nature portfolio

Peer Review File

Luminal breast epithelial cells of BRCA1 or BRCA2 mutation carriers and non-carriers harbor common breast cancer copy number alterations

Corresponding Author: Professor Samuel Aparicio

This manuscript has been previously reviewed at another journal. This document only contains reviewer comments, rebuttal and decision letters for versions considered at Nature Genetics.

Version 0:

Decision Letter:

17th Jul 2024

Dear Professor Aparicio,

Your Article, "Luminal breast epithelial cells of wildtype and BRCA mutation carriers harbor copy number alterations commonly associated with breast cancer" has now been seen by 3 referees. You will see from their comments below that while they find your work of interest, some important points are raised. We are interested in the possibility of publishing your study in Nature Genetics, but would like to consider your response to these concerns in the form of a revised manuscript before we make a final decision on publication.

We therefore invite you to revise your manuscript taking into account all reviewer and editor comments. Please highlight all changes in the manuscript text file. At this stage we will need you to upload a copy of the manuscript in MS Word .docx or similar editable format.

We are committed to providing a fair and constructive peer-review process. Do not hesitate to contact us if there are specific requests from the reviewers that you believe are technically impossible or unlikely to yield a meaningful outcome.

When revising your manuscript:

*1) Include a “Response to referees” document detailing, point-by-point, how you addressed each referee comment. If no action was taken to address a point, you must provide a compelling argument. This response will be sent back to the referees along with the revised manuscript.

*2) If you have not done so already please begin to revise your manuscript so that it conforms to our Article format instructions, available
here.
Refer also to any guidelines provided in this letter.

*3) Include a revised version of any required Reporting Summary: https://www.nature.com/documents/nr-reporting-summary.pdf
It will be available to referees (and, potentially, statisticians) to aid in their evaluation if the manuscript goes back for peer review.
A revised checklist is essential for re-review of the paper.

Please be aware of our guidelines on digital image standards.

Please use the link below to submit your revised manuscript and related files:

Link Redacted

Note: This URL links to your confidential home page and associated information about manuscripts you may have submitted, or that you are reviewing for us. If you wish to forward this email to co-authors, please delete the link to your homepage.

We hope to receive your revised manuscript within four to eight weeks. If you cannot send it within this time, please let us know.

Please do not hesitate to contact me if you have any questions or would like to discuss these revisions further.

Nature Genetics is committed to improving transparency in authorship. As part of our efforts in this direction, we are now requesting that all authors identified as 'corresponding author' on published papers create and link their Open Researcher and Contributor Identifier (ORCID) with their account on the Manuscript Tracking System (MTS), prior to acceptance. ORCID helps the scientific community achieve unambiguous attribution of all scholarly contributions. You can create and link your ORCID from the home page of the MTS by clicking on 'Modify my Springer Nature account'. For more information please visit please visit www.springernature.com/orcid.

We look forward to seeing the revised manuscript and thank you for the opportunity to review your work.

Sincerely,

Safia Danovi, PhD
Senior Editor, Nature Genetics
ORCiD: 0009-0007-7822-5479

Referee expertise:

Referee #1: scDNA seq

Referee #2: breast, somatic evolution

Referee #3: cancer genomics

Reviewers' Comments:

Reviewer #1:
Remarks to the Author:
Williams et al present a single-cell DNA analysis of pre-malignant BRCA mutant/WT tissues. They depth of the data enabled analysis of CNAs, separated by luminal and basal myoepithelial cells. Key results include the presence of stereotypical CNA patterns in pre-malignant lesions that are also seen in invasive carcinomas. The analysis is robust. The results are largely supported by the presented analyses. Limitations of the study are noted.

This study contributed to a body of work enhancing our knowledge about mosaic mutations/CNAs. An additional layer of novelty is the higher frequency of such CNAs in luminal over basal myoepithelial cells.

A weakness of the study is the small patient sample size.

Another potential limitation is the presence of CNAs in several patients with prior exposure to chemotherapy. This is not sufficiently discussed, but probably should be.

It is interesting that the authors find CNAs present in pre-malignant/normal tissues that are not present in tumor tissues. How do they interpret this? Are cells with these CNA patterns preferentially immunoedited? Or do they have a fitness disadvantage of some other sort?

Reviewer #2:
Remarks to the Author:
In their manuscript, Williams et al., performed single cell-based copy number alterations (CNAs) for a total of ~42,756 luminal and basal ductal cells that were obtained from normal breast epithelial tissues from 20 donors with and without a BRCA1/2 germline pathogenic allele using DLP+scWGS. Through the analysis, they demonstrated:

Summary of the key results
1. Overall, 2.69% of cells tested (0.1-5.9%) contained 1-4 aneuploid chromosome arms, including recurrent alterations, which most frequently affected 1q (gain), followed by 16q, 10q, 22q, and 7q (losses).
2. Aneuploid cells were age-related and more prevalent in luminal than basal cells (3.6% vs. 1.4%) and in BRCA carriers

than WT donors.
3. Luminal and basal cells showed distinct distributions of CNAs; gains of 1q, losses of 16q, 10q, 22q, and 7q were largely restricted to luminal cells and also found in breast cancer samples, where not enriched in TERT-immortalized cell line derived from WT breast tissues, suggesting that they are the results of tissue-specific selection.
4. The most common combinations of CNAs included 1q gain/16q loss and 1q gain/10q loss, which were enriched in luminal cells.
5. Allele-specific CNA analysis revealed that multiple independent events for these recurrent CNAs, suggesting strong positive selection of these CNAs in luminal cells.
6. Extreme aneuploids, defined by ≥9 aneuploid chromosome arms, were rare but were present in 0.1% of cells, of which 23 cells had a CNA profile similar to that of breast cancer samples, accompanied by loss of BRCA1 or BRCA1 allele and the TP53 locus at 17p.
7. These results support recent findings on the role of 1q gain in breast carcinogenesis and provide new insights into the early history of breast cancer development, particularly Luminal A cancers.

Originality and significance etc.
This is a very important study and provides invaluable knowledge to understand the evolution of breast cancer, especially during its early development. The results are solid, highly robust and reliable, based on the technology that enables high-throughput single-cell copy number analysis. The manuscript is clearly written. Statistics are sound. I enjoyed reading this manuscript. This reviewer has only a few minor questions.

Specific comments
1) Regarding the recurrent combination of 1q gain/16q loss and 1q gain/10q loss, is it possible to demonstrate the presence of the der(1;16) and der(1;10) chromosomes in the samples harboring these lesions by FISH analysis? Or are there any reports describing the unbalanced translocations resulting in gain of 1q and loss of 10q, 22q and 7q?
2) According to the recent report (ref#32), 1q gain/16q losses cells widely distributed in the large non-cancer area of the affected breast in patients with der(1;16)(+) cancer. Meanwhile, the frequency of cells harboring 1q gain/16q loss was only 0.23% on average. Therefore, it would be of interest to see the frequency of cells harboring this combination in non-cancer cells from the breast carrying der(1;16)(+) breast cancer/DCIS. This is related to the question of whether the development of breast cancer is limited to those individuals who carry widely expanded der(1;16) clones, even though der(1;16) clones could develop at low frequency in many women.
3) The authors showed a higher frequency of CNAs in BRCA1 and BRCA2 carriers. Is there evidence that only haploid loss of BRCA1/2 can lead to errors in double-strand break repair? In this regard, a recent study reported a role for glycolytic metabolic bypasses in impaired BRCA2 functon(DOI: 10.1016/j.cell.2024.03.006)? What about BRCA1? Also, der(1;16) is enriched in luminal A type cancer, in which BRCA1/2 germline mutations are less common. Can the authors made any comments on the apparent discrepancy between histology and the impact of BRCA1/2 mutations on CNAs including der(1;16)?

Reviewer #3:
Remarks to the Author:

Williams, Oliphant, Au et al present a very interesting study on the analysis of aneuploidy rates in normal epithelial cells from individuals carrying pathogenic germline variants in BRCA1/2 as well as those with wild-type BRCA1/2 alleles. Specifically, the authors have leveraged the single-cell WGS platform they have developed over the last few years, which although delivering low sequencing depth per cell, permit the detection of large copy number aberrations at the clone level -- and with less accuracy, at the single cell level also.

This study is timely and addresses one of the major gaps in our understanding in somatic evolution. Specifically, while there has been substantial effort devoted to elucidating the rates of small mutations (point mutations and to a lesser extent indels) in normal tissues, we still lack critical knowledge about the rates and patterns of somatic copy number aberrations and structural variants in normal tissues.

