy number signatures”, and **Supplementary Fig. 8a+d**.

In short, we tested the definitions and activities of each pair of signatures (CIN signatures + OV signature) with a permutation test and corrected the cosine similarity with the highest random cosine similarity from the permutation test (**Supplementary Fig. 8**). We chose such a strong correction of the cosine similarities in order to only detect the strongest links between the two sets of signatures. In supplementary figure 8c and f, the tiles with white or red colours are comparisons with highly significant p-values.

5) The coverage at genomic positions corresponding to SNPs on the SNP6 array should be, on average, about 40x; whereas the coverage of the shallow WGS (sWGS) conducted in Macintyre et al. 2018 was ~0.1x. **Could some events not have been detected by sWGS?**

The influence of different high-throughput methods on the signatures is indeed very important. We agree with the reviewer that genomic resolution can be very variable and we do see some resolution limitations in our signatures. CX12 is defined by the shortest segment sizes (around 30kb) and we cannot find a single association with it despite the signature having a well-defined pattern. We think this might be due to a lower limit in resolution and needs to be picked up by future studies.

However, every signature has been detected at least once using another sequencing platform or analysis setting according to the robustness test (**Extended Data Fig. 6**), showing that the different resolutions only have limited impact on signature definition. We can recommend the use of WGS and WES data with our signatures, as well as for data sets that have limitations, such as cell lines that have no matched normal samples.

6) Related to the previous point, **are the ovarian cancer copy number signatures from the previous study not cancer-specific signatures** then, as they showed some degree of similarity with many different pan-cancer CIN signatures in this study?;

In the original submission, the labels “pan-cancer” or “cancer-specific” did not relate to the activity of a signature but to their origin (i.e. which data set they were derived from, either all cancer types jointly, or each individually). We have now renamed the two categories of signatures to make the distinction between them clearer.

In the resubmission, the compendium contains (a) 10 pan-cancer signatures, which were derived from jointly analysing all samples across all cancer types, and (b) 7 cancer-type enriched signatures, which were not identified on the pan-cancer level but were found in an analysis of individual cancer types (**Manuscript page 3, Figure 1 and Extended Data Fig. 3**).

The details of signature derivation are described in **Methods, pages 13-14**.

7) The **definition of cancer-specific CIN signatures is unclear** (p.23, line 680): are these signatures specific to one cancer type? To a small subset? From Figure 2a, it seems that none of the 17 signatures have activity in only one, or very few, TCGA cohorts. In addition, why are the 17 signatures termed “pan-cancer” (e.g. p.4, line 109) when the authors stated in the methods that “therefore the 10 pan-cancer signatures were combined with the 7 cancer-specific signatures to yield a set of 17 signatures of CIN in human cancers” (p.24, lines 709-710)?;

Covered by point 6)

8) Activity estimation is referred to as a heuristic process, there is mention to an activity matrix, on it not necessarily being strictly dependent on the number of CNA events detected, etc. But **could the authors clearly state in the main text what they mean precisely by signature “activity”?**

The matrix of signature activities is one of the two matrices derived by NMF, together with the signature definitions.

In the context of other mutational signatures, for example the original COSMIC signatures (Nik-Zainal et al., Cell, 2012), these values are often called “exposures”. We did not use the word “exposure” here, because it is often confused with epidemiological exposures, whereas we only find cell intrinsic causes.

9) In p.5, lines 134-135, “we numbered the signatures by their prevalence across the TCGA cohort. Signatures 1 and 2 show pervasive activity patterns”, **is it linked to prevalence?** In addition, could it be shown that there is **no linear relationship** between CNA numbers and signature activity in a given cancer type?

Because our signatures were constructed using copy number features that may not involve a one-to-one mapping between feature and copy number change (e.g. lengths of oscillating changes) the activity does not necessarily represent the number of CNAs generated by the CIN-type but rather the strength of that CIN-type in a tumour.

Despite this, we did find a linear relationship between signature activity and the number of CNAs attributed to the activity of each signature. However this relationship differs for each signature as they tend to be influenced by different features. Therefore, to estimate the number of CNAs for each signature, we fitted a linear model on the raw signature values with

respect to the number of CNAs in each sample. We set the intercept to zero, to only allow positive coefficients. Interaction terms between the signatures were avoided as they resulted in negative coefficients, indicating that the model overfitted. From the model we extracted the coefficients and multiplied with the raw signature values for each sample to get the estimates of CNAs produced by a signature (**Methods page 38; Supplementary Fig. 15**). This analysis facilitated an understanding between the activity of different signatures and their manifestation in the genome which we address in the discussion section of the manuscript.

10) Signature 17 seems intriguing: from the heatmap in panel 2b, it seems like there were only a few CNAs called in PCPG samples (>1%, short segment size), the rest of the heatmap column appears in white, so I would assume 0% CNAs detected in every other cohort. However, activity was detected in multiple cancer types, as shown in panel 2a. In relation to point 2), could this signature, or even signature 6 that does not seem to be associated with any biological covariates either (Table 1), be detected in **other independent, validation cohorts**?

Signature CS17 is now signature CX12. It remains one of the few signatures without evidence for an aetiology and needs to be elucidated in future research.

11) Signature 1 is said to be possibly related to a molecular clock-like mutational process and include events acquired sporadically through the normal ageing process (p.6, lines 181-192). In that sense, from only association studies as carried out here, it is difficult to understand its real value/confidence in relation to specific cancer-associated processes. I wonder if this same signature would be detectable in **normal tissue samples, as a control**?

We have toned down this interpretation of CX1.

We still find CX1 to be correlated significantly with SBS1, an age-related signature, suggesting that errors linked to CX1 might also be mediated via a natural ageing process such as age-related telomere attrition.

Studies on normal samples find almost no copy number changes and they are generally not discussed in the literature (Martincorena et al, 2015), making it impossible to use them as a control for detecting copy number signatures.

Referee #2

Drews et al. describe a new pan-cancer signature method for chromosomal instability (CIN). They used TCGA data SNP array profiles to expand CIN signatures they previously reported for ovarian cancer. The features they used to define signatures are segment size, copy number changes, breakpoints per 10 Mb, breakpoints per chromosome arm, and the number of switching between two copy number states. They found 17 CIN signatures. For each signature, the authors report:

- 1) The type of features that contribute to the signatures;
- 2) The similarity to ovarian CIN signatures;
- 3) Correlations with gene mutations;
- 4) Present evidence that these associated genes may contribute to DNA repair deficiency.

The authors focus on two signatures CS1 and CS2, that present in nearly all samples with CIN. They report that CS1 is correlated with shortening telomeres and indicate evidence of a negative correlation between telomere activity with TERT and TERC. CS2 was associated with HRD.

The paper is well written and presents results with potential importance, particularly the association of CS signatures with mutations in DNA repair genes that have not been demonstrated before. The authors focused on establishing CS1 and CS2 signature validity with established signatures such as telomere length or HRD signatures (SBS3). Moreover, the authors associate genes such as BRCA1/2 and TERT/TERC with these signatures.

However, there are multiple biomarkers for mutations in these genes. It might be worth focusing more on **novel gene-signature associations** that will demonstrate better the CIN signatures' **usefulness**.

This is an excellent point. **Extended Data Fig. 8** now gives an overview of gene-signature associations, including novel ones.

Additionally, the resubmission contains a broad analysis of associations between signatures and drug targets, which also demonstrates the usefulness of the signatures.

Major revisions:

1) The survival analysis was done on tumor type. Although the regression analysis found a significant association, it seems, as the author mentioned, that a small subset of tumors with a low proportion of CIN drive this association. Tumor types with high CIN, as the author said, have variable OS. I suggest **showing the association of survival with CIN within tumor type** to convince the reader that CIN is truly associated with OS.

This survival analysis is no longer part of the manuscript.

2) While the association of CS2 with BRCA1/2 is convincing, 70% of the samples identified as CS2 lack any event. None of the previous studies showed a high fraction of unexplained HRD cases (when methylation was performed). This may be concerning. **I would like the authors to compare CS2 to other copy number based HRD signatures such as HRD scores and HRDetect.**

In the resubmission we are now discussing three signatures of impaired HR, called CX2, CX3 and CX5 and have done a wide-ranging analysis to support their aetiologies and clinical usefulness:

Manuscript page 5: "Several lines of evidence support the link between these signatures and HR: increases in CX2, CX3 and CX5 activity across germline mutated *BRCA1* carriers (and *BRCA2* carriers for CX3); higher activity in cases with methylated *RAD51C*¹¹ (**Extended Data Fig. 9a**); correlation with tandem duplication scores⁴⁷ (**Supplementary Fig. 5**), rearrangement signatures 1, 3 and 5⁴⁸ (**Supplementary Fig. 6**), SBS signature 3 and ID signature 6⁴⁹ (**Supplementary Fig. 4**), centrosome amplification score⁵⁰ (**Supplementary Fig. 7**), ovarian CNA signatures 3 and 7³¹ (**Supplementary Fig. 8**); association with loss of heterozygosity⁵¹, chromothripsis⁵² and kataegis (**Supplementary Fig. 2**); increased utilisation of NHEJ,

theta-mediated end joining (TMEJ), and single strand annealing (SSA) backup repair pathways visible as micro-homologies at breakpoints⁵³ (**Supplementary Fig. 9**), as well as seven HRD metrics^{28,51,54–56} (**Extended Data Fig. 9d**)."

The seven HRD metrics are the Myriad myChoice score, *TP53* inactivation score, telomeric imbalances score, large-scale state transition score, LOH score, DNA repair proficiency score, and a protein expression score for 23 DNA-damage repair genes. Details in **Methods, page 32**, section "Seven HRD metrics".

3) For a subset of ovarian and breast TCGA samples, WGS is included in PCAWG. Hence, **HRDetect can be calculated**. HRD scores (PMID: 26957554) can be calculated from SNP arrays.

This is a very good suggestion. We now compare our classifier and two other clinical classifiers - including HRDetect - on five cancer cohorts from the TCGA, PCAWG and ICGC projects. The results highlight the strengths of the CX3/CX2 classifier we propose in the resubmission (**Extended Data Fig. 10, Supplementary Figs. 17-20**).

4) Recent WGS studies have produced a very detailed description of somatic copy number alterations at higher resolution than SNP arrays. **Can the authors compare the presence of CSs to the PCAWG copy number alterations profiles** that were recently published and briefly mentioned in the text?

Unfortunately one of the limitations of our approach is that the signatures are quantified at a whole-genome level and it is not yet possible to quantify signature activity at specific locations in the genome. Therefore, it was not possible for us to compare our signatures directly to these highly curated structural events in PCAWG samples. We have highlighted this limitation in the discussion section of the manuscript.

5) Since the biological interpretation of CIN signatures is of paramount importance, how will these CSs be identified based on **WGS**? The ability to detect these is tricky since the number of segments and their sizes will be different.

The reviewer is completely correct and we have benchmarked our approach across technology platforms (see above) and identified some limitations, which we have explicitly mentioned. For example, CX12 is defined by the shortest segment sizes (around 30kb) and we cannot find a single association with it despite the signature having a well-defined pattern. The other signatures are less affected and can be derived accurately by different technologies.

6) The methodology for the signature derivation is not fully clear. Please clarify the method, maybe by using an illustration or a flow chart.

We have added figures to give an overview of feature derivation and matrix decomposition (**Extended Data Fig. 2**) as well as merging the signatures into a compendium (**Extended Data Fig. 3**).

Referee #3

Dreus et al. present an attempt to quantify the pan-cancer landscape of copy number instability using a mutational signature approach. They apply their previously published (Macintyre et al.) computational method (originally applied to the TCGA ovarian cancers) to the rest of the tumor types in the TCGA dataset, and claim that the sources of chromosomal instability (CIN) are mitotic errors and impaired HR.

Although the idea that CNV signatures can be informative for understanding CIN is promising, the approach described is **not sufficiently novel or informative, and the findings lack solid supporting evidence**. Here are some major issues:

1.i) Too many of authors' **conclusions are based on correlation coefficients**. As the authors are undoubtedly aware, when there are many data points, even weak relationships can become statistically significant. But if the effect size is small, the relationship is often not meaningful.

Our process never depends on individual correlation coefficients, but draws support from multiple sources of evidence to guarantee that the relationships are meaningful.

We have improved the presentation of the process we used to derive signatures and added **Extended Data Fig. 7** for more clarity. In the manuscript we now more clearly list all sources of evidence for each aetiology.

For example for the IHR signatures (**Manuscript, page 5**): "Several lines of evidence support the link between these signatures and HR: increases in CX2, CX3 and CX5 activity across germline mutated *BRCA1* carriers (and *BRCA2* carriers for CX3); higher activity in cases with methylated *RAD51C*¹¹ (**Extended Data Fig. 9a**); correlation with tandem duplication scores⁴⁷ (**Supplementary Fig. 5**), rearrangement signatures 1, 3 and 5⁴⁸ (**Supplementary Fig. 6**), SBS signature 3 and ID signature 6⁴⁹ (**Supplementary Fig. 4**), centrosome amplification score⁵⁰ (**Supplementary Fig. 7**), ovarian CNA signatures 3 and 7³¹ (**Supplementary Fig. 8**); association with loss of heterozygosity⁵¹, chromothripsis⁵² and kataegis (**Supplementary Fig. 2**); increased utilisation of NHEJ, theta-mediated end joining (TMEJ), and single strand annealing (SSA) backup repair pathways visible as microhomologies at breakpoints⁵³ (**Supplementary Fig. 9**), as well as seven HRD metrics^{28,51,54–56} (**Extended Data Fig. 9d**).

Finally, our star-rating (**Fig. 2**) clearly shows how much evidence there was for each aetiology, thus providing a very transparent view of the level of supporting evidence.

ii) For instance, the paper claims that there is a correlation between 5-years survival and CIN. But it can be plainly seen from Fig 1b that this **correlation is driven by a small number of outliers**, and I am surprised that the authors are claiming based on this figure that CIN has association with survival. Cancers of different tumor types have vastly different outcomes not because of CIN but more likely because of several other factors, e.g., testis and liver cancers have similar values for the median number of CINs and proportions of patients with CIN, but the five-year survival of the patients with these tumor types are on the two ends of the spectrum.

This survival analysis is no longer part of the manuscript.

iii) Similarly, the conclusion that the majority of CINs arise due to errors during mitosis or impaired homologous recombination also is based on the co-occurrence of signatures with gene alterations or other signatures. But co-occurrence of minor CIN signatures with CS1 and C2 given that these two are present in 97.4% and 71.4% of samples is **not conclusive in determining their source**.

Co-occurrence analysis is no longer part of the manuscript.

iv) The telomere shortening is discussed in depth, but the **effect is prevalent mostly in the top 5% of samples** with CS1, making it difficult to claim that CS1 is directly connected with telomere shortening.

Association with telomere length is still one of the pieces of evidence supporting the aetiologies of mitotic signatures CX1, 6, and 14 (**Manuscript, page 4**), but the analysis has been improved and we no longer restrict it to the top 5%. Rather, we have unified our approach to robust linear models and slope estimation for all signatures. The analysis of telomere length is shown in **Supplementary Fig. 3** and discussed in detail in **Methods, pages 30-31**.

2. The study fails to **describe how the present paper is different from** and adds to other papers on the same topic, especially given that the SNP data are very old and have been analyzed previously, e.g., Zack et al (Nat Gen 2013), which analyzed a large subset (~5000 samples) of the data analyzed in the present paper.

There are major conceptual differences to previous studies, including Zack et al. and the other studies cited in the Introduction. Previous work described consequences of CIN, whereas the copy number signature approach allows us to understand the causes leading to different types of CIN.

We have rewritten the introduction to make the main advance clearer and added a paragraph to the Discussion (**Manuscript, page 10**): “Our approach complements previous landscape studies of the genetic consequences of CIN^{9–12}, which generally focused on recurrent somatic copy number events at individual loci that are likely to reflect selective pressures¹². In contrast, copy number signatures^{32,33} uncover mechanistic biases in the patterns of alterations across all chromosomes.”

3. For impaired HR signature, the length of copy number segments (4.9-30.8 Mb) seems to be **much larger** than what is reported by the SV signatures (1-10 kb) (Nik Zainal et al. Nature 2016), ...

The previously reported SV signatures were derived using WGS which has a higher resolution compared to SNP arrays. 1-10kb copy number changes are typically not detected by SNP arrays and therefore do not form part of our signature definitions.

Conversely, the SV signature approach groups together everything above 10 Mb whereas we differentiate with fine detail above 10Mb (11 components) allowing us to detect long copy

number segments called large-scale transitions (>10 Mb) which are predictive of *BRCA1* mutation status and HRD (Popova et al., Cancer Research, 2012).

Additionally, it is important to note that our signatures are defined by four other features (not just segment length) and there is extensive evidence to support the association of CX2, CX3 and CX5 with impaired HR, partly cited above in point 1.i and also in **Extended Data Fig. 7c**.

... and it is unexpected for HR deficiency to be as prevalent as reported in this study. Authors claim that nine distinct signatures are related to homologous recombination deficiency, but they **don't provide any mechanistic explanation**. It is puzzling why such a large variety of patterns was not identified by studies carrying out SV analysis in whole genomes. We do not see why the authors did not extend their analysis to WGS and demonstrate that their findings are reproducible and consistent with the literature.

We have refined the analysis of IHR signatures (**Main text**, quoted above in point 1.i, **Ext Data Fig. 9**) and now have the evidence to support a mechanistic explanation (**Fig. 4a**):

IHR alone gives rise to CX2, a signature of small copy number changes indicative of tandem duplication. IHR plus replication stress causes CX5, involving larger CNAs. Finally, IHR plus replication stress, impaired damage sensing and impaired NER, gives rise to CX3 with the largest CNAs that are strongly associated with LOH. As a result, only CX3 associations are reminiscent of what has been termed "BRCAness" or "homologous recombination deficiency" (HRD), while CX5, and especially CX2, appear to represent more moderate impairment of HR. In this model, widespread IHR does not contradict more limited HRD.

We support this notion by developing a clinical classifier exploiting this exact mechanistic explanation: If samples have a higher signature CX3 activity than CX2, they are predicted to be sensitive to platinum treatment, which indicates therapeutically relevant levels of impaired HR. Across the whole ovarian cancer cohort, 27% of samples fall in this category, a much smaller number than overall IHR.

4. To determine whether CIN is present in a sample or not, authors use a somewhat **arbitrary 20 CNA threshold** which may inflate the fraction of samples with actual genomic instability. Again, the same samples have been analyzed before (not just SNP arrays but also WGS), but there is no comparison or **examination of robustness** of the conclusions on these parameters.

The choice of threshold has been covered above (see Reviewer 1, point i). 20 CNAs is a rigorous threshold and our dCIN estimates are in line with previous studies.

5. **Possible overfitting and false positive assignments of signatures** should be investigated further. Authors only tested the necessity of applying a minimum threshold on signature activity for two of their signatures. However, the two signatures they tested are those that associate with whole chromosome or chromosome arm copy number changes, which are easier to detect. Signatures that have high probabilities of generating smaller size CNAs and/or that have distributions that overlap with other signatures are likely to suffer more from false-positive assignments. In their supplementary table for sample specific assignments (SM_7_Estimated_CNAs_per_signatu), there are nonzero assignments with very low values. For example 88% of nonzero exposures for Signature 1, 82% of nonzero exposures for

Signature 7, 80% of nonzero exposures for Signature 6 are less than 5. A signature of 43 dimensions cannot be assigned accurately with such **low values of exposures**. The modelling with Gaussian/Poisson distributions can also aggravate this issue by increasing the overlap of probability distributions across different signatures. If the signature assignments are inflated due to overfitting, the co-occurrence based association of various signatures with HRD would lead to misleading conclusions.

This is indeed an important observation and we have addressed it in this resubmission by estimating signature-specific activity thresholds using Monte Carlo simulations summarised in Extended Data Figure 5.

We perturbed the copy number features of the samples and observed the resulting variance in signature activities, fitted a Gaussian distribution to the results coming from 0-activity samples and used the quantile 0.95, to shrink all values below this to 0. This procedure results in activity thresholds adapted to each signature's noise levels, directly addressing the reviewer's point.

Additionally, co-occurrence no longer plays a role in the definition of aetiologies.

-End-

Reviewer Reports on the First Revision:

Referees' comments:

Referee #1 (Remarks to the Author):

Our concerns have been satisfactorily addressed and the new version of the manuscript is substantially improved. However, the signatures identified remain correlative with disease aetiology, as establishing a direct link would require in vitro validation of specific gene mutations (using for example CRISPR engineered cell lines) as drivers for individual signatures.

Minor points:

1. P. 2: "We did not include a feature representing the copy number of a segment to avoid redundancy if signatures for the same aetiology appear across different ploidy backgrounds". While the reasoning for excluding the absolute copy number as a feature (p. 31) is acceptable, the authors should discuss whether changes in the ploidy status (except CX4: whole-genome duplication) may constitute additional CIN signatures. If so, the authors should discuss whether aetiologies for the changes in the ploidy status are identifiable.
2. The "Workflow for determining signature aetiology and confidence rating" figures (p. 17 and p. 50) can be placed as either Method Fig. 3 or Extended Data Fig. 7, but it should not be duplicated.
3. Fig. 2: The methodology for determining the "pan-cancer prevalence" should be discussed. As it is unclear whether the frequency was calculated independently of cancer type, by pooling data for all cancers together.
4. P. 4: "Finally, CX1 positively correlated with the "clock-like" SBS1 signature (Extended Data Fig. 8, Supplementary Fig. 4) suggesting these errors might also be mediated via a natural ageing process such as age-related telomere attrition". If data are available, it would be informative to test whether CX1 is significantly enriched in older patients.
5. P. 7: "Correlation of CX13 (also associated with replication stress) with response to CDK4 inhibition, reinforced the interplay between replication stress and immune system activation". It is incorrect to state that the susceptibility of CX13-classified cancer cells to CDK4 inhibition highlights the interplay between replication stress and immune response. CX13 may, however, "potentially represent a biomarker-led approach for improving CDK4/6 inhibitor mediated tumour sensitisation to immune checkpoint blockade" (p. 7).
6. For each of the signature associated to pathway deregulation (e.g. CX2, CX3, CX5 for HR), it would be informative to specify which fraction of the classified samples was associated with known gene mutations, for example BRCA1 or BRCA2 mutations.

Referee #2 (Remarks to the Author):

The manuscript has greatly improved since the first submission. It is thorough and impressive regarding data visualization. The methods section makes the work transparent and useful for the reproducibility of the study. The authors also showed the concordance of the SNP array-derived CIN signatures with multiple other sequencing platforms. Extended Figure 7 clearly shows how the etiology is derived for the signatures. The association with genes may reveal new biomarkers for

these functions and better explain their role. In addition, the authors performed extensive work in comparison with other published signatures. Lastly, the authors added a section on the correlation of cells with specific CIN signatures with drugs and synthetic lethality which if validated offers a way to target such tumors.

However, since this paper was completely rewritten, I have a few more comments:

1) Can the authors visualize the signatures using bar plots? I do not find heatmaps in Extended Data Figure 5 useful. For example, something like the following figure will be useful
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7048622/figure/F6/>

If space is limited, the authors can include it in the supplementary data.

2) TERT and TERC amplification were negatively associated with signatures associated with chromosomal arm changes of CX6 and CX1. Therefore, the authors suggested that shortening telomere length is associated with arm-level chromosomal changes.

But TERT and TERC amplification events are highly common in cervical cancers which are hallmarks in many chromosomal arm changes. Can the authors analyze these CIN signatures in the context of cervical tumors and HPV+ Head-and-Neck?

3) The association of genes with signatures is important. But two major events, CCNE1 amplification and CDK12 mutation (both common in ovarian cancer) which are associated with large copy Tandem Duplicator Phenotype (TDP) alterations were not included whereas FBXW7 mutation that is linked to TDP2 and TDP3 was. Were these genes not associated with CIN signatures?

4) The authors did not include BRCA1 methylation in Extended Data Fig. 9. In addition, the authors did not find an association between somatic mutations in BRCA1/2 with IHR signatures. Did somatic mutations in BRCA1/2 only include deleterious mutations or did they also include somatic missense mutations?

5) The results on the association of CX2,3,5 with BRCA1/2 and RAD51C are important and the relationship with response to platinum might provide a new means to stratify patients. HRDetect has excellent accuracy in classifying tumors as HRD but the Impairment HR (IHR) signatures may capture which cancer types have the tendency to become cancer. I would like the authors to add yet another figure showing Cx2, CX3, and CX5 in breast (TNBC, non-TNBC) and ovarian which are negative for BRCA1/2 and RAD51C as the results may indicate why some tumor types tend to be affected more by HRD.

6) Although this paper is already very dense, it is important that the authors shed light on the question regarding etiological differences in CIN between racial groups. Especially in Black patients shown to have higher HRD in lung cancer which might be interesting since CX5 is enriched with HR copy number alterations

Referee #3 (Remarks to the Author):

The authors have done extensive work since the first submission, including analysis of several additional cohorts of human cancers and cell line models. For example, they developed a classifier based on CIN signatures and a model for prediction of platinum response within the BRCA1-mutated ovarian group (Figure 4d). This is an interesting observation, although, due to the low number of cases with WGS, one cannot conclude whether HRDetect could achieve this (Extended Data Fig. 10). The authors also compared different sequencing platforms to assess the robustness of their CIN signature approach (Extended Data Fig. 6), although the comparisons are based on down-sampled WGS/WES data and it is unclear whether the same level of robustness still holds when their method is applied to the original WGS/WES data.

Despite a great deal of additional work, we do not find the major claims of the new version convincing. As detailed below, there are serious methodological issues, and we find the strong statements on etiology of signatures speculative. As a result, we do not find this manuscript suitable for publication at Nature.

Major issues

1. Flaws with the design of the signature discovery method: The choice of features results in one copy number alteration event contributing to multiple channels in the signature definition, unlike previous signature analysis based on point mutations or structural variations, where each variant is only counted once. This signature definition leads to two non-intuitive effects. First, signature exposures can no longer be interpreted as the number of alterations contributed by a mutational process. Yet these exposures are used to draw correlations (e.g., Supplementary Fig 11, 12) for inferring etiologies or predictive purposes. Second, the signature definition makes NMF an unsuitable algorithm for decomposition, since NMF has no awareness of the inherent correlations among different channels introduced by the definition. As a result, the NMF-derived signatures are often ill-defined. For example, CX13 has a high chance of causing large copy number changes but has no corresponding segment length or CNAs per 10Mb. This should not happen since the CNAs would have to belong to a segment length group. This methodological issue is not obvious and had escaped us in the initial review, but it is a major problem.

