nature portfolio

Peer Review File

Aneuploidy selects for the acquisition of driver genes in breast cancer

Corresponding Author: Dr Daniel Schramek

Any redactions in this file are there to maintain patient confidentiality, the confidentiality of unpublished data, or to remove third-party material.

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

Version 0:

Reviewer comments:

Referee #1

(Remarks to the Author)
In this study the authors focus on a basal-like breast cancer aiming to look at the selective advantage of specific CNA's and their underlying cancer driving genes. Methodology includes a novel, simple and valuable technique that could be used across several fields. They make use of CRISPR to knock out and/or activate genes in BLBC. The authors show that BLBC has increased CNA, and identify the cancer driver genes and consequent in vivo tumor initiation/rate of survival caused by these genes. They support this data by comparing mouse and human genomes from mammary tumors, where they found selection of the recurring aneuploidies in human and mouse models. Most importantly, this manuscript proposes that activation of key driver genes is sufficient to render the acquisition of CNA not necessary for tumor progression. They also describe Plgrkt as a novel oncogene in breast tumors. While the concepts are compelling, the conclusions are not fully supported by the data and the manuscript would benefit from additional experiments and analyses.

Main issues

Paper title is misleading— Their rationale for the title is that they pick a type of cancer that is high in CNA and look at the gains/losses and the genes that drive tumor initiation in their experiments. The title is tooo broad as it refers to all cancers while the work shown is mainly in breast cancer. They should have 'in breast cancer' in the title.

Do the sg/dgNTC bind the human genome? is it a library of negative controls or just a few sgRNAs?
They should use a library of sgRNA that target non coding (possibly introns) of genes to make sure the difference between targeting and control non targeting sgRNA is not the fact that sgRNA do cut while the controls do not (of course this is different in CRISPR A).

The authors need to do a histological characterization of the tumors obtained for all this experiments to see which breast cancer subtypes they obtain.

Most importantly for figure 3F that claims that CNA are not needed in the presence of strong cancer driver genes, I would like to see the WGS profile of all the tumors; they should be added to suppl and be reviewed. Can they also calculate the chromosome-level or arm-level aneuploidy in addition to fraction of genome altered?

In figure 4H can they show that this effect on oxygen consumption is also true for the mutant forms of Plgrkt?
More in general, Plgrkt was the major hit on chr9p, which is what this figure is based on. Fig 4.a shows that CRISPR KOALA is a robust system. However, Figure 4 could use more experiments showing the link between collagen, PGLK1, and mitochondrial dysfunction. They could supplement this with more extended figures.

Notably, the system used has several caveats with modeling and interpreting CNAs:
they are modeling not aneuploidy (more subtle dosage change) but complete KO or high over-expression; thus the conclusions are true only for drastic gene dosage changes of single genes

Minor issues

Manuscript could be improved in terms of language and description; there is no proper flow.

Can they check by qPCR or WB the extent of KO or overexpression for their candidates in figure 2?

Figure 4a: is it skin cancer or mammary cancer? legend is not correct

Did the authors estimate the “MOI” in vivo?

Referee #2

(Remarks to the Author)
In this manuscript, Al-Zahrani et al. describe the development and broad application of CRISPR-KOALA, a method that allows for the inactivation and activation of libraries of candidate genes in autochthonous mouse models of cancer using CRISPR-KO and CRISPR-A. They use this technique to study the impact of over 3000 genes that are found in common chromosomal arm deletions and amplifications in BLBC. This is an important system that allows the analysis of a large number of genes in parallel in an immunocompetent setting. The authors rightly point out that many of their hits are not found in cell line studies, underscoring the unique value of these types of in vivo (i.e. non-cell line) approaches.

In their parallel inactivation and activation study they start with two models that recapitulate genetic alterations that are common in human breast cancer, namely P53/PTEN and P53/RB1. Through their screen they identify 90 cancer driver genes, most of which do not have previously known roles in BLBC progression. They convincingly show that 5 hits from their screen act as tumor suppressors or oncogenes in vivo in single-gene validation experiments. Perhaps interestingly, alteration of these genes did not drive the tumors to converge on some aggressive cell state but rather each genotype generated unique effects. They also show (somewhat unexpectedly) that genetic modulation of any of these 5 genes reduces the selection for aneuploidy. They delve into the mechanism of tumorigenesis of one novel oncogene (Plgrkt) and show that Plgrkt upregulation is associated with mitochondrial reprogramming, specifically an inhibition of oxidative phosphorylation, in in vitro models. Their approach of inactivating and activating libraries of genes in a high-throughput and physiological system is novel and could be applied to the study of driver genes in other cancer types.

Major Points

1. It was not well described what they consider a tumor. Are these individual masses that form? I presume that they are as they only inject 20 mice per library = 80 mammary glands. Let’s image that is not quite the case and some tumors are not clearly clonal masses. From the methods, they only count the dominant sgRNA and any sgRNA that has at least 90% as many reads. This seems like a peculiar decision as an sgRNA that is within 90% likely means that the clonal tumor arose from an initial cell that got transduced with two lentiviral vectors (this also make sense that those might gain the most advantage). Meanwhile a second clonal tumors in that mass which has 50% or 10% as many reads to ignored. While their hits will likely be hits regardless of what arbitrary cutoffs they use, I would expect that more information could be glean with a more thoughtful analysis approach. They also entire lacks statistical in this analysis. The sgRNAs have >100x differences in representation in the pool (from their Sup Fig), I would have expected this representation to factor into a null expectation for sgRNA representation in tumors. It would have likely been helpful if they had sg/sgNTC within all of their pools, rather than separate (as most people do for in vitro CRISPR screens)

2. They only present merged data from their Pten/p53 and Rb1/p53 screens. These results should also be shown separately. Perhaps their data has insufficient resolution to compare across the Pten/p53 and Rb1/p53 screens but the data should still be presented.

3. Their approach suffers from only identifying the very largest driver genes as these greatly increase tumor growth and their sequencing-based analysis only focus on the largest clonal growth. There are likely many other genes in these regions that have more moderate effects that they are not uncovering. This might be something that they should acknowledge along with any other limitation, along with potential solutions. For example, would repeating these library screens without sgRNA/dgRNAs with the largest effects could help identify others with more moderate effects? dgRNA induced genes are not necessarily induced at the correct levels, many could be too low and that could lead to false negative in their screen etc.

4. Their phrasing of how inactivation of their new cancer genes (i.e. tumor suppressor and oncogene hits) leads to an overall reduction in aneuploidy is a little imprecise. They suggest on line 237-239 that alteration of a gene of interest would alleviate the pressure to accumulate that CNA. That make perfect sense and has been shown for before Pten in a mouse SCLC model (McFadden et al, Cell.). But then my interpretation, which might also be theirs, is that when they generate tumors with one of these genotypes of interest that tumor is fit enough on its own that cells that arise with aneuploidy have insufficiency selective advantage to emerge. Otherwise, it is difficult to image how many different single genes “suppress” aneuploidy.

5. In Figure 4, the authors convincingly disprove the hypothesis that Plgrkt drives tumor initiation through plasminogen activation and hypothesize that Plgrkt increases tumorigenesis through mitochondrial reprogramming. The authors show

separately that Plgrkt increases mammary tumor development in vivo and that 4T1 cells in vitro exhibit mitochondrial reprogramming but it would be more convincing if they could provide a link between these two observations (i.e MT disfunction in cancer cells from the autochthonous tumors and/or increased in vivo growth of 4T1 cells with Plgrkt over expression) .

6. While the screen is elegant, it would be of value if the authors could provide some additional mechanistic insights into how Plgrkt changes mitochondrial function and/or additional evidence that this gene is an oncogene in human tumors/cell lines.

7. While the authors do their screens in both the Pten/p53 and Rb1/p53 models they do not do any validation or follow-up studies in the Rb1/p53 model. Would this change any of their ultimate conclusions, I am not sure. It would seem reasonable to look at at least their top hit in the Rb1/p53 model and its effect on aneuploidy.

Minor points

1. Their screening platform allows gene inactivation and gene activation, however they don't really employ this as a "bidirectional" screen. For me that would mean that they inactivation and activate the same genes, and get reciprocal effects. Line 134 claims "simultaneous"(line 134) which is not exactly correct. They weren't in the same mice.

2. Line 250 I don't think that "discrepancy is the correct word here

3. I think that they should be more careful claiming 1000x coverage. Most cells have no ability to become a tumor at all ever. So it seems disingenuous to count those.

4. Line 179, what do they mean by tumor penetrance, it seems that essentially all mice get tumors at some point

5. I am not sure what Figure 4f is showing at all. Is there a fractionation control in there? Are the mutants behaving as expected or something?

6. While not the direct point of their study, the results of whole arm aneuploidies (or other large CNAs) is the change in expression of many genes and it's the summation of those that needs to lead to an overall change in fitness. If a loss contains a TSG but also many essential genes then that might not be selected for. Not sure if they can better articulate that their finding represent some of the major drivers but that they overall effect on fitness will be influences by many other genes in the same region. Perhaps altering combination of genes is a future direction for the field.

7. Line 358, "extent" not "extend"

Referee #3

(Remarks to the Author)
Genome instability and aneuploidy are important hallmarks of cancer. Specific arm level copy number alterations across tumor types and focal amplifications driving classic oncogenes have long suggested that these events play important roles in cancer formation and increasing fitness. Until recently however, deconvolution of target genes on arm-level CNA has been challenging due to the large number of affected genes.

This manuscript functionally characterizes oncogenic and tumor suppressor genes in the context of basal-like breast cancer (BLBC). The authors focus on 10 arm-level CNA events that are both specific to and prevalent in BLBC. They deploy their previously described CRISPR KOALA technique that multiplexes guides allowing for gene activation and inactivation using a single CRISPR Cas9 system by intraductal mammary injection in two genetic backgrounds (TP53/PTEN and RB/TP53). The sg/dgRNA libraries targeting 3,752 genes yielded 90 hits across 1,478 mouse tumors. The authors functionally characterize one of these hits, Plgrkt, in the murine ecoderm of PIKC3CA-H1047R animals and describe induced squamous cell carcinomas and changes of mitochondrial reprogramming.