The manuscript presents interesting data in a very relevant system. However, there are multiple issues to be addressed before considering the manuscript ready for publication:

Major comments:

The authors observe recurrent copy number aberrations (e.g. chr1q loss). A recent study by the Kops lab reported that the rate of chromosome missegregation is correlated with nuclear chromosome locations [PMID: 35831506]. The per chromosome rates of chromothripsis reported for various cancer types also correlated with the estimated chromosome missegregation frequencies [PMID: 35831506]. I wonder if the authors tested whether the number of whole chromosome or chromosome arm aneuploidies detected in breast epithelial cells correlates with either chromosome missegregation rates reported by the Kops lab or the per-chromosome rates of complex genomic rearrangements in breast cancers. Was this the case? Due to the confounding effect of positive selection, this analysis might be most informative if performed using all aneuploidy events detected first, and then focusing on those non-recurrent ones. A discussion on these analyses would be an important addition to the manuscript to contextualize the findings reported further.

Related to the above, how do the authors interpret that the rate of gains in wild-type immortalized breast tissue cell line (hTERT cells) is uniform across chromosomes given the experimental data suggesting that chromosome missegregation errors are not uniform across chromosomes? I am referring to, for example, the following statement (lines 146-48):
“This suggests that chromosome arms have a relatively uniform susceptibility to CNAs and that the higher prevalence of CNAs within certain chromosomes in normal breast epithelium is a tissue- and cell type-specific process”

Gains of 1q in breast cancers have been recently linked with activation of the Notch pathway [PMID: 38388848]. These results and how the present study relates to previous work should be discussed further.

The effect size in Supplementary Figure 4 is minimal. Could the authors test if the higher similarity with breast cancer holds after performing eg. Bootstrap analysis?

Overall, the main text lacks information about which statistical tests were used to assess significance for the comparisons analysed. This information should be provided.

A discussion on the resolution that the data generated affords for copy number analysis should be included for readers not too familiar with the DLP+ method to follow the results and limitations of the study. This is critical for the interpretation of many results presented in the manuscript, such as the statements and analyses about the cancer-like copy number profile of individual cells. For example, some breast cancers are driven by high-level amplifications of oncogenes, would such events be detectable using the DLP data? Similarly, some complex karyotypes with minimal DNA loss arising from e.g. balanced chromothripsis would lead to complex karyotypes – would derivative chromosomes arising in this way be detectable by DLP? Similarly (and in relation in particular to the analysis of punctuated vs gradual evolution in Figure 4), what is the resolution of DLP+ to detect complex events, such as canonical chromothripsis events, characterized by small deletions? Such events are common in breast cancers and might play a critical role in cancer evolution.

In addition, what is the sensitivity and specificity for the detection of copy number gains and losses in single cells given the low sequencing depth per cell? Showing data in this regard for the technology but also in relation to the data specifically generated and presented in this study is critical to assess the conclusions of the manuscript.

Line 182: “Furthermore, this also suggests alterations on either allele have similar phenotypic effects.“
Is this conclusion supported by other studies or any functional data? The loss of multiple copies of the same chr in independent events might also be explained by a high background rate of losses of such chr with non-functional consequences. Could the authors comment on this?

The legend for Hom and Gained in Figure 3 is very hard to read as the colour shades used for purple and green are very similar to each other. A brief description on how the allele-specific copy number analysis was performed should be included in the main text.

Lines 215-219. An alternative hypothesis could be that only cells with a specific karyotypic configuration (even if very complex) leads to a clonal expansion of a size big enough to be sampled by the authors. This would be compatible with the gradual accumulation of events hypothesis. Could the authors comment on this?

Line 288: what previous studies support the following claim? “Alternatively, intermediate states may be more susceptible to immune surveillance leading to rapid elimination or require additional alterations to overcome LOH and undergo transformation.”

How do the results of this study relate to recent pan-tissue aneuploidy analyses generated using other technologies, such as PMID: 37904053?

Minor comments:
The legend in Figure 1a for cancer history is hard to read.

Information about the statistical tests used in Figure 1 is missing in the figure legend.

Figure 4. The sentence in the caption: “Fraction of the aneuploid cells that have X aneuploid arms“ should be corrected, as X is misleading (eg could be interpreted as chr X).

Gene names are not uniformly italicized across the manuscript or figures (e.g. Supplementary Fig 8).

Version 1:

Decision Letter:

Our ref: NG-A65451R

2nd Oct 2024

Dear Sam,

Thank you for submitting your revised manuscript "Luminal breast epithelial cells of wildtype and BRCA mutation carriers harbor copy number alterations commonly associated with breast cancer" (NG-A65451R). It has now been seen by the original referees and their comments are below. The reviewers find that the paper has improved in revision, and therefore we'll be happy in principle to publish it in Nature Genetics, pending minor revisions to satisfy our editorial and formatting guidelines.

As we're keen to move forward with acceptance such that we can co-ordinate publication with Nature, I've prepared some requests (see below) for you to start working on. In a few days, you'll receive a copy of these with some additional ones, but I think it would make sense for you to get started now.

If the current version of your manuscript is in a PDF format, please email us a copy of the file in an editable format (Microsoft Word or LaTex)-- we can not proceed with PDFs at this stage.

We are now performing detailed checks on your paper and will send you a checklist detailing our editorial and formatting requirements soon. Please do not upload the final materials and make any revisions until you receive this additional information from us.

Thank you again for your interest in Nature Genetics Please do not hesitate to contact me if you have any questions.

Sincerely,

Safia Danovi, PhD
Senior Editor, Nature Genetics
ORCiD: 0009-0007-7822-5479

Reviewer #1 (Remarks to the Author):

The authors have sufficiently addressed my previous comments with both clarifications, additional discussion of the normal CNA patterns, and new data expanding the cohort size with respect to patients and genomes.

I would suggest to provide additional references in the revised manuscript in particular as it relates to mosaic CNAs that are not shared with subsequent cancers.

Other than that, I find that the work provides a novel contribution to the field.

Reviewer #2 (Remarks to the Author):

In their revised manuscript, the authors have fully addressed the issues raised. This reviewer has no further concerns. Thank you for giving me an opportunity to review this elegant study.

Reviewer #3 (Remarks to the Author):

The authors have addressed my questions satisfactorily. I have not additional concerns. I welcome very much the inclusion of additional patients, which has made the study stronger overall.

EDITORIAL NOTES

Title
* We would suggest the following title:
Luminal breast epithelial cells of BRCA1 or BRCA2 mutation carriers and non-carriers harbor common breast cancer copy number alterations

If this is OK, please ensure that you update in the revised manuscript and also on the submission portal

Abstract
* Please cut your Abstract to 150 words (max)

Editorial summary
* We would suggest the following to accompany the paper - please let us know if it is correct and OK:
Single-cell DNA sequencing identifies recurrent copy number changes in healthy breast tissue from women with or without germline BRCA1 or BRCA2 mutations.

Word count
* Please ensure that you do not exceed 4400 words. Any additional text may be housed in the Supplementary note which should be cited at least once in the main text

Article structure
* You may have up to 8 main display items. Each should be provided as a separate files and cited at least once in the main text (in the right order). Please include all figure legends at the end of the main manuscript file

* You presently have 10 Supplementary figures. We'd suggest that you provide these as Extended data figures (you may include up to 10) which should be included as separate files and cited at least once in the main text (in the right order). Legends to Extended data figures should be provided in the Inventory of Supplementary Information

* Please provide Supplementary tables as a single Excel sheet with multiple tabs. Please ensure that each table has a title and is cited at least once in the main text

* Please ensure that you update all in-text callouts to reflect these changes

Main text
* Please reduce results subheadings to 60 characters

* Please define all abbreviations/acronyms on first mention

* Please remove claims of primacy (largest/biggest/first etc.) Novelty should be clear from context

* Please check gene nomenclature throughout

* Please reformat all in-text callouts as follows: Fig. 1a, Table 1, Extended data fig. 2c, * Supplementary table 3, Supplementary Fig. 5, Supplementary note

* Please change ‘gender’ to ‘sex’ throughout

* Please avoid 2-letter abbreviations - these should be spelled out in full

* Please check that your use of population descriptors/definitions is appropriate. For more information, please refer to https://www.nature.com/articles/s41588-024-01708-8 for more information

Figures, tables, and legends

* Please change all red/green colour contrasts

* Please define all elements in legend (abbreviations, acronyms, colours, symbols, lines etc.)