2. Strong claims based on correlations. In the previous version, we raised a concern about the use of correlations for deriving strong conclusions, e.g., correlation to survival. In response, the authors simply removed this part and now present other results, stating that the conclusions are based on multiple correlations. However, we believe the same issue still remains--having multiple correlations

does not necessarily justify the final conclusions stated in the manuscript. Correlations of signature exposure vs driver/gene dependency still play an important role in the manuscript (Figure 1, 2, 3, Supplementary Fig 11,12,14, 22). From these associations, strong conclusions are made about the origin of the signature (Fig 2, 4a, 'mitotic signatures' section). I would interpret such correlations carefully, as they may be confounded, e.g., by cancer type.

At times, even sequential and causal relationships without further supporting evidence are concluded. Below we give a few examples, but this is a general theme in the manuscript:

a. The fact that CX2 and CX5 are enriched in tumors with mutations on homologous recombination pathway does not mean that these signatures are specific or due to impaired homologous recombination. These signatures can be related to a process that is enriched with impairment of HR but is still present in other samples as well. These signatures also correlate strongly with chromothripsis. Although chromothripsis may occur more often in tumors with impaired HR, it does not mean that their presence reflects impaired HR, because of other mechanisms that can lead to these signatures.

b. The authors go even one step further and claim a specific relationship among signatures CX2, CX5 and CX3, stating that samples with these signatures have different levels of HR impairment and that the signatures develop in a stepwise manner. However, there is no evidence showing that the CX2 precedes CX5 and CX5 precedes CX3 in developmental stages of tumors.

c. The authors state: "Significantly higher CX8 activity in the context of amplification and overexpression of U2AF1, ERBB3 and MAPK1 implied R-loop formation, increased origin firing, premature replication fork termination, or asymmetric forks in combination with reliance on (error prone) NHEJ over HR as a putative mechanism (demonstrated via correlation with ID6 and ID8)". All these are deduced from U2AF1, ERBB3 and MAPK1 amplification and overexpression, but how does the enrichment in amplifications or overexpression of these genes compare to other genes? What other evidence exists for the R-loop formation?

These are only a few examples of a general trend of over-statements we observe in the paper.

3. Although IHR signatures are supposed to be the most clinically relevant signatures and are the most extensively discussed in the manuscript, the conclusions on IHR can be at times contradictory or misleading.

a. This distinction between correlation and causal relationships is important as authors quote 69%, 63%, 56% prevalence of these signatures pan-cancer. Does this mean that patients with these signatures will benefit from treatments targeting homologous recombination deficiency? If yes, the percentage is much higher than what is reported in literature. If not, a better threshold for selecting HRD tumors should be defined.

b. The authors suggest that CX5 is associated with PARPi sensitivity, but at the same time with platinum resistance in ovarian cancers.

c. Why do authors use CX2 and CX3 to define sensitive and resistant samples rather than CX5, which is associated with platinum resistance and was quoted to be related to IHR as well?

d. Are the authors suggesting their method will improve the detection of HRD, for example compared to HRDetect? It is not convincing from the existing results that it performs in a superior manner.

A pan-cancer compendium of chromosomal instability

Drews et al, 2020-07-14327C-Z

Point to point response

This document provides a detailed point-to-point response to all reviewer comments.

In the main text and the methods, minor text changes are not tracked, but all major text changes are highlighted in [blue](#).

Additionally, here is a list of all new figures we added to the manuscript:

1. Associations of signature activities and sample ploidy (new Supplementary Fig. 30)
2. Barplot of signature definitions (new Supplementary Fig. 31)
3. Associations of signature activities with patient ethnicity (new Supplementary Fig. 32)
4. Assessment of signature quantification and identification using whole-genome and whole-exome data (new Supplementary Fig. 33)
5. Segment size distributions across technology platforms and algorithms (new Supplementary Fig. 34)
6. Impaired HR signatures in different tumour types (new Supplementary Fig. 35)
7. Associations of signature activities and sample ploidy (new Method Fig. 6)

And new tables:

1. Testing for differences in signature activity in CCNE1 mutants and non-mutants (new Methods Table 3)
2. Results of a multivariate regression testing for IHR signatures and BRCA1/2 mutants (new Method Table 4)
3. List of mutations per gene and sample (new Supplementary Table 35, Supplementary Material 4)
4. Summary of mutations types per gene (new Supplementary Table 36, Supplementary Material 4)
5. Mutations per pathway associated in signature aetiologies (new Supplementary Table 38, Supplementary Material 4)
6. Associations between high-confidence genes and signatures (corrected) (new Supplementary Table 39, Supplementary Material 4)

Reviewer 1

Our concerns have been satisfactorily addressed and the new version of the manuscript is substantially improved.

Response: Thank you for this positive assessment.

However, the signatures identified remain correlative with disease aetiology, as establishing a direct link would require in vitro validation of specific gene mutations (using for example CRISPR engineered cell lines) as drivers for individual signatures.

Response: We completely agree. Experimental validation has become a very active research area for the SNV signatures introduced 10 years ago and it will be as important for copy number signatures. The strong evidence we have for most signatures across multiple independent data sets will guide these follow-up studies.

Minor points:

1. P. 2: *"We did not include a feature representing the copy number of a segment to avoid redundancy if signatures for the same aetiology appear across different ploidy backgrounds". While the reasoning for excluding the absolute copy number as a feature (p. 31) is acceptable, the authors should discuss whether changes in the ploidy status (except CX4: whole-genome duplication) may constitute additional CIN signatures. If so, the authors should discuss whether aetiologies for the changes in the ploidy status are identifiable.*

Response: Thank you for highlighting this important point. There are generally two ways a tumour cell can change its ploidy status (where ploidy is measured as the average copy number state observed across the genome):

1. whole-genome duplication which doubles the ploidy status; or
2. an accumulation of copy number gains or losses which increases or decreases the overall ploidy status.

Number 1 is represented by CX4. The reviewer is correct that Number 2 is not directly read out by our signatures, but it is very likely captured by our existing signatures.

To explore this question, we analysed the relationship between signature activity and tumour ploidy by means of a multivariate regression analysis correcting for CX4 and WGD status (see Figure R1 below). This analysis showed that CX6 and CX8 are associated with an overall increase in ploidy state, whereas CX2 causes a decrease. This is consistent with the proposed aetiologies and we have added this to the manuscript (Methods page 33, Supplementary Fig. 30).

Thus, in summary our signatures already capture the two main causes of ploidy change.

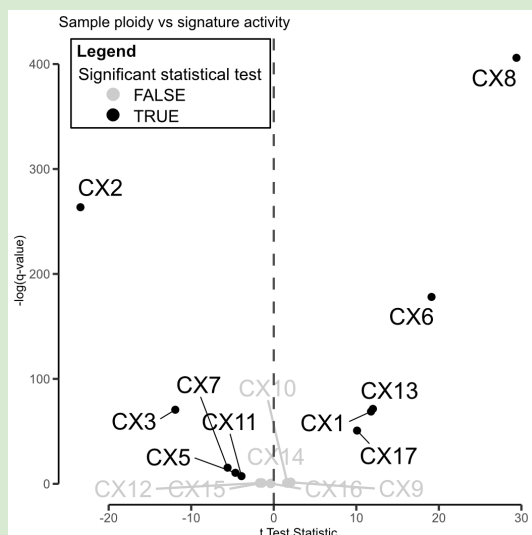


Fig. R1: Associations of signature activities and sample ploidy

Volcano plot of the t-Test statistic and $-\log(q\text{-value})$. CX4 is omitted as it was used for correction in a multivariate regression model.

This figure is now Supplementary Fig. 30.

2. The “Workflow for determining signature aetiology and confidence rating” figures (p. 17 and p. 50) can be placed as either Method Fig. 3 or Extended Data Fig. 7, but it should not be duplicated.

Response: We agree and have removed the figure from the methods section.

3. Fig. 2: The methodology for determining the “pan-cancer prevalence” should be discussed. As it is it seems unclear whether the frequency was calculated independently of cancer type, by pooling data for all cancers together.

Response: We added a sentence to Methods (section “Definition of signature-specific thresholds”, page 15): “To calculate pan-cancer prevalence we counted in how many samples each signature activity was above its specific threshold. This approach gives identical results for pooled data or for merging prevalence per cancer type.”

4. P. 4: “Finally, CX1 positively correlated with the “clock-like” SBS1 signature (Extended Data Fig. 8, Supplementary Fig. 4) suggesting these errors might also be mediated via a natural ageing process such as age-related telomere attrition”. If data are available, it would be informative to test whether CX1 is significantly enriched in older patients.

Our response: The reviewer points out a very interesting observation for which we have added more detail. We tested the correlation between CX1 and the age of patients by using classic and robust linear models correcting for cancer type. Pan-cancer, we found no significant relationship (linear model: p-value for CX1 and age 0.81; robust linear model: p-value 0.88). This is in line with previous publications which found very few CNAs genomewide in normal tissue, suggesting negative selection (Martincorena *et al*, Science, 2018; PMID: 30337457).

However, we performed the same analysis for each cancer type according to the method of Alexandrov *et al*. (Nature, 2020; PMID: 32025018) and found that there was a positive correlation between age and the number of estimated CNAs for CX1 for ovarian and lung squamous cell cancers (OV: F-statistic=5.9, q-value=0.048; LUSC: F-statistic=6.8, q-value=0.044). Given that ovarian cancer has one of the longest latencies (over 20 years, Gerstung *et al*. Nature 2020; PMID: 32025013), it might be that extreme latency periods lead to an accumulation of mitotic errors. We have added these results to the Methods page 32.

5. P. 7: “Correlation of CX13 (also associated with replication stress) with response to CDK4 inhibition, reinforced the interplay between replication stress and immune system activation”. It is incorrect to state that the susceptibility of CX13-classified cancer cells to CDK4 inhibition highlights the interplay between replication stress and immune response. CX13 may, however, “potentially represent a biomarker-led approach for improving CDK4/6 inhibitor mediated tumour sensitisation to immune checkpoint blockade” (p. 7).

Response: We agree with the reviewer that an association between CX13 and CDK4 inhibition supports interplay between replication stress and immune system activation is an overstatement. Therefore we have removed this statement from the manuscript.

6. For each of the signature associated to pathway deregulation (e.g. CX2, CX3, CX5 for HR), it would be informative to specify which fraction of the classified samples was associated with known gene mutations, for example BRCA1 or BRCA2 mutations.

Response: This is an excellent suggestion and we have added this information to the revised manuscript. The number of mutations underlying the correlations is listed in Supplementary Material 4, supplementary tables 36 and 37.

We also added a second table to present this information in detail (Supplementary Material 4, Supplementary Table 38). Below is a screenshot of the head of the table with the examples of CX3 and CX4 to show how we organised the information. For each signature we list the identified key genes for that pathway (for CX3 those are BRCA1, BRCA2 and RAD51C) as well as the union of these genes (row “Pathway”). For each of these we list how many mutants appear in samples with/without the signature (columns “Inactive” and “Active”) and stratify the activity further into deciles (columns “Activity deciles”). The constant rise of mutant proportions across the deciles shows that genes are associated with high activity.

CX3			Activity deciles									
Gene	Inactive	Active	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th
BRCA1	5	186	3	7	5	4	7	13	10	20	27	90
BRCA2	15	150	5	6	8	9	13	17	15	24	25	28
RAD51C	3	26	1	1	1	1	1	1	3	2	5	10
Pathway	23	362	9	14	14	14	21	31	28	46	57	128
WT	1384	4566	484	479	479	478	472	462	464	447	436	365
Mutant proportion	2%	7%	2%	3%	3%	3%	4%	6%	6%	9%	12%	26%

CX4			Activity deciles									
Gene	Inactive	Active	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th
PIK3R2	5	99	2	4	4	7	6	10	15	11	16	24
AKT1	12	105	3	5	9	9	11	10	10	12	17	19
PIK3CA	181	776	70	67	65	81	81	91	93	73	82	73
Pathway	194	922	75	75	76	95	92	103	110	89	103	104
WT	1549	3670	385	384	383	364	367	356	349	370	356	356
Mutant proportion	11%	20%	16%	16%	17%	21%	20%	22%	24%	19%	22%	23%

Fig R2: Number and proportion of mutated samples per signature activity decile

Referee #2

The manuscript has greatly improved since the first submission. It is thorough and impressive regarding data visualization. The methods section makes the work transparent and useful for the reproducibility of the study. The authors also showed the concordance of the SNP array-derived CIN signatures with multiple other sequencing platforms. Extended Figure 7 clearly shows how the etiology is derived for the signatures. The association with genes may reveal new biomarkers for these functions and better explain their role. In addition, the authors performed extensive work in comparison with other published signatures. Lastly, the authors added a section on the correlation of cells with specific CIN signatures with drugs and synthetic lethality which if validated offers a way to target such tumors.

Response: Thank you for this very positive assessment.

However, since this paper was completely rewritten, I have a few more comments:

1) Can the authors visualize the signatures using bar plots? I do not find heatmaps in Extended Data Figure 5 useful. For example, something like the following figure will be useful <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7048622/figure/F6/>

If space is limited, the authors can include it in the supplementary data.

Response: This is a very good suggestion and the figure example the reviewer provides is an excellent example of concise data visualisation. We have added the following supplementary figure and reference in the main text:

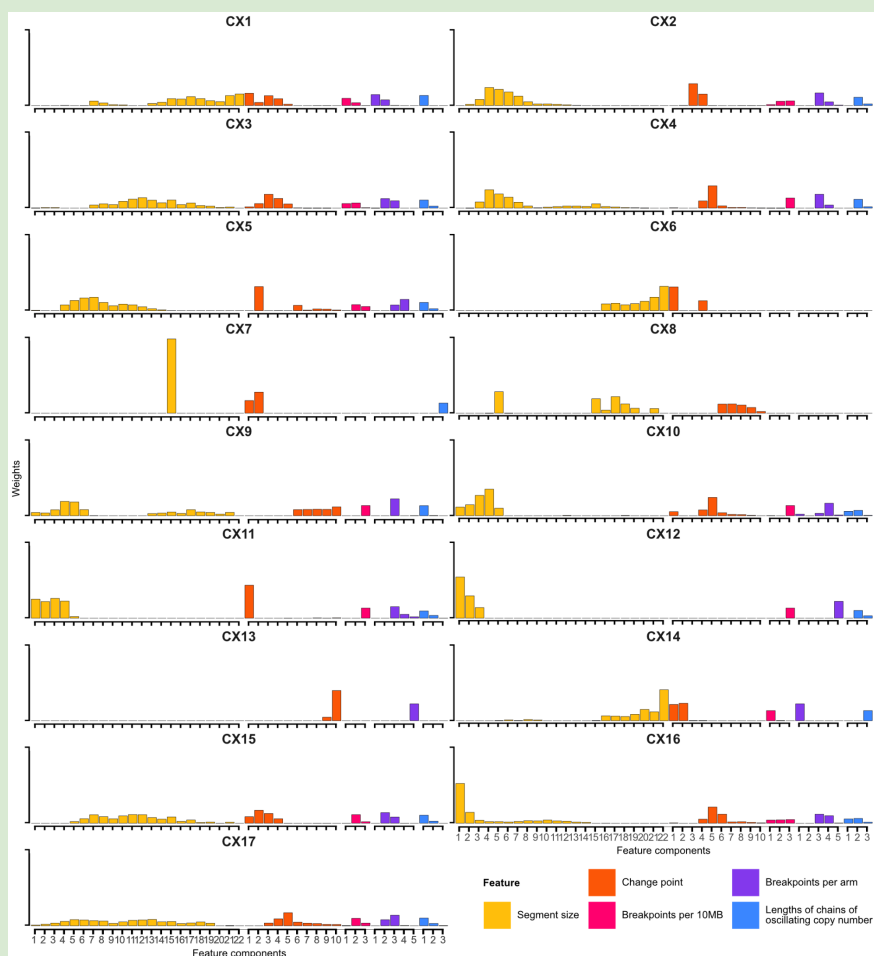


Fig. R3: Barplot of signature definitions.

Displayed on the x-axis are the 43 components of the five features. Components go from small to large means from left to right (see Extended Data Fig. 4 or Supplementary Material 2, Supplementary Table 2 for component means). The heatmap bar height shows the adjusted weights for each signature. Weights were normalised per signature and adjusted for each feature by the number of components. Now Supplementary Fig. 31.

2) *TERT* and *TERC* amplification were negatively associated with signatures associated with chromosomal arm changes of CX6 and CX1. Therefore, the authors suggested that shortening telomere length is associated with arm-level chromosomal changes. But *TERT* and *TERC* amplification events are highly common in cervical cancers which are hallmarks in many chromosomal arm changes. Can the authors analyze these CIN signatures in the context of cervical tumors and HPV+ Head-and-Neck?

Response: This is indeed an interesting observation and may suggest that this subset of tumours has a different relationship between *TERT*/*TERC* amplification and chromosomal instability.

Table R1 shows a break-down of the sample numbers available. CESC with *TERC*/*TERT* amplification (n=16) and HNSC HPV+ with *TERT*/*TERC* amplification (n=13) are only a small part of the data set, but large enough to perform exploratory analyses.

Group	n
Pan-cancer No amp	4486
Pan-cancer Amp	542
CESC No amp	49
CESC Amp	16
HNSC HPV+ No amp	47
HNSC HPV+ Amp	13
HNSC HPV- No amp	203
HNSC HPV- Amp	57

Table R1

We compared numbers of whole arm-level events across CESC and HNSC HPV+ tumours (n=29) with *TERT*/*TERC* amplifications, to those without amplifications, and to all tumours with and without amplifications (Figure R4a). As expected, the pan-cancer samples with *TERC*/*TERT* amplification showed fewer arm-level events on average. Overall, the CESC and HNSC tumours showed high levels of arm-level events, however, we did not see a significant drop in the number of arm-level events when *TERC*/*TERT* was amplified in HPV+ tumours. This suggests that these tumours either acquire amplification late in the evolution of the tumour, after a large number of arm level events have been acquired, or have a different relationship between amplification of *TERT*/*TERC* and chromosomal instability compared to the pan-cancer setting.

To explore this further, we determined if CESC and HNSC HPV+ tumours with *TERT*/*TERC* amplifications presented higher activity levels of a specific signature compared to pan-cancer samples with amplifications (Fig. R4b). We found that CESC and HNSC HPV+ with amplification showed a large increase of CX1 activity compared to other cancer types with *TERT*/*TERC* amplification.

Finally we explored any differences in CX1 activity between pan-cancer samples with *TERC*/*TERT* amplification and CESC and HNSC HPV+ tumours with and without *TERT*/*TERC* amplifications (Fig. R4c). Both amplified and non-amplified cases showed increased CX1 activity, suggesting that late acquisition of *TERT*/*TERC* amplification may be the likely hypothesis, however, more work needs to be done to confirm this.

While the results presented here are potentially interesting, the small sample size of the CESC and HNSC HPV+ amplification group (n=29) precludes us from drawing any robust conclusions and as such, we decided not to include this cancer-specific analysis in the manuscript.

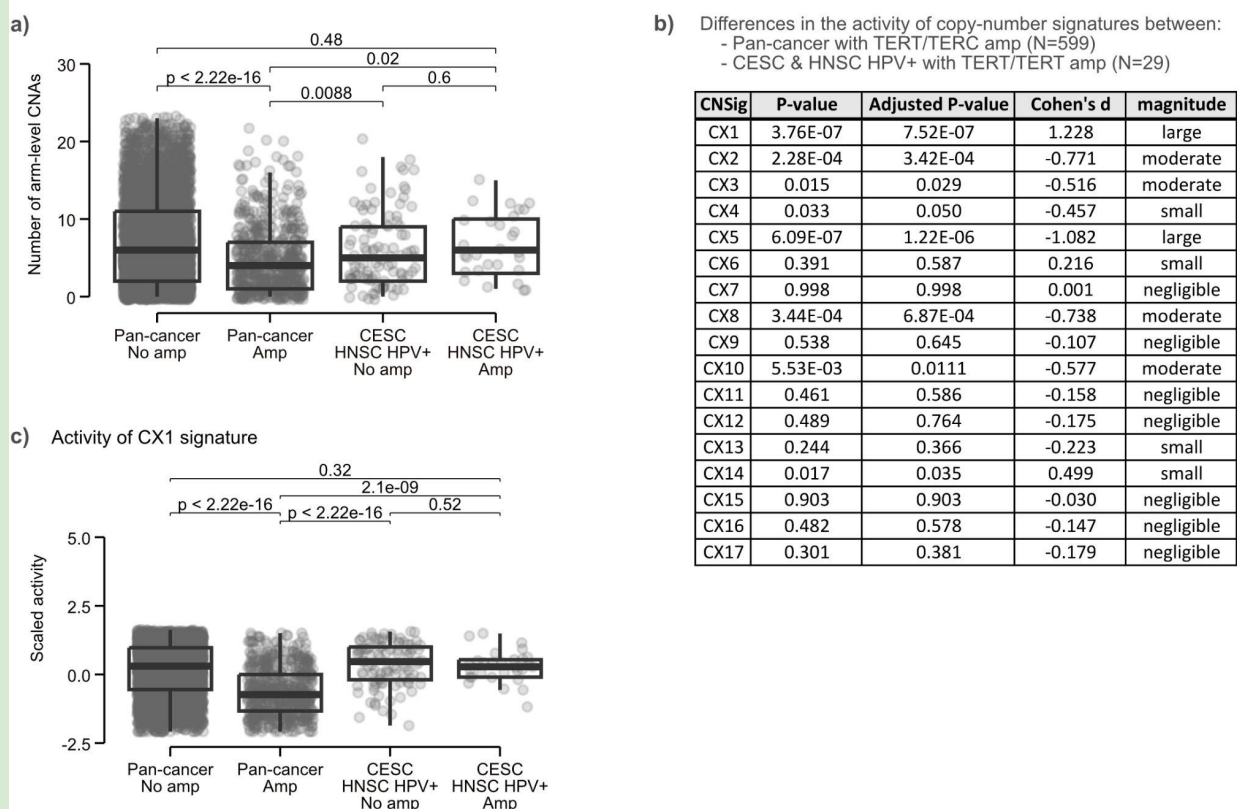


Figure R4

3) The association of genes with signatures is important. But two major events, *CCNE1* amplification and *CDK12* mutation (both common in ovarian cancer) which are associated with large copy Tandem Duplicator Phenotype (TDP) alterations were not included whereas *FBXW7* mutation that is linked to TDP2 and TDP3 was. Were these genes not associated with CIN signatures?

Response: We thank the reviewer for the opportunity to clarify this point. In fact, *CDK12* showed strong associations with CIN signatures CX2 and CX5 but our multiplicity correction (for 19,000 tests) scaled their q-values to 0.06, above our q-value threshold of 0.005. However, those trends support the notion that CX2 and CX5 are associated with class 2 and 3 tandem duplications (Extended Data Fig. 8, Supplementary Fig. 5).

CCNE1 was not identified as a driver gene in the TCGA landmark study (Bailey et al., Cell, 2018), which is why we did not include it in the initial manuscript. In additional analyses we found that, classified as pan-cancer oncogene, *CCNE1* is significantly associated with eight signatures (CX2, CX3, CX4, CX5, CX8, CX9, CX10). We have added these results to the revised manuscript (Table R2 below, now Methods page 29, Methods Table 3).

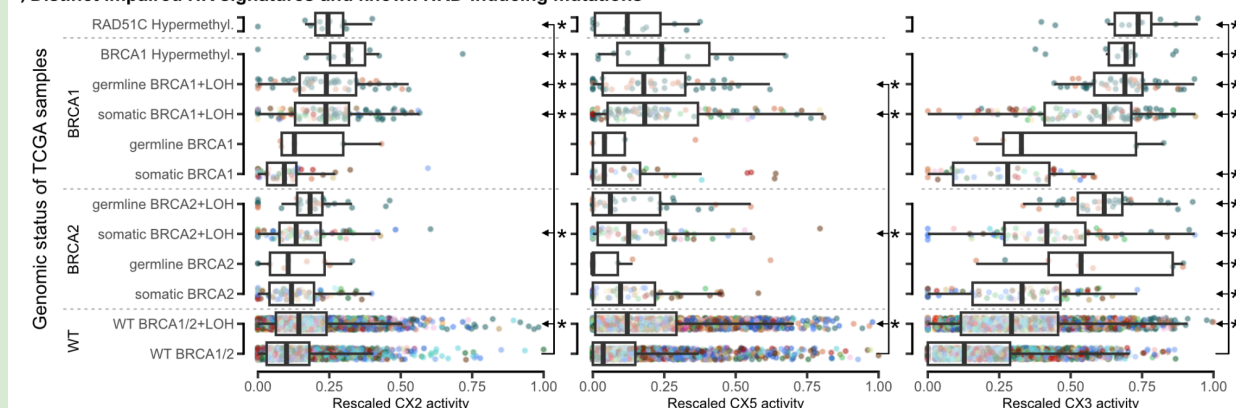
Table R2: Testing for differences in signature activity in *CCNE1* mutants and non-mutants
(Now Methods Table 3 in the manuscript)

Signature	# Mutants	# Non-mutants	Mean Mutants	Mean non-mut.	Diff.	Adj. p-val.
CX1	222	5003	-0.8948	0.02564	-0.9204	3.31E-34
CX2	222	5003	0.5389	-0.02311	0.562	8.38E-06
CX3	222	5003	0.7946	-0.03026	0.8249	2.29E-29
CX4	222	5003	0.775	0.01175	0.7632	3.13E-08
CX5	222	5003	0.8564	0.003766	0.8526	9.38E-16
CX6	222	5003	-0.1438	0.04042	-0.1842	0.03818669
CX7	222	5003	0.08714	0.01124	0.0759	0.79330761
CX8	222	5003	0.7787	0.01299	0.7657	4.06E-10
CX9	222	5003	0.7212	0.01558	0.7056	4.54E-10
CX10	222	5003	0.6581	0.005778	0.6523	8.48E-06
CX11	222	5003	0.04352	0.01781	0.02571	0.93504243
CX12	222	5003	-0.01747	-0.01413	-0.00334	0.99471727
CX13	222	5003	0.2693	0.02006	0.2492	0.00036043
CX14	222	5003	-0.2866	9.25E-05	-0.2867	3.23E-11
CX15	222	5003	-0.2824	0.002981	-0.2854	5.41E-12
CX16	222	5003	0.04237	-0.01184	0.05421	0.45347171
CX17	222	5003	0.3179	0.008652	0.3092	0.01508555

4) The authors did not include *BRCA1* methylation in Extended Data Fig. 9.