Overall, this research paper is a solid body of work describing an approach to interrogate genes affected by copy number alterations in vivo and has several preliminary, yet interesting conclusions. However, considering that the technology (KOALA) and approach (in vivo positive selection functional genomic screens to validate genes targeted by copy number events) have been previously reported, the paper's ultimate impact is predicated on the depth of discovery and mechanistic follow-up, generalizability and therapeutic relevance.

These elements are more preliminary in the current manuscript, tempering enthusiasm.

Major points:

1. Specificity of the hits and models to individual CN events: for each of the 90 oncogenic/tumor suppressor hits, it would be helpful to see the enrichment of each targeting guide across each of the individual tumors and genetic backgrounds analyzed. Each individual tumor and genetic background harbors different arm level copy number gains and we cannot fully evaluate the specificity of the hits in regards to the central hypothesis (mutual exclusivity for copy number change), beyond the one example in Figure 3 (MDM4). It seems somewhat strange that tumors without the oncogene (e.g. METTL14) are

highly aneuploid while those with the oncogene are stable. This does not appear to fully validate the hypothesis that the target specifically removes the selective pressure for one specific chromosomal event. In a reciprocal manner, evaluating the arm-level CNAs at baseline across many individual tumors in each murine genetic background would help us understand the relevance of each murine model to the human arm-level data (e.g. a version of extended data 2b could be shown in main figure for both human and murine data).

2. Relevance of hits to human BLBC: Even though the authors provide a nice comparison between CNA in human and mouse BLBC samples, a more in-depth and side by side association to human BLBC samples (e.g. TCGA) would increase the relevance of this work and evaluation of hits. For instance, comparing transcriptional signatures derived in this paper to those of patients with similar gains and losses of human homologs could be helpful.

3. Translational relevance of hits: the authors convincingly demonstrate that this approach can identify novel genes that when perturbed can accelerate tumorigenesis. It could be powerful to demonstrate that modulation of these hits once a tumor is established can have the opposite effect on growth dynamics.

4. Depth of mechanistic follow-up. The follow-up data on the Plgrkt gene is performed in a distinct system than the screen and produces SCC tumors (not BLBCs). The data is interesting but preliminary for a paper in the intended journal. A comprehensive evaluation of Plgrkt oncogenesis and chr9p gain across cancer types, in human and murine data, and relating these changes to mitochondrial phenotypes is needed.

Referee #4

(Remarks to the Author)
Nature is committed to facilitate training in peer-review and to ensure that everyone involved in our peer review process is appropriately recognised. This reviewer co-reviewed one of the listed reports.

Version 1:

Reviewer comments:

Referee #1

(Remarks to the Author)
The authors have thoroughly addressed all of my previous concerns. The addition of non-genic control libraries, comprehensive WGS profiling with arm-level aneuploidy calculations, and the substantially expanded mechanistic characterization of PLGRKT considerably strengthen the manuscript. The cross-species transcriptional analyses and demonstration that PLGRKT protein levels in mouse tumors match those in human BLBC are particularly compelling. I am satisfied with the revised title and the acknowledgment of limitations regarding dosage sensitivity.
The authors should confirm that CRISPR-KOALA plasmids have been deposited at Addgene, the WGS data will be released upon publication and all raw sequencing data from the screens are deposited in an appropriate repository.

Referee #2

(Remarks to the Author)
The manuscript it greatly improved, both in scope and precision (of data and language). They provide new analysis and data from new pools, both of which were request quite specifically by the Reviewers. The provide additional single gene validation. They have provided sufficient additional discussion of the caveats of the approach. On one hand, they have done a huge amount of work; on the other hand, some experiments (e.g. having Safe targeting control sgRNAs) should have been included in the first place without the reviewers asking. Overall, this paper has employed novel and interesting in vivo approaches to uncover new cancer drivers.

Referee #3

(Remarks to the Author)
We would like to thank the authors for their thorough and thoughtful response to our, and other reviewers' questions, generating an impressive and substantial body of work. Demonstrating the technical feasibility of screening many hundreds of genes in vivo is certainly an important advance and the amount of new data, results and figures is quite vast.

The authors provided data and analyses to begin to address the five major points we raised, which we feel were partially addressed in the manuscript revisions. Additional analyses using MAGeCK, expanded data generation in multiple genetic background and whole-genome sequencing data sufficiently addressed concerns regarding the specificity and penetrance of the identified hits. Moreover, the penetrance of PLGRKT oncogene activity was also demonstrated in additional models and backgrounds.

The rebuttal clarifies the authors' belief that the technical approach described in this paper is a fundamentally novel technical advance beyond previous work. While we do not dispute this argument, we believe that CRISPR-KOALA is best categorized as an important iteration on existing methods. Of course, there are other (both experimental and computational) methods for deconvoluting driver genes in copy number regions as well and it is unclear what other approaches might have

discovered.

We think the paper is stronger than before. It is a tour-de-force that should be published but we remain unsure of its suitability for the current journal.

Please find our response to the rebuttal, these numbers correspond to our initial list of major points:

1. Reducing selective pressure for CN events. The revised data nicely demonstrate that the latency of tumor formation is an important factor in determining whether sufficient time has elapsed for subclones to become clonal (and therefore detectable via bulk sequencing). Given that each of the genes induces different biology (Ext Fig 7i) we would expect, for example, that the presence of PLGRKT reduces selective pressure on Chr19 events, but has no impact on other CN events (unless all possible events are functionally the same). Based on the updated Figure ED11a, certain CNAs do co-occur—for example, Chr19 and Chr15—with Chr19 hosting PLGRKT – so this would suggest there is likely to be complementary biology here. We simply remain unsure whether the observations here are best explained by (1) highly tumorigenic cells produce polyclonal tumors requiring less time for selection or (2) there is unique biology that obviates CN selection in meaningful ways. The analysis remains somewhat premature to distinguish.

2. Relevance to human BLBC: We appreciate and recognize the authors' efforts to demonstrate that the GEMMs used in this study closely resemble the transcriptional profiles of human BLBC in TCGA data. Additional cross-species data demonstrating in vivo transformation of high-penetrance hits in MCF10A cells further strengthens their findings. However, we find the genomics data in 3d and 3e preliminary.

3. Tumor maintenance. While it is true that a systematic validation of all driver genes as targets might be challenging, we believe it is possible to perform experiments to ascertain PLGRKT as a cancer target. Extended Data Fig 6L emphasizes that the magnitude of the PLGRKT effect may be somewhat modest in human cells, compared to other hits. This further emphasizes the importance of being very sure this is a robust human-relevant result.

4. Depth of mechanistic follow up: An expanded figure has been added to further deconvolute the mechanism by which PLGRKT enhances tumor formation, as described in earlier figures. However, this work stops short of elucidating the precise mechanism by which mitochondrial activity is increased; instead, it documents additional models in which this phenotype is observed, the co-localization of PLGRKT to mitochondria, and several assays used to measure mitochondrial activity.

5.Systematic analysis of PLGRKT in human cancer. The authors address oncogene penetrance in additional tumor types, noting that fully resolving this question would require substantial additional data generation and may fall outside the scope of this manuscript. We agree with this assessment and appreciate the inclusion of PLGRKT amplification data across additional tumor types, noting that several exhibit significantly higher gain/amplification frequencies than BLBC. The recurrent mutation data could be interpreted as somewhat counterintuitive, with a recurring frameshift mutation most likely resulting in loss of function of PLGRKT.

Referee #4

(Remarks to the Author)
I co-reviewed this manuscript with one of the reviewers who provided the listed reports.

Version 2:

Reviewer comments:

Referee #3

(Remarks to the Author)
Thank you for the comprehensive response.

Clearly a great deal of thought and work has gone into the mechanistic confirmation of PLGRKT - this concern is generally addressed.

On the question of whether the hypothesis that enhanced cellular fitness decreases selective pressure for certain/all copy number events has been definitely proven, we are neutral. We recognize and appreciate the authors thinking here.

On the question of whether the manuscript definitively describes how PLGRKT relates to human breast cancer, we would imagine this involving analyses of publically available datasets like TCGA and HTAN data, describing overexpression, and amplification of the PLGRKT gene and locus, quantification of the frequency of such events, delineation of the boundaries of the chromosomal alterations and mutual exclusivity analyses. The hypothesis the authors are making is that human tumors that overexpress PLGRKT are more aggressive and have fewer copy number events overall.

Without this type of data, the relevance seems moving in the right direction but somewhat preliminary. Overall, we are neutral on whether this is required for acceptance.

Referee #4

(Remarks to the Author)
I co-reviewed this manuscript with one of the reviewers who provided the listed reports.

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/

Point-by-point Response to Referees' Comments (2024-08-17371)

We sincerely thank the four reviewers for their thorough analyses and valuable suggestions, which have greatly guided us in improving our study. Over the past twelve months, we have gathered substantial new experimental evidence, including 27 new main figure panels and 59 new extended data figure panels - bringing the total to 41 main figure panels and 121 extended data figure panels - to further strengthen our work. We hope that, after reviewing our detailed point-by-point responses below, the editor and reviewers will agree that all concerns have been fully addressed. We believe the revised manuscript is now well-suited for publication as a research article in Nature.