* Please check gene/protein nomenclature in all figures and legends

* Please add heatmap colour scales/keys to all relevant figures

* Please remove borders/shading/gridlines from all figures (unless they convey specific information)

* Please add datapoints to bar charts/violin plots/box-and-whisker plots

* Please ensure that all graphs have clearly labelled axes

* Please check that all legends have a title and that all figure panels have a legend

* Please check that all display items are cited correctly in the text (in the right order)

* Please provide tables in editable format

* Please increase font size to 6 pts (min) in all figures

* Please add scale bars to all images of tissues or cells. The corresponding length should be defined in the figure legend

* Please add chromosomal co-ordinates to all sequencing/browser tracks

* Please add MWM markers to blots/gels

* Please supply uncropped blots as source data

* Please ensure that data presented in a plot, chart or other visual representation format shows data distribution clearly (e.g.

dot plots, box-and-whisker plots, violin plots). When using bar charts, please overlay the corresponding data points (as dot plots) whenever possible and always for n ≤ 10. All box-plot elements (center line, limits, whiskers, points) should be defined in the legends accompanied by precise n numbers

* "Wherever statistics have been derived (e.g. error bars, box plots, statistical significance) the legend needs to provide and define the n number (i.e. the sample size used to derive statistics) as a precise value (not a range), using the wording “n=X biologically independent samples/animals/cells/independent experiments/n= X cells examined over Y independent experiments” etc. as applicable

* We strongly discourage deriving statistics from technical replicates, unless there is a clear scientific justification for why providing this information is important. Conflating technical and biological variability, e.g., by pooling technically replicates samples across independent experiments is strongly discouraged.

* Statistics, such as error bars, sigificance and p-values, cannot be derived from n<3 and must be removed from all such cases

* All error bars need to be defined in the legends (e.g. SD, SEM) together with a measure of centre (e.g. mean, median). For example, the legends should state something along the lines of “Data are presented as mean values +/- SEM” as appropriate. All box plots need to be defined in the legends in terms of minima, maxima, centre, bounds of box and whiskers and percentile

* The figure legends must indicate the statistical test used. Where appropriate, please indicate in the figure legends whether the statistical tests were one-sided or two-sided and whether adjustments were made for multiple comparisons. For null hypothesis testing, please indicate the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P values noted. Please provide the test results (e.g. P values) as exact values whenever possible and with confidence intervals note

* Please state in the legends how many times each experiment was repeated independently with similar results. This is needed for all experiments, but is particularly important wherever results from representative experiments (such as micrographs) are shown. If space in the legends is limiting, this information can be included in a section titled “Statistics and Reproducibility” in the methods section

* Please provide an Extended Data figure to graphically account for all FACS sequential gating/sorting strategies, or provide gating/sorting strategies in-figure. If the former, please be sure to indicate, in the legend, which gating panel(s) correspond to which FACS data panel(s) in the manuscript figures.(For an example, please see https://www.nature.com/articles/ncomms15067#supplementary-information)

Data availability
* Please note that EGAS00001007716 is not working

Code availability
* Please deposit all reproducibility code for the study

* Please mint a DOI for the code which should be cited in the Methods-only reference

Methods
* Please ensure that this is is ~3000 words. Any additional text should be housed in the Supplementary note which should be cited at least once in the Methods

* Please provide an Ethics statement that clearly affirms that you have obtained all necessary approval (the IRB should be named in full). If you did not need ethical approval, this must be stated together with the reason why

* Please provide a statistics and reproducibility section with general information on study design, how the statistical analyses of the data were conducted and general information on the reproducibility of experiments. Please include statements on how sample sizes were chosen, on whether any data were excluded (including why), whether randomization and blinding was used. If statements are negative please include the following statements: “No statistical method was used to predetermine sample size. “No data were excluded from the analyses”; "The experiments were not randomized"; "The Investigators were not blinded to allocation during experiments and outcome assessment”.

* Please add version numbers to all mentions of software packages

* Please provide sufficient details of the experiments in the Methods section such that they could be reproduced without reference to published papers. Use of the term "as described previously" is not encouraged

References
* Please ensure that all reference lists are formatted in the Nature style

* Please check if any preprints have been published and update your reference list accordingly

* Please remember to reformat your references before resubmission

Reporting summary
* Please ensure that this is updated to reflect/match all the above changes

Open Access This Peer Review File is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.
In cases where reviewers are anonymous, credit should be given to 'Anonymous Referee' and the source.
The images or other third party material in this Peer Review File are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.
To view a copy of this license, visit https://creativecommons.org/licenses/by/4.0/

We thank the reviewers for their enthusiastic feedback on our paper, please find below a point by point response. Major changes include the addition of 8 new samples, a comparison with missegregation rates published in the literature and an investigation of high-level amplifications and chromothripsis in our dataset.

Changes to figures and new figures included in this revision:

- All figures updated to include new data.
- **Figure 1e** includes heatmap from new dataset added for this revision (B1-6139)
- **Supplementary Figure 4:** Comparison between CNA frequencies and TCGA cancer subtypes updated to show distributions across bootstrapped samples.
- **Supplementary Figure 5**: Comparison of measured CNA frequencies with missegregation rates from Klaasen *et al.*
- **Supplementary Figure 7a:** Plot showing that allele bias measurements per chromosome arm are as expected for chromosome arm gains and losses.
- **Supplementary Figure 9:** Examples of high-level amplifications in single cells.

We include both a tracked changes version and a clean version. Line numbers in response refer to the tracked changes version.

Reviewer #1: 1
Reviewer #2: 5
Reviewer #3: 9

Reviewer #1:

Remarks to the Author:
Williams et al present a single-cell DNA analysis of pre-malignant BRCA mutant/WT tissues. They depth of the data enabled analysis of CNAs, separated by luminal and basal myoepithelial cells. Key results include the presence of stereotypical CNA patterns in pre-malignant lesions that are also seen in invasive carcinomas. The analysis is robust. The results are largely supported by the presented analyses. Limitations of the study are noted.

This study contributed to a body of work enhancing our knowledge about mosaic mutations/CNAs. An additional layer of novelty is the higher frequency of such CNAs in luminal over basal myoepithelial cells.

We thank the reviewer for their positive comments on our work.

A weakness of the study is the small patient sample size.

We have generated additional data from 8 donor samples (3 WT, 4 BRCA1 carriers and 1 BRCA2 carrier) as well as supplementing existing samples with additional cells. In total we have included an additional 5,678 genomes. All our main findings are robust to these additions. We observe enrichment of CNAs on chromosomes 1,7,10,16 and 22 that are largely restricted to luminal cells. See below for CNA heatmaps for these additional data (included in **Supplementary Figure 1**). Our dataset now consists of 48,434 genomes from 28 donors, up from 42,756 from 20 donors in our original submission.