Response: We have added the *BRCA1* hypermethylation data to Extended Data Fig. 9a.

a) Distinct impaired HR signatures and known HRD-inducing mutations



Extended Data Fig. 9 (part of): Impaired HR signatures and their associations a: The boxplots summarise signature activities of different patient groups (rows) defined by their driver gene mutation status. Ovarian samples are coloured in dark green and breast in orange. The stars on the right hand side of each plot show where there was a significant difference between the WT BRCA1/2 population and mutant (WT = wild type; LOH = loss of heterozygosity). Panels **b-d** remained unchanged are not shown here.

In addition, the authors did not find an association between somatic mutations in BRCA1/2 with IHR signatures.

Response: We thank the reviewer for the opportunity to clarify this point. In fact, we did find a significant shift in signature activity between mutated and non-mutated *BRCA1/2* samples for CX2, CX3 and CX5 when analysing all TCGA samples together (Extended Data Fig. 9). However, when splitting the TCGA samples into two groups based on their LOH status (Maxwell *et al.*, Nature Communications, 2017), the samples with mutated *BRCA1/2* but without LOH at the respective loci had no significant shift in either CX2, CX3 or CX5 activity.

Did somatic mutations in BRCA1/2 only include deleterious mutations or did they also include somatic missense mutations?

Response: We filtered the set of SNV mutations for synonymous and ambiguous SNVs defined by having received the following variant classification: Silent, Intron, 3'Flank, 3'UTR, 5'Flank, 5'UTR. In addition, we only kept SNVs that passed the TCGA filter criteria. For tumour-suppressor genes specifically, such as *BRCA1/2*, we used deletions and all of the remaining types of point mutations.

We added Supplementary Table 36 (Supplementary Material 4) to indicate which type of mutations were identified for which gene. This screenshot shows how the information is organised:

Gene	AMP	DEL	Frame_Shift_Del	Frame_Shift_Ins	In_Frame_Del	In_Frame_Ins	Missense_Mutation	Nonsense_Mutation	Nonstop_Mutation	RNA	Splice_Site	Translation_Start_Site
AAAS	0	0	3	0	1	0	24	2	0	0	0	0
ABL1	0	0	4	1	1	0	64	5	0	0	1	0
ACD	0	0	1	1	0	0	30	6	0	0	0	0
ACTB	0	1	2	1	3	0	64	1	0	0	2	1
ACTL6A	0	0	0	0	1	0	26	2	0	0	2	0
ACTL6B	0	0	2	3	1	0	35	2	0	0	4	0
ACTR1A	0	1	1	0	0	0	20	0	0	0	1	0
ACTR5	0	0	3	0	0	0	30	1	0	0	1	0
ACTR8	0	1	2	0	0	0	27	3	0	0	1	0
ACVR1B	0	1	2	0	0	0	61	6	1	0	3	0
ACVR2A	0	0	46	2	2	1	44	9	0	0	4	1
ADPRHL2	0	0	1	0	0	0	12	0	0	0	0	0
AEBP2	0	0	1	0	0	0	19	1	0	0	0	0
AHCTF1	0	0	7	4	0	0	143	12	0	0	5	0
AJUBA	0	0	12	8	2	0	30	13	0	0	3	0
AKAP9	0	0	20	6	2	0	235	30	0	0	4	0
AKT1	79	0	0	0	0	0	77	0	0	0	0	0
AKT2	0	0	0	0	0	0	40	2	0	0	1	0

5) The results on the association of CX2,3,5 with BRCA1/2 and RAD51C are important and the relationship with response to platinum might provide a new means to stratify patients.

Response: We agree. The greater nuance of our IHR measures compared to previous HRD measures definitely has this potential, but follow-up work needs to validate these observations.

HRDetect has excellent accuracy in classifying tumors as HRD but the Impairment HR (IHR) signatures may capture which cancer types have the tendency to become cancer. I would like the authors to add yet another figure showing Cx2, CX3, and CX5 in breast (TNBC, non-TNBC) and ovarian which are negative for BRCA1/2 and RAD51C as the results may indicate why some tumor types tend to be affected more by HRD.

Response: We thank the reviewer for this important point. For each of the three signatures, we now display the activity in breast (TNBC and non-TNBC) and ovarian (BRCA1/2 mutant and WT) and a control group consisting of other WT BRCA1/2 and RAD51C methylation negative samples in the TCGA cohort (Fig. R5; now Supplementary Fig. 35, Methods page 39). Sample sizes for OV WT (n=17) and TNBC (n=7) were small compared to OV mutant (n=528), non-TNBC (n=229) and control (n=2916), which limits the power of any comparison analyses. However, we do find that TNBC samples appear to have activities of CX2 and CX5 matching those of mutant ovarian cancer suggesting these may have a higher propensity to be affected by HRD than non-TNBC, which have activities similar to non-mutant OV and pan-cancer tumours.

Given the sample size, however, our analyses are only exploratory and we do not emphasize them in the main manuscript.

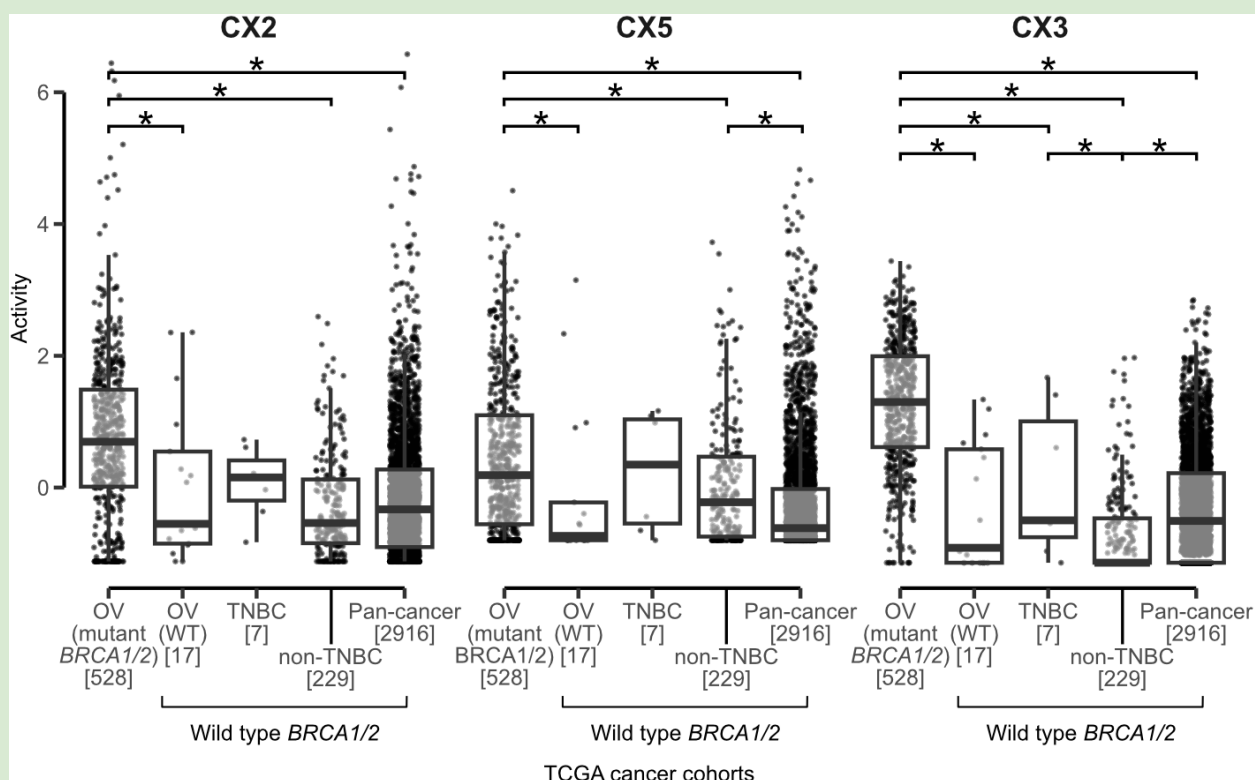


Fig R5: IHR signatures in different tumour types (Now Supplementary Fig. 35)

Boxplots showing the signature activity of impaired HR signatures (CX2, CX5, CX3, ordered by increased severity) for TCGA samples with wild type *BRCA1/2* genes (also without loss of heterozygosity at either loci). Samples from ovarian cancer (OV with and without mutant *BRCA1/2*), positive triple-negative breast cancers (TNBC) and negative triple-negative breast cancers (non-TNBC) were highlighted. Differences in mean activity was tested with Welch's t-test and corrected for false discovery rate (q-value < 0.05).

6) Although this paper is already very dense, it is important that the authors shed light on the question regarding etiological differences in CIN between racial groups. Especially in Black patients shown to have higher HRD in lung cancer which might be interesting since CX5 is enriched with HR copy number alterations

<https://www.nature.com/articles/s43018-019-0009-7?proof=t>

African ancestry can be derived from J Carrot-Zhang · 2020

Response: We agree that showing differences between racial groups is very important and therefore have added an analysis based on the clinical information already supplied by the TCGA. We found that there are substantial differences in signature activities with respect to the

patients' ethnicities (Fig. R6 below, Supplementary Fig. 32, Methods page 37). We also found that black patients have significantly higher activities in CX3 and CX5, both of which are implicated in impaired HR. A more comprehensive analysis will be the focus of a future manuscript.

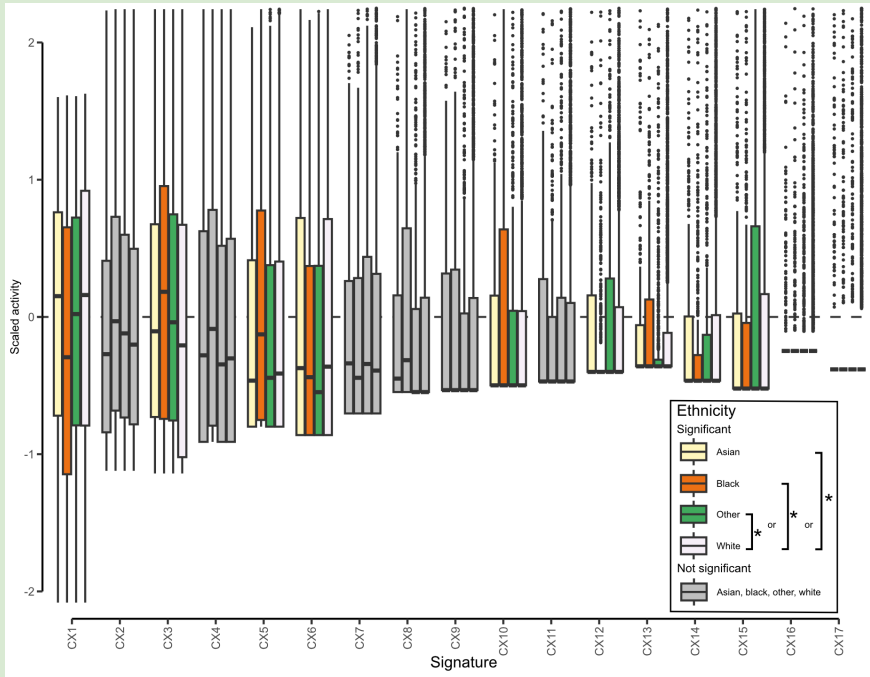


Fig. R6: Associations of signature activities with patient ethnicity

Boxplots comparing signature activity between groups of samples split by their ethnicity. Statistical significance was tested by Welch's t-test and coloured when there were significant differences between any of the three comparisons with the white patient population (q-value <0.001).

This is now Supplementary Fig. 32 of the paper.

The study referenced by the reviewer used the OncoScan platform, which is not as widely used as the technologies we compared in the manuscript and will need technical adaptations and a separate application to dbGap. Because of this we could not include it in the revised version but will do in our future study.

Referee #3

The authors have done extensive work since the first submission, including analysis of several additional cohorts of human cancers and cell line models. For example, they developed a classifier based on CIN signatures and a model for prediction of platinum response within the BRCA1-mutated ovarian group (Figure 4d). This is an interesting observation, although, due to the low number of cases with WGS, one cannot conclude whether HRDetect could achieve this (Extended Data Fig. 10).

Response: We agree with the reviewer that this is an interesting observation that requires further exploration with prospectively collected data. Indeed, the available sample numbers for HRDetect (n=3) were low on this subset of germline BRCA1 mutant ovarian patients, which may lead to misleading conclusions. Therefore, we have removed HRDetect from the comparison. However, as HRDetect uses BRCA1/2 mutation status as a positive label during training of the algorithm, we do not expect it to be able to separate this BRCA1 germline mutant group without modifications to the algorithm.

The authors also compared different sequencing platforms to assess the robustness of their CIN signature approach (Extended Data Fig. 6), although the comparisons are based on down-sampled WGS/WES data and it is unclear whether the same level of robustness still holds when their method is applied to the original WGS/WES data.

Response: This is an important point. We developed our signature approach using TCGA SNP6 data because this represents by far the largest pan-cancer collection of high-resolution copy number profiled tumours.

Nevertheless, the resolution of copy number calling from SNP6 data is lower than whole-genome sequence data, which is a limitation of our approach. In our original cross-technology platform comparison, we normalised all data to match the resolution of SNP6 arrays and showed that our signature methodology is robust across technology platforms at this resolution (Extended Data Figure 6).

If the resolution of copy number calling is increased (e.g. by using base-pair resolution whole-genome sequencing derived copy number), the underlying feature distributions change (Fig. R7, now Supplementary Figure 35). This change can have two interpretations:

1. it may be possible to identify new signatures that were not detected at SNP6 resolution;
2. the SNP6 resolution signatures may not be fit for purpose applied to high resolution data.

For 1), as more pancancer WGS data becomes available, it may be possible to redefine our signatures at higher resolution (if needed) but this is out of scope of the present study.

For 2) we tested this by applying our SNP6 signatures to higher resolution WGS and to WES data processed by multiple different copy number callers (Fig. R8, now Supplementary Fig. 34). As expected, we see a drop in similarity scores for the activities and signature definitions. In some cases (such as ABSOLUTE on WGS), these reductions are minor and it could be possible to use our signatures “out-of-the-box” on these data.

To make our approach applicable to higher resolution data, we have provided a guide in the methods on how to match the resolution of these data to SNP6 in order to obtain robust copy

number signature estimates (Method, section “Recommendations for WGS and WES data”, page 21).

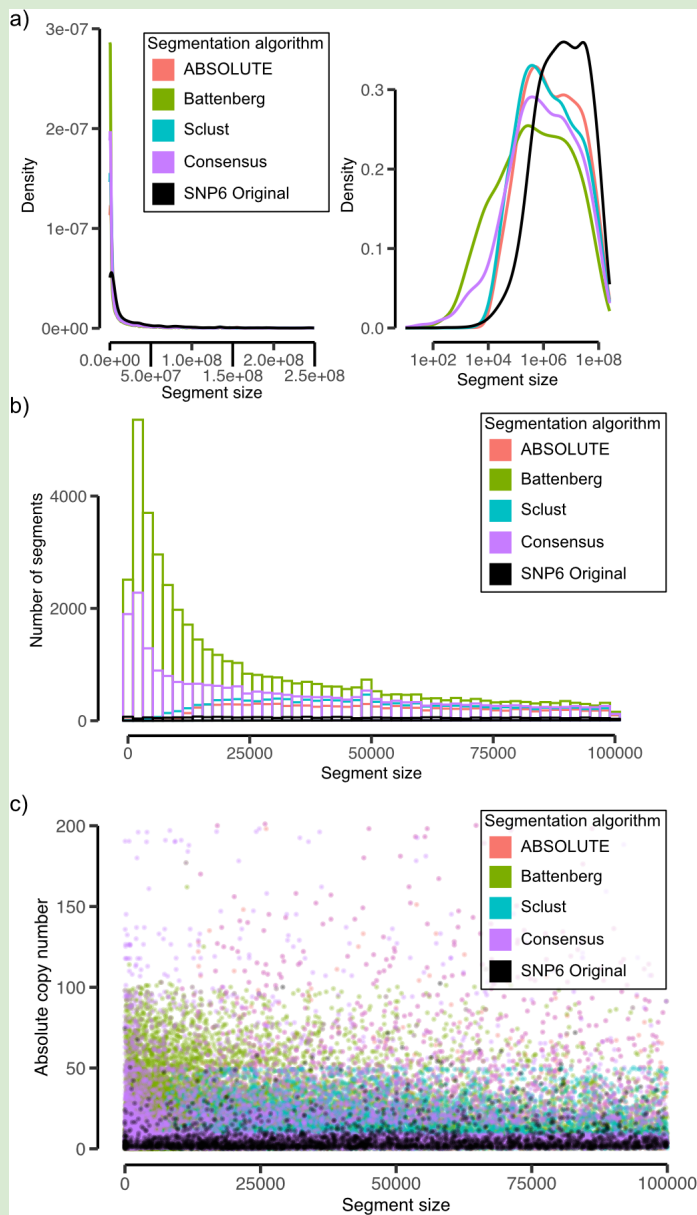


Fig. R7: Segment size distributions (<100kb) across technology platforms and algorithms

a: Density plots of the number of copy number segments for 451 samples that were present for all segmentation algorithms.

b: Histogram of segment sizes for segments smaller than 100kb.

c: Scatterplot of segments by their segment size (smaller than 100kb) and their absolute copy number for the 451 TCGA/PCAWG samples.

This is now Supplementary Fig. 35.

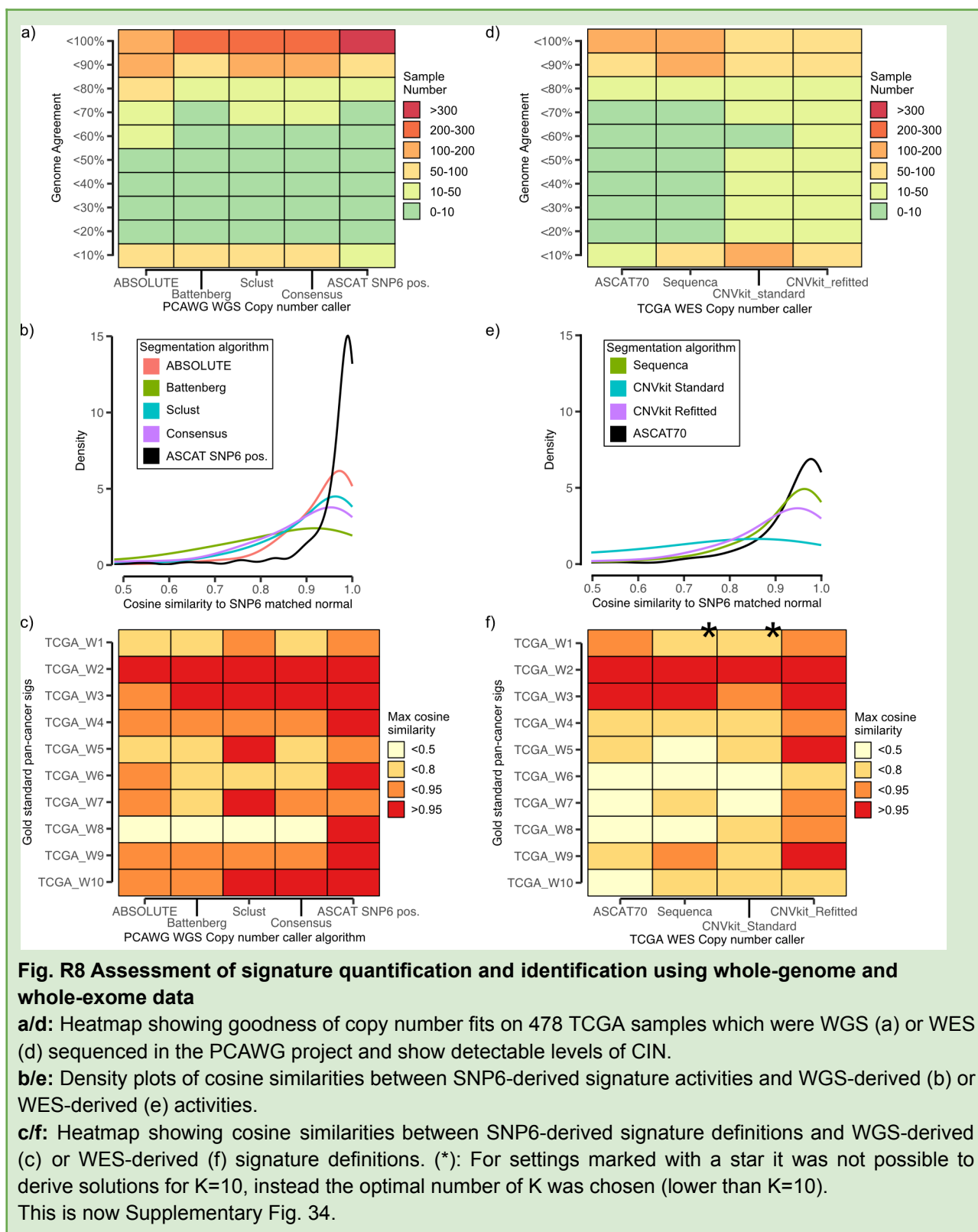


Fig. R8 Assessment of signature quantification and identification using whole-genome and whole-exome data

a/d: Heatmap showing goodness of copy number fits on 478 TCGA samples which were WGS (a) or WES (d) sequenced in the PCAWG project and show detectable levels of CIN.

b/e: Density plots of cosine similarities between SNP6-derived signature activities and WGS-derived (b) or WES-derived (e) activities.

c/f: Heatmap showing cosine similarities between SNP6-derived signature definitions and WGS-derived (c) or WES-derived (f) signature definitions. (*): For settings marked with a star it was not possible to derive solutions for K=10, instead the optimal number of K was chosen (lower than K=10).

This is now Supplementary Fig. 34.

Despite a great deal of additional work, we do not find the major claims of the new version convincing. As detailed below, there are serious methodological issues, and we find the strong statements on etiology of signatures speculative. As a result, we do not find this manuscript suitable for publication at Nature.

We thank the reviewer for their constructive comments. Below we address them point by point.

There are two general themes: First, the reviewer views our copy number approach through the lens of previous SNV/SV approaches. As a result, some of our design choices look unsuitable to them. In response, we clarify the unique challenges of copy number analysis which support our design decisions. Second, we clarify how different types of evidence (based on genomic patterns, gene associations and additional data) come together to support putative aetiologies.

Major issues

1. Flaws with the design of the signature discovery method: The choice of features results in one copy number alteration event contributing to multiple channels in the signature definition, unlike previous signature analysis based on point mutations or structural variations, where each variant is only counted once.

Response: We thank the reviewer for highlighting this very important point. The reviewer is correct that in our framework a CNA contributes to more than one feature, which is different to the way SNV and SV signatures are designed. This was a conscious design choice -as the nature of copy number data demands a fundamentally different approach- and constitutes a major conceptual advance.

Modelling the effects of different types of chromosomal instability presents a unique challenge compared to SNV signatures. This challenge stems from the fact that for CNA analyses we need to distinguish between a genome-changing event and the number of CNAs it produces. For some types of chromosomal instability, a single event causes a single copy number change (e.g. whole chromosome missegregation), while for other types a single event can cause many changes (e.g. chromothripsis). Therefore, to adequately model the activity of different types of CIN, the number of CNAs needs to be decoupled from the number of “events”. This is different from SNV signatures, where the number of SNVs is assumed to be a good measure of the activity of a mutational process.

Developers of SV signatures faced a similar problem and Li *et al.* (Nature, 2020, PCAWG SV study) addressed it by performing extensive clustering of SVs into mutually exclusive “event” types, ranging from events that encompassed a single deletion of a certain size, to events that involved large numbers of SVs as part of chromoplexy chains. This facilitated the counting of these events in a tumour and the application of NMF, similar to what had been previously done for SNVs. The approach of Li *et al.* thus allowed SV signatures to be interpreted in a similar manner to SNV signatures.

Copy number data, however, presents unique challenges, precluding the use of a similar approach. CNAs cannot be easily clustered into single “events” as is the case for SVs, as they lack the information required to relate one CNA to another. Further confounding this problem is the fact that copy number changes are quantified with respect to reference genome coordinates. For example, the loss of a “chimeric” chromosome (an already rearranged chromosome containing parts of many different chromosomes), is a single “event” in the tumour cell. However, mapped to a reference genome, this can manifest as many copy number changes across multiple chromosomes. Therefore, it is not possible to generate a vector of “event” counts and thus approaches similar to those used in the SNV and SV setting cannot be applied.

Modelling copy number signatures therefore demands a completely new and different approach. As originally published in Macintyre et al (Nature Genetics, 2018) and substantially enhanced here, we used the published literature to define a set of copy number related features representing known outcomes of different types of CIN including chromothripsis, tandem duplication, whole-genome duplication and focal amplification (Methods pages 8-10 and Methods Table 2). This formed the basis of our set of 5 features.

As the reviewer points out, embedding CNAs observed in a tumour into this feature space, means single CNAs can add weight to multiple features. This actually has the significant advantage of allowing “events” to be encoded that correspond to low or high numbers of CNAs, and thus overcomes the significant challenges outlined above of working with copy number data. However, this approach does have some potential caveats when interpreting the signature definitions, which we discuss below.

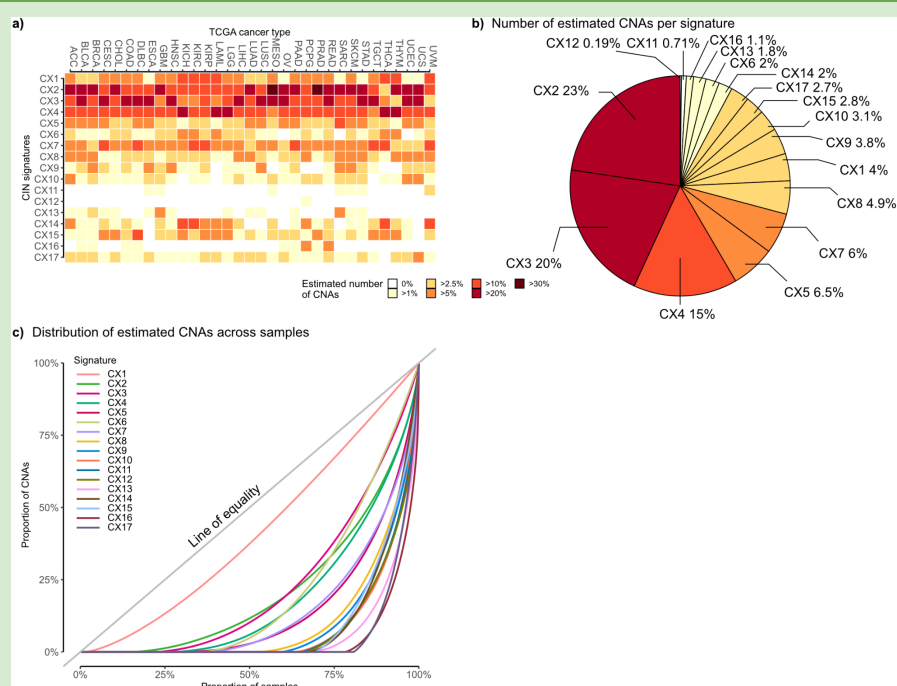
We have added the explanations provided here to the revised manuscript (Methods, section “Design choices and signature interpretation”, page 22).