There were three common comments that all reviewers and the editor raised:

1. Solidify the robustness of the screen methodology and analysis.

1.1. As suggested by the first reviewers, we re-analysed the screening data using a MAGeCK-based statistical model, which compares the gRNA abundance across all tumors to the initial viral library. This approach confirmed our high penetrance hits, and as hypothesized, its increased sensitivity to subclonal gRNA representation and potential cooperative effects between gRNAs identified 78 candidate lower penetrance driver genes. A key advantage of this approach is the detection of significantly depleted gRNAs, which showed the expected depletion of common essential genes as defined by DepMap. Underscoring the importance of performing these screens *in vivo*, this analysis also revealed that several genes classified as common essential in cell culture, such as *Ube2n, Csnk1a1, Mettl14, and Ino80*, instead function as potent tumor suppressors in the mammary gland. Please see new **Fig. 2d and Extended Data Fig. 6c and d and Supplementary Table 3**.

1.2. We also included new dgRNA and sgRNA control libraries targeting non-genic control loci to control for CRISPR-Cas9 effects. We performed these experiments in the Pten/p53- and Rb1/p53-mutant backgrounds (35 animals), which was the rate-limiting step of the revision as the median latency for the Rb1/p53-model is 427 days (not counting library design/cloning/viral preparation and breeding the experimental colony). Notably, these additional control libraries exhibited the same latency and tumor multiplicity as the original non-targeting control libraries, further corroborating our findings. Please see new **Fig. 1c.**

1.3. To rule out the possibility that the high-penetrance hits masked the emergence of additional, lower-penetrance ones, we cloned and re-screened the chr5q library excluding the major hit, *Csnk1a1*. This screen did not identify any new hits, showing that our primary screens were performed at or near saturation and confirming the *bona fide* identity of the lower penetrant tumor suppressors. Please see new **Extended Data Fig. 6e-h.**

1.4. We performed whole genome sequencing on another 19 tumors from the RB1/p53-mutant model (bringing our WGS dataset to a total of 89 samples), which nicely confirmed that engineering any of the identified driver genes significantly reduces aneuploidy within the resulting tumors. This further supports our findings that the loss or gain of these genes alone suffices to trigger tumorigenesis in both Pten/p53- and Rb1/p53-mutant backgrounds, identifying them as the true drivers selected for in aneuploid tumors. please see new **Fig. 4d and e and Extended Data Fig. 11e-j.**

1.5. We also provided additional KOALA experiments in *vitro*, in allograft and in GEMM models, showing that this technology allows for simultaneous screening of gene ablation and overexpression. Additionally, we have prepared all the plasmids required for *in vitro* and *in vivo* CRISPR-KOALA screening and have begun the process of depositing them at Addgene to serve as a public resource. Please see new **Extended Data Fig. 4h, Extended Data Fig. 6i-l and Supplementary Table 3.**

2. Provide more extensive validation and follow-up.

2.1. As suggested by the reviewers, we now successfully validated all hits (*Fbxw7*, *Mettl14*, *Csnk1a1, Plgrkt* and *Mrps18b*) not only in the Pten/p53 but also in the Rb1/p53 mouse model. Please see new **Fig. 3a.**

This was also a rate-limiting step, as the latency of this model is 427 days, and we needed to harvest frank tumors to be able to perform whole-genome sequencing on those (see below).

2.2. We also now successfully validated lower penetrance hits (*Adam10*, *Trim60*, *Fryl* and *Sox5)* and a new high penetrance hit (*Mrps18b)* for CRISPRa in the Pten/p53 mouse model. Please see new **Fig. 3a and b.**

2.3. We also assayed the 90 hits from our mouse KOALA screen in human MCF10A mammary epithelial cells in a simultaneous, tightly controlled KOALA allograft and *in vitro* screens. Next-generation sequencing of the MCF10A '*hit-containing*' tumors and cell lines revealed strong enrichment of several hits including *FBXW7, APC, SPRED1* and *PLGRKT* whose oncogenic potential only manifest *in vivo*. Interestingly, several common essential genes, including *CSNK1A1, INO80, METTL14 and UBE2N* were significantly depleted in MCF10A cell lines and/or tumors, but score as strong tumor suppressors in our autochthonous mouse models *in vivo*, confirming the context-dependency of our screens. Please see new **Fig. 2e and f and Extended Data Fig 6i-l.**

2.4. We added a whole new cross-species analysis, comparing the transcriptional profiles elicited by 4 of our top hits in our mouse models to the profiles of human TCGA basal-like breast cancer with the corresponding alterations - losses of *CSNK1A1, METTL14* and *FBXW7*, or gains of *PLGRKT*. This revealed a highly significant, cross-species conservation supporting the clinical relevance of our findings and indicates that these CNA driver genes help shape the transcriptional state of human tumors harboring the respective alteration. Please see new **Fig. 3f-i and Extended Data Fig. 10a-d**.

2.5. Further, we extensively validated and mechanistically studied one of our top hits, PLGRKT, to support human relevance of our findings and provide deeper mechanistic insights into PLGRKT's oncogenic function. We have now added a wealth of data that PLGRKT is in the mitochondrial intermembrane space, that it regulates mitochondrial dynamics under stressed conditions, leading to significantly better proliferation, and that survival under stressed conditions is all routed in increased ROS detoxification. Please see new **Fig. 5** and **Extended Data Fig. 12-14**, and **Supplementary Video 1-10**.

3. Contextualization regarding the driving roles of single genes versus cumulative roles of multiple genes.

We have added a lot of discussion about the driver genes, proved the existence of higher and lower-penetrance genes within the same chromosome arm and carefully contextualized how they overcome the need of aneuploidies. We also discuss that the extent to which these cancer driver genes cooperate and lead to cumulative effects within a given chromosome arm remains to be determined (likely using developing similar CRISPR KOALA tools that allow for expression of a larger number of gRNAs from a single array - e.g. Cas12 modeling).

In the following point-by-point response document, reviewer comments are presented in regular font, followed by our responses in green text.

Reply to Reviewer #1:

In this study the authors focus on a basal-like breast cancer aiming to look at the selective advantage of specific CNA's and their underlying cancer driving genes. Methodology includes a novel, simple and valuable technique that could be used across several fields. They make use of CRISPR to knock out and/or activate genes in BLBC. The authors show that BLBC has increased CNA, and identify the cancer driver genes and consequent in vivo tumor initiation/rate of survival caused by these genes. They support this data by comparing mouse and human genomes from mammary tumors, where they found selection of the recurring aneuploidies in human and mouse models. Most importantly, this manuscript proposes that activation of key driver genes is sufficient to render the acquisition of CNA not necessary for tumor progression. They also describe Plgrkt as a novel oncogene in breast tumors. While the concepts are compelling, the conclusions are not fully supported by the data and the manuscript would benefit from additional experiments and analyses.

Reply: We thank the reviewer for their insightful comments and positive assessment of our work. We are pleased they recognize CRISPR-KOALA's novelty and its broad application to identify key driver genes in recurrent CNA regions, determine how they drive breast cancer biology and - importantly - how those driver genes influence CNA acquisition. We also very much appreciate the constructive criticism and the many important suggestions below on how to improve our data to strengthen our major conclusions. We are pleased that we could address them all with (1) additional CRISPR-KOALA library controls, (2) additional WGS and histology from all our tumor groups, including the addition of the *Rb1*;*Trp53* knockout cohort, and importantly (3) a completely reworked section providing new data demonstrating and confirming the oncogenic role of PLGRKT as a regulator of several mitochondrial functions and ROS detoxification in both human and mouse models.

Main Points:

1. Paper title is misleading— Their rationale for the title is that they pick a type of cancer that is high in CNA and look at the gains/losses and the genes that drive tumor initiation in their experiments. The title is too broad as it refers to all cancers while the work shown is mainly in breast cancer. They should have 'in breast cancer' in the title.

Reply: We thank the reviewer for this comment and agree that our title, based on the data provided in the manuscript, should be focused on breast cancer. As such, we have changed the title, which now reads "*Aneuploidy selects for the acquisition of breast cancer-specific driver genes*".

2. Do the sg/dgNTC bind the human genome? is it a library of negative controls or just a few sgRNAs? They should use a library of sgRNA that target non coding (possibly introns) of genes to make sure the difference between targeting and control non targeting sgRNA is not the fact that sgRNA do cut while the controls do not (of course this is different in CRISPR A).

Reply: The reviewer raises an excellent point. We agree that controlling for the effects of Cas9-induced double-stranded break in our controls is a key parameter of CRISPR screening that we had initially overlooked. As the reviewer correctly points out, our sg/dgNTC library does not bind the genome. For transparency, we have updated the manuscript to read "*and control libraries harboring sgRNAs that do not bind the mammalian genome (3,778 sg/dgNTCs; Non-Targeting Controls)*" on Lines 145-146.