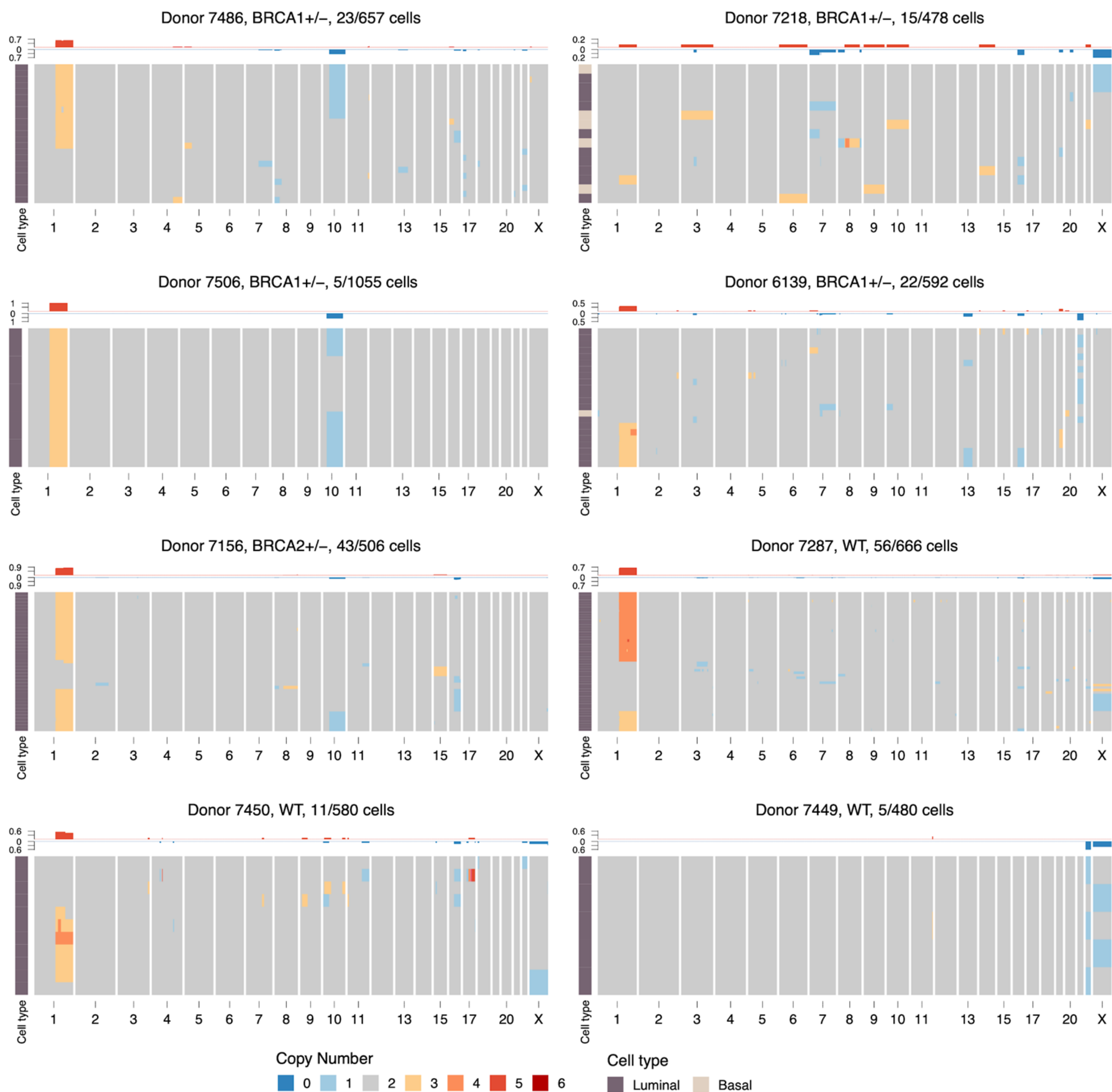

Figure R1 *Heatmaps of aneuploid cells from new data added during revision. title shows donor name, genotype and number of aneuploid cells out of total number of cells*

Another potential limitation is the presence of CNAs in several patients with prior exposure to chemotherapy. This is not sufficiently discussed, but probably should be.

While we observed no difference between CNA rate in donors with and without prior chemotherapy exposure (see **Figure R2** below) we agree this is important context and have added the following sentence (line 254):

> *"We note that we did not observe any enrichment of aneuploid cells in the small subset (n=4/28) of the donors that received chemotherapy due to prior cancer history."*

We also note that none of the new donor samples added for this revision have prior history of cancer or exposure to chemotherapy and show consistent patterns.

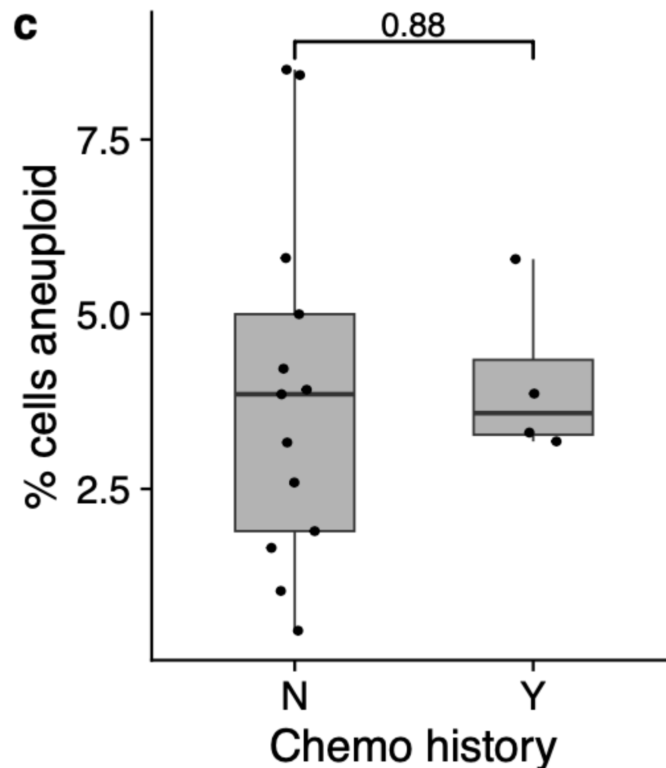

Figure R2 % cells aneuploid in donor samples with and without prior exposure to chemotherapy. p-value from Wilcoxon rank-sum test is shown at the top.

It is interesting that the authors find CNAs present in pre-malignant/normal tissues that are not present in tumor tissues. How do they interpret this? Are cells with these CNA patterns preferentially immunoedited? Or do they have a fitness disadvantage of some other sort?

We interpret the distribution of non-clonally expanded CNAs as a continuous foreground of mutational processes which affect replicating normal cells, but at a low frequency. This has been shown to be the case for other forms of mutation, for example point mutations, but has been understudied for CNAs. We have begun to address the question of whether the chromosomal pattern observed is due to chromosome specific mutational tendency, fitness effects, or both in this study. However the question of quantifying CNA fitness requires longitudinal measurements, ie. serial sampling, one which we hope to follow up on in subsequent studies. We note that overrepresentation of mutations in normal tissue relative to cancer has also been observed in other studies, most notably NOTCH1 in the esophagus and we have made reference to this in the discussion (line 637):

> *"We speculate that 7q loss may impede tumorigenesis, analogous to the recent observation that NOTCH1 mutations in the esophagus are positively selected in normal epithelium, but–underrepresented in esophageal cancers relative to normal epithelium[47]."*

We have also expanded our discussion on the possibility of immunoediting of cells carrying CNAs, see reply to Reviewer #3.

Reviewer #2:

Remarks to the Author:
In their manuscript, Williams et al., performed single cell-based copy number alterations (CNAs) for a total of ~42,756 luminal and basal ductal cells that were obtained from normal breast epithelial tissues from 20 donors with and without a BRCA1/2 germline pathogenic allele using DLP+scWGS. Through the analysis, they demonstrated:

Summary of the key results
1. Overall, 2.69% of cells tested (0.1-5.9%) contained 1-4 aneuploid chromosome arms, including recurrent alterations, which most frequently affected 1q (gain), followed by 16q, 10q, 22q, and 7q (losses).
2. Aneuploid cells were age-related and more prevalent in luminal than basal cells (3.6% vs. 1.4%) and in BRCA carriers than WT donors.
3. Luminal and basal cells showed distinct distributions of CNAs; gains of 1q, losses of 16q, 10q, 22q, and 7q were largely restricted to luminal cells and also found in breast cancer samples, where not enriched in TERT-immortalized cell line derived from WT breast tissues, suggesting that they are the results of tissue-specific selection.
4. The most common combinations of CNAs included 1q gain/16q loss and 1q gain/10q loss, which were enriched in luminal cells.
5. Allele-specific CNA analysis revealed that multiple independent events for these recurrent CNAs, suggesting strong positive selection of these CNAs in luminal cells.
6. Extreme aneuploids, defined by ≥9 aneuploid chromosome arms, were rare but were present in 0.1% of cells, of which 23 cells had a CNA profile similar to that of breast cancer samples, accompanied by loss of BRCA1 or BRCA1 allele and the TP53 locus at 17p.
7. These results support recent findings on the role of 1q gain in breast carcinogenesis and provide new insights into the early history of breast cancer development, particularly Luminal A cancers.

Originality and significance etc.
This is a very important study and provides invaluable knowledge to understand the evolution of breast cancer, especially during its early development. The results are solid, highly robust and reliable, based on the technology that enables high-throughput single-cell copy number analysis. The manuscript is clearly written. Statistics are sound. I enjoyed reading this manuscript. This reviewer has only a few minor questions.

We are grateful for the reviewer's positive assessment of our work.

Specific comments
1) Regarding the recurrent combination of 1q gain/16q loss and 1q gain/10q loss, is it possible to demonstrate the presence of the der(1;16) and der(1;10) chromosomes in the samples harboring these lesions by FISH analysis? Or are there any reports describing the unbalanced translocations resulting in gain of 1q and loss of 10q, 22q and 7q?