This signature definition leads to two non-intuitive effects. First, signature exposures can no longer be interpreted as the number of alterations contributed by a mutational process. Yet these exposures are used to draw correlations (e.g., Supplementary Fig 11, 12) for inferring etiologies or predictive purposes.

Response: The reviewer is correct that CNA signature activity cannot be interpreted as the number of changes (in contrast to SNV signatures, where it can be). However, rather than being a flaw, this is exactly the right way to model CNA data and one of our major methodological advances (as discussed above). As a result, our signatures represent a readout of the activity of a CIN process, irrespective of how many copy number changes that process generates in a single “event”. This makes them very well suited to infer aetiologies and to be used as predictive biomarkers.

We do, however, provide an estimate of how many copy number changes are represented by each signature in a tumour. We analysed the relationship between individual signatures and the number of CNAs and found a strong correlation, which allowed us to estimate how many CNAs each signature produced in each sample (Supplementary Figure 15 and Methods page 43; for clarity reproduced below). As expected, each signature had its own linear relationship with the number of CNAs it generated.

Methods, page 43: We fitted a linear model on the raw signature values to the number of CNAs in each sample. We set the intercept to zero, to only allow positive coefficients. Interaction terms between the signatures were avoided as they resulted in negative coefficients, indicating the presence of overfitting. From the model we extracted the coefficients and multiplied with the raw signature values for each sample to get the estimates of CNAs produced by a signature (**Supplementary Fig. 15**, see below).



Supplementary Fig. 15:
Estimated number of CNAs produced by each signature

a: Heatmap showing estimated number of CNAs produced by the 17 signatures split by the 33 TCGA cancer types.

b: Pie chart of the estimated number of CNAs produced by the 17 signatures.

c: Lorenz curve of inequality showing the distribution of estimated number of CNAs produced by each signature across the 6,335 TCGA samples.

Second, the signature definition makes NMF an unsuitable algorithm for decomposition, since NMF has no awareness of the inherent correlations among different channels introduced by the definition. As a result, the NMF-derived signatures are often ill-defined. For example, CX13 has a high chance of causing large copy number changes but has no corresponding segment length or CNAs per 10Mb. This should not happen since the CNAs would have to belong to a segment length group. This methodological issue is not obvious and had escaped us in the initial review, but it is a major problem.

Response: As the reviewer points out, our feature encoding introduces substructure in the input matrix that is different to that of a typical SNV count matrix. However, this structure does not violate any assumptions of NMF (the only assumption of basic NMF is that the matrix does not contain negative numbers), so NMF can still be applied. However, there are indeed some caveats to interpreting the resulting matrices. Our choice of NMF was again by design and here we provide our justification, which we have added to Methods (section “Design choices and signature interpretation”, page 22).

When faced with deciding on which approach to use for matrix decomposition, we identified two possibilities:

1. A bespoke approach, similar to the Hierarchical Dirichlet Process (HDP) model with latent variables that was used in Li *et al.* (Nature, 2020, PCAWG SV study), where we specifically take into account the substructure of our feature encoding; or
2. the widely used NMF approach, which is agnostic to the substructure.

The advantage of the HDP-like approach is that we can impose prior knowledge that certain features should behave in certain, coordinated ways. Using the example pointed out by the reviewer, we could ensure that signatures of focal amplifications have both high change-point weights and short segment size weights. However, as the feature encoding itself already

strongly imposes prior knowledge, this additional constraint may suppress any ability to perform data-driven discovery and could be overly biased by our prior knowledge.

The advantage of using NMF is that it is robust and widely used across many scientific domains, including face recognition, astronomical image analysis, and text mining. It has well defined behaviour and provides a data-driven approach to signature identification across our feature space. However, as the reviewer rightly points out, NMF may result in signature weights that do not reflect the expectations of our prior knowledge.

In this work, we aimed for a balance between using prior knowledge and data-driven discovery: we encoded prior knowledge in the selection of copy number features (which were designed to capture known patterns of CIN) but used a widely used, well-tested, and robust “off the shelf” method for deconvolution.

We interpreted the signature definitions with caution. A significant strength of our approach is that it is easy to identify potentially erroneous signature definitions as these have strong weights on a single component. In this case, the single defining weight is so strong that other related weights (which we might expect to see) get shrunk to 0 (this is part of the NMF regularisation to ensure a unique solution). Fortunately, CX13 is the only signature that exhibits such behaviour. Our low to medium amplification signatures (CX8 and CX9) have weights on high change-points but also have weights on the expected segment sizes (both short and long). However, in the case of CX13, the high changepoint component is such a strong defining feature that no weight is given to segment size components. While this may yield a slightly unexpected signature definition, this does not necessarily imply that this signature is not robust in representing the process generating high-level focal amplification. Indeed, *post hoc* analysis of segments with the defining change point value (at least 10 copies) yields a median segment size of 790kB, with an interquartile range between 280kb and 2.7Mb which is consistent with estimates for ecDNA sizes of 1-3 MB (Liao *et al*, BBA Reviews on Cancer 2020).

While we do acknowledge that this effect may be more subtly manifested across other signatures as well, the replication of all pan-cancer signatures during our cancer-type enriched signature analysis suggests that our signatures are not ill-defined, but rather are robust readouts of biological signals.

2. Strong claims based on correlations. In the previous version, we raised a concern about the use of correlations for deriving strong conclusions, e.g., correlation to survival. In response, the authors simply removed this part and now present other results, stating that the conclusions are based on multiple correlations. However, we believe the same issue still remains--having multiple correlations does not necessarily justify the final conclusions stated in the manuscript.

Correlations of signature exposure vs driver/gene dependency still play an important role in the manuscript (Figure 1, 2, 3, Supplementary Fig 11,12,14, 22). From these associations, strong conclusions are made about the origin of the signature (Fig 2, 4a, ‘mitotic signatures’ section). I would interpret such correlations carefully, as they may be confounded, e.g., by cancer type.

Response: We thank the reviewer for the opportunity to clarify this very important point. Correlation between signature activity and individual genes is only one of the pieces of evidence supporting aetiologies. These correlations are important, but for most signatures they are

backed up by several pieces of orthogonal evidence. We first examine the pattern defining each signature to formulate a hypothesis, which we then test via gene interactions and finally in additional data sources (overview in Extended Data Figure 7; details of all evidence supporting each signature in Supplementary Materials 5, Supplementary Tables 40–41).

A strength of requiring these multiple lines of orthogonal evidence (patterns + genes + additional data) is that our approach is robust to most confounders such as cancer type (because they might affect the signature activities, but not the signature definitions themselves which are imprints of basic processes).

The mitotic signatures mentioned by the reviewer are excellent examples of how this process works: CX1, CX6 and CX14 all encode patterns related to whole-arm or whole-chromosome changes and significantly correlated with direct counts of whole-chromosome changes. This suggests putative causes resulting in chromosome missegregation during mitosis underlying these signatures. This hypothesis agrees with associated gene mutations and independent data like telomere length and *TERT* expression, among others. Not all signatures show this strength of evidence and we explicitly presented these differences in the star rating shown in Figure 2 and explained in Ext Data Fig. 7.

However, we agree that this level of nuance might have been lost in the main text. For example, we wrote sentences like “The signature is associated with biological process X via mutations in gene Y” and have now changed them to a more careful phrasing “The signature is associated with mutations in gene Y, potentially implying process X in its aetiology”.

We have carefully examined all such phrases in the manuscript, and to mitigate any risk of overstatement we retain only those which have multiple gene mutation associations to support the biological process link, and have thus removed more specific biological associations such as mutations in *U2AF1* associated with R-loop formation. Our main insights and high-level interpretations (presented in Fig. 2) are not affected by these changes.

Finally, we agree that cancer type may be a confounder in our association analysis between driver gene mutation and signature activity. An obvious example is *VHL* mutation, which is nearly exclusively limited to renal cancer. However, having a single tumour type enriched for a specific gene mutation and type of chromosomal instability does not mean that the association is not putatively causal (there is good support in the literature for *VHL*’s role in regulating mitosis, e.g. Jonasch et al, Nat Rev Nephrol. 2021; PMID: 33144689). Therefore, when performing this type of analysis, we face a conundrum: an analysis without correction potentially reports confounded results, and an analysis with correction potentially excludes strong and real biology (e.g. *BRCA2* is mutated in only a small number of tumour types and therefore may not pass correction). To deal with this in the revised manuscript, we opted to combine both approaches. We used an uncorrected approach to propose putative aetiologies, then used a corrected approach to see if these aetiologies continued to have support. The original list of genes associated with signatures can be found in Supplementary Table 37 and the corrected genelist in Supplementary Table 39 (both Supplementary Material 4). Nearly all genes in the uncorrected analysis (n=24/32) appeared in the corrected list. The following gene/signature associations appeared in the original list, but did not appear in the corrected list:

- CX1: *IDH1*, *CDKN2C*, *SOX9*, *PBRM1*, *VHL*
- CX3: *NFE2L2*, *PIK3R2*, *PPP2R1A*
- CX10: *PPP2R1A*

These associations have the potential to be confounded and therefore aetiologies based on these need to be treated with caution. Of these, the only ones that pose significant concern are *VHL* and *PBRM1*, as these are used to link CX1 with mitotic errors. However, this concern is mitigated by the fact that in the corrected list, tumours with inactivating mutations in *CDH1*, a key regulator of mitosis, are significantly higher in CX1 activity, supporting the putative aetiology of mitotic errors. These results demonstrate that our proposed aetiologies, which use driver gene mutation associations, are robust to potential confounding effects of tumour type. In the revised manuscript we have included a similar description to what is mentioned here (Methods page 30) and the additional corrected gene list as Supplementary Table 39.

At times, even sequential and causal relationships without further supporting evidence are concluded. Below we give a few examples, but this is a general theme in the manuscript:

a. The fact that CX2 and CX5 are enriched in tumors with mutations on homologous recombination pathway does not mean that these signatures are specific or due to impaired homologous recombination. These signatures can be related to a process that is enriched with impairment of HR but is still present in other samples as well. These signatures also correlate strongly with chromothripsis. Although chromothripsis may occur more often in tumors with impaired HR, it does not mean that their presence reflects impaired HR, because of other mechanisms that can lead to these signatures.

Response: We present three signatures of IHR: CX2, CX3 and CX5. Their putative aetiologies were all derived in the same way, accumulating evidence from the signature pattern, gene associations and additional data. In particular, CX2 and CX5, like CX3, are not just associated with mutations in HR genes but also have copy number patterns indicative of defects in HR, which argues that they are specific markers of this process.

The reviewer does not mention CX3 here (the strongest form of IHR) and therefore we assume they are convinced by our analysis for this signature. The question then is: Why are there three signatures and not just one? Are CX2 and CX5 really independent forms of IHR, or are their associations driven by co-occurrence with CX3?

To address this question, we performed an association analysis between *BRCA* mutant cases and the three signatures (CX2, CX3, CX5), using a multivariate regression model which explicitly corrects for the effects of association with the other two signatures, cancer types and *BRCA* mutant cases (see Table R3 below). The results of this analysis demonstrates that both CX2 and CX5 have significantly higher activities in *BRCA* mutant patients independent of CX3 activity levels. This provides evidence that CX2 and CX5 are likely directly associated with *BRCA* mutant status and therefore potentially a result of impaired HR. We have added these results to our revised manuscript (Methods page 38).

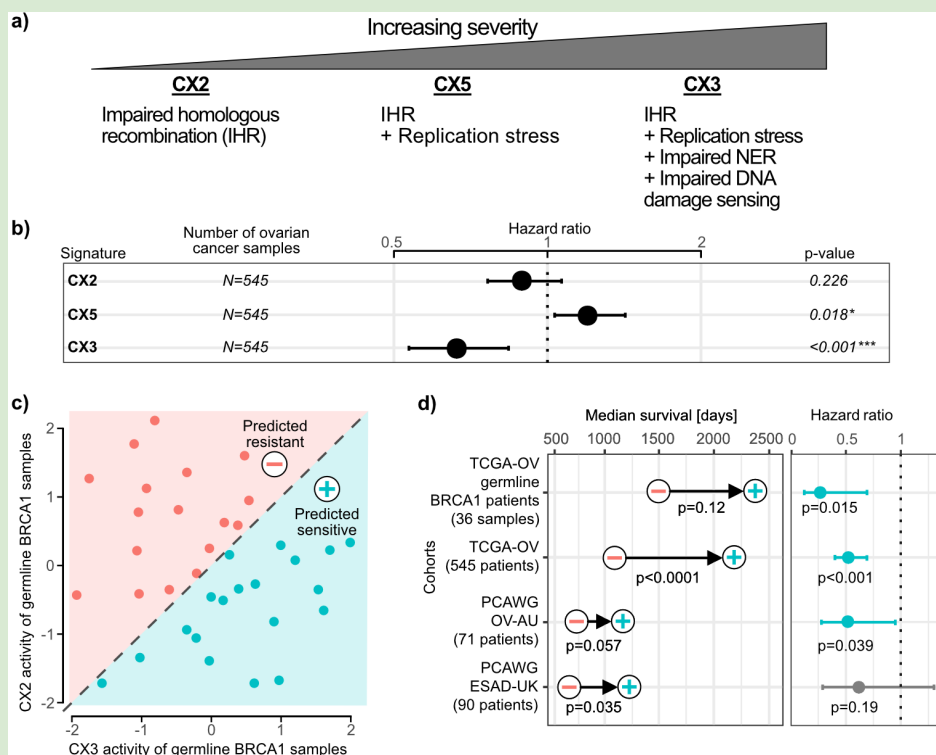
Table R3: Results of a multivariate regression testing the association between BRCA1/2 mutation status and signature activity for CX2, CX3 and CX5 while correcting in each analysis for two of the signatures and cancer type

	CX2 (corrected for CX3 and CX5)		CX5 (corrected for CX2 and CX3)		CX3 (corrected for CX2 and CX5)	
	t	p-value	t	p-value	t	p-value
gBRCA1+LOH	2.2	0.031 *	3.6	0.00002 ***	14.3	<2*10⁻¹⁶ ***
gBRCA2+LOH	-0.3	0.80	-0.2	0.81	10.8	3.0*10⁻⁰⁹ ***
sBRCA1+LOH	2.9	0.00392 **	6.5	8.8*10⁻¹¹ ***	14.2	<2*10⁻¹⁶ ***
sBRCA2+LOH	1.6	0.11	-0.2	0.85	8.7	<2*10⁻¹⁶ ***
BRCA1 hyperm.	1.3	0.19	3.5	0.0004***	6.2	6.6*10⁻¹⁰***
RAD51C hyperm.	1.1	0.28	-1.4	0.15	6.4	1.7*10⁻¹⁰ ***
WT BRCA1/2	= benchmark against which we compared					

b. The authors go even one step further and claim a specific relationship among signatures CX2, CX5 and CX3, stating that samples with these signatures have different levels of HR impairment and that the signatures develop in a stepwise manner. However, there is no evidence showing that the CX2 precedes CX5 and CX5 precedes CX3 in developmental stages of tumors.

Response: We apologise for being unclear. We are not claiming that the signatures “develop in a stepwise manner”. Rather, we present robust evidence that the three IHR signatures represent different levels of impairment (CX3 > CX5 > CX2) and we do not speculate how they developed.

We acknowledge that the arrows in our proposed model in Figure 4a could be interpreted as representing a developmental relationship, rather than an increase in complexity. To prevent misinterpretation, we have removed these arrows and improved the accompanying main text.



Updated Figure 4

c. The authors state: “Significantly higher CX8 activity in the context of amplification and overexpression of *U2AF1*, *ERBB3* and *MAPK1* implied R-loop formation, increased origin firing, premature replication fork termination, or asymmetric forks in combination with reliance on (error prone) NHEJ over HR as a putative mechanism (demonstrated via correlation with ID6 and ID8”. All these are deduced from *U2AF1*, *ERBB3* and *MAPK1* amplification and overexpression, but how does the enrichment in amplifications or overexpression of these genes compare to other genes? What other evidence exists for the R-loop formation? These are only a few examples of a general trend of over-statements we observe in the paper.

Response: (also see first response to point 2 above). In this instance, the *U2AF1*, *ERBB3* and *MAPK1* mutation associations support a potential role for different types of replication stress underlying CX8. Our intention here in specifically enumerating the different types of possible replication stress (e.g. R-loop formation) was not to claim definitely that they are involved, but rather provide more detailed guidance for future followup experiments. We agree that we do not have additional evidence supporting e.g. R-loop formation. Therefore, we have removed all such detailed aetiology claims in the manuscript. Our main insights and high-level interpretations (presented in Figure 2) are not affected by these changes.

3. Although IHR signatures are supposed to be the most clinically relevant signatures and are the most extensively discussed in the manuscript, the conclusions on IHR can be at times contradictory or misleading.

a. This distinction between correlation and causal relationships is important as authors quote 69%, 63%, 56% prevalence of these signatures pan-cancer. Does this mean that patients with these signatures will benefit from treatments targeting homologous recombination deficiency? If yes, the

percentage is much higher than what is reported in literature. If not, a better threshold for selecting HRD tumors should be defined.

Response: Thank you for raising this important point. The term “homologous recombination deficiency (HRD)” is widely used to describe a clinical phenotype where patients benefit from platinum or PARPi treatments. We have no evidence that all patients with CX2, CX3 and CX5 activity could benefit from these treatments and do not claim this in the manuscript.

Rather, we developed a clinical classifier based on CX2 and CX3, which estimates that 27% of the TCGA OV cohort might be platinum sensitive. Extending this classifier pan-cancer, we estimate that at most 8% of all tumours might be sensitive. We have added this result to the revised manuscript.

Relating these numbers to HRD percentages in the literature is not straightforward, because different studies propose widely varying HRD rates, depending on the methods and cohorts used. A few examples: A computational classifier found 52% in ovarian cancer [Ref 1 below]; Olaparib response rates are reported as 33% in recurrent ovarian cancer [2]; platinum-response has been reported as 51% in germline *BRCA* cases [3]; and PARP response 41% in germline mutants and 24% in samples without mutations [4]. This range of values shows that discussion of the best definition of HRD and efficient ways to measure it, is not yet concluded.

Our results bring an important new dimension to this ongoing debate. Signatures CX2, CX3 and CX5 present a nuanced view of different levels of impairment of the HR machinery and other repair pathways (Figure 4a). They are not ‘HRD signatures’ and we were careful not to label them in this way. Rather, we called them ‘IHR signatures’ (IHR = impaired HR) to make the difference explicit.

We explicitly state this in the manuscript: “These CX3 associations are reminiscent of what has been termed BRCAness³⁷ or HRD⁶¹. However, CX5, and especially CX2, appear to represent a more moderate impairment of HR. Therefore we use the term impaired homologous recombination (IHR) for the aetiology underlying all three signatures rather than HRD.”

In this way, our approach decouples an analysis of impaired DNA repair from a prediction of therapy success, which is often conflated when the term ‘HRD tumor’ is used synonymously for ‘platinum/PARPi sensitive’. Our results show that the number of samples with evidence of impaired HR is much larger than the number of samples where we would expect HRD treatments to work.

References:

1. <https://www.nature.com/articles/s41467-020-19406-4#Fig4>
2. [https://www.thelancet.com/journals/lancet/article/PIIS0140-6736\(10\)60893-8/fulltext](https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(10)60893-8/fulltext)
3. <https://www.astrazeneca.com/media-centre/press-releases/2019/lynparza-achieved-a-72-objective-response-rate-in-patients-with-relapsed-germline-brca-mutated-advanced-ovarian-cancer-03062019.html#>
4. [https://www.thelancet.com/journals/lanonc/article/PIIS1470-2045\(11\)70214-5/fulltext](https://www.thelancet.com/journals/lanonc/article/PIIS1470-2045(11)70214-5/fulltext)

b. The authors suggest that CX5 is associated with PARPi sensitivity, but at the same time with platinum resistance in ovarian cancers.

Response: This is an intriguing observation, which indeed appears counterintuitive as the current accepted paradigm is that platinum sensitivity also implies PARPi sensitivity. CX5 appears to suggest that some platinum resistant tumours can be sensitive to PARPi too. This has some support in the literature, where germline *BRCA* mutant ovarian cancer resistant to platinum show sensitivity to PARPi (Agarwal et al, JCO Glob Onc 2021; PMID: 33852339). However, we decided we did not have enough data to explore this question further and did not emphasize it in the manuscript.

c. Why do authors use CX2 and CX3 to define sensitive and resistant samples rather than CX5, which is associated with platinum resistance and was quoted to be related to IHR as well?

Response: We did, in fact, consider all 3 signatures in our analysis and compared different models, but the addition of CX5 did not improve performance over the CX3/CX2 classifier (Method, section “Building a clinical classifier based on signature activities”, page 41).

We focussed on CX3/CX2 because it allowed a very simple interpretation: CX2 measures baseline genomic changes due to impaired HR which, in the absence of other mutated pathways, is not enough to develop sensitivity. For CX3 (the most severely impaired HR) to predict sensitivity, its activity must surpass this baseline level.

This is a very simple, easily interpretable model, which can be measured even on shallow WGS data, which increases its clinical usefulness compared to methods depending on expensive deep WGS.

We have emphasized this point in the manuscript:

“Given that the IHR signatures were able to dissect platinum response, we further hypothesised that they could be used in combination to provide better predictors of platinum sensitivity. Given that CX2 was not predictive, we used it as a baseline for measuring IHR-related genomic changes, and required CX3 activity to exceed it, resulting in a simple classification rule: “if CX3 activity is greater than CX2 activity, then predict sensitivity” (**Fig. 4c**). This interpretable classifier was able to distinguish significant overall survival separation within *BRCA1* germline mutant ovarian cancers, across the entire TCGA ovarian cohort, an independent validation cohort, and across an oesophageal cancer cohort (also routinely treated with platinum-based chemotherapy) (**Fig. 4d, Ext Data Fig. 10, Suppl Figs. 17-20**). Other classifiers using all 3 IHR signatures, including more complex machine learning methods, did not outperform this decision rule (**Supplementary Fig. 21**). Furthermore, this simple classifier had comparable performance to more complex state-of-the-art HRD predictors, which rely on additional data beyond copy number, across ovarian, oesophageal and breast cancer cohorts (**Ext Data Fig. 10c-d**).”

d. Are the authors suggesting their method will improve the detection of HRD, for example compared to HRDetect? It is not convincing from the existing results that it performs in a superior manner.

Response: As also detailed in our response to point 3a above, we do not make claims about the superior detection of HRD compared to other methods.

The comparisons of our CX3/CX2 classifier to HRDetect were not designed to demonstrate superior performance, but rather to demonstrate that our method, based on low resolution copy number data alone, had comparable performance to a state-of-the-art approach that uses high resolution WGS and other multi-omic data. See text cited above.

- END -

Reviewer Reports on the Second Revision:

Referees' comments:

Referee #1 (Remarks to the Author):

All our concerns have been adequately addressed. We now recommend publication of this manuscript.

Referee #2 (Remarks to the Author):

The authors addressed my previous comment in a satisfactory manner.

Comments:

1) The authors mentioned that 528 of the ovarian cancers have BRCA1/2 mutations and only 17 are wild type (pages 38 and 39 in the Supplementary Methods file). This seems to be unreasonable. I could not find in the methods section how they derive this high number. What classes of mutations are included in the BRCA1/2 cases? I checked in the Ovarian TCGA data set and there are about 40% of cases with BRCA1/2 somatic or germline mutations + BRCA1/RAD51C promoter methylation.

2)Supplementary Figure 2 is missing d-f labels

Referee #2 comments on Referee #3 assessment:

This reviewer agrees with Reviewer 3, that the authors should demonstrate that by simulation of the genome they can reconstruct the input signatures using their method. The authors already introduced noise into the CNA profiles and found stability of the solutions. It will be useful to show the average profiles of signatures that were retrieved after the noising process. This reviewer agrees with the authors that NMF assumptions are not violated, however, the results might not reflect a biological mutational process. I would suggest that they module a genome using the CIN signatures (including the inter-channels correlation) and then run NMF and see how the obtained signatures are similar to the input signatures.

Referee #3 (Remarks to the Author):

Review of Drews et al, 2nd revision.

I find it surprising and disappointing that the authors failed to consider our critiques more carefully. We pointed out a critical flaw in the central formulation of their signature decomposition approach. If they

do not find our critique credible, they should have performed methodological studies to investigate. Instead, they simply insist that they have made "a major conceptual advance" and claim, with no demonstration

whatsoever, that their method can resolve issues encountered in previous approaches. In addition, they seem to be unaware of previous literature in the field nor have sufficient understanding of the mathematical framework they are employing.

Furthermore, in each round of reviews, we point out only the most critical issues, giving an example where possible. The documents (manuscripts/supplement/response) are very long and we cannot make an exhaustive list of issues or examples, as doing so would take an inordinate amount of time. The authors should not think that they have addressed the main problem by addressing only the specific example (e.g., by deleting that part). In their first round—as an example—they claimed that their survival analysis was a major finding. When we pointed out problems, they simply deleted that part and came up with new claims. That has happened again in this round.

Here, we make three points. Given that this is a second appeal with no substantive methodological change, I am unwilling to make an exhaustive list of problems.

Point A:

In response to our comment:

1. Flaws with the design of the signature discovery method: The choice of features results in one copy number alteration event contributing to multiple channels in the signature definition, unlike previous signature analysis based on point mutations or structural variations, where each variant is only counted once.

Authors responded (in part):

...

Copy number data, however, presents unique challenges, precluding the use of a similar approach. CNAs cannot be easily clustered into single “events” as is the case for SVs, as they lack the information required to relate one CNA to another. Further confounding this problem is the fact that copy number changes are quantified with respect to reference genome coordinates. For example, the loss of a “chimeric” chromosome (an already rearranged chromosome containing parts of many different

chromosomes), is a single “event” in the tumour cell. However, mapped to a reference genome, this can manifest as many copy number changes across multiple chromosomes. Therefore, it is not possible to generate a vector of “event” counts and thus approaches similar to those used in the SNV and SV setting cannot be applied.