Furthermore, as suggested by the reviewer, we have now controlled for potential CRISPR-mediated DNA damage by creating a new control library targeting 2,600 non-genic regions of the genome. These non-genic regions were selected at random and evenly distributed throughout the genome based on previous work designed to measure off-target Cas9-mediated toxicity in high-throughput screens (Morgens, Wainberg, *et al., Nature Communications* 2017). Furthermore, the reviewer's comment prompted us to consider the optimal controls for our CRISPR-A screen. While there is no need to control for DNA damage-inducing events, recruitment of strong transcription complexes to the genome should be controlled for. To be thorough and address this potential concern, we also generated 2,600 dgRNAs targeting the same non-genic regions. We have added this line to the manuscript when defining what control libraries were used in the screen to read "*or target non-genic regions (2,600 loci; 5,200 sg/dgRNAs; Non-Genic Controls)*" on Lines 146-147. The results from these new controls strongly support our conclusions. We found no statistical differences in tumor onset or penetrance between these non-genic controls and our original non-targeting controls. Most importantly, all of our results regarding the chromosome libraries remained significant when compared against the more stringent

non-genic controls in both the *Pten*$^{-/-}$;*Trp53*$^{-/-}$ and the *Rb1*$^{-/-}$;*Trp53*$^{-/-}$ models. Please see new **Fig. 1c and Extended Fig 5f and g**:

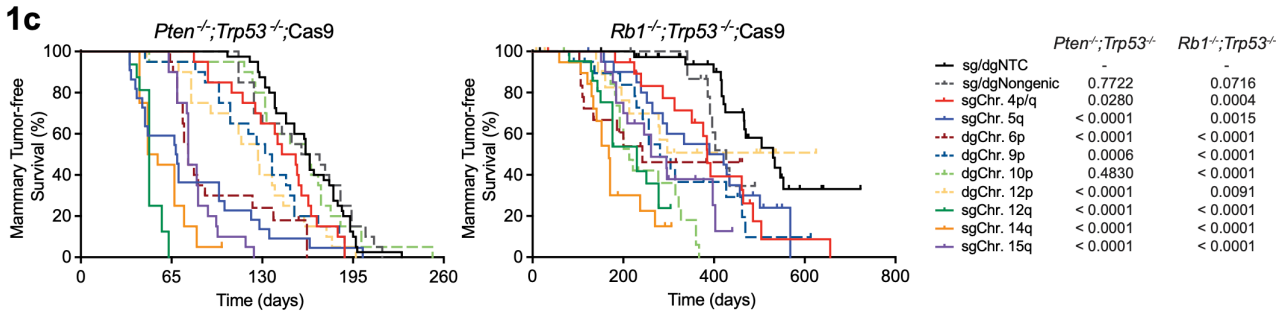
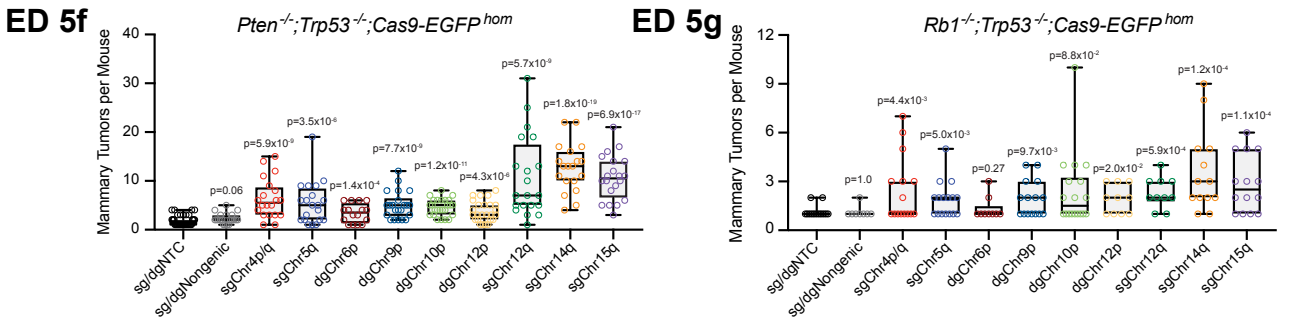


Fig. 1c: *Intraductal injections of the CRISPR-KOALA libraries targeting the mouse orthologs of the top ten specific basal-like breast cancer copy number alterations significantly accelerate tumor onset in both Pten*$^{fl/fl}$*;Trp53*$^{fl/fl}$*;Cas9 and Rb1*$^{fl/fl}$*;Trp53*$^{fl/fl}$*;Cas9 mice (n=20 mice/80 mammary glands per background), compared to control libraries (n=40 mice/160 mammary glands per background for sg/dgNTC libraries and n=20 mice/80 mammary glands in the Pten*$^{fl/fl}$*;Trp53*$^{fl/fl}$*;Cas9 or n=15 mice/60 mammary glands in the Rb1*$^{fl/fl}$*;Trp53*$^{fl/fl}$*;Cas9 background for sg/dgNongenic libraries) with the exception of the dgChr10p library in the Pten*$^{fl/fl}$*;Trp53*$^{fl/fl}$*;Cas9 mice. Statistical analysis was performed using the Gehan-Breslow-Wilcoxon test for each individual chromosome library compared to the sg/dgNTC control library within a genotype.*



Extended Data Fig. 5f and g: *Quantification of the number of mammary tumors per mouse for each arm-level CRISPR-KOALA library in both Pten*$^{-/-}$*;Trp53*$^{-/-}$ (**f**) *or Rb1*$^{-/-}$*;Trp53*$^{-/-}$ (**g**) *models. Each dot represents an individual mouse. Statistical analysis was performed using two-tailed unpaired t-tests compared to sg/dgNTC controls. Data are mean with the box extending from the 25*th *to 75*th *percentile with whiskers from the minimum to maximum datapoint.*

Of note, the long latency of the *Rb1*$^{-/-}$;*Trp53*$^{-/-}$ model explains the relatively long revision period (median survival of 427 days - not counting library design/cloning/viral preparation and breeding the experimental colony), needed for the cohort to mature with the new non-genic control libraries. Ultimately, these new data provide robust evidence that chromosome arm level CNAs contain distinct driver genes that significantly accelerate mammary tumor development in basal-like breast cancer mouse models.

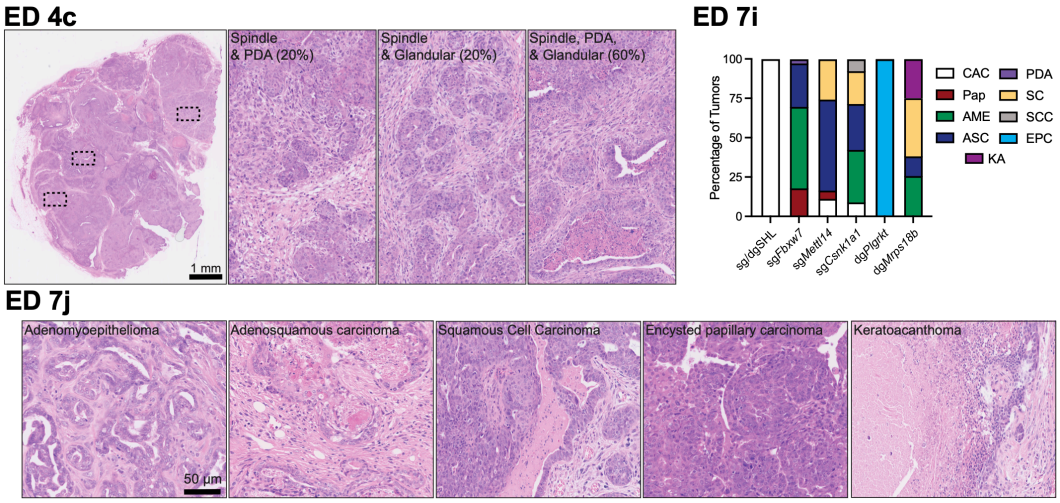
Finally, we wish to highlight that all of our mechanistic and validation experiments were performed with a distinct and highly rigorous set of controls. For these follow-up studies, we used sg/dgRNAs targeting two well-characterized safe-harbour loci, *Tigre* and *Hipp11* (collectively sg/dgSHL). Since these controls induce DNA double-strand break at a defined, neutral genomic locus, they effectively rule out any confounding effects from DNA damage response itself. This ensures that the phenotypes observed in our validation experiments are due specifically to the loss of our target gene, not a potential experimental artifact. We have highlighted this point in the manuscript: "*compared to mice transduced with sgRNAs targeting two different safe harbor loci, Tigre or Hipp11 (hereafter collectively referred to as sgSHL)*" on Line 126-127.

3. The authors need to do a histological characterization of the tumors obtained for all this experiments to see which breast cancer subtypes they obtain.

Reply: We would like to highlight both **Extended Data Fig. 4c**, **Extended Data Fig. 7i** and **j**, and **Extended Data Fig. 8c,** where we present detailed histological analysis of the tumors obtained from all our experiments (133 tumors in total: 29 sg*Fbxw7*, 57 tumors sg*Mettl14*, 24 sg*Csnk1a1*, 8 dg*Plgrkt*, 8 dg*Mrps18b* and 7 sg/dgSHL tumors). In addition, during the revisions, we also tested and added the histological analysis of another hit, *Mrps18b,* on the amplified chromosome arm 6p, which also dramatically accelerates tumor development, and shows the same heterogeneity in histological subtypes observed with the other CNA cancer

genes. These findings are described in the text as: "*As expected, Rb1 ablation significantly accelerated tumor development compared to control mice transduced with sgRNAs targeting two different safe harbor loci, Tigre or Hipp11 (hereafter collectively referred to as sgSHL;* ***Extended Data Fig. 4b****) and these tumors displayed features of complex adenocarcinoma reminiscent of mixed basal-like human BLBC tumors (****Extended Data Fig. 4c****).*" on Lines 124-128. Importantly, reference 29 carefully explores the histology of mouse and human BLBC and corroborates that the histology of our mouse tumors does reflect the mouse manifestation of BLBC.

We further described in the text the detailed histological abalysis of all tumors from a given genotype: "*Next, we molecularly characterized several high penetrance hits. While sg/dgSHL tumors exclusively presented as complex adenocarcinomas with pockets of distinct histology, the CNA cancer gene tumors displayed a variety of adenomyoepithelial, adenosquamous and papillary morphologies in different proportions, recapitulating features of human BLBC histologies*[29] *(****Extended Data Fig. 7i and j****)… Furthermore, all four CNA cancer gene tumors retained BLBC markers (****Extended Data Fig. 8c****)."* on Lines 235-245.