To clarify, the co-association of the gains and losses in single cells does not explicitly confirm these to be linked by translocation. However, as the reviewer suggests, our cells with 1q gain/16q loss might contain at least some der(1;16) given that these events have been reported in the literature as founder events especially in low grade ER+ breast cancers. With regards to

unbalanced translocations of chromosomes in combination with 1q-gain other than chr 16, we could not find any reports of recurrent alterations in cancer genomes. We note that we never observed 10q-loss without 1q-gain which does suggest that this is also the result of a derivative chromosome. FISH analysis would be very hard to deploy for this type of quantification because in contrast to the analysis of Nishimura *et al*, large clonal expansions of cells containing derivative chromosomes are likely rare; only 0.2% of cells in our study contain 1q-gain/16q-loss. Given the very low frequency of observed single cell events, observing sufficient single cells would require a substantial scan of tissues and patients, with the technical issue of section thickness creating hazards for false positive and negative detection of FISH spots. We have referred to the work of Nishimura et al and it is possible the 1q gain/16 loss cells in particular may harbour some der(1:16) as well as independent gains of 1 and 16 in single cells.

We have therefore added the following sentences to the discussion commenting on the likelihood of our aneuploid cells containing derivative chromosomes (line 610):

> *"Due to limitations in the resolution of our sequencing data, we were unable to conclusively determine whether 1q-gain/16q-loss events in our dataset were a result of der(1;16), although a statistical enrichment of reads with 1:16 split mapping was noted (**Supplementary Table 3**). It seems plausible that at least a subset of cells carry der(1;16) given the frequency of der(1;16) in breast cancers and the recent report on der(1;16) founder clones[37]."*

While our study was squarely focused on CNAs, we nevertheless investigated whether the sequencing data could be used to identify translocation breakpoints of putative derivative chromosomes in response to the reviewer's comment. Due to the rarity of these cells and the low coverage per cell, most cells lack the translocation breakpoint and furthermore, standard structural variant calling approaches cannot be used. However, we reasoned that cells with 1q gain/16q loss or 1q gain/10q loss may be enriched for split alignment reads where a portion of the read aligns to chromosome 1 and a portion to chromosome 16. To test this, we searched for such split alignments in cells harbouring 1q-16q and compared them to a set of randomly sampled diploid cells (n=634) that closely matched the coverage of the 1q-16q cells. We restricted our search to alignments (MAPQ>20) in regions in between the copy number transitions in each chromosome, these are in the vicinity of the centromere, where previous studies have reported that these translocations typically reside. Split alignment reads may be the consequence of chimeric sequence reads that arise during library preparation or be due to sequencing or mapping errors; we therefore expect a background rate of such alignments and thus tested for enrichment in the 1q-16q cells vs diploid. We did observe a modest enrichment of split alignment reads in 1q-16q cells that was borderline statistically significant (p=0.04675). Given the caveats with such analysis, notably that read mapping close to the centromere is error-prone we do not believe this is conclusive but nonetheless supports the notion that these cells contain derivative 1q-16q chromosomes. On the other hand, we did not observe any enrichment for split alignments in 1q-10q cells. See tables below for the statistics (these are provided in **Supplementary Table 3** in the manuscript).

Technological developments such as single-cell long read sequencing may be required to fully address this question.

	Number of cells	**Number of cells with 1q/16q split alignments**	**Proportion of cells with 1q/16q split alignments**
1q-16q cells	98	13	0.133
Diploid cells	654	50	0.0765
		p-value (one-sided proportions test)	**0.04675**

	Number of cells	**Number of cells with 1q/10q split alignments**	**Proportion of cells with 1q/10q split alignments**
1q-10q cells	129	1	0.008
Diploid cells	654	18	0.023
		p-value (one-sided proportions test)	**0.7498**

2) According to the recent report (ref#32), 1q gain/16q losses cells widely distributed in the large non-cancer area of the affected breast in patients with der(1;16)(+) cancer. Meanwhile, the frequency of cells harboring 1q gain/16q loss was only 0.23% on average. Therefore, it would be of interest to see the frequency of cells harboring this combination in non-cancer cells from the breast carrying der(1;16)(+) breast cancer/DCIS. This is related to the question of whether the development of breast cancer is limited to those individuals who carry widely expanded der(1;16) clones, even though der(1;16) clones could develop at low frequency in many women.

We agree this is an interesting question. However, investigating associations with cancer incidence would require a very large cohort of cancer and non-cancer patients age matched and with relevant patient material with associated mapping of der1:16, in conjunction. We note that Nishimura *et al*, reported in a small number of individuals clonal expansions of der1:16 in non-malignant breast tissue associated with patients with cancer (a different situation than sampling individuals who are cancer free). Their phylogenetic mutation timing analysis suggested in some cases the der1:16 preceded the cancer. However establishing a conclusive relationship lies well beyond the scope of this initial study on the prevalence and lineage distribution of rare CNAs in the breast.

3) The authors showed a higher frequency of CNAs in BRCA1 and BRCA2 carriers. Is there evidence that only haploid loss of BRCA1/2 can lead to errors in double-strand break repair? In this regard, a recent study reported a role for glycolytic metabolic bypasses in impaired BRCA2 functon(DOI: 10.1016/j.cell.2024.03.006)? What about BRCA1?

This is an interesting point. It has been suggested that single allelic loss of BRCA1 leads to intermediate phenotypes of telomere shortening, consistent with a published role for HRD in

telomere maintenance (for example, https://www.nature.com/articles/ncomms8505). We have added commentary to highlight this in the discussion.

Also, der(1;16) is enriched in luminal A type cancer, in which BRCA1/2 germline mutations are less common. Can the authors made any comments on the apparent discrepancy between histology and the impact of BRCA1/2 mutations on CNAs including der(1;16)?

We note that while BRCA1 mutations are indeed uncommon in ER+ cancers, about 50% of BRCA2 cancers are ER+. The relationship between BRCA1 status and mutation prevalence in our dataset is weak - we have revised the text to reflect that the nature of any possible differences in aneuploidy rates will require a much larger cohort. Our cohort was designed to estimate the signals present in women undergoing prophylactic mastectomy. Testing breast cancer associations in the general population would require a specific epidemiological cohort design that would be an entirely new study. In the discussion we reference literature showing that HRD haploinsufficiency leads to intermediate phenotypes of telomere shortening, which might be expected to contribute to higher rates of chromosome arm loss and/or fusions.

The results sections now reads (line 249):

> *"Aneuploid cells overall were more prevalent in luminal cells compared to basal cells (3.73% vs. 1.38%, p=0.001, Wilcoxon rank-sum test, Fig. 1h), and trended higher in BRCA carrier donors compared to WT: 3.63% in BRCA1 and 3.65% in BRCA2 compared with 2.45% in WT donors (p=0.12 and p=0.11 respectively, Wilcoxon rank-sum test, Fig. 1i)."*

We have also added this to the discussion (line 642):

> *"It has been suggested that haploinsufficiency in BRCA1/BRCA2 results in intermediate phenotypes of telomere erosion, and metabolic alterations which can promote aneuploidy[50,51]. While we observed a trend toward increased rates of aneuploidy in BRCA1/2 carriers, suggesting haploinsufficiency of BRCA1/2 may contribute to acquisition of CNAs, larger cohorts of samples will be needed to definitely demonstrate this."*

Reviewer #3:

Remarks to the Author:

Williams, Oliphant, Au et al present a very interesting study on the analysis of aneuploidy rates in normal epithelial cells from individuals carrying pathogenic germline variants in BRCA1/2 as well as those with wild-type BRCA1/2 alleles. Specifically, the authors have leveraged the single-cell WGS platform they have developed over the last few years, which although delivering low sequencing depth per cell, permit the detection of large copy number aberrations at the clone level -- and with less accuracy, at the single cell level also.

This study is timely and addresses one of the major gaps in our understanding in somatic evolution. Specifically, while there has been substantial effort devoted to elucidating the rates of small mutations (point mutations and to a lesser extent indels) in normal tissues, we still lack critical knowledge about the rates and patterns of somatic copy number aberrations and structural variants in normal tissues.

The manuscript presents interesting data in a very relevant system. However, there are multiple issues to be addressed before considering the manuscript ready for publication:

We appreciate the reviewers' positive feedback.