1. Authors mentioned the challenges of analyzing CNA data and imply that their method can resolve them. But they do not demonstrate that their method is able to deal with those challenges. For example, they mentioned that loss of a chimeric chromosome can manifest as many copy number changes across multiple chromosomes. Can their method resolve that? A chimeric chromosome can still be counted as multiple SVs just as CNAs do. This point does not represent a CNA-specific challenge and does not relate to the issue we raised.
2. There are other ways of approaching CNA signature analysis that do not suffer from the issue we raised:
 - a) Copy number signatures have been constructed in a way that does not use one CNA in multiple channels by <https://www.biorxiv.org/content/10.1101/2021.04.30.441940v1.full>, despite the challenges mentioned by authors.
 - b) If authors want to use the same CNA in multiple channels as they reiterate in several of their responses, they could still formulate the solution in a better way to handle different feature categories. Potential methods based on NMF include joint NMF and integrative NMF, which have been used in the computational genomics field for some time such as in integrative multi-omics tools. For example, please see: <https://pubmed.ncbi.nlm.nih.gov/26377073/>.
3. The advantage of using a CNA in multiple features, as claimed in the author responses on several occasions, is not supported by any simulation or concrete analysis.

To summarize the above points, it is not obvious how the unique challenges authors mentioned are solved by their approach. Furthermore, there are alternative ways to capture the information they are trying to include in the calculation without running into the issue we raised, so it is not clear why the proposed method is a major advance.

Point B:

In response to our question:

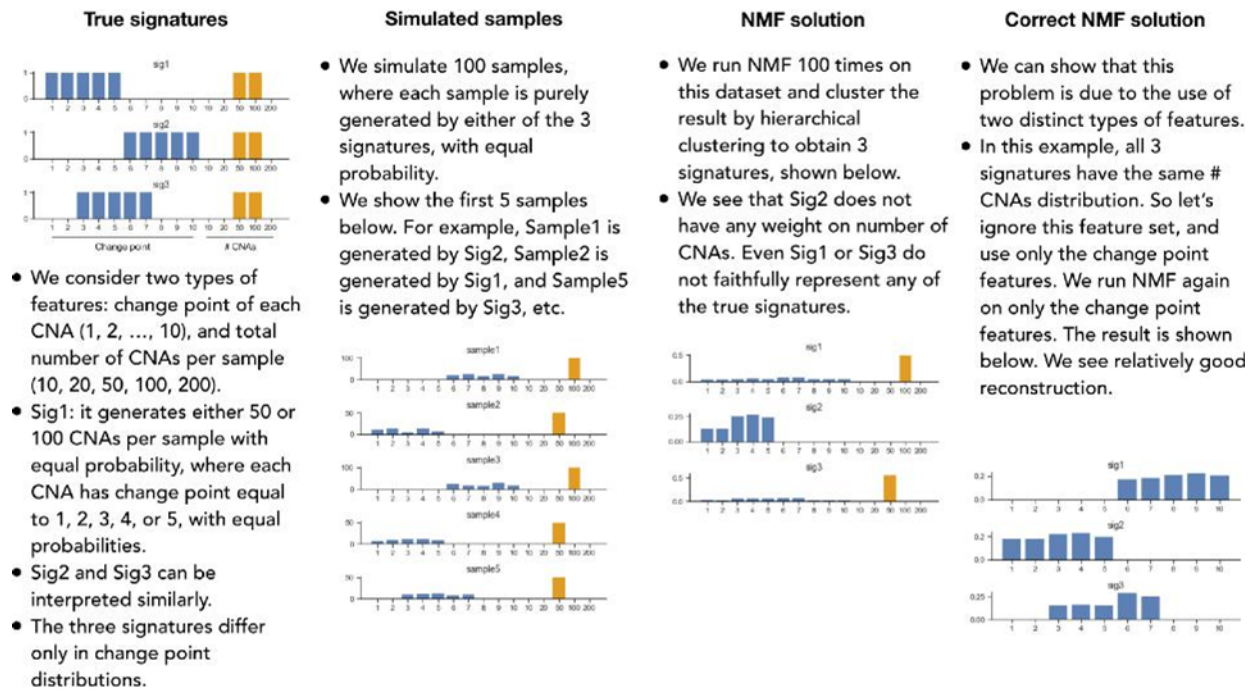
Second, the signature definition makes NMF an unsuitable algorithm for decomposition, since NMF has no awareness of the inherent correlations among different channels introduced by the definition. As a result, the NMF-derived signatures are often ill-defined. For example, CX13 has a high chance of causing large copy number changes

but has no corresponding segment length or CNAs per 10Mb. This should not happen since the CNAs would have to belong to a segment length group. This methodological issue is not obvious and had escaped us in the initial review, but it is a major problem.

Authors responded:

... Fortunately, CX13 is the only signature that exhibits such behaviour. Our low to medium amplification signatures (CX8 and CX9) have weights on high change-points but also have weights on the expected segment sizes (both short and long). However, in the case of CX13, the high changepoint component is such a strong defining feature that no weight is given to segment size components. While this may yield a slightly unexpected signature definition, this does not necessarily imply that this signature is not robust in representing the process generating high-level focal amplification. ...the replication of all pan-cancer signatures during our cancer-type enriched signature analysis suggests that our signatures are not ill-defined, but rather are robust readouts of biological signals.

1. Authors claimed that applying NMF is reasonable because the assumption of NMF is not violated, i.e., nonnegativity. However, the assumption of NMF is not only nonnegativity. Yes, one could apply NMF to any nonnegative matrix, but it will only produce desired interpretation when the data is generated from the generative model that NMF assumes. In fact, NMF with KL divergence/Frobenius norm can be derived from MLE of Poisson/Gaussian distributed data.
2. Robustness does not mean it is not ill-defined. One can get reproducible (thus robust) but non-interpretable and meaningless results.
3. Authors claimed that “CX13 is the only signature that exhibits such behavior”, but how could the authors possibly know that CX13 is the only one? There is no methodological reason that all the other signatures should not suffer from the same problem. In fact, many other signatures are missing weights in one or more feature categories, e.g., CX6, 7, 8, 12, 14. Missing weights is an extreme demonstration of the issue we raised, and we only picked CX13 as an example in the last round of review. More importantly, the same issue will also lead to significant biases and errors in the other signatures.
4. Since the authors did not, we have performed a simulation study below to demonstrate the problem that the proposed method has. We believe that this issue is also present in the earlier publication Macintyre et al., where it may be more difficult to catch because there were not as many obviously problematic signature distributions. See Point A for possible solutions that would address this issue.



Point C:

Authors state they evaluated all their correlations carefully, but several of the correlations for which we asked for further support were removed due to lack of support. We only mentioned selected examples to make the point that correlations are not always well supported. This does not mean that our prior list was exhaustive. The fact that a large portion of our questions resulted in removal of statements from the paper is a contradiction with the claim that the correlations stated are well supported.

Authors Reports on the Second Revision:

Roadmap of key steps in the revision process.

This has been a lengthy but rewarding review process that substantially improved our work. We now present a very rich paper and navigating the path that led to this revision is a daunting task. To make it easier we highlight here some of the major developments throughout the review process of this paper. This might be particularly helpful for the new **Reviewer 4**, who was not involved in previous revision rounds.

There were four main steps in the revision process:

1. The initial submission focussed on the derivation of the signatures and their biological interpretation. Our method extends the one previously described for ovarian cancer by Macintyre *et al.* (Nature Genetics 2018) and we reported methodological advances to scale it to pan-cancer analysis and the evidence we found for the aetiologies of the 17 signatures we discovered. Reviewers asked us to perform a major revision of the paper strengthening the aetiologies and adding a description of the clinical impact of the signatures. The editor advised that substantial re-writing and re-focusing were needed.
2. The second submission contained no changes to the method and the number of signatures, but added substantial evidence to the aetiologies, and -importantly- added a focus on drug targets and therapy response prediction using both cell lines and patients. Reviewers 1 and 2 were very supportive, but Reviewer 3 challenged the underlying methodology. Reviewer 3 viewed our method through the lense of previous SNV/SV approaches and in response we pointed out the unique challenges of copy number analysis.
3. After the third submission, Reviewer 3 crystallised their methodological concern into a simulation study. This was very helpful, because it clearly communicated the issue they thought they had spotted.
4. In this response we show that Reviewer 3's simulation has a crucial design flaw and does not represent our approach (**page 21**) and we present our own comprehensive series of simulations to settle this technical point:
 - a. We simulate how well-known mutational processes shape cancer genomes and show that our approach recovers them (**page 4, Methods page 13**);
 - b. We extend our existing Monte Carlo simulation for signature activities and now also show the robustness of our signature definitions (**page 6, Methods page 22**);
 - c. We compare single-feature encoding of CNAs to our multi-feature encoding and show that our approach recovers mutational processes more accurately (**page 9, Methods page 17**).

The simulations show the robustness of our approach. However, they also demonstrate that shrinkage in signature definitions is a universal feature of NMF (**page 16**), even for widely used single base substitution signatures. To counteract this bias, we now present an extended signature interpretation approach that uses signature activity for feature refinement.

In this version, the main manuscript remains almost unchanged, but we have substantially restructured the methods section to improve readability and navigation. Sections with major improvements are highlighted in green.

The topics we discuss with the Reviewers have become more detailed and more technical over time, and it is important not to lose sight of the bigger picture.

Our study synthesises an extremely large and comprehensive amount of genomic data into a coherent framework to measure chromosomal instability. This framework provides a significant step change for the deeper understanding of CIN in a field that has only seen piecemeal progress over the last decades

Referee #1

All our concerns have been adequately addressed. We now recommend publication of this manuscript.

Thank you!

Referee #2

The authors addressed my previous comment in a satisfactory manner.

Thank you!

Comments:

1) The authors mentioned that 528 of the ovarian cancers have BRCA1/2 mutations and only 17 are wild type (pages 38 and 39 in the Supplementary Methods file). This seems to be unreasonable. I could not find in the methods section how they derive this high number. What classes of mutations are included in the BRCA1/2 cases? I checked in the Ovarian TCGA data set and there are about 40% of cases with BRCA1/2 somatic or germline mutations + BRCA1/RAD51C promoter methylation.

We agree that our description of BRCA status lacked detail.

Indeed, only 27% of the ovarian cancers in our study have a confirmed mutation or promoter hypermethylation of BRCA1/2 or RAD51C. However, a large fraction of non-mutant cases have LOH of BRCA1/2, which has been shown to be important to convey a BRCA1/2 phenotype¹. This is where the discrepancy arises, because we considered WT for this analysis to be an absence of both mutation and LOH.

We have now added further explanation in the figure captions and relevant methods section (**updated Supplementary Fig. 35, Methods pages 32 and 44**) to clarify this point.

2) Supplementary Figure 2 is missing d-f labels

We fixed this issue.

Referee #2 comments on Referee #3 assessment:

This Reviewer agrees with Reviewer 3, that the authors should demonstrate that by simulation of the genome they can reconstruct the input signatures using their method. The authors already introduced noise into the CNA profiles and found stability of the solutions. It will be useful to show the average profiles of signatures that were retrieved after the noising process.

This Reviewer agrees with the authors that NMF assumptions are not violated, however, the results might not reflect a biological mutational process. I would suggest that they module a genome using the CIN signatures (including the inter-channels correlation) and then run NMF and see how the obtained signatures are similar to the input signatures.

We agree with the Reviewer and added the requested simulations to the manuscript and provide a summary below.

- In Part 1 of our response we describe a simulation producing cancer genome copy number profiles and show that our method can reconstruct the input signatures.
- In Part 2 we expand our previous Monte Carlo simulation estimating the variance in signature activities to also estimate the variance in signature definitions. We find that the signature definitions are highly robust.
- In Part 3 we tested the inter-channel correlations. In particular, we focused on two of our features and simulated three signatures. We recovered all three signatures as well as a correct assignment of the features to each other. For more details, please see our response to Reviewer 3 Point B subpoint 4 (**page 21**).

PART 1 SIMULATION OF CANCER GENOMES

Full details of the simulation are described in section “Simulating CIN” (**Methods page 13, Supplementary Figs. 44-47**). Here we provide a brief overview of the simulations and results.

To reduce bias, we did not use our signature definitions for the simulations, but rather simulated types of CIN previously described in the literature:

- chromosome missegregation via mitotic errors (CHR)²,
- large-scale state transitions via homologous recombination deficiency (LST)³,
- focal amplification via ecDNA circularisation and amplification (ecDNA)⁴,
- early whole-genome duplication via cytokinesis failure (early-WGD)⁵, and
- late whole-genome duplication via endoreduplication (late-WGD)⁵.

For any given sample, three of the five mutational processes were active (**Methods Table 3**). We modelled WGD by setting a tetraploid background before inserting other CNAs (early-WGD) or by multiplying all copy number states after all CNAs were placed (late-WGD). We drew the number of affected chromosomes by a mutational process as well as the number of CNAs from a Poisson distribution, allowing for a wide range of variability to be included (**Method Tables 3-5, Supplementary Fig. 44**). We specifically allowed for the overlap of CNAs to include additional biological complexity in the simulation. This way of simulating gave us a gold-standard to compare our derived signatures against (**Supplementary Fig. 44 and Fig. R1**).

We simulated 240 samples in this way and then applied our feature encoding and signature identification methodology in the same manner in which it was applied to the TCGA data. This process yielded 5 signatures, which we then compared to the gold standard (**Supplementary Fig. 45 and Fig. R2**). Reassuringly this showed a good match between the simulated CIN types and the identified signatures: Average cosine similarity was at 0.64 (**Methods Table 6**). All permutations tests based on the cosine similarities

were significant except for the CHR signature, which can be explained by additional small weights on other components, decreasing the similarity value.

We also assessed the accuracy of the signature activities by observing matches between the simulated and observed dominant signatures (**Methods Table 3, Supplementary Figs. 46-47**). The overall accuracy was 78.3% (329/420 dominant samples).

Overall the results of this simulation shows that our approach is able to both accurately identify signatures of mutational processes and quantify signature activities.

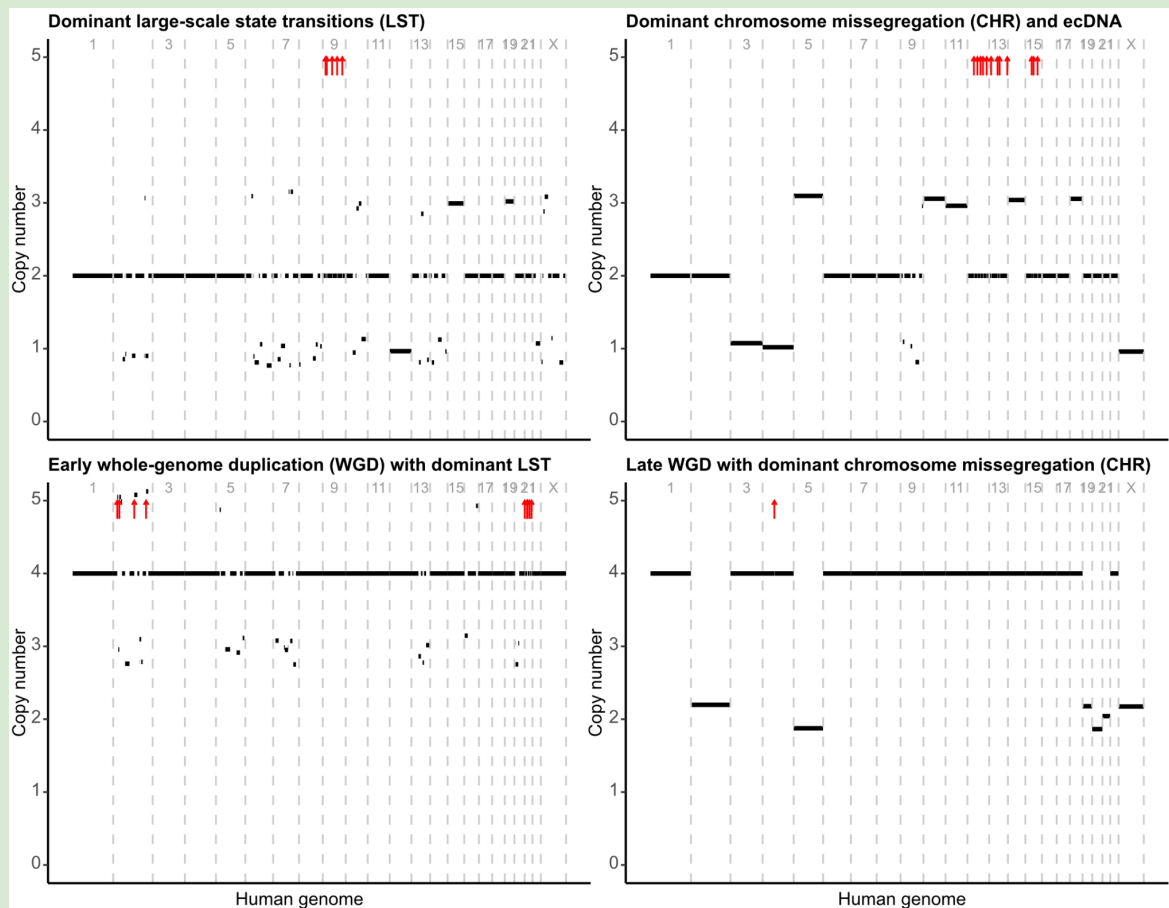


Fig. R1: Example copy number profiles of the simulation producing genomes exposed to mutational processes

The titles of each plot state the dominant mutational processes that acted on this sample. The x-axis denotes the length of the human genome, with vertical, grey dashed lines showing chromosome boundaries. Y-axis shows the absolute copy number. Red arrows denote copy number aberrations with a copy number higher than 5.

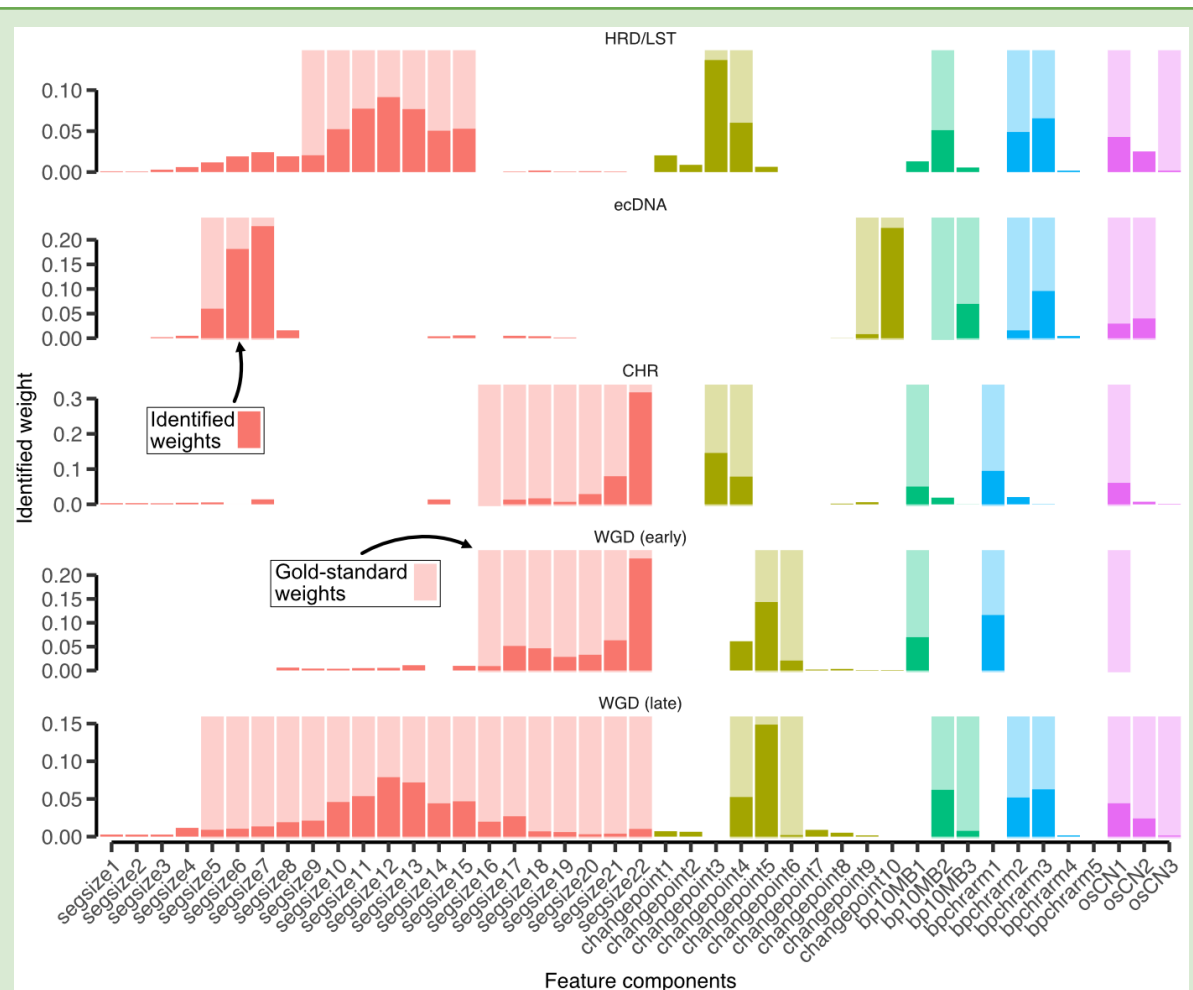


Fig. R2: Gold-standard and identified copy numbers signatures of five mutational processes modelled for the simulation of genomes. Shaded bars: Components received weight depending on the importance of this component to reflect the mutational process. Fully coloured bars: Identified signature weights.

PART 2 ROBUSTNESS OF SIGNATURE DEFINITIONS

One of our previous simulations (“Signature activity stability”, **Methods page 21, Extended Data Fig. 5**) showed the stability of signature activities and estimated signature-specific thresholds by introducing noise into the input matrix. The reviewer suggested also observing the effects of copy number calling noise on the signature definition matrix (signature by component). We have performed this analysis (full description in **Methods page 22**, summary below).

We applied our signature identification methodology to 100 noisy input matrices and observed how many of the signatures matched the 10 original pan-cancer signatures (cosine >0.85, **Fig. R3, Supplementary Fig. 42a**). 7 of the 10 pan-cancer signatures had a match across all 100 noisy matrices, with the remaining 3 signatures matching 98 or 99 out of 100 times.

As suggested by the Reviewer we visualised the noise in the signature definitions by adding error bars to the barplot representation of our signatures (**Fig. R4, Supplementary Fig. 42b**). Error bars are very small, often hardly visible.

Overall these show our pan-cancer signatures are extremely robust to noise in copy number calling.



Fig. R3 - Heatmap of signature definition robustness. Heatmap of signature definition robustness. Grey denotes that all 100 noisy matrices matched with cosine similarity > 0.85). Three signatures had mismatches of up to 1-2% of cases.

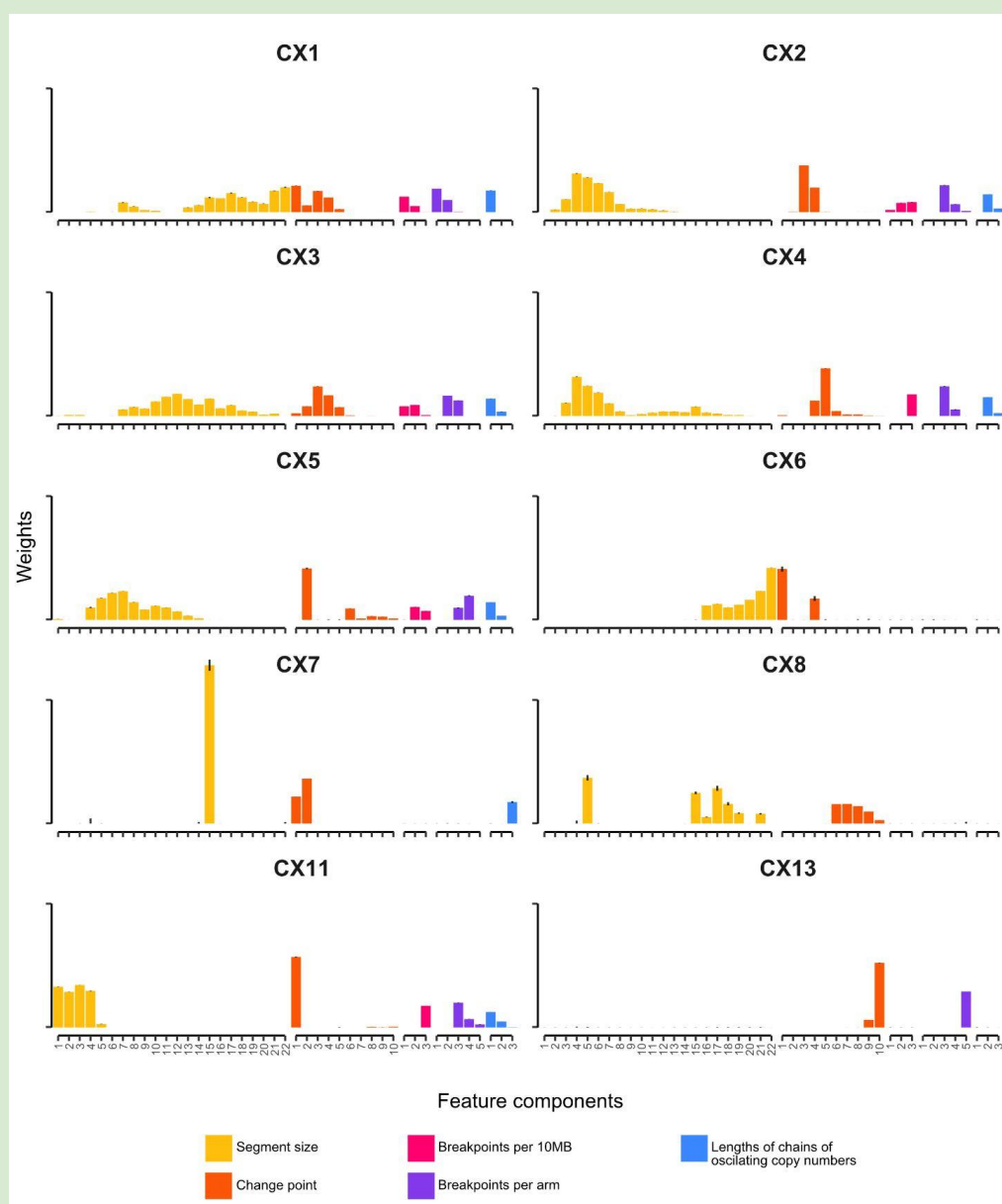


Fig. R4. Barplots showing the variability of signature definitions with added noise. Displayed on the x-axis are the 43 components of the five features. The bar height shows the adjusted weights for each signature. Weights were normalised per signature. Error bars represent the standard error of weights for signatures estimated from 100 noisy simulations.