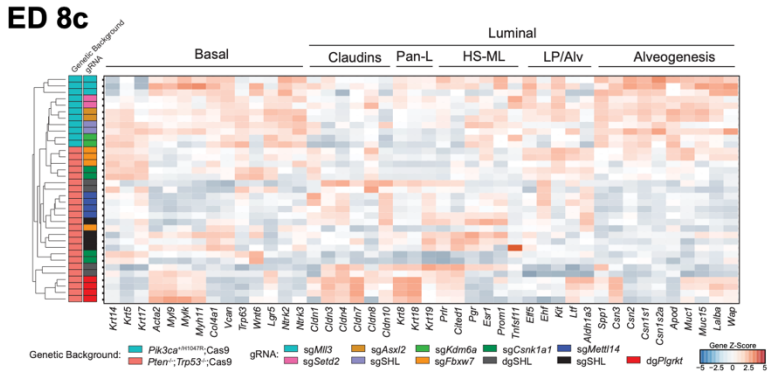


Extended Data Fig. 4c (top left): *Representative haematoxylin and eosin staining from Pten$^{-/-}$;Trp53$^{-/-}$ tumors (n=7 tumors). Insets show three areas of different morphologies. The percentage of Pten$^{-/-}$;Trp53$^{-/-}$ tumors presenting with each morphology is labelled within the inset. Scale bar, 1mm.*

Extended Data Fig. 7i (top right): *Distribution of histological subtypes within CNA cancer gene tumors (sg/dgSHL; n=7 tumors, sgFbxw7; n=29 tumors, sgMettl14; n=57 tumors, sgCsnk1a1; n=24 tumors, dgPlgrkt; n=8 tumors, and dgMrps18b; n=8 tumors). Histological subtypes include CAC: Complex Adenocarcinoma; Pap: Papillary Adenocarcinoma; AME: Adenomyoepitheloima; ASC: Adenosquamous Carcinoma; PDA: Poorly Differentiated Adenocarcinoma; SC: Squamous Cyst; SCC: Squamous Cell Carcinoma; EPC: Encysted Papillary Carcinoma; and KA: Keratoacanthoma.*

Extended Data Fig. 7j (bottom): *Representative haematoxylin and eosin staining from CNA cancer gene tumors. The corresponding histological subtype is labelled inside the micrograph. Scale bar, 50 μm.*

Importantly, we also performed extensive transcriptional analysis (n=4 tumors per genotype), which is confirmed expression of basal markers and the basal-like subtype of our mouse tumors – please see **Extended Data Fig. 8c**:



Extended Data Fig. 8c: *Unsupervised hierarchical clustering of an integrated transcriptional analysis from CNA driver gene tumors (basal-like breast cancer subtype) and previously identified 'Epidriver' tumors (Luminal subtype)*[25] *based on lineage markers of the mammary epithelial hierarchy.*

Beyond the histological comparison of human and mouse tumors and to further ascertain that these mouse tumors recapitulate the biology seen in human BLBC, we have now performed a cross-species analysis, which shows that baseline *Pten;P53* (sg/dgSHL) tumors as well as all genetically engineered tumors that were

transcriptionally profiled (*Fbxw7*, *Mettl14*, *Csnk1a1*, and *Plgrkt*) closely cluster with human BLBC and not Luminal or HER2+ breast cancers. Please see new **Fig. 3d and e**:

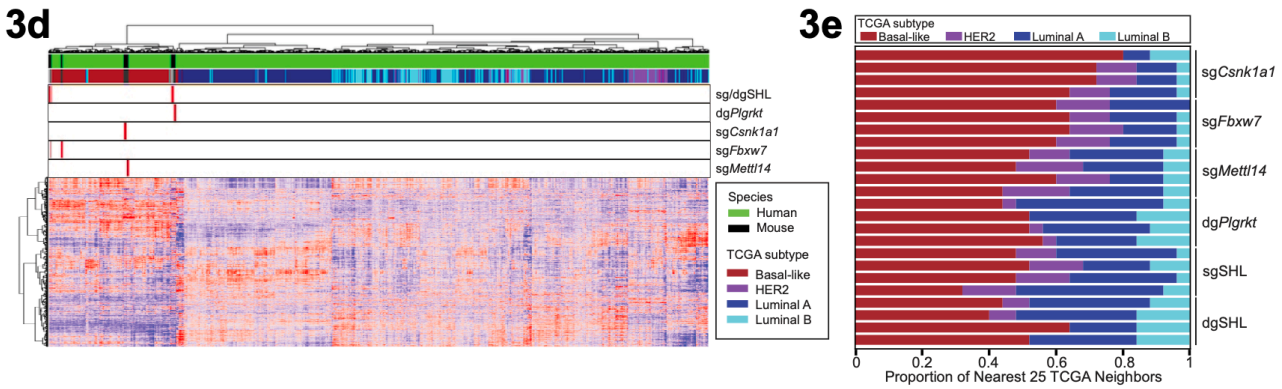


Fig. 3d (left): *Integration of human breast cancer and mouse CNA driver gene transcriptomes followed by unsupervised hierarchical clustering reveals that all mouse tumors cluster within the human basal-like breast subtype.*
Fig. 3e (right): *Nearest neighbor analysis for each mouse CNA driver gene tumor (n=4 tumors per gene) showing that the majority of the top 25 nearest human breast cancer samples belong to the human basal-like breast subtype.*

In addition, further analysis revealed a highly significant, cross-species conservation of transcriptional programs driven by all four driver genes further supporting the clinical relevance of our findings and indicating that these CNA driver genes help shape the transcriptional state of human tumors harboring the respective alteration. Please see new lines 256-267 and new **Fig. 3f-i**.

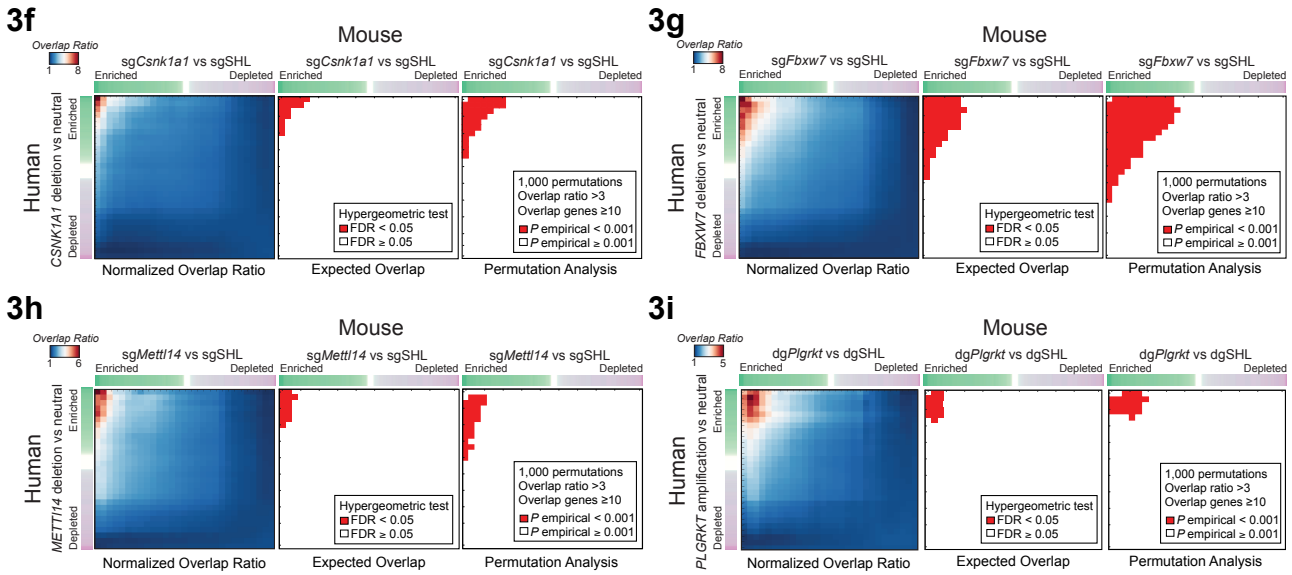


Fig. 3f-i: *One-sided hypergeometric tests were performed across thresholds (every 100 genes up to 2,000; every 500 between 2,500–5,000; and every 2,000 thereafter) for concordant upregulated genes comparing mouse CNA driver gene tumors and their corresponding human basal-like breast cancer sample harboring the given aneuploidy (CSNK1A1 loss; **f**, FBXW7 loss; **g**, METTl14 loss; **h**, and PLGRKT gain; **i**). Enrichment was quantified as $-\log_{10}$ FDR-adjusted p values and normalized overlap ratios (left), with $-\log_{10}$ p values FDR-adjusted normalized to the expected overlap at each threshold (middle). Alternatively, overlap significance was evaluated by permutation testing, where the mouse gene ranking was randomly permuted 1,000 times to derive a null distribution of overlap scores, and empirical p values were computed as the proportion of permutations yielding overlaps ≥ the observed one. Only overlaps satisfying ratio > 3 and ≥ 10 overlapping genes were retained as significant (right).*

4. Most importantly for figure 3F that claims that CNA are not needed in the presence of strong cancer driver genes, I would like to see the WGS profile of all the tumors; they should be added to suppl and be reviewed. Can they also calculate the chromosome-level or arm-level aneuploidy in addition to fraction of genome altered?

Reply: To be thorough, we have now performed whole genome sequencing (WGS) on 23 new tumors, including four new *Pten*;*Trp53* tumors overexpressing *Mrps18b* (a new hit we have added during revisions) and a larger

cohort of 19 *Rb1*;*Trp53* tumors with various driver genes (*Fbxw7* (n=2), *Mettl14* (n=1), *Csnk1a1* (n=3), *Plgrkt* (n=1) and *Mrps18b* (n=2)) and controls (sg/dgSHL; n=10), for a total of 89 WGS profiles. As requested by the reviewer, the raw data for all 89 individual WGS profiles are provided in the Sequencing Reads Archive under accession PRJNA1146341 (available for revision at the following secure link: https://dataview.ncbi.nlm.nih.gov/object/PRJNA1146341?reviewer=5r6mdq8n4j59gqq4a2km2v7rju) and the copy number profiles of each individual tumor and control spleens have now been included and are available for review as new **Supplementary Figure 6**), as suggested by the reviewer. A reference to this has been added to the Methods sections as follows "*Individual plots for each sequenced mammary tumor are provided in Supplementary Figure 6*)." on Lines 1120-1121.