Major comments:

The authors observe recurrent copy number aberrations (e.g. chr1q loss). A recent study by the Kops lab reported that the rate of chromosome missegregation is correlated with nuclear chromosome locations [PMID: 35831506]. The per chromosome rates of chromothripsis reported for various cancer types also correlated with the estimated chromosome missegregation frequencies [PMID: 35831506]. I wonder if the authors tested whether the number of whole chromosome or chromosome arm aneuploidies detected in breast epithelial cells correlates with either chromosome missegregation rates reported by the Kops lab or the per-chromosome rates of complex genomic rearrangements in breast cancers. Was this the case? Due to the confounding effect of positive selection, this analysis might be most informative if performed using all aneuploidy events detected first, and then focusing on those non-recurrent ones. A discussion on these analyses would be an important addition to the manuscript to contextualize the findings reported further.

Related to the above, how do the authors interpret that the rate of gains in wild-type immortalized breast tissue cell line (hTERT cells) is uniform across chromosomes given the experimental data suggesting that chromosome missegregation errors are not uniform across chromosomes? I am referring to, for example, the following statement (lines 146-48):
"This suggests that chromosome arms have a relatively uniform susceptibility to CNAs and that the higher prevalence of CNAs within certain chromosomes in normal breast epithelium is a tissue- and cell type-specific process"

We thank the reviewer for reminding us of this interesting paper. We have performed the analysis suggested and do not find any correlation between CNA rates reported by Klaasen et al. (PMID: 35831506) and CNA rates in our study, see **Figure R3** below. We note that the Klassen et al study reported mostly from intestinal epithelium and retinal pigment epithelium. It is becoming apparent that both the pattern and rate of somatic chromosomal aneuploidy can be lineage specific. Lineage specific differences in nuclear chromosome location and lineage specific differences in cell polarity could account for the differences between cell types. We have therefore rephrased the highlighted text to take into account the observation that the CNA patterns we observe are different to both the cell-line we sequenced and previous reports of biased missegregation rates (line 419):

> *"The prevalence and pattern of alterations was also different to a recent study that showed that missegregation rates are influenced by nuclear chromosome locations[36] (**Supplementary Figure 5**). This suggests that the higher prevalence of CNAs within certain chromosomes in normal breast epithelium is a tissue- and cell type-specific process, potentially linked to lineage differentiation and/or epithelial cell orientation within a tissue context[37]."*

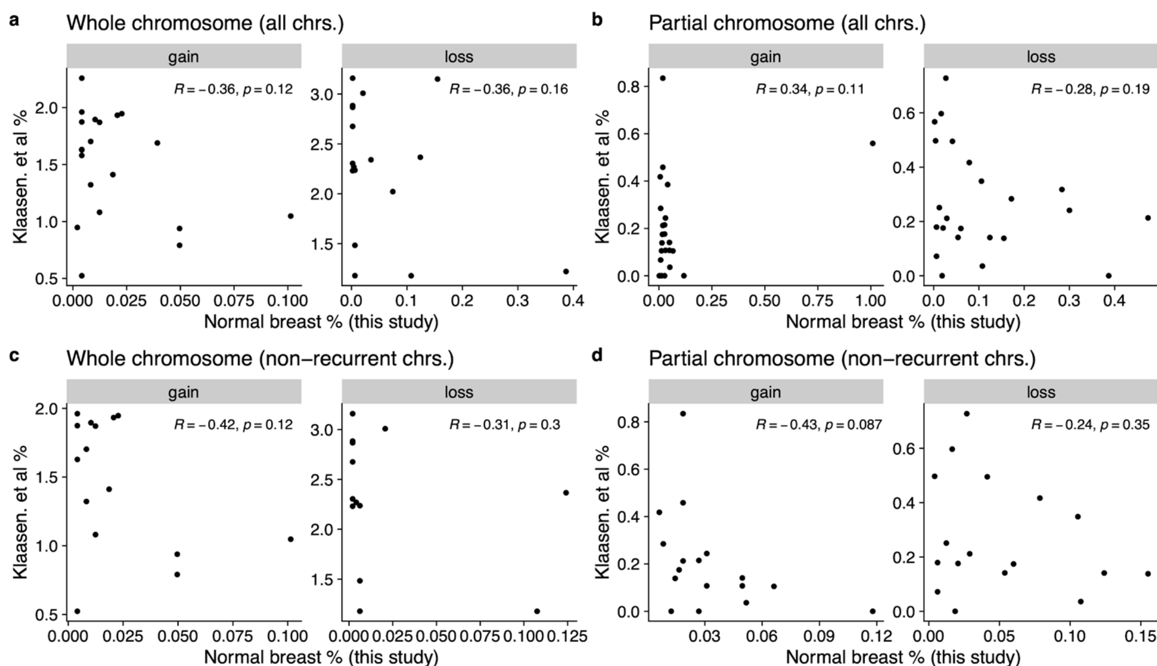

Figure R3
Aneuploidy rates per chromosome reported in Klassen et al. vs this study. Analysis was performed separately for: a) whole chromosome events across all chromosome b) partial chromosome events across all chromosomes c) whole chromosome events across non-recurrent chromosomes d) partial chromosome events across all non recurrent chromosomes. Non-recurrent chromosomes are all chromosomes after remove chromosomes 1,7,10,16,22 and X. Each plot shows the pearson correlation coefficient and associated p-value. Normal breast % are from all cells (luminal and basal cell populations).

Gains of 1q in breast cancers have been recently linked with activation of the Notch pathway [PMID: 38388848]. These results and how the present study relates to previous work should be discussed further.

We have added the following sentence and citation to the relevant section of the discussion (line 654):

“Activation of NOTCH signalling has also been suggested as a possible driver of 1q gains[56].”

The effect size in Supplementary Figure 4 is minimal. Could the authors test if the higher similarity with breast cancer holds after performing eg. Bootstrap analysis?

We agree the effect size here is minimal and bootstrapping to increase confidence is an excellent suggestion. To do this we sampled without replacement 80% of TCGA samples per cancer type and 80% of cells included in the study 25 times each and then compared the cosine similarity between all subsampled datasets. Our new figure (shown below as **Figure R4**) shows the distribution of cosine similarity across these subsampled datasets. We have replaced **Supplementary Figure 4** with this newer version. This again shows that breast cancer is the most similar cancer type, however we recognize that the effect size remains minimal hence our choice to only include this result as a supplementary figure and to caveat appropriately.

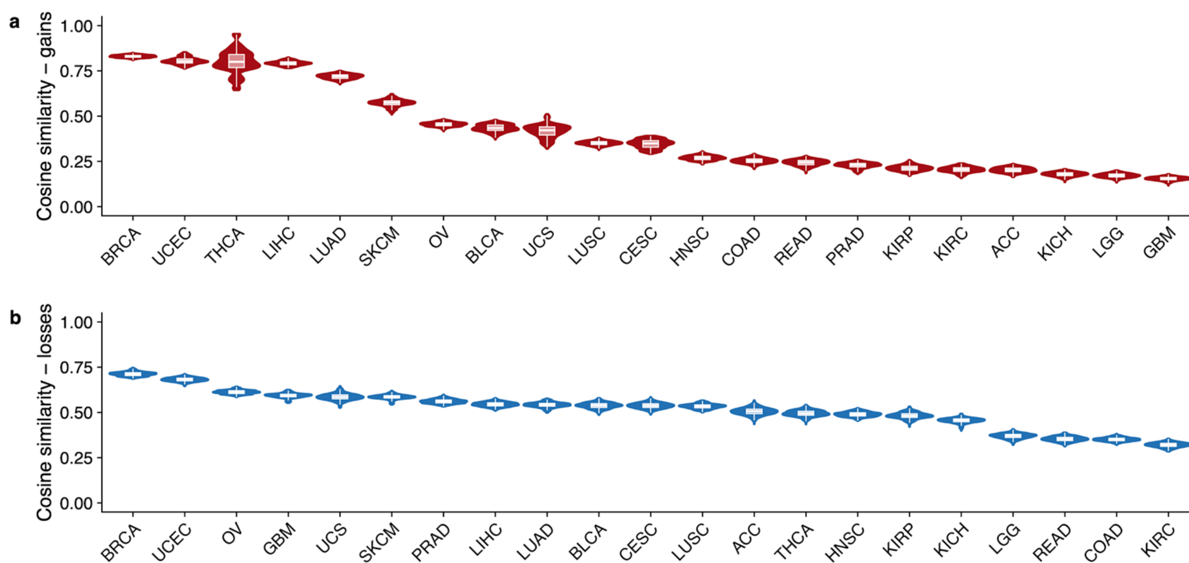

Figure R4
Cosine similarity between landscape of CNAs in scWGS of normal breast epithelia and TCGA subtypes for gains a) and losses b). Distribution is over subsampled datasets (80% of cells/samples were sampled with replacement and then cosine similarity computed across all combinations.