Referee #3

I find it surprising and disappointing that the authors failed to consider our critiques more carefully. We pointed out a critical flaw in the central formulation of their signature decomposition approach. If they do not find our critique credible, they should have performed methodological studies to investigate. Instead, they simply insist that they have made "a major conceptual advance" and claim, with no demonstration whatsoever, that their method can resolve issues encountered in previous approaches. In addition, they seem to be unaware of previous literature in the field nor have sufficient understanding of the mathematical framework they are employing.

Furthermore, in each round of reviews, we point out only the most critical issues, giving an example where possible. The documents (manuscripts/supplement/response) are very long and we cannot make an exhaustive list of issues or examples, as doing so would take an inordinate amount of time. The authors should not think that they have addressed the main problem by addressing only the specific example (e.g., by deleting that part). In their first round—as an example—they claimed that their survival analysis was a major finding. When we pointed out problems, they simply deleted that part and came up with new claims. That has happened again in this round.

Here, we make three points. Given that this is a second appeal with no substantive methodological change, I am unwilling to make an exhaustive list of problems.

Point A:

In response to our comment:

1. Flaws with the design of the signature discovery method: The choice of features results in one copy number alteration event contributing to multiple channels in the signature definition, unlike previous signature analysis based on point mutations or structural variations, where each variant is only counted once.

Authors responded (in part):

... Copy number data, however, presents unique challenges, precluding the use of a similar approach. CNAs cannot be easily clustered into single "events" as is the case for SVs, as they lack the information required to relate one CNA to another. Further confounding this problem is the fact that copy number changes are quantified with respect to reference genome coordinates. For example, the loss of a "chimeric" chromosome (an already rearranged chromosome containing parts of many different chromosomes), is a single "event" in the tumour cell. However, mapped to a reference genome, this can manifest as many copy number changes across multiple chromosomes. Therefore, it is not possible to generate a vector of "event" counts and thus approaches similar to those used in the SNV and SV setting cannot be applied.

1. Authors mentioned the challenges of analyzing CNA data and imply that their method can resolve them. But they do not demonstrate that their method is able to deal with those challenges. For example, they mentioned that loss of a chimeric chromosome

can manifest as many copy number changes across multiple chromosomes. Can their method resolve that? A chimeric chromosome can still be counted as multiple SVs just as CNAs do. This point does not represent a CNA-specific challenge and does not relate to the issue we raised.

Response: The reviewer challenges us to better demonstrate why our method tackles some of the major hurdles in modelling CIN via CNAs. In this revision, we present extensive simulation studies to demonstrate this (some already described above and repeated below for completeness).

The concrete example of chimeric chromosomes clearly demonstrates a challenge specific to CNAs because CNAs do not provide the level of detail that structural variants (SV) do. The loss of a chimeric chromosome results in multiple SVs but these can be tied together via split or spanning reads allowing reconstruction of the chimeric chromosome. For SVs it would not make sense to ignore this information and in fact, the PCAWG consortium used it to cluster SVs into complex rearrangements⁶, which is not possible for CNAs.

The question remains whether our approach can appropriately detect loss or gain of chimeric chromosomes. We therefore added these events to our simulated genomes (described in “Simulating CIN”, **Methods page 13**) and tested whether any of the signatures representing whole-chromosome change were perturbed. This analysis showed that CX14, a signature of whole chromosome changes, significantly increased upon addition of chimeric chromosomes suggesting our signatures can indeed encode these events correctly (Welch’s t-test with FDR correction; q-value=2.54*10⁻⁰⁶).

A more general evaluation of our multi-feature method follows below.

2. There are other ways of approaching CNA signature analysis that do not suffer from the issue we raised:
 - a. Copy number signatures have been constructed in a way that does not use one CNA in multiple channels by <https://www.biorxiv.org/content/10.1101/2021.04.30.441940v1.full>, despite the challenges mentioned by authors.

Response: The cited approach encodes CNAs in terms of the copy number of different sized segments split by heterozygosity status. Here, we compared this single-feature approach to our multi-feature encoding using simulated genomes.

The simulation is the same as the one above in our response to Reviewer 2 (see Part 1, **page 4**). For completeness we summarize it here again (full description in **Methods page 13**). We simulated cancer genomes which were exposed to five well-studied mutational processes:

- chromosome missegregation via mitotic errors (CHR),
- large-scale state transitions via homologous recombination deficiency (LST),

- focal amplification via ecDNA circularisation and amplification (ECDNA),
- early whole-genome duplication via cytokinesis failure (early-WGD),
- late whole-genome duplication via endoreduplication (late-WGD)

Multi-feature encoding results: These are fully described above (Reviewer 2, response part 1, **page 4**). In summary, the multi-feature encoding enables the identification of five signatures that closely match the expected signature definitions and activities (**Figs. R2-R3, Supplementary Figs. 44-47, Methods page 13**).

Single-feature encoding results: We adapted the approach outlined in the preprint the reviewer cites to our simulated data by classifying each copy number segment into one of 28 categories (**Methods page 17**). These categories were designed to match those found in the preprint as closely as possible. As our approach uses total copy number (rather than allele-specific) the match could not be 100%, but was close enough to expect results comparable to the original method.

We used the same approach as the preprint to identify signatures (SigProfilerExtractor), which identified 4 signatures (the true number of signatures is 5) (**Fig. R5, Supplementary Fig. 49**).

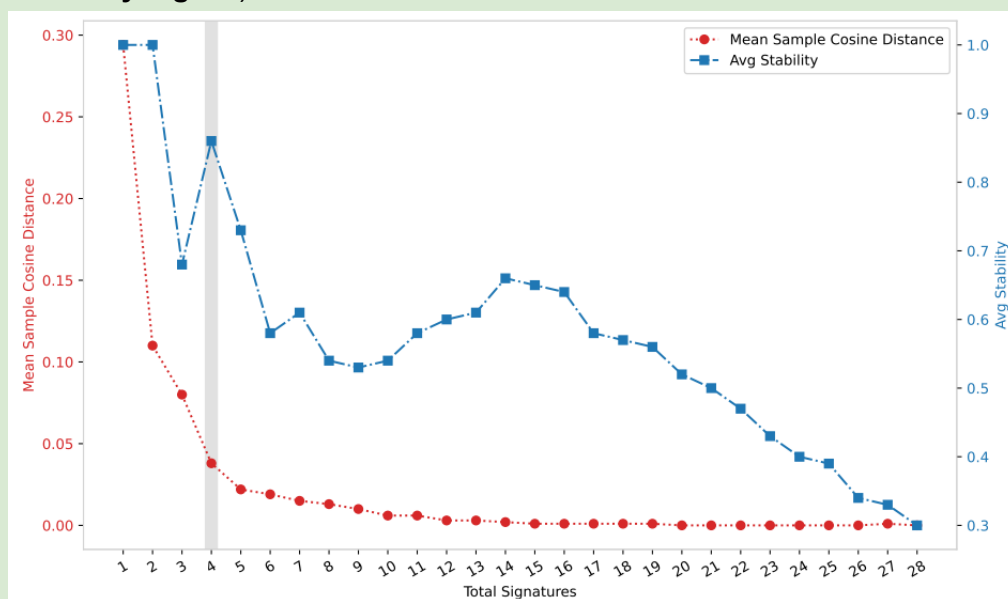


Fig. R5. Optimal number of signatures according to SigProfilerExtractor. Plot shows the stability and mean sample cosine distance for different solutions. Grey bar denotes the optimal solution.

We constructed 5 gold-standard signature definitions using this encoding for LSTs, ecDNA, CHR, WGD-early and WGD-late, and compared these to the 4 identified signatures (**Fig. R6, Supplementary Fig. 50**). Of the 5 gold-standard signatures, 2 were identified correctly, 1 could be inferred from combinations of the signatures, and 2 were not detected (**Table R1**). Signatures A and B represented LSTs on different ploidy backgrounds. Signatures C and D represented ecDNA on different ploidy backgrounds. While there was no signature directly representing WGD, this could be inferred via the different ploidy backgrounds. Signatures for chromosome missegregation events and early-WGD were not detected.

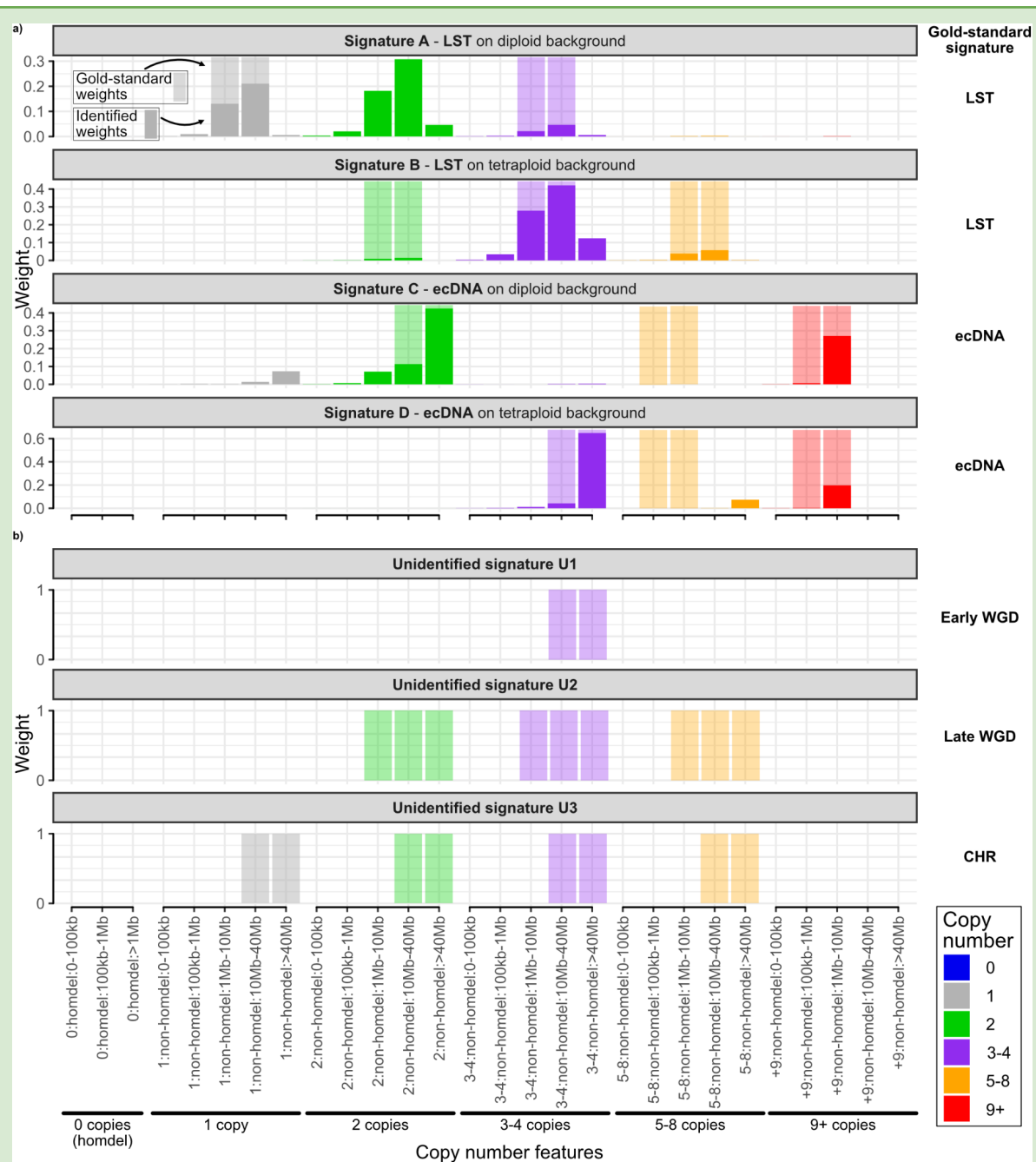


Fig. R6. Barplot showing the proportion of the mutations for each signature identified from simulated genomes using the alternative approach. Gold-standard signatures of the five mutational processes are shown with shaded bars. Their patterns reflect how the mutational processes were designed and the background ploidy status is taken into account. a) The four signatures identified by the single-feature encoding with SigProfilerExtractor. b) The three signatures not identified by the single-feature encoding method.

As the single-feature encoding uses absolute copy number, it is expected that the gold-standard signatures will be artificially split across different ploidy backgrounds. This is precisely why we removed copy number from our multi-feature encoding.

CHR, WGD-early and WGD-late are captured by multi-feature encoding due to the ability to represent many biological properties of CNAs across the feature space. For example, early and late WGD albeit having similar copy number change points, can be separated by the segment size feature (components 16-22 vs 5-22) as well as the breakpoint per 10MB

(components 1 vs 2-3) and breakpoints per chromosome arm features (components 1 vs 2-3).

In summary, using single-feature encoding does not allow the NMF to identify the underlying mutational processes, while the multi-feature encoding we propose allows to study CIN in much greater detail and precision.

Table R1: Comparisons of different encoding approaches for the discovery of CIN producing mechanisms

Mutational process	Identified by...	
	Multi-feature encoding (our approach)	Single feature encoding (comparable to preprint)
Correct number of signatures	Yes	No
Early WGD	Yes	No
Late WGD	Yes	Indirectly
Large-scale state transition	Yes	Yes
Chromosome missegregation	Yes	No
ecDNA	Yes	Yes

(Note: as the approach in the preprint is unpublished, we have not performed an extensive benchmarking exercise for inclusion in our manuscript, but have included the above results in the Methods on **Methods page 17**).

- b. If authors want to use the same CNA in multiple channels as they reiterate in several of their responses, they could still formulate the solution in a better way to handle different feature categories. Potential methods based on NMF include joint NMF and integrative NMF, which have been used in the computational genomics field for some time such as in integrative multi-omics tools. For example, please see: <https://pubmed.ncbi.nlm.nih.gov/26377073/>.

Response: Indeed, many matrix decomposition methods exist, which could in principle be applied to our data. However, without extensive methodological changes, the approaches suggested by the Reviewer are impractical given the structure of our data.

For example, one of the constraints of joint and integrative NMF is that the number of possible identifiable signatures must be less than the minimum number of components observed for each feature (or sub-matrix)^{7,8}. Thus, in our data, the maximum number of potential signatures would be 2, which obviously underestimates the complexity of CIN.

Given this methodological limitation and the facts that (1) we do not violate any of the assumptions of standard NMF and (2) we get highly robust and meaningful signatures using standard NMF, there is no strong argument for altering our approach at this stage.

To strengthen our discussion on the tradeoff between using regular NMF and other methods, we have added further discussion in the Methods section “Deriving signatures” (**Methods pages 19**) where we articulate the limitations of using joint and integrative NMF.

3. The advantage of using a CNA in multiple features, as claimed in the author responses on several occasions, is not supported by any simulation or concrete analysis.

Response: We have clearly demonstrated the advantage of our multi-feature encoding over single-feature encoding (see above response to Point A, subpoint 2a, **page 9**).

To summarize the above points, it is not obvious how the unique challenges authors mentioned are solved by their approach. Furthermore, there are alternative ways to capture the information they are trying to include in the calculation without running into the issue we raised, so it is not clear why the proposed method is a major advance.

Response: The Reviewer challenges us to better present our approach and its advantages. We have articulated the unique challenges of using CNAs to identify and detect signatures of CIN in the Methods section “Copy number feature encoding” (**Methods page 8**).

Via simulation we have shown that our method is able to overcome these challenges. Compared to alternative ways to encode copy number signatures, we have demonstrated that our multi-feature encoding is superior to single-feature encoding approaches.

Thus, in summary, several lines of evidence support that our approach is a major advance.

Point B:

In response to our question:

Second, the signature definition makes NMF an unsuitable algorithm for decomposition, since NMF has no awareness of the inherent correlations among different channels introduced by the definition. As a result, the NMF-derived signatures are often illdefined. For example, CX13 has a high chance of causing large copy number changes but has no corresponding segment length or CNAs per 10Mb. This should not happen since the CNAs would have to belong to a segment length group. This methodological issue is not obvious and had escaped us in the initial review, but it is a major problem.

Authors responded:

... Fortunately, CX13 is the only signature that exhibits such behaviour. Our low to medium amplification signatures (CX8 and CX9) have weights on high change-points but also have weights on the expected segment sizes (both short and long). However, in the case of CX13, the high changepoint component is such a strong

defining feature that no weight is given to segment size components. While this may yield a slightly unexpected signature definition, this does not necessarily imply that this signature is not robust in representing the process generating high-level focal amplification. ...the replication of all pan-cancer signatures during our cancer-type enriched signature analysis suggests that our signatures are not ill-defined, but rather are robust readouts of biological signals.

Response: The reviewer cites selected parts of our previous response, which leaves out some of its nuance and context.

In particular, we did not state that CX13 was *the only* signature that exhibited zero weights for a feature, rather CX13 was the only signature with *many features with zero weights*. We also acknowledged that the observed effect may be present across all signatures, albeit at lower levels. Relevant sentences from our previous response are below:

“We interpreted the signature definitions with caution. A significant strength of our approach is that it is easy to identify potentially erroneous signature definitions as these have strong weights on a single component. In this case, the single defining weight is so strong that other related weights (which we might expect to see) get shrunk to 0 (this is part of the NMF regularisation to ensure a unique solution). Fortunately, CX13 is the only signature that exhibits such behaviour.”

“While we do acknowledge that this effect may be more subtly manifested across other signatures as well, the replication of all pan-cancer signatures during our cancer-type enriched signature analysis suggests that our signatures are not ill-defined, but rather are robust readouts of biological signals.”

1. Authors claimed that applying NMF is reasonable because the assumption of NMF is not violated, i.e., nonnegativity. However, the assumption of NMF is not only nonnegativity. Yes, one could apply NMF to any nonnegative matrix, but it will only produce desired interpretation when the data is generated from the generative model that NMF assumes. In fact, NMF with KL divergence/Frobenius norm can be derived from MLE of Poisson/Gaussian distributed data.

Response: We assume the Reviewer is referring to whether our choice of cost function (KL divergence) is well matched to how our data (the input matrix) is distributed.

Our NMF implementation⁹ provided the ability to choose from the β -divergence family of cost functions (where $\beta=0$ assumes gamma distributed noise, $\beta=1$ assumes Poisson, and $\beta=2$ assumes Gaussian) and we had to choose an appropriate cost function for our data. Our input matrix contains sum-of-posterior probabilities and we found through distribution fitting that 35 out of the 42 components were better fit by a Poisson (as for count data) than a Gaussian or gamma distribution. Therefore we chose $\beta=1$ which is equivalent to the KL-divergence cost function.

2. Robustness does not mean it is not ill-defined. One can get reproducible (thus robust) but non-interpretable and meaningless results.

Response: We agree with this general statement. This is exactly why we spent so much time and effort on a comprehensive characterisation of our signatures and extensively testing their usefulness in clinical applications. For example, we show strong evidence that 12 of the 17 signatures are biomarkers of mutational processes and that they can be used as clinical predictors of chemotherapy response. These results show very clearly that our signatures are not only robust and reproducible, but also interpretable and meaningful.

3. Authors claimed that “CX13 is the only signature that exhibits such behavior”, but how could the authors possibly know that CX13 is the only one? There is no methodological reason that all the other signatures should not suffer from the same problem. In fact, many other signatures are missing weights in one or more feature categories, e.g., CX6, 7, 8, 12, 14. Missing weights is an extreme demonstration of the issue we raised, and we only picked CX13 as an example in the last round of review. More importantly, the same issue will also lead to significant biases and errors in the other signatures.

Response: We are using signatures as markers of mutational processes. Being an accurate marker does not require that *all* biological details are resolved.

The successful identification of the simulated signatures (above), in combination with the aetiological and clinical associations reported in the manuscript, demonstrates that having signatures with zero-weight features does not negatively impact the utility of our pan-cancer signatures. This shows it is not necessary to have weight assigned to each feature for the purpose of accurate and robust quantification of signature activity.

However, we agree with the Reviewer that for the purposes of interpreting the copy number pattern represented by a signature, it is ideal to have weights defined for each feature. The Reviewer makes an astute observation that there are zero signature weights for some features where intuitively we would expect non-zero weights given what we know about the biological properties of the feature. For example, every signature should have some weight on at least one segment size component given that each copy number event has a segment size.

One solution to this problem, as the Reviewer suggests, could be to use a different NMF which forces weight on each feature. However, as discussed in Methods section “Deriving signatures” (**Methods page 19**) this may bias the resulting signatures. Furthermore, even if a different NMF is used, this may still result in zero-weight features as weight shrinkage is a general property of NMF (**Methods page 27, Fig. R7; PART I below**).

Therefore, to improve the interpretation of our signatures and counteract feature shrinkage, we developed a procedure to generate a “signature interpretation matrix” allowing us to better study the copy number pattern associated with each mutational process. This signature interpretation matrix uses the activity matrix to “fill the gaps” and ensure that each signature has weights on each feature. The procedure is outlined and

validated below and has been added to the **Methods (Methods page 29, Fig. R8, Extended Data Fig. 4; PART II below)**.

PART I: Signature weight shrinkage is a common consequence of NMF

Here we hypothesised that the unexpected zero-weight features observed for some of our copy number signatures, might actually be the result of normal NMF behavior, rather than a quirk of our method. If so, we would also expect to see evidence of artificial weight shrinkage in other situations where NMF is used, for example SNV signatures. Looking closely at our signatures, we noticed that when two or more signatures were defined by the same feature, quite often one signature “stole” the weight from the other signature.

In order to study this phenomenon we simulated mixes of COSMIC SBS signatures which are well-studied single-feature signatures of single base substitutions^{10–12}. We chose three closely related signatures which only differed in the strength of a few components. We created 500 samples from those signatures with differing proportions of signature activities. We then applied a common NMF algorithm and compared the resulting signature definitions with the original definitions. We ran this procedure 250 times to establish a reliable variance.

We did this simulation for two sets of SBS signatures ({SBS10a, SBS10b, SBS10c} and {SBS5, SBS16, SBS92}) and for two of the most common NMF algorithms (NMF with Brunet algorithm, R package NMF¹³, v0.21.0; SigProfilerExtractor, v1.1.4¹⁰).

Starting with the Brunet NMF, in the first example we tested SBS10a, SBS10c and SBS10d (**Fig. R7a; Supplementary Fig. 43a**). Weights of T[C>A]A and T[C>A]T significantly decreased in the SBS10c signature. Specifically, the weight of T[C>A]A mutations drops by an order of magnitude from 0.06 to 0.0071 [9.67e-07, 0.031]; while the weight of T[C>A]T decreases from 0.19 to 0.059 [0.0035, 0.13]. In the second example (**Fig. R7b; Supplementary Fig. 43b**), we combined SBS5, SBS16 and SBS92. In this case, the weight of A[T>C]A significantly decreased from 0.046 to 8.83e-06 [1.76e-12, 1.86e-04] in SBS5.

This property can also be observed when using SigProfilerExtractor which was used for the PCAWG studies. Results of both examples are consistent with those obtained with the Brunet algorithm (**Supplementary Fig. 51**).

In summary, we established that shrinkage of certain features is a common phenomenon when using NMF. We observed that the missing weights were generally weak or moderate signals.

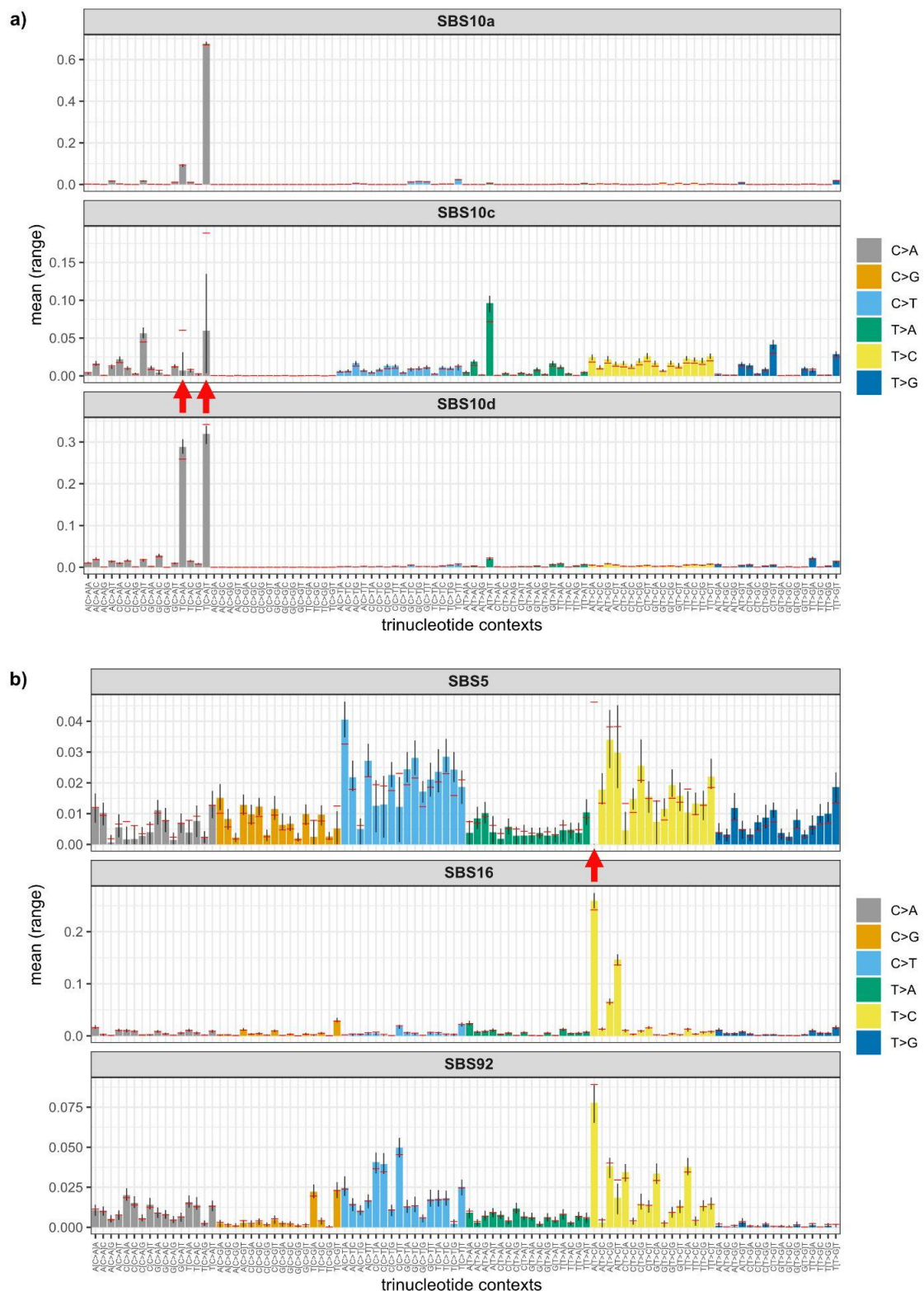


Figure R7. Bar plots showing SBS signatures identified from simulated data. Coloured bars show the median signature weight obtained by 250 simulations with each including 500 samples. Error bars depict the minimum and maximum values obtained across all 250 simulations. Horizontal red lines show the original component weights and the red arrows depict the components which decreased the strongest compared to the original definitions. The Brunet algorithm was used to decompose signatures. **a:** Combination of SBS10a, SBS10c and SBS10d. **b:** Combination of SBS5, SBS16 and SBS92.