Importantly, the additional data from the *Pten*;*Trp53*;dg*Mrps18b* and the entire *Rb1*;*Trp53* cohort confirm our central hypothesis that manipulating these driver genes overcomes the need to accumulate aneuploidy. In fact, *Rb1*/*Trp53* tumors actually show significantly higher baseline aneuploidy ($p = 2.09x10^{-5}$) compared to *Pten*;*Trp53* tumors, likely due to the increased genetic instability upon loss of Rb1 and the significantly longer latency of this model. Still, *Rb1*/*Trp53* tumors with our driver hits have significantly less aneuploidy and a stochastic pattern of gains and losses, providing powerful evidence of our paper's main conclusion. Please see new **Fig 4d** and **Extended Data Fig. 11e and h**):

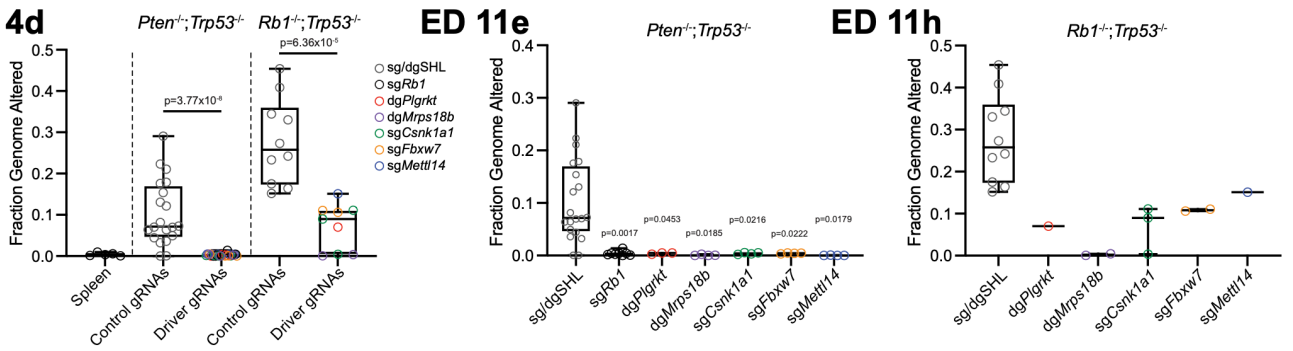
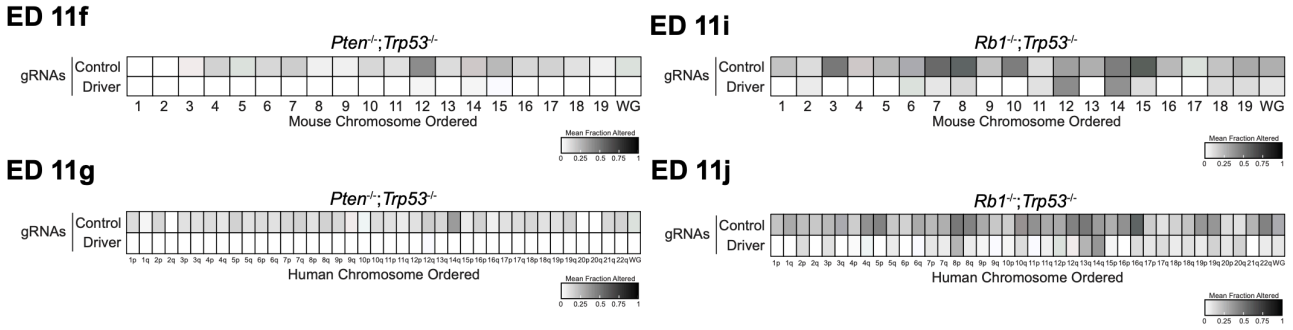


Fig. 4e (left): *Fraction genome altered from spleens (n=5), Pten$^{-/-}$;Trp53$^{-/-}$ tumors transduced with Control gRNAs (sg/dgSHL; n=20) or CNA driver gene gRNAs (n=27), and Rb1$^{-/-}$;Trp53$^{-/-}$ tumors transduced with Control gRNAs (sg/dgSHL; n=9) or CNA driver gene gRNAs (n=8) Individual driver genes are represented by the coloured circles within the boxplot. Statistical analysis was performed using a two-tailed unpaired t-tests. Data are mean with the box extending from the 25th to 75th percentile with whiskers from the minimum to maximum datapoint pooled from across four independent whole genome sequencing runs.*
Extended Data 11e (middle): *Fraction genome altered from Pten$^{-/-}$;Trp53$^{-/-}$ tumors transduced with sg/dgSHL (n=20), sgRb1 (n=8), sgFbxw7 (n=4), sgMettl14 (n=4), sgCsnk1a1 (n=4), dgMrps18b (n=4) or dgPlgrkt (n=3). Statistical analysis was performed using a two-tailed unpaired t-tests compared to sg/dgSHL. Data are mean with the box extending from the 25th to 75th percentile with whiskers from the minimum to maximum datapoint pooled from across four independent whole genome sequencing runs.*
Extended Data 11h (right): *Fraction genome altered from Rb1$^{-/-}$;Trp53$^{-/-}$ tumors transduced with sg/dgSHL (n=10), sgFbxw7 (n=2), sgMettl14 (n=1), sgCsnk1a1 (n=3), dgMrps18b (n=2) or dgPlgrkt (n=1). Statistical analysis was performed using a two-tailed unpaired t-tests compared to sg/dgSHL. Data are mean with the box extending from the 25th to 75th percentile with whiskers from the minimum to maximum datapoint pooled from across four independent whole genome sequencing runs.*

Lastly, as requested by the reviewer, we have also calculated both the chromosome-level and arm-level aneuploidy across our entire WGS dataset in both mouse and human chromosome ordering as new **Extended Fig. 11f and g, l and j**. Importantly, we observed that almost all chromosome arms show a significant reduction in fraction alteration when comparing driver gene to control tumors (with the exception of mouse chromosome 12 which maps to a syntenic region of human chromosome 14q).



Extended Data Fig. 11f and g (left): *Individual chromosome or chromosome-arm level alterations from Pten$^{-/-}$;Trp53$^{-/-}$ in mouse (**f**) or human (**g**) chromosome ordering with samples stratified as control (n=20) or driver (n=27) gRNAs. WG represents the whole genome.*

Extended Data Fig. 11i and j (right): *Individual chromosome or chromosome-arm level alterations from $Rb1^{-/-};Trp53^{-/-}$ in mouse (**f**) or human (**g**) chromosome ordering with samples stratified as control (n=20) or driver (n=27) gRNAs. WG represents the whole genome.*

5. In figure 4H can they show that this effect on oxygen consumption is also true for the mutant forms of Plgrkt? More in general, Plgrkt was the major hit on chr9p, which is what this figure is based on. Fig 4.a shows that CRISPR KOALA is a robust system. However, Figure 4 could use more experiments showing the link between collagen, PGLK1, and mitochondrial dysfunction. They could supplement this with more extended figures.

Reply: We thank the reviewer for this comment and have now substantially extended the characterization of PLGRKT and its function in the mitochondria and OXPHOS in mouse and human tumors. The original **Fig. 4** (now **Fig. 5**) was revised to include many more experiments, which are now also presented in a total of 3 Extended Data Figures (**Extended Data Fig. 12-14**) and 10 Supplementary Videos. In this direct response to the reviewer, we focus our comments on our data linking PLGRKT and mitochondrial dysfunction, but would like to highlight that additional data validating the localization of PLGRKT within the mitochondrial intermembrane space and of the role of PLGRKT as a human breast cancer oncogene are also provided within these figures.

Firstly, we have confirmed that the dramatic effects on cell proliferation and oxygen consumption under stress that is observed with wildtype PLGRKT is also observed with the plasminogen-binding deficient PLGRKT mutants (K147R and Δ139-147). These data have now been included as **Fig. 5k and Extended Data Fig. 13k-n**, which is described on Lines 394-404:

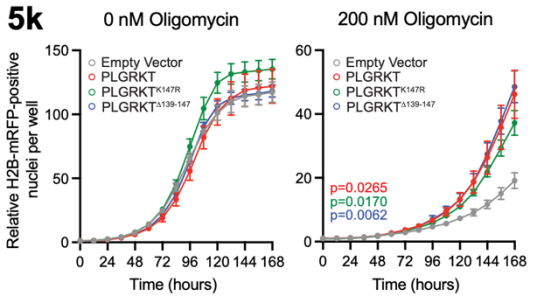
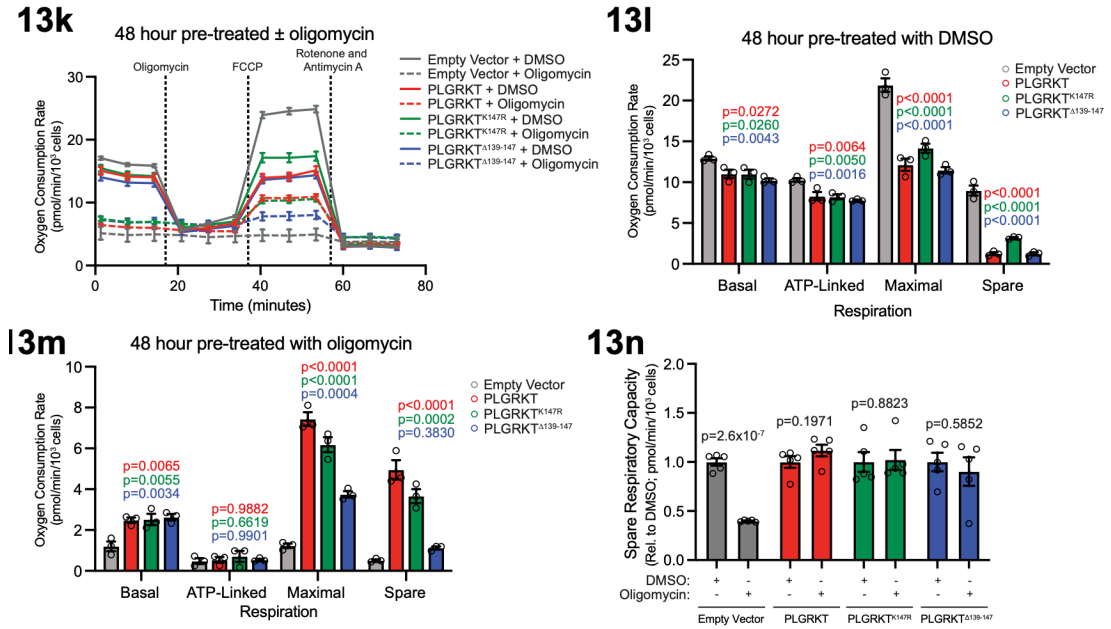