Overall, the main text lacks information about which statistical tests were used to assess significance for the comparisons analysed. This information should be provided.

We have now added information on the statistical test performed in all relevant areas of the text.

A discussion on the resolution that the data generated affords for copy number analysis should be included for readers not too familiar with the DLP+ method to follow the results and limitations of the study. This is critical for the interpretation of many results presented in the manuscript, such as the statements and analyses about the cancer-like copy number profile of individual cells. For example, some breast cancers are driven by high-level amplifications of oncogenes, would such events be detectable using the DLP data?

Similarly, some complex karyotypes with minimal DNA loss arising from e.g. balanced chromothripsis would lead to complex karyotypes – would derivative chromosomes arising in this way be detectable by DLP? Similarly (and in relation in particular to the analysis of punctuated vs gradual evolution in Figure 4), what is the resolution of DLP+ to detect complex events, such as canonical chromothripsis events, characterized by small deletions? Such events are common in breast cancers and might play a critical role in cancer evolution.

In addition, what is the sensitivity and specificity for the detection of copy number gains and losses in single cells given the low sequencing depth per cell? Showing data in this regard for the technology but also in relation to the data specifically generated and presented in this study is critical to assess the conclusions of the manuscript.

We thank the reviewer for this comment and appreciate that many readers will be unfamiliar with DLP+ and its capabilities. To address this comment we have therefore modified the text and included additional citations, QC'ed the total copy number calls and finally provided a deeper exploration of high-level amplifications and chromothripsis in single cells. We address these points in turn below.

1. Changes to text

We are able to detect copy number gains and losses that are of the order 1Mb; our standard bin size in our copy number calling pipeline is 0.5Mb. We have added a sentence to the beginning of the results section (line 193) outlining the resolution and the types of events that can be detected using it.

> *"At the sequencing depth utilized in this study, DLP+ detects copy number variations at 500kb to megabase-scale resolution, enabling the identification of whole chromosome and chromosome-arm aneuploidies, high-level amplifications, and complex genome rearrangements in single cells[25,31–34]"*

The above statement is supported by a number of papers which we reference and provide here:

1. Laks, Emma, et al. "Clonal decomposition and DNA replication states defined by scaled single-cell genome sequencing." *Cell* 179.5 (2019): 1207-1221.
Original DLP+ publication, showing it has superior QC properties to previous methods and can be used to estimate missegregation rates and identify high-level amplifications.
2. Salehi, Sohrab, et al. "Clonal fitness inferred from time-series modelling of single-cell cancer genomes." *Nature* 595.7868 (2021): 585-590.
3. Funnell, Tyler, et al. "Single-cell genomic variation induced by mutational processes in cancer." *Nature* 612.7938 (2022): 106-115.
Contains investigation of high-level amplifications using DLP
4. Williams, Marc J., et al. "Tracking clonal evolution of drug resistance in ovarian cancer patients by exploiting structural variants in cfDNA." *bioRxiv* (2024): 2024-08.
Contains investigation of high-level amplifications including chromothripsis using DLP
5. McPherson, Andrew W., et al. "Ongoing genome doubling promotes evolvability and immune dysregulation in ovarian cancer." *bioRxiv* (2024): 2024-07.
Used DLP+ to measure missegregation rates and genome doubling in ovarian cancer
6. Ng, Alvin Wei Tian, et al. "Disentangling oncogenic amplicons in esophageal adenocarcinoma." Nature Communications 15.1 (2024): 4074.

Uses DLP+ to investigate ecDNA amplicons in esophageal cancer

7. Kim, Minsoo, et al. "Single-cell mtDNA dynamics in tumors is driven by coregulation of nuclear and mitochondrial genomes." Nature Genetics (2024): 1-11.

2. QC of total copy number calls

In our initial investigation of this data we noted that most CNAs affected whole chromosome arms, so this became the focus of our analysis. We define an aneuploid cell as a cell with at least 1 altered chromosome arm (majority of the arm in a non-diploid copy number state). Given our resolution is of the order 1Mb and chromosome arms are much larger than this, we expect any errors to be negligible.

In order to confirm this and QC our results we have now used our allele specific copy number inference to show that allelic distributions aggregated across chromosome arms are strongly skewed in individual cells as expected (**Figure R5**). This is included as **Supplementary Figure 7a**. We note that this allelic information is not used to initially infer total copy number. Our allele specific inference assumes total copy number is correct and then attempts to infer the most likely allele specific state based on B-Allele frequencies derived from phased SNPs genotyped in single cells. We also note that Funnell et al. (PMID: 36289342) includes extensive benchmarking of allele specific copy number.

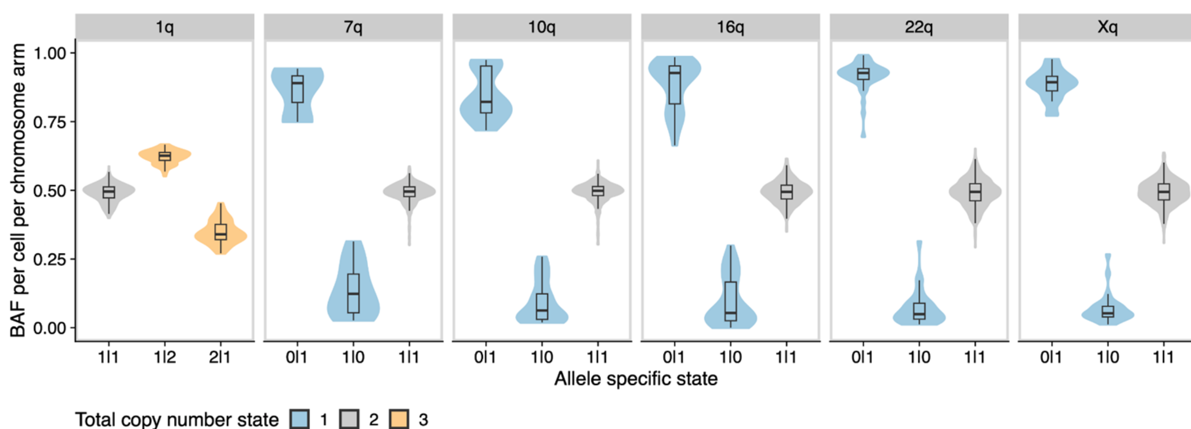

Figure R5
BAF distribution in chromosome arms across cells stratified by allele specific state. Non-diploid states are strongly skewed towards either 0.0 or 1.0 depending on which allele is gained/lost thus supporting the total copy number calls.

3. High-level amplifications

With regard to high-level amplifications, we agree that these events likely play a critical role in cancer evolution. This comment motivated us to look in more detail at high-level amplifications, specifically identifying any cells that had isolated high-level amplifications with limited additional copy number alterations (**Figure R6** and **Supplementary Figure 9** in the manuscript). A high-level amplification was defined as any segment greater than 2Mb in size but smaller than a chromosome arm that had copy number ≥5. These were very rare but we observed multiple examples across different samples including: 10 cells from sample B1-6548 that had the same complex amplification in chromosomes 11 and 19 in 13 cells, a cell from WT-7450 with

amplification of 17q23 and a cell from sample B2-21 that had an amplification of 6q2. Amplifications of *CCND1*, 17q23 and 6q21 are events commonly seen in breast cancers. This figure has been included in the manuscript as Supplementary Figure 9 and the following text added (line 559).