PART II: Generating the signature interpretation matrix

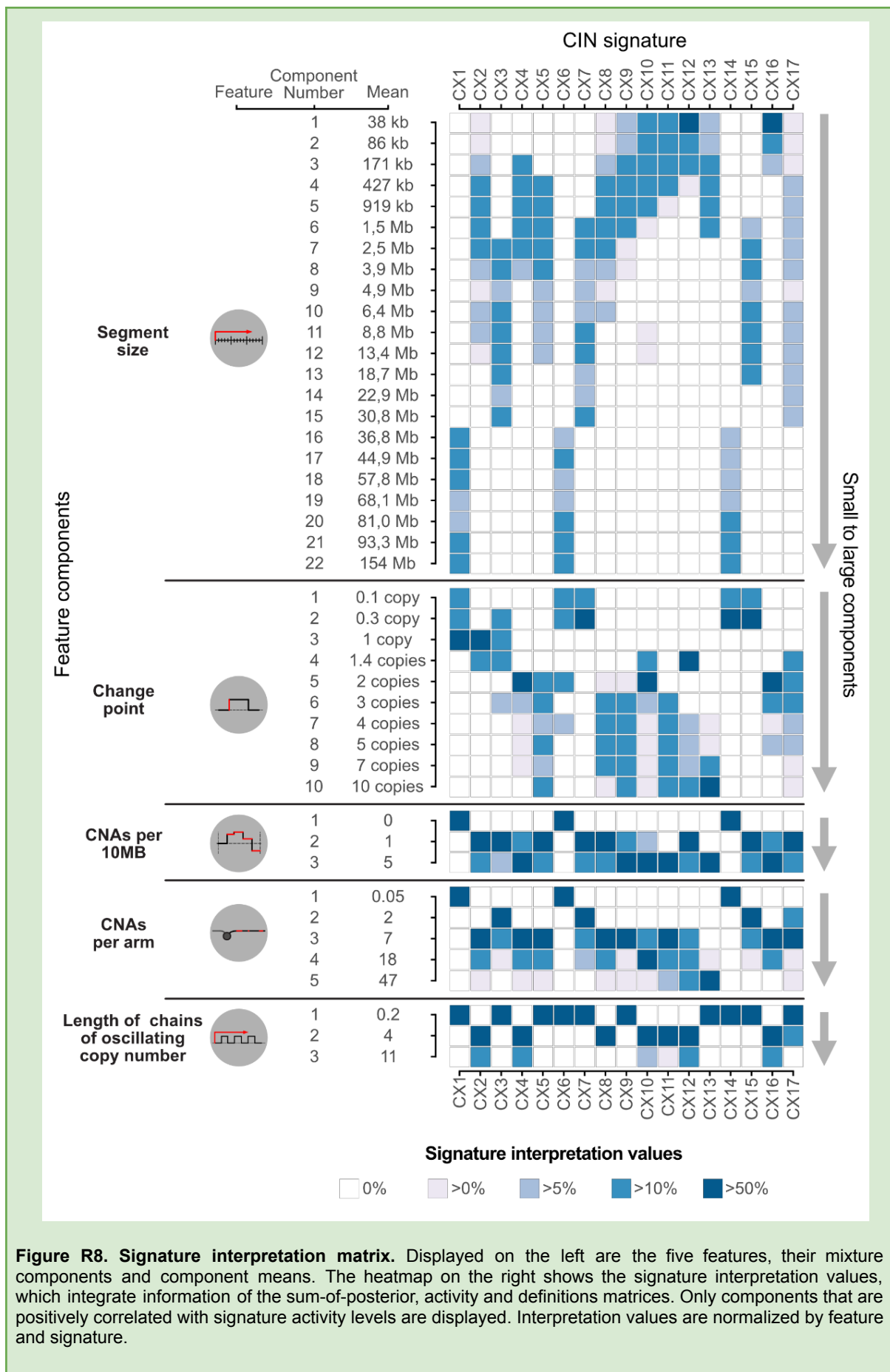
As established above, the copy number signatures are accurate biomarkers of mutational processes, however not all signatures are defined across all five features (**Supplementary Fig. 48**).

Therefore, to overcome this problem we computed a signature interpretation matrix (**Methods page 29**). The matrix is calculated using information from the input matrix (sum-of-posterior matrix) scaled by the signature activity matrix, and then combined with the original signature definition matrix (**Fig. R8, Extended Data Fig. 4**). This allows us to have every feature defined for the 17 signatures, and provide more complete descriptions of the patterns underlying each signature (**main manuscript Fig. 2**).

We validated this procedure using simulated signatures (same as above). We calculated a signature interpretation matrix for the 5 signatures identified from the simulated genomes, then compared this to the gold-standard signatures (**Supplementary Fig. 52, Fig. R9**). This showed an increase in the average cosine similarity compared to the definition matrix from 0.64 to 0.75 (**Methods Table 8, Method page 30**). In addition, the standard deviation reduced down from 0.13 to 0.09, suggesting more consolidated results and all cosine similarities were significant (Permutation test with 10,000 iterations of cosine similarities with FDR-correction, **Methods Table 8**).

These results support the use of a signature interpretation approach. The interpretation matrix also demonstrates results consistent with the literature. For example, CX13 shows weight on the segment size feature components 4-6 (300kb to 2MB) which overlaps strongly with literature estimates for ecDNA^{14,15}. Its other features (bp10MB and bpchrm) suggest highly clustered mutations, in line with biological expectations of this process.

In summary, the use of the signature interpretation matrix allows us to better understand the mutational spectra caused by the mutational processes which are captured by the copy number signatures.



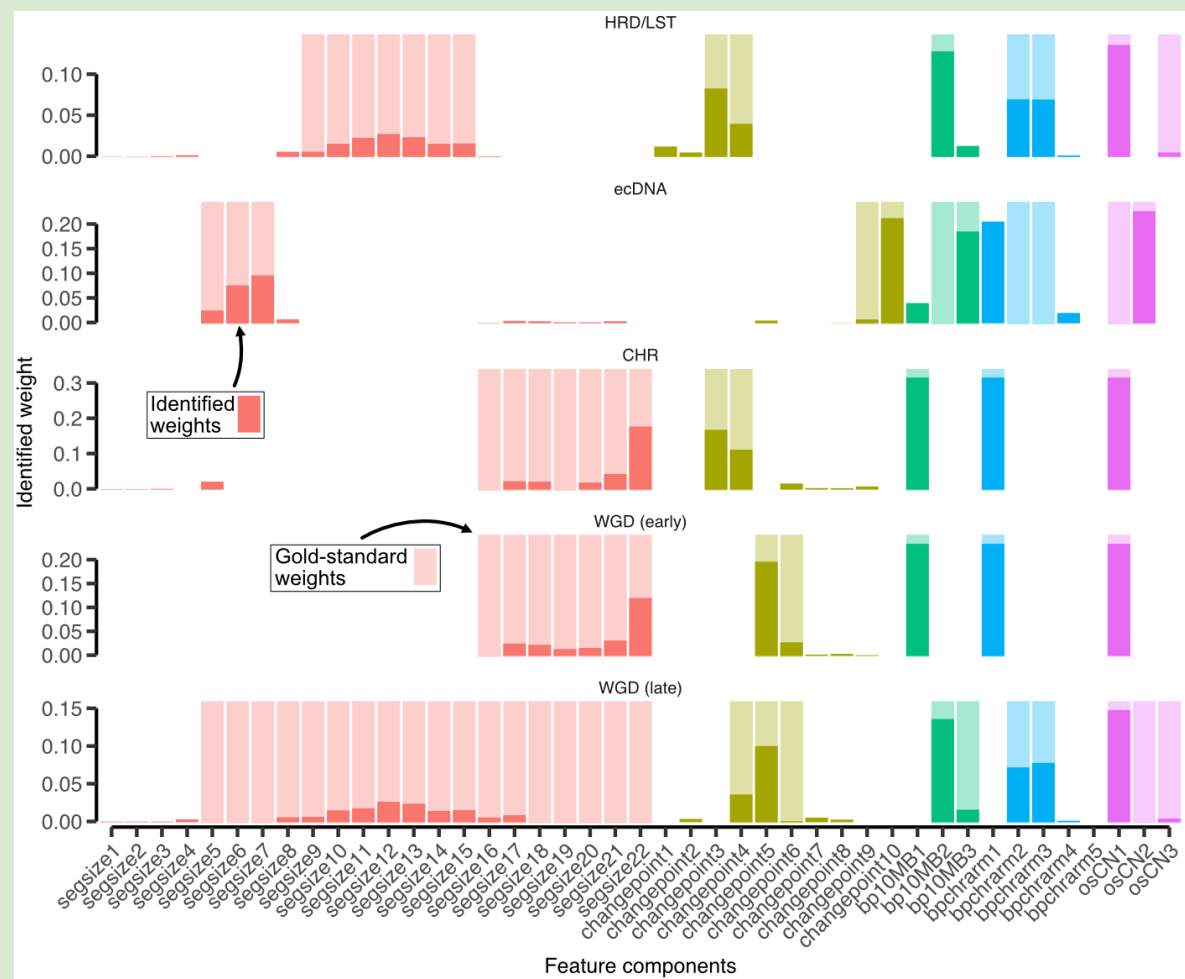
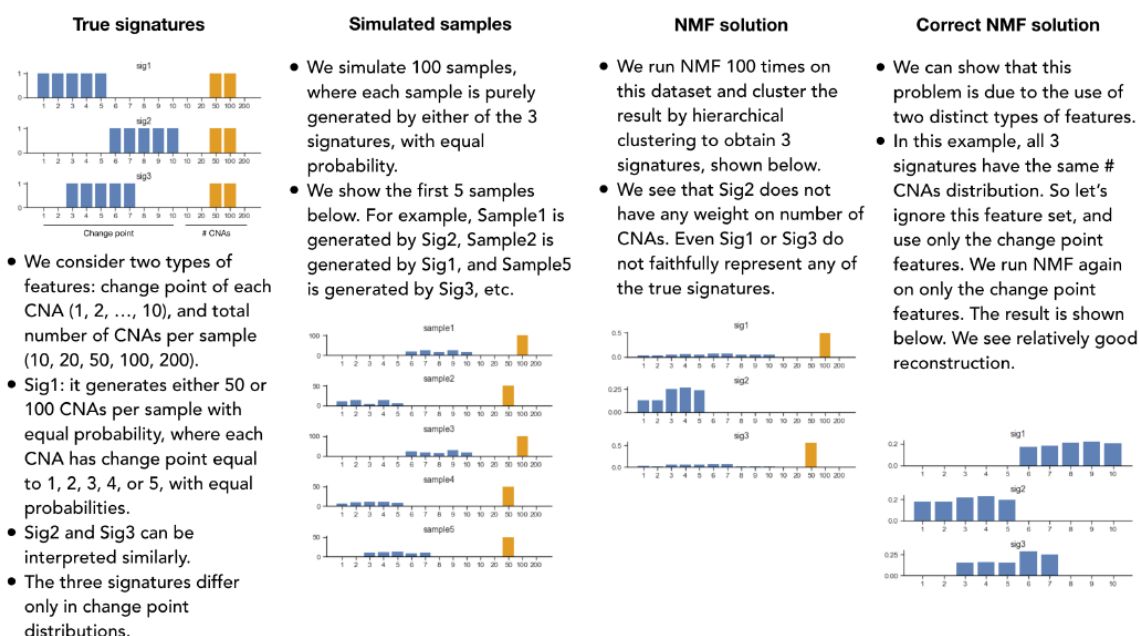


Fig. R9: Gold-standard and signatures from the signature interpretation matrix of five the mutational processes modelled for the simulation of genomes. Shaded bars: Components received either a full or half weight depending on the importance of this component to reflect the mutational process. Fully coloured bars: Signature interpretation values.

4. Since the authors did not, we have performed a simulation study below to demonstrate the problem that the proposed method has. We believe that this issue is also present in the earlier publication Macintyre et al., where it may be more difficult to catch because there were not as many obviously problematic signature distributions. See Point A for possible solutions that would address this issue.



Response: The reviewer presents a simulation example where NMF fails to reconstruct input signatures and claims this to be a serious problem of our approach. However, below we show that their simulation is not a valid representation of our approach and suffers from a fatal design flaw that artificially causes the NMF to fail.

The problem with the reviewer's simulation is an incorrect feature encoding. They encode the number of copy number events as an ordinal feature (categories with an order) not a count-based or count-like feature (like our sum-of-posterior probabilities).

NMF cannot decompose an ordinal feature without extensions (see <https://arxiv.org/abs/2006.01034> for more information) and as such the signatures cannot be identified. We show that by converting the problematic feature to a count-based encoding, it can be used in conjunction with the changepoint feature and signatures can be correctly identified (**Figs. R10-12**). For full transparency we have attached to our submission an Rmarkdown file containing all code and analysis for this simulation.

Summary of simulation results

The Reviewer encoded CNA counts as either 50 or 100 (two yellow bars further highlighted by red arrow in **Fig. R10** left panel). This ordinal embedding is unorthodox, does not represent our method, and artificially causes the NMF to fail. Counts of 50 or 100 are mutually exclusive and no sample shows two yellow bars (**Fig R10** middle panel). The

NMF cannot reconstruct what it has never seen, and thus never has two yellow bars in a signature (**Fig R10** right panel) - a problem caused by the Reviewer's design, not our approach.

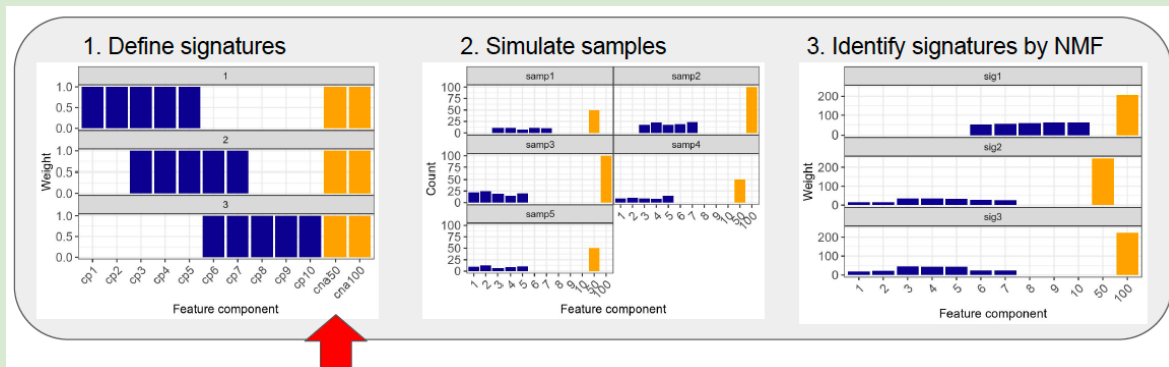


Fig. R10: Reviewer's simulation of a multi-feature approach

To fix the Reviewer's simulation, we use a single bar to represent CNA counts (one yellow bar and red arrow in **Fig. R11** left panel). The height of this bar varies from sample to sample (**Fig. R11** middle panel). The NMF picks up this variation and reconstructs the signatures (**Fig. R11** right panel).

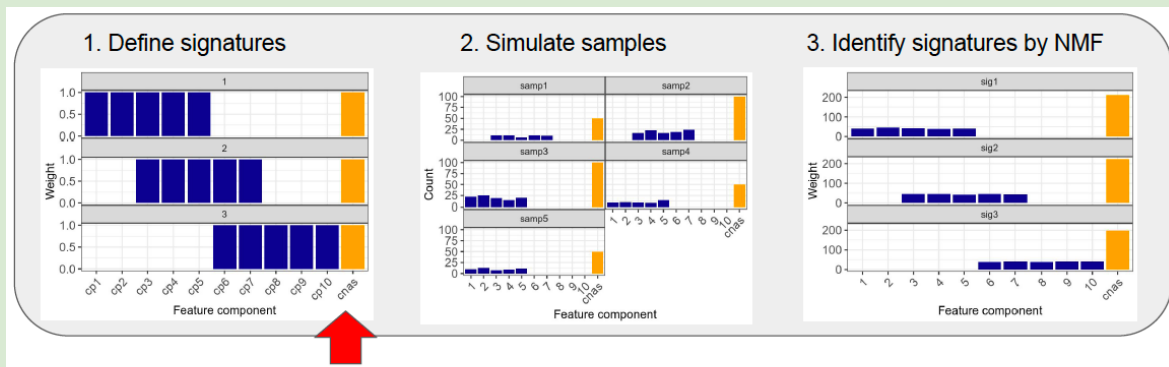


Fig. R11: Reviewer's simulation with a corrected feature encoding

This fixes the Reviewer's simulation, but is not yet representative of our approach (where each feature has more than 1 component).

Thus, we performed a more representative simulation with two features (changepoint and number of breakpoints per arm), each with multiple components (cyan and red bars in **Fig. R12** left panel). Samples now have weights on multiple components, allowing the NMF to see all kinds of combinations (**Fig. R12** middle panel). As a result, our approach reconstructs the input signatures well (**Fig. R12** right panel).

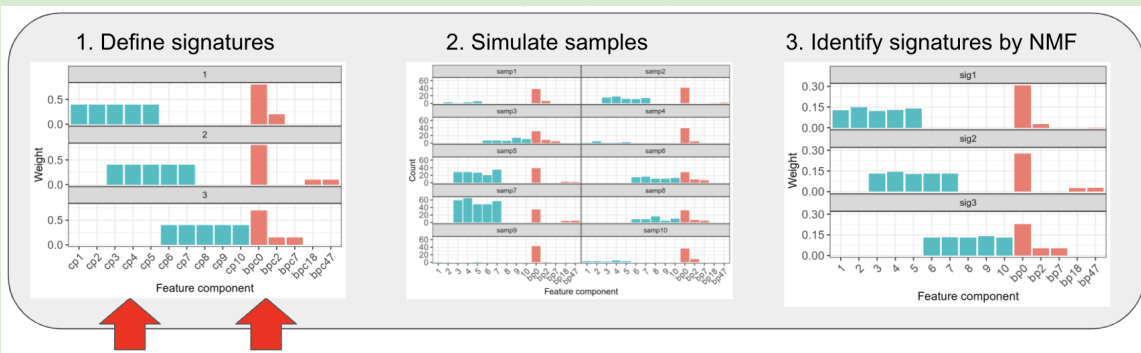


Fig. R12: Results of a version of the Reviewer's simulation representative of our approach

In addition, to better understand our approach on biologically known types of CIN, we added a more complex simulation of cancer genomes (as Reviewer 2 requested) to this document (**page 5**) and the manuscript (**Methods page 13**), which have all been described above.

Point C:

Authors state they evaluated all their correlations carefully, but several of the correlations for which we asked for further support were removed due to lack of support. We only mentioned selected examples to make the point that correlations are not always well supported. This does not mean that our prior list was exhaustive. The fact that a large portion of our questions resulted in removal of statements from the paper is a contradiction with the claim that the correlations stated are well supported

Response to Point C: This is an unusual and surprising criticism. At all stages we have engaged closely with all Reviewers' comments and used their feedback to improve the paper. By necessity, taking Reviewer comments on board will often result in changes to the paper. This is exactly how constructive peer review should work and hardly a weakness of our manuscript.

There are indeed a very small number of results, which only appeared in the very first submission, before the reviewers requested a major revision and the editor advised a complete refocusing of the manuscript. The best example is the survival analysis of detectable CIN, which was replaced by a survival analysis using the full set of pan-cancer signatures and is now much closer aligned with the focus of the paper (now section 4 of the manuscript focussed on platinum sensitivity).

However, in general the Reviewer's comment misrepresents the continuity of our key results: since the first submission, the methodology, the number of signatures, and their biological interpretation have stayed very stable.

Over time we improved the emphasis, tone and specificity with which we describe our results. One example is the association of signatures with replication stress. In the previous version of our manuscript we wanted to be as specific as possible with the potential type of replication stress associated with our signatures, e.g. R-loop formation. The Reviewer correctly pointed out that this specific association was only supported by a

single gene, and in response we toned down the specificity of the claim by only mentioning general replication stress, for which we have multiple gene associations. The link to R-loop formation has not been removed, however, and is still reported in the supplementary material (**Supplementary Material 5, Table 41**).

We cite here the relevant sentences from the 2nd and 3rd submission:

- 2nd submission:

“All three signatures were associated with increasing cell cycle score⁸⁶ (Supplementary Fig. 11) and APOBEC mutagenesis (SBS2 and SBS13, Supplementary Fig. 4a), suggesting replication stress as a putative cause^{87,88}. Significantly higher CX8 activity in the context of amplification and overexpression of U2AF1, ERBB3 and MAPK1 implied R-loop formation^{71,72}, increased origin firing^{67,68}, premature replication fork termination⁶⁸, or asymmetric forks⁶⁸....”

- 3rd submission:

“All three signatures were associated with increased cell cycle score⁸⁶ (Supplementary Fig. 11) and APOBEC mutagenesis (SBS2 and SBS13, Supplementary Fig. 4a), and had higher activity in the context of amplification and overexpression of U2AF1, ERBB3 and MAPK1, suggesting replication stress as a putative cause^{67,68,71,72,87,88}.”

In summary, we present a substantive and well supported piece of work that synthesises an extremely large and comprehensive amount of genomic data into a coherent framework to measure chromosomal instability.

References

1. Maxwell, K. N. *et al.* BRCA locus-specific loss of heterozygosity in germline BRCA1 and BRCA2 carriers. *Nat. Commun.* **8**, 319 (2017).
2. Jallepalli, P. V. & Lengauer, C. Chromosome segregation and cancer: cutting through the mystery. *Nat. Rev. Cancer* **1**, 109–117 (2001).
3. Popova, T. *et al.* Ploidy and large-scale genomic instability consistently identify basal-like breast carcinomas with BRCA1/2 inactivation. *Cancer Res.* **72**, 5454–5462 (2012).
4. Kim, H. *et al.* Extrachromosomal DNA is associated with oncogene amplification and poor outcome across multiple cancers. *Nat. Genet.* **52**, 891–897 (2020).
5. Davoli, T. & de Lange, T. The causes and consequences of polyploidy in normal development and cancer. *Annu. Rev. Cell Dev. Biol.* **27**, 585–610 (2011).
6. Li, Y. *et al.* Patterns of somatic structural variation in human cancer genomes. *Nature* **578**, 112–121 (2020).
7. Zhang, S. *et al.* Discovery of multi-dimensional modules by integrative analysis of cancer genomic data. *Nucleic Acids Res.* **40**, 9379–9391 (2012).
8. Yang, Z. & Michailidis, G. A non-negative matrix factorization method for detecting modules in heterogeneous omics multi-modal data. *Bioinformatics* **32**, 1–8 (2016).
9. Tan, V. Y. F. & Févotte, C. Automatic relevance determination in nonnegative matrix factorization with the (β)-divergence. *IEEE Trans. Pattern Anal. Mach. Intell.* **35**, 1592–1605 (2013).
10. Alexandrov, L. B. *et al.* The repertoire of mutational signatures in human cancer. *Nature* **578**, 94–101 (2020).
11. Alexandrov, L. B. *et al.* Signatures of mutational processes in human cancer. *Nature* **500**, 415–421 (2013).
12. Nik-Zainal, S. *et al.* Mutational processes molding the genomes of 21 breast cancers. *Cell* vol. 149 979–993 (2012).
13. Gaujoux, R. & Seoighe, C. A flexible R package for nonnegative matrix factorization. *BMC Bioinformatics* **11**, 367–1471 (2010).
14. Liao, Z. *et al.* Classification of extrachromosomal circular DNA with a focus on the role of extrachromosomal DNA (ecDNA) in tumor heterogeneity and progression. *Biochim. Biophys. Acta Rev. Cancer* **1874**, 188392 (2020).
15. Hung, K. L. *et al.* ecDNA hubs drive cooperative intermolecular oncogene expression. *Nature* (2021) doi:10.1038/s41586-021-04116-8.

- End of document -

Reviewer Reports on the Third Revision:

Referees' comments:

Referee #2 (Remarks to the Author):

The authors addressed my previous comments in a satisfactory manner (again).

Referee #4 (Remarks to the Author):

Summary

The aim of the study is to elucidate the extent, diversity and origin of chromosomal instability. The new CIN signatures have been brought into context with homologous recombination deficiency, mutations in driver genes etc and survival. The signatures can also be used to predict chemotherapy response and to possibly establish new drug targets.

The authors present an impressive body of work. And as a late reviewer, I have focussed on a few remaining issues of clarity. In my view, the authors have adequately addressed the reviewers comments from the previous rounds. I think this is an important manuscript, and a welcome addition to the field.

Comments:

A cartoon explaining how the interpretation matrix was constructed would be useful as it is difficult to follow how this was done and therefore the common signature weights might be more easily interpretable by more people

It is not obvious right away that segments with total copy number 2 are excluded from the feature composition. This is stated in the figure legend of Extended Data figure 2 but it might be good to also include in the main text. The explanation why this was done is very comprehensive.

SNV mutational processes largely reflect passenger alterations, and, as such, can be considered to more directly reflect the processes generating mutations, with less direct influence of selection. With regards to SCNA signatures, this is less clear - e.g. a process that generates occasionally generates homozygous deletions will not be clear as it is likely there will be strong negative selection against homozygous deletions. This should be made clear.

The authors do not introduce the nomenclature in the main text. On page 4, they begin talking about CX1, CX6 and CX14 - but this isn't very clear what they are referring to.

Methods page 6: It states that the method requires unrounded copy number. I'm unclear why unrounded copy number is required here?