Fig. 5k: *Representative growth curves for 4T1 cells expressing empty vector, PLGRKT, $PLGRKT^{K147R}$ and $PLGRKT^{\Delta139-147}$ treated with DMSO or oligomycin. Statistical analysis was performed using two-tailed unpaired t-tests compared to empty vector controls. Data are means ± s.e.m.*



Extended Data Fig. 13k (top left): *Oxygen consumption rate, determined by Seahorse bioanalyzer of 4T1 cells overexpressing empty vector, PLGRKT, $PLGRKT^{K147R}$ or $PLGRKT^{\Delta139-147}$ following pre-treatment for 48 hours with DMSO or oligomycin (n=11-12 wells per group from 3 biological replicates). Statistical analysis was performed using a two-tailed unpaired t-test. Data are mean ± s.e.m.*
Extended Data Fig. 13l (top right) and m (bottom left): *Quantification of key oxygen consumption rate parameters, determined by Seahorse bioanalyzer of 4T1 cells overexpressing empty vector PLGRKT, $PLGRKT^{K147R}$ or $PLGRKT^{\Delta139-147}$ following pre-treatment for 48 hours with DMSO (**l**) or oligomycin (**m**), at baseline and after exposure to oligomycin, FCCP and rotenone/antimycin A (n=3 biological*

replicates). Statistical analysis was performed using an ordinary one-way ANOVA with Tukey's correction for multiple testing. Data are mean ± s.e.m.

Extended Data Fig. 13n (bottom right): *Relative spare respiratory capacity of 4T1 cells overexpressing empty vector, PLGRKT, PLGRKT^{K147R} or PLGRKT$^{\Delta 139\text{-}147}$ cells (n=5 wells per group) between cells pre-treated with oligomycin or vehicle. Statistical analysis was performed using a two-tailed unpaired t-test. Data are mean ± s.e.m.*

To extend our analysis connecting PLGRKT overexpression with mitochondrial metabolic dysfunction, we have now also quantified the proportion of cellular ATP produced through mitochondrial respiration and glycolysis in PLGRKT-overexpressing mouse 4T1 and human MCF10A cell lines. This analysis revealed a shift towards increased glycolytic ATP production following oligomycin treatment, further supporting the increased proliferative capacity of PLGRKT-overexpressing cell lines observed under stress. This is provided as the new **Fig. 5j:**

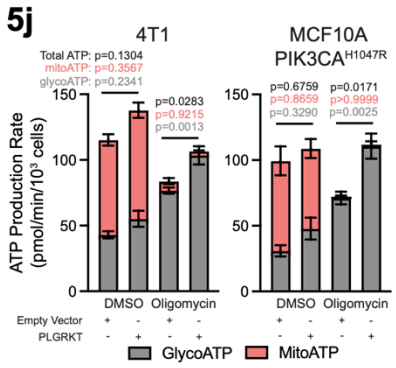
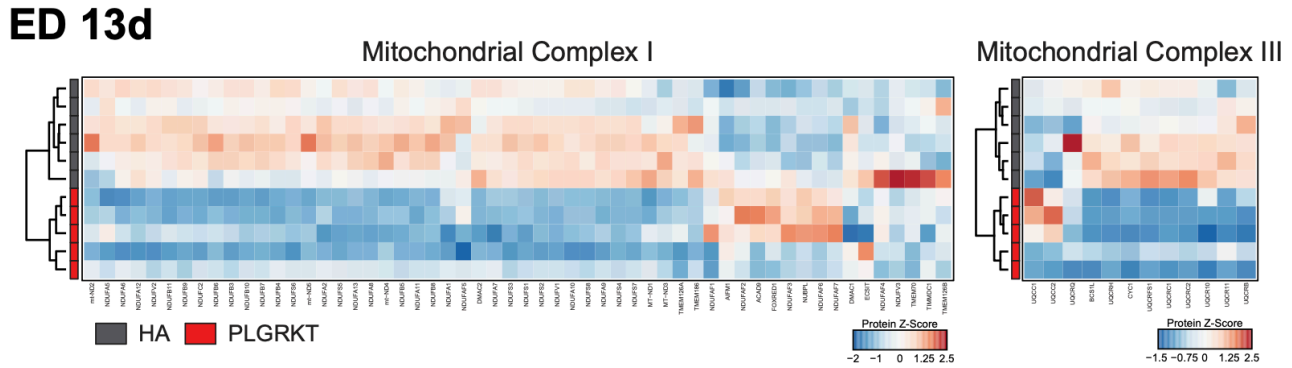


Fig. 5j: *ATP production rate of 4T1 (left) and MCF10A PIK3CA^{H1047R} (right) cells overexpressing empty vector or PLGRKT and treated with DMSO or oligomycin using the Seahorse bioanalyzer (n=4 replicates per group). Statistical analysis was performed a two-tailed unpaired t-tests compared to empty vector controls Data are mean ± s.e.m.*

In support of the decreased oxygen consumption and increased glycolytic capacity of PLGRKT-overexpressing cells, we performed mass spectrometry on freshly isolated mitochondria from 4T1 xenografts and found that they had a significant reduction in Complexes I and III of the electron transport chain. This is provided as the new **Extended Data Fig. 13d**).



Extended Data Fig. 13d: Unsupervised hierarchical clustering of mitochondrial proteins from empty vector and PLGRKT 4T1 xenografts mammary tumors based on all detected proteins belonging to Mitochondrial Complexes I (left) and III (right).

In addition to the effects on oxygen consumption and ATP production, we also found that PLGRKT overexpression enhances mitochondrial motility in stressed mouse and human cells through Enhanced Super Radial Fluctuation (eSRRF) and live cell imaging. Please see Lines 377-386, the new **Fig. 5f-i and Supplementary Videos 1-10**.

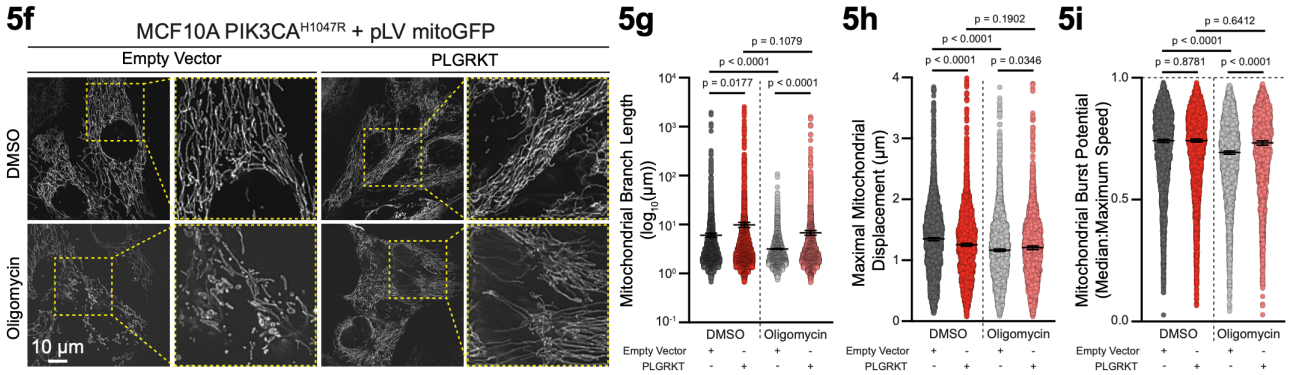
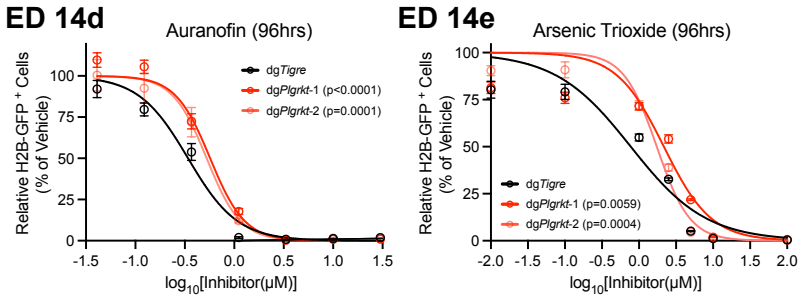


Fig. 5f (left): *Representative images of mitoGFP from enhanced super resolution-radial fluctuation (eSRRF) imaging of MCF10A PIK3CA^{H1047R} cells overexpressing empty vector or PLGRKT and treated with DMSO or oligomycin. Scale bar, 10 μm.*
Fig. 5g-i (right): *Quantification of mitochondrial branch length from eSRRF imaging (**g**), and maximal mitochondrial displacement (**h**) or mitochondrial burst potential (**i**) from mitochondrial live imaging of MCF10A PIK3CA^{H1047R} cells overexpressing empty vector or PLGRKT and treated with DMSO or oligomycin (n=24 fields of view per condition collected over three independent biological replicates). Statistical analysis was performed using or an ordinary one-way ANOVA with Tukey's correction for multiple testing. Data are median ± 95% confidence interval.*

Next, we sought to connect the mitochondrial dysfunction to a cellular mechanism and found that PLGRKT not only conferred a significant growth advantage to oligomycin, but to several other ROS-inducing chemotherapies including auranofin and arsenic trioxide. Please see the new **Extended Data Fig. 14d and e**).