> *"We also found a rare subset of cells with high level amplifications (copy number > 5) with minimal additional CNAs including cells with gains of 17q23 and 6q21 (**Supplementary Figure 9**), common alterations found in breast cancers[40,41]."*

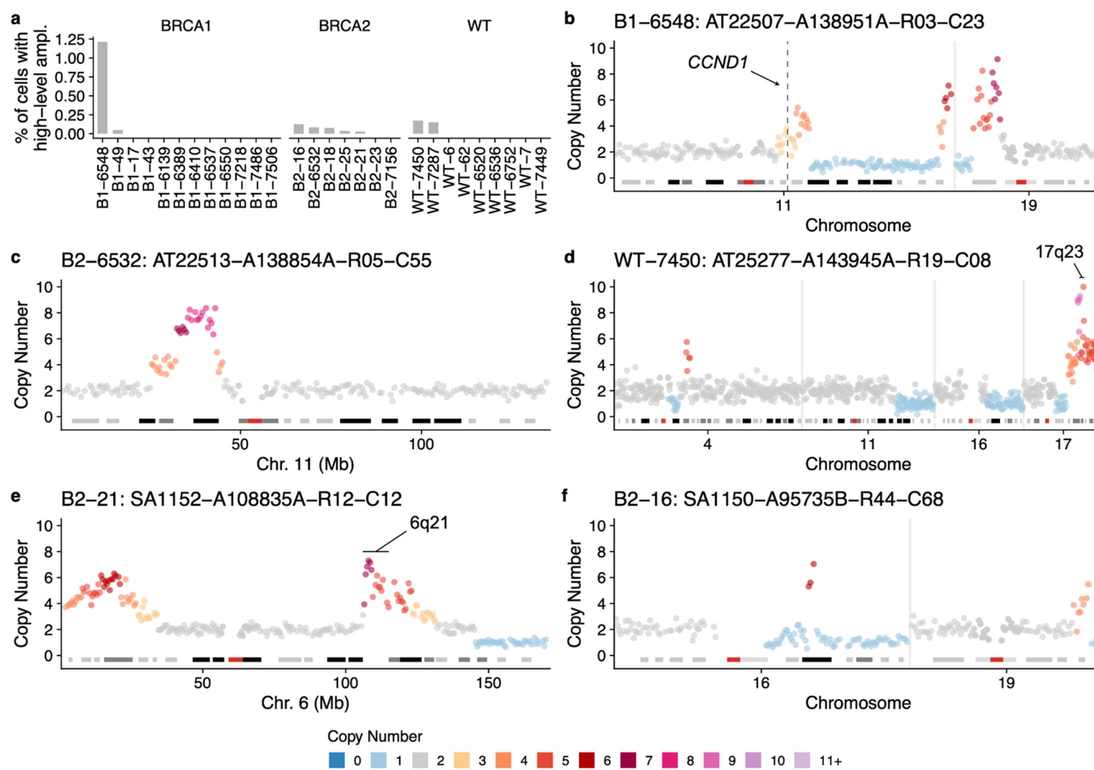

Figure R6
a) % of cells that have high level amplifications across all samples split by genotype. b)-f) 5 example cells with high level amplifications. Shown are chromosomes that have an amplification or some other CNA, all other chromosomes not shown are diploid. Arrows highlight interesting regions that are known to be drivers in breast cancer.

4. Chromothripsis

We did not find any evidence of chromothripsis in our dataset. We have used DLP+ to identify chromothripsis in an ovarian cancer dataset (see Williams *et. al.* Biorxiv), however given the resolution of DLP+, deletions smaller than 1Mb may not be captured. We have added the following sentence to the manuscript (line 562):

> *"We did not find any evidence of cells harboring chromothripsis, another a common event in breast cancer[42], although chromothripsis patterns characterized by small deletions (<100kb) would be difficult to detect in single cell genomes without clonal amplification using standard resolution DLP+."*

Line 182: "Furthermore, this also suggests alterations on either allele have similar phenotypic effects."

Is this conclusion supported by other studies or any functional data? The loss of multiple copies of the same chr in independent events might also be explained by a high background rate of losses of such chr with non-functional consequences. Could the authors comment on this?

This is a good point and we have now added a caveat to this statement (line 496):

> *"This suggests that maternal and paternal allele alterations may have similar fitness and phenotypic consequences resulting in convergence due to equivalent fitness or neutral effects."*

The legend for Hom and Gained in Figure 3 is very hard to read as the colour shades used for purple and green are very similar to each other. A brief description on how the allele-specific copy number analysis was performed should be included in the main text.

We have added the following sentence to the main text (line 436) and modified the colours to be more distinguishable.

> *"SIGNALS is a hidden Markov Model that uses a measure of allelic imbalance derived from phased germline SNPs that are genotyped in single-cells to infer the most likely allele specific profile given a cell's total copy number profile."*

Lines 215-219. An alternative hypothesis could be that only cells with a specific karyotypic configuration (even if very complex) leads to a clonal expansion of a size big enough to be sampled by the authors. This would be compatible with the gradual accumulation of events hypothesis. Could the authors comment on this?

We agree that this is another reasonable hypothesis we have modified the text as follows (line 698):

> *"Alternatively, copy number evolution may proceed in a stepwise gradual way but that intermediate states never reach a large enough size to be sampled in our study."*

Line 288: what previous studies support the following claim? "Alternatively, intermediate states may be more susceptible to immune surveillance leading to rapid elimination or require additional alterations to overcome LOH and undergo transformation."

We have now modified the text and added two citations (line 550):

> *"Alternatively, intermediate states may be more susceptible to immune surveillance leading to rapid elimination or require additional alterations to overcome LOH and undergo transformation, although it is unclear in general how large clones[61] need to be and the degree to which CNAs stimulate immune surveillance[62]."*

61. Gejman, Ron S., et al. "Rejection of immunogenic tumor clones is limited by clonal fraction." *Elife* 7 (2018): e41090.
62. Davoli, Teresa, et al. "Tumor aneuploidy correlates with markers of immune evasion and with reduced response to immunotherapy." *Science* 355.6322 (2017): eaaf8399.

How do the results of this study relate to recent pan-tissue aneuploidy analyses generated using other technologies, such as PMID: 37904053?

Several studies taking advantage of GWAS or GTEX studies have attempted to survey somatic CIN rates, we referenced in the original manuscript including Gao et al noted above. These studies have either not addressed breast epithelium or have been restricted in resolution (limited to clones with frequency > 20% in Gao et al) because of the need to infer CNA differences based only on bulk RNA-seq analysis. Gao et al only detected 8 events for breast tissue, we suggest this is underpowered. This latter study did observe patterns in esophagus, pituitary and adrenal with sufficient events to recognise the landscape background in tumours emerging from those lineages. Nishimura et al conducted targeted and deep WGS of clonal organoids and breast cancers and observe, similar to our study, gains in 1q and losses in 16 as relatively common events. All studies have emphasised that mutation rates are, in general, strongly cell/lineage dependent, as we have also noted for the breast.

In the introduction we have expanded on our discussion of Gao *et al* and other studies (line 149):

> *"A few studies have attempted to discover somatic CNAs in normal tissues[20–24] by reanalyzing bulk sequencing data but have been limited to blood or to detecting large clonally expanded populations carrying CNAs. For example, Gao et al. identified 8 breast samples harboring CNAs using bulk RNA-seq, which only permits detecting high-frequency alterations, thus precluding the ability to define the underlying generative process of CNAs in individual cells."*

Minor comments:
The legend in Figure 1a for cancer history is hard to read.

Legend has been moved to the bottom of the panel.

Information about the statistical tests used in Figure 1 is missing in the figure legend.

These have been added to the legend and throughout the text.

Figure 4. The sentence in the caption: "Fraction of the aneuploid cells that have X aneuploid arms" should be corrected, as X is misleading (eg could be interpreted as chr X).

Changed to fraction of the aneuploid cells that have **N** aneuploid arms

Gene names are not uniformly italicized across the manuscript or figures (e.g. Supplementary Fig 8).

All gene names have now been italicised.

Reviewer #1:
Remarks to the Author:
The authors have sufficiently addressed my previous comments with both clarifications, additional discussion of the normal CNA patterns, and new data expanding the cohort size with respect to patients and genomes.

I would suggest to provide additional references in the revised manuscript in particular as it relates to mosaic CNAs that are not shared with subsequent cancers.

Other than that, I find that the work provides a novel contribution to the field.

Reviewer #2:
Remarks to the Author:
In their revised manuscript, the authors have fully addressed the issues raised. This reviewer has no further concerns.
Thank you for giving me an opportunity to review this elegant study.

Reviewer #3:
Remarks to the Author:
The authors have addressed my questions satisfactorily. I have not additional concerns. I welcome very much the inclusion of additional patients, which has made the study stronger overall.

We thank the reviewers for their positive assessment of our work and helping us to strengthen the manuscript. We are not aware of any papers that focus on describing mosaic CNAs that are not shared with subsequent cancers. We believe we have discussed the most relevant studies that describe CNAs in normal tissue in the introduction and discussion.