Methods Page 14 - "The average number of affected chromosomes and number of CNAs produced

by a mutational process, as well as the segment size and induced copy number change, were taken from literature where possible and otherwise were based on personal experience” - this doesn’t sound like the most reproducible approach. Further elaboration would be good here, as this doesn’t sound terribly scientific.

Page 5 bottom: It is stated that CX2, CX3 and CX5 are associated with chromothripsis, but the association with signature CX3 is not significant.

Extended Data Figure 8b: A legend is visible saying S Short M Medium L Long, this part of the legend might not be necessary

Methods Page 4: The link to the GitHub repository is not working
<https://github.com/Martingales/CINsignatures>

Methods Page 6: “For 815 multisample patients with 2,021 samples” - in Extended Data Figure 1 it says 1255 samples from 526 multisample patients

Methods Page 11: “we set signature activities above 5% to 0 and multiplied the activity matrix by the signature definition matrix to derive a matrix with undetected signal” - “above” should be “below”?

Methods Page 8: “number of CNAs needs to be decoupled from the number of “events” - It would be helpful to define what is meant by an “event” here.

Methods Page 9 - Change point: Diploid segments were removed for the feature encoding but here diploid segments were assumed for the change point from the start of a chromosome. Furthermore it is clearly stated in a later section that diploid segments were skipped for the change point feature, but if this is the case some segments would not have neighbours. It would be good to clarify which segment was then used to calculate the change point.

Methods Page 20 - “This graph showed 12 independent clusters” - where is this graph?

Methods Page 29 - “A mutational process is a faulty biological process that produces a certain type of mutation.” - consider re-wording, mutations will reflect mutation and lack of repair, not solely a faulty biological process.

Methods Page 40 Paragraph 5 - it should be Supplementary Figure 3f

Supplementary Figure 2 - How was the cutoff chosen on which signatures to include in the table?

Supplementary Figure 37 - change “thesis” to paper/study/work etc.

Methods Page 15 - “We allowed for multiple mutational processes to be active on the same chromosome by nesting CNAs (smaller CNAs fully enclosed by larger CNAs), but did not allow CNAs to overlap by only one breakpoint.” - I think overlap would be easier to understand than ‘enclose’.

Methods Page 18 - “In summary, using only a single feature encoding does not allow the NMF to identify the underlying mutational processes. Multi-feature encoding allows us to study CIN in much greater detail and precision.” - This is quite a strong statement, given that the list of features used for single feature encoding was by no means exhaustive.

Author Rebuttals to Third Revision:

Referee #2 (Remarks to the Author):

The authors addressed my previous comments in a satisfactory manner (again).

Thank you!

Referee #4 (Remarks to the Author):

Summary

The aim of the study is to elucidate the extent, diversity and origin of chromosomal instability. The new CIN signatures have been brought into context with homologous recombination deficiency, mutations in driver genes etc and survival. The signatures can also be used to predict chemotherapy response and to possibly establish new drug targets.

The authors present an impressive body of work. And as a late reviewer, I have focussed on a few remaining issues of clarity. In my view, the authors have adequately addressed the reviewers comments from the previous rounds. I think this is an important manuscript, and a welcome addition to the field.

Thank you! We greatly appreciate your detailed assessment of our study and your suggested changes have greatly improved our clarity of presentation.

Comments:

A cartoon explaining how the interpretation matrix was constructed would be useful as it is difficult to follow how this was done and therefore the common signature weights might be more easily interpretable by more people

Thank you for this important suggestion. We improved the description of how we derive the interpretation matrix in two ways:

1. We rewrote the methods section describing the procedure to make it clearer (Methods page 30),
2. We added a figure summarising the different steps needed for constructing it (Figure R1, Supplementary Figure 54).

For convenience these changes are listed here:

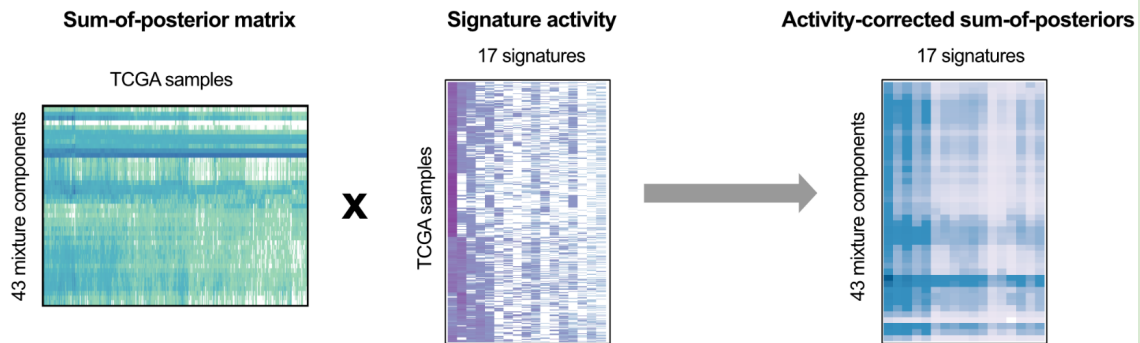
Text in the methods section: *"In detail, we multiplied the sum-of-posterior input matrix by the activity matrix, resulting in a signature by component matrix representing the activity-corrected sum-of-posteriors for each signature. We normalised the matrix to sum to 1, then added it to the original signature definition matrix, also normalised to sum to 1, to generate the unfiltered signature interpretation matrix (Supplementary Fig. 54).*

We then filtered the interpretation matrix to only display non-zero values where there was

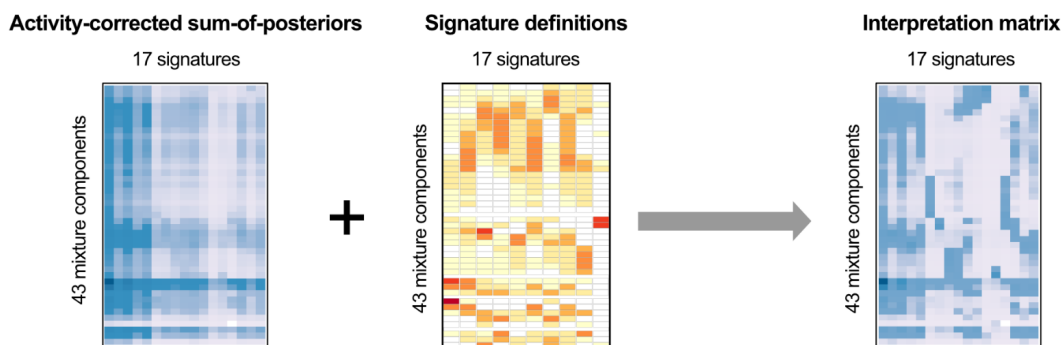
a positive, significant correlation between signature activity and component sum-of-posterior probability. This was determined by first selecting samples in which the signature activity levels were greater than 0, then performing a linear correlation between signature activity levels and the sum-of-posteriors of each component. In the interpretation matrix, we only reported components positively correlated with activity levels ($\beta > 0$ and Benjamini–Hochberg corrected p -value < 0.05)."

Derivation of signature interpretation matrix

a) **Step 1:** Calculate activity-corrected sum-of-posteriors per signature



b) **Step 2:** Get the unfiltered interpretation matrix



c) **Step 3:** Select components positively correlated with activity levels

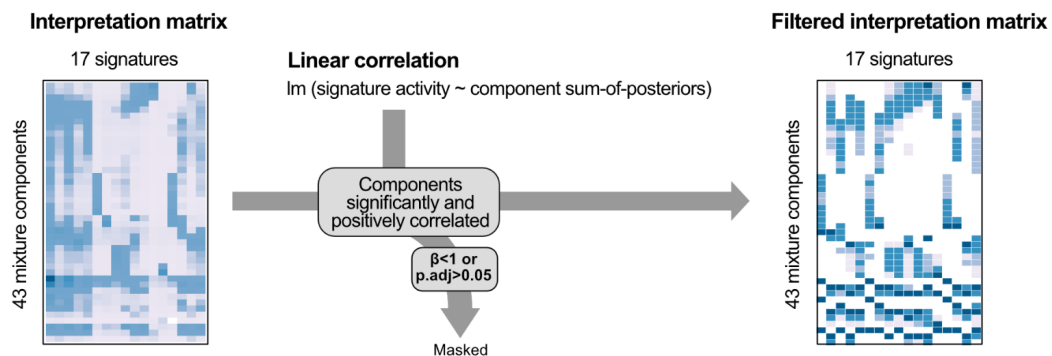


Figure R1: Overview of the signature interpretation matrix derivation

a: The sum-of-posterior input matrix is multiplied by the signature activity matrix to get an activity-corrected sum-of-posteriors matrix.

b: Both the activity-corrected sum-of-posteriors and signature definition matrix are normalised to sum to 1, guaranteeing they have equal contributions, and are then summed to generate the interpretation matrix.

c: To remove any signal from the sum-of-posterior contribution that does not positively correlate with signature activity, we performed a linear correlation between signature activity and component sum-of-posteriors and only retained weights which showed significant, positive correlation ($\beta > 0$ and Benjamini–Hochberg corrected p-value < 0.05) generating the final filtered interpretation matrix.

It is not obvious right away that segments with total copy number 2 are excluded from the feature composition. This is stated in the figure legend of Extended Data figure 2 but it might be good to also include in the main text. The explanation why this was done is very comprehensive.

This is a good suggestion. We added the following sentence to the paragraph about the feature composition on manuscript page 2:

“Only segments which deviated from the normal, diploid state were considered for the segment size and changepoint features.”

SNV mutational processes largely reflect passenger alterations, and, as such, can be considered to more directly reflect the processes generating mutations, with less direct influence of selection. With regards to SCNA signatures, this is less clear - e.g. a process that occasionally generates homozygous deletions will not be clear as it is likely there will be strong negative selection against homozygous deletions. This should be made clear.

The reviewer raises an important point here and we have added it to the Discussion section.

“In its current form, the signature methodology cannot account for selection pressures on CNAs. For SNV signatures, passenger mutations provide strong signals for detection. However, for CNAs the distinction between driver and passenger mutations is less clear. For example, large homozygous deletions are likely to be subject to strong negative selection, whereas other CNAs can be subject to strong positive selection. This has implications for the ability to detect signatures of CIN. Those processes that generate CNAs under positive selection will be easier to detect than those that generate CNAs under negative selection. Quantitatively, the relationship between signature detection and selection is not yet well understood and will depend on genomic background. For example, negative selection will be weaker in whole-genome duplicated samples (~50% of tumours) and in tumours which have lost their ability to sense DNA damage (e.g. via TP53 mutation).”

The authors do not introduce the nomenclature in the main text. On page 4, they begin talking about CX1, CX6 and CX14 - but this isn't very clear what they are referring to.

We added the following text in parentheses when first mentioning the 17 signatures on manuscript page 4 under “Putative causes underlying each signature”:

*“To determine putative causes underlying each of the 17 signatures (**named CX1 to CX17**), ...”*

Methods page 6: It states that the method requires unrounded copy number. I'm unclear why unrounded copy number is required here?

Rounding copy number values is a technique used to remove noise, but one of its drawbacks is that we lose the ability to see subclonal events. In our analysis, we carefully removed noise and thus all unrounded values that remain are likely to be indicative of subclonal changes. Keeping them in the analysis thus provides a much more comprehensive picture of the copy number landscape. By rounding copy number, neighbouring subclonal copy number segments may be artificially merged and therefore affect the ability to derive copy number signatures. For instance, segments of 1.5 and 2.4 copies would be rounded to 2 but most likely represent different copy number states and potentially biologies. Thus, our analysis of unrounded copy number avoids removing potentially important signal.

In order to clarify this point better, we added the following sentence to Methods page 6: *"We decided to use unrounded copy number states in order to avoid artificially removing signal (e.g. subclonal copy number changes) by rounding the copy number to the nearest integer state".*

Methods Page 14 - "The average number of affected chromosomes and number of CNAs produced by a mutational process, as well as the segment size and induced copy number change, were taken from literature where possible and otherwise were based on personal experience" - this doesn't sound like the most reproducible approach. Further elaboration would be good here, as this doesn't sound terribly scientific.

We agree with the reviewer that this was poorly worded and does not properly reflect the principles we used. We have added the following explanation to the relevant Methods section (Methods page 15):

"We could not find evidence in the literature for two parameters for our simulation: the changepoint for the LST/HRD process and the average number of CNAs per chromosome for the focal amplifications. Fortunately, our previous work on large numbers of high-grade serous ovarian cancer genomes (120 published¹, ~2000 unpublished) allowed us to qualitatively choose these parameters as means of Poisson distributions.

Firstly, the changepoint for the LST process was not mentioned in the original paper². In the ovarian genomes we observed that most genomes coming from BRCA1/2 germline deficient samples tended to have single copy number changes. Therefore, we assumed a single copy loss / gain changepoint, which has also been qualitatively reported in the literature³.

Secondly, for the average number of CNAs per chromosome produced by focal amplifications, Deshpande et al.⁴ stated that on average 5 rearrangements were involved per amplicon using deep WGS. However, in the ovarian genomes, where we used shallow WGS, we qualitatively observed that the number of events per amplicon was lower. This is expected given the lower resolution of the technology. The SNP6 array technology used in this study also has lower resolution than deep WGS, thus we reduced the number to 3."

Page 5 bottom: It is stated that CX2, CX3 and CX5 are associated with chromothripsis, but the association with signature CX3 is not significant.

The reviewer is correct. Data Fig. 8 and Supplementary Fig. 2 show the correct results but the main text was not well worded. Therefore, we (1) changed the statement in the main text by adding a clarifying text in parentheses and (2) added the following text to the Methods page 39:

“The link between impaired HR and chromothripsis is currently being studied^{5,6}. In our data, we find that IHR signatures CX2 and CX5 are correlated with chromothripsis but not CX3. This may indicate a mutual exclusivity between chromothripsis and HRD, however, further investigation is required to understand the complex relationships between CIN types.”

Extended Data Figure 8b: A legend is visible saying S Short M Medium L Long, this part of the legend might not be necessary

We have corrected the legend and removed the unnecessary elements.

Methods Page 4: The link to the GitHub repository is not working
<https://github.com/Martingales/CINsignatures>

The current address links to a private repo used for the revision process. The new and public repository will be located here: <https://github.com/markowetzlab/Drews2022>
This repository will contain all the scripts necessary to reproduce the signatures, figures and analyses presented in the paper. It will be opened at the publication date of the paper.

Methods Page 6: “For 815 multisample patients with 2,021 samples” - in Extended Data Figure 1 it says 1255 samples from 526 multisample patients

Both sets of numbers are correct and reflect the effect of different filtering steps.

Before excluding noisy samples, there were 815 multisamples patients with 2,021 samples (as previously mentioned on Methods page 6). After filtering, there were 526 multisample patients with 1,255 samples (as mentioned in Extended Data Fig. 1a).

We have simplified the description on Methods page 6 to match Extended Data Fig. 1a.

Methods Page 11: “we set signature activities above 5% to 0 and multiplied the activity matrix by the signature definition matrix to derive a matrix with undetected signal” - “above” should be “below”?

“Above” is actually correct. The reason is that this particular analysis looks at residual CNAs that can **not** be assigned well to a signature.

We have improved the method description to make this more clear (Methods page 7):

“The motivation here was to use our previous ovarian cancer signature method to estimate which CNAs in a sample can be robustly assigned as coming from a CIN signature and which CNAs cannot. The number of these residual, unassigned CNAs in a sample, can be used as an indicator of the minimum number of CNAs required to detect a signature of CIN, and thus establish an empirical threshold for detectable CIN. We started by calling ovarian copy number signatures on 539 high-grade serous ovarian cancer samples using the method described in Macintyre et al. to generate the input matrix, and YAPSA (R package, v1.12.0) and the signature definition matrix from Macintyre et al. to call signatures.

We considered signatures with activity above 5% in these samples to be robustly detected signatures of CIN and considered signatures below 5% as potentially representing residual CNAs that could not provide robust signal. To convert these <5% signature activities into an estimate of the number of assigned CNAs, we set signature activities above 5% in the activity matrix to 0 (leaving only the signal from the signatures with less than 5% activity), and multiplied this matrix by the signature definition matrix to derive an input matrix containing signal representing residual CNAs which could not be robustly assigned to any signature. For each sample and feature, we summed the remaining probabilities to estimate the number of residual CNAs. For each feature, we defined the quantile 0.95 of the distribution of residual CNAs. From this distribution of quantiles 0.95, we took the 3rd quartile (which is less volatile and more robust than the maximum) as the empirical threshold of the detectability of CIN which was 20 CNAs (Extended Data Fig. 1b).”

Methods Page 8: “number of CNAs needs to be decoupled from the number of “events” - It would be helpful to define what is meant by an “event” here.

In this section “event” does indeed have a very particular meaning. We have therefore improved our definition of this term.

On Methods page 8 we have added the following text:

“We consider an event to be the result of a single error in a regular biological process which leads to chromosomal instability (e.g an error during mitosis). The resulting event may involve one or more copy number changes depending on the type of error. For instance, a whole chromosome missegregation event of a so far unchanged chromosome results in a single copy number change. Whereas, a chromothripsis event, will result in many CNAs. Therefore, to adequately model the activity of different types of CIN, we should model activity in terms of numbers of events, not numbers of CNAs. Therefore the number of CNAs needs to be decoupled from the number of events.”

Methods Page 9 - Changepoint: Diploid segments were removed for the feature encoding but here diploid segments were assumed for the changepoint from the start of a chromosome. Furthermore it is clearly stated in a later section that diploid segments were skipped for the changepoint feature, but if this is the case some segments would not have neighbours. It would be good to clarify which segment was then used to calculate the changepoint.

We clarified the use of diploid segments for the changepoint feature on Methods page 10, both for the “Changepoint” and the “Additional details” section.

The paragraph in the “Changepoint” section now reads:

“The changepoint feature captures the relative copy number change or, more precisely, the difference in absolute copy number compared to the left neighbouring segment. For diploid segments no changepoint value is extracted. For non-diploid segments, the changepoint value is calculated with respect to the segment to the left (even if the neighbouring segment is diploid). If the first segment of a chromosome is a CNA, then an artificial diploid segment is used to calculate a changepoint value for this CNA, i.e. two is subtracted from the absolute copy number value.”

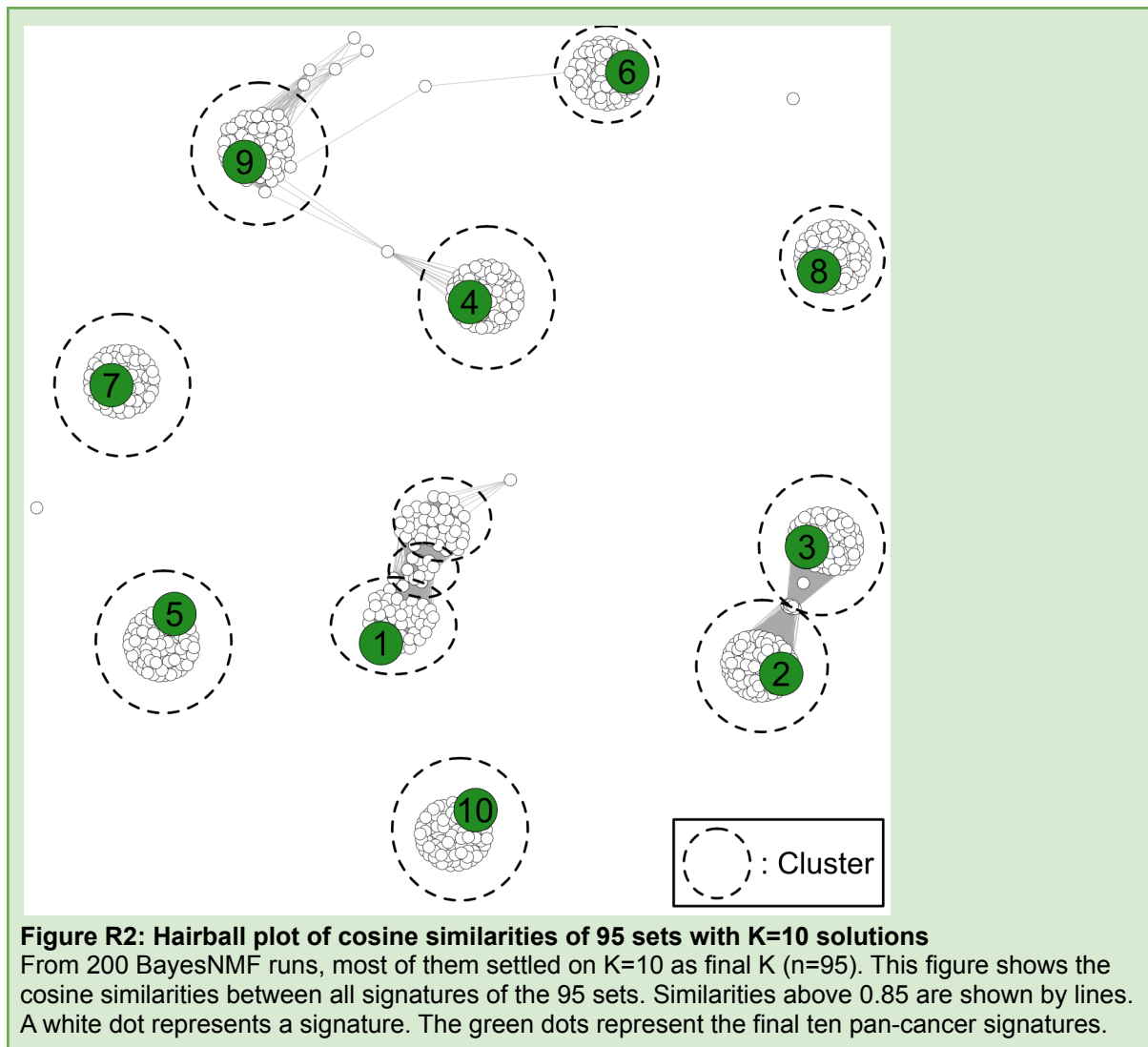
The paragraph in the “Additional details” section now reads:

“We collapsed and merged near diploid segments to a diploid state ($1.9 < x < 2.1$). During feature extraction, we then ignored these diploid segments for the segment size and the changepoint feature to avoid inflating these distributions in genomes with moderate CIN.”

Methods Page 20 - “This graph showed 12 independent clusters” - where is this graph?

This plot was accidentally omitted and we have now added the plot as Supplementary Fig. 53 and down below as Figure R2.

This plot shows the cosine similarities between the 95 sets of 10 pan-cancer signatures resulting from the BayesNMF. Every signature of the 95 sets is represented as a white dot, the final pan-cancer signatures as green dots. A line was drawn between two dots if the similarity was above 0.85, a value obtained from the literature⁷. Clusters are highlighted by circles.



Methods Page 29 - “A mutational process is a faulty biological process that produces a certain type of mutation.” - consider re-wording, mutations will reflect mutation and lack of repair, not solely a faulty biological process.

The reviewer is correct and we changed the sentence to:

“A mutational process is a faulty biological process that either produces a certain type of genomic change and/or prevents the genomic change from being repaired^{8,9}.”

Methods Page 40 Paragraph 5 - it should be Supplementary Figure 3f

Fixed.

Supplementary Figure 2 - How was the cutoff chosen on which signatures to include in the table?

The cut-off was chosen for each covariate based on how the results clustered in terms of significance and effect-size. Each set of results displayed two clusters - those which had strong significance and large effect-size, and those that had weak or no significance, and low effect-size. For display purposes we highlighted those results that (1) were significant and (2) had large effect sizes. We noticed that some of the significant results clustered with the small effect sizes and de-emphasized them. The cutoffs were different for different covariates because the distributions varied widely.

We have added this comment to Methods page 39.

Supplementary Figure 37 - change “thesis” to paper/study/work etc.

Wording changed to “study”.

Methods Page 15 - “We allowed for multiple mutational processes to be active on the same chromosome by nesting CNAs (smaller CNAs fully enclosed by larger CNAs), but did not allow CNAs to overlap by only one breakpoint.” - I think overlap would be easier to understand than ‘enclose’.

Wording changed to “overlapped”.

Methods Page 18 - “In summary, using only a single feature encoding does not allow the NMF to identify the underlying mutational processes. Multi-feature encoding allows us to study CIN in much greater detail and precision.” - This is quite a strong statement, given that the list of features used for single feature encoding was by no means exhaustive.

We agree with the reviewer and (1) have toned down the statement and (2) now emphasise that our conclusion depends on the set of features chosen (Methods page 19):

“In summary, using this set of features, a single feature encoding does not allow the NMF to identify the underlying mutational processes, whereas the multi-feature encoding does.”

References

1. Macintyre, G. *et al.* Copy number signatures and mutational processes in ovarian carcinoma. *Nat. Genet.* **50**, 1 (2018).
2. Popova, T. *et al.* Ploidy and large-scale genomic instability consistently identify basal-like breast carcinomas with BRCA1/2 inactivation. *Cancer Res.* **72**, 5454–5462 (2012).
3. Abkevich, V. *et al.* Patterns of genomic loss of heterozygosity predict homologous recombination repair defects in epithelial ovarian cancer. *Br. J. Cancer* **107**, 1776–1782 (2012).
4. Deshpande, V. *et al.* Exploring the landscape of focal amplifications in cancer using AmpliconArchitect. *Nat. Commun.* **10**, 1–14 (2019).

5. Piazza, A. & Heyer, W.-D. Homologous Recombination and the Formation of Complex Genomic Rearrangements. *Trends Cell Biol.* **29**, 135–149 (2019).
6. Nazaryan-Petersen, L., Bjerregaard, V. A., Nielsen, F. C., Tommerup, N. & Tümer, Z. Chromothripsis and DNA Repair Disorders. *J. Clin. Med. Res.* **9**, (2020).
7. Alexandrov, L. B. *et al.* The repertoire of mutational signatures in human cancer. *Nature* **578**, 94–101 (2020).
8. Volkova, N. V. *et al.* Mutational signatures are jointly shaped by DNA damage and repair. *Nat. Commun.* **11**, 2169 (2020).
9. Frigola, J., Sabarinathan, R., Gonzalez-Perez, A. & Lopez-Bigas, N. Variable interplay of UV-induced DNA damage and repair at transcription factor binding sites. *Nucleic Acids Res.* **49**, 891–901 (2021).

- End of document -

Reviewer Reports on the Fourth Revision:

Referees' comments:

Referee #4 (Remarks to the Author):

The authors have done an excellent point addressing all my comments. The only remaining issue is that that github is still 'private' - which means I can't really evaluate whether the figures can be reproduced etc.