DataExtended Data Fig. 14d and e: *Relative viability of 4T1 cells expressing dgTigre, dgPlgrkt-1 or dgPlgrkt-2 treated with Auranofin (**d**) or Arsenic Trioxide (**e**) for 96 hours. Statistical analysis was performed by fitting normalized response curve with a variable slope and comparing differences in fits. Data are mean ± s.e.m*

This led us to speculate that PLGRKT overexpression regulates resistance and promotes cell survival when challenged with elevated levels of ROS. We now show that PLGRKT itself regulates mitochondrial superoxide and total cellular ROS detoxification. Please see the new **Fig. 5l and m and Extended Data Fig 14 a-c.**

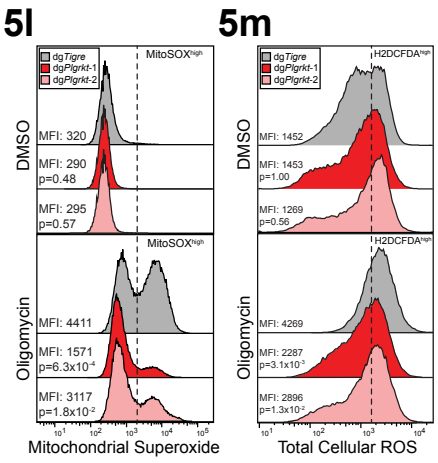
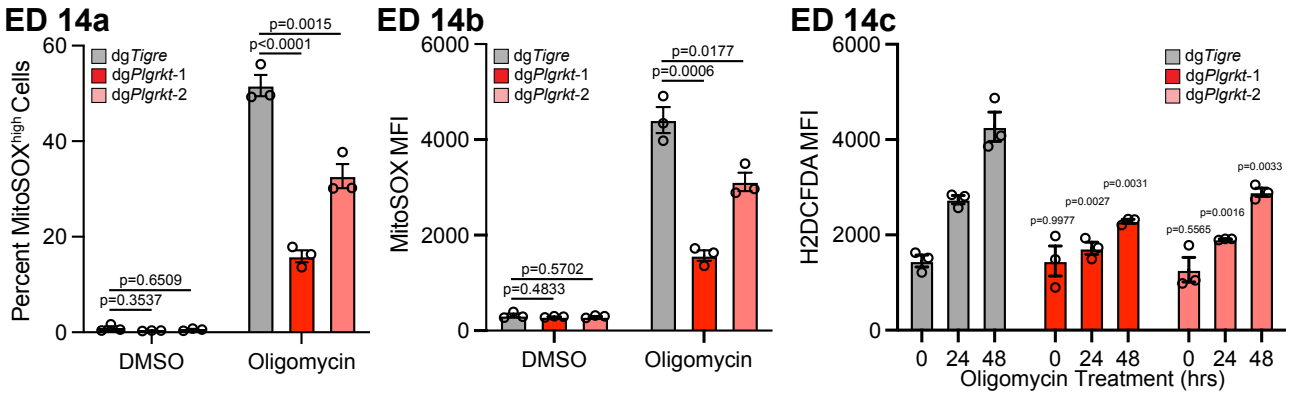


Fig. 5l and m: *Representative histograms of the mean fluorescence intensity (MFI) for mitochondrial superoxide (mitoSOX; **l**) or total cellular ROS (H2DCFDA; **m**) as measured by flow cytometry in 4T1 cells expressing dgTigre, dgPlgrkt-1 or dgPlgrkt-2 and treated with DMSO or oligomycin (n=3 replicates per condition). Statistical analysis was performed two-tailed unpaired t-tests compared to empty vector controls.*



Extended Data Fig. 14a-c: *Quantification of the percentage of MitoSOX^{+high} (**a**), MitoSOX Mean Fluorescence Intensity (MFI; **b**), and H2DCFDA MFI (**c**) in 4T1 cells expressing dgSHL, dgPlgrkt-1 or dgPlgrkt-2 and treated for 48 hours with DMSO or Oligomycin (**a** and **b**) or across a 48 hour time course (**c**). Statistical analysis was performed using a two-tailed unpaired t-test Data are mean ± s.e.m.*

Lastly, treatment of PLGRKT overexpressing cell lines with oligomycin or hydrogen peroxide to induce ROS resulted in a significant elevated expression of antioxidant genes, which we confirmed by qRT-PCR. Please see the new **Fig. 5o and Extended Data Fig. 14g**.

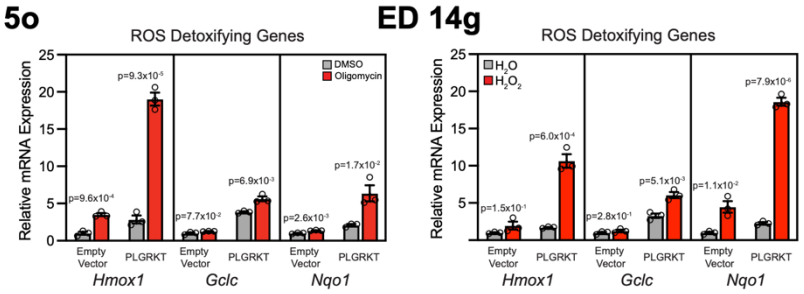


Fig. 5o and Extended Data Fig. 14g*: qRT-PCR analysis for several antioxidant target genes from 4T1 cells overexpressing empty vector or PLGRKT at baseline or treated with oligomycin (**5o, left**) or H_2O_2 (**ED14g, right**) for 12 hours (n=3 samples per group). Statistical analysis was performed using a two-tailed unpaired t-test Data are mean ± s.e.m.*

Importantly, we also found that human BLBC tumors with high PLGRKT expression show significant enrichment of gene sets associated with oxidative damage and stress responses, which now extends our findings to the human setting, which we think is critically important. Please see new **Fig. 5n**:

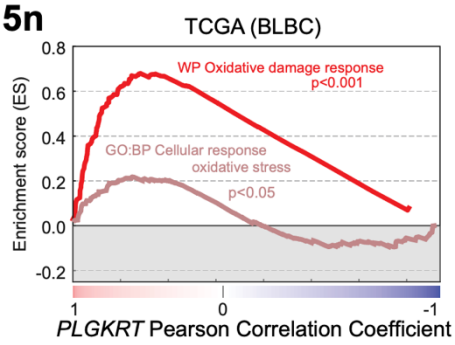


Fig. 5n: *Gene-set enrichment analysis of oxidative response pathways in human BLBC tumors. Genes were ranked by their Pearson correlation coefficient with PLGRKT expression.*

Overall, we believe that we have significant enhanced our molecular understanding of how PLGRKT functions in breast cancer independent of its known role as a plasminogen binding protein. Together, our new data suggests that PLGRKT is a potent proto-oncogene on chr9p, that drives tumorigenesis through its increased localization to the mitochondria, dysregulation of the electron transport chain, elevated ROS detoxification, and promoting healthier, stress-resistant mitochondria. Please see the updated summary schematic of the functions of PLGRKT as the new **Fig. 5p**:

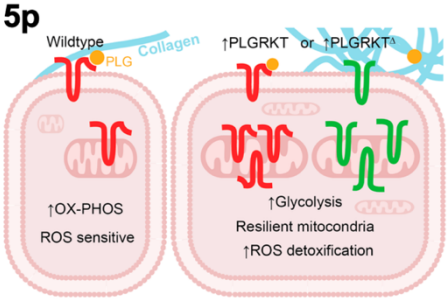
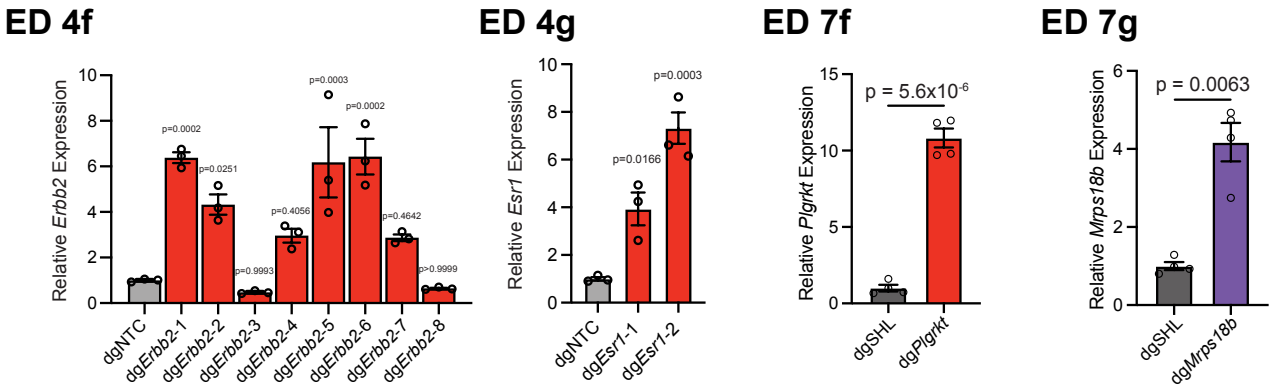


Fig. 5p: *Schematic of the canonical and non-canonical functions for PLGRKT in promoting tumorigenesis.*

6. Notably, the system used has several caveats with modeling and interpreting CNAs: they are modeling not aneuploidy (more subtle dosage change) but complete KO or high over-expression; thus the conclusions are true only for drastic gene dosage changes of single genes.

Reply: We thank the reviewer for the important point that CRISPR-A and CRISPR-KO have the potential of confound the interpretation of subtle dosage changes by inducing dramatic changes in gene expression.

Firstly, to address gene dosage for activation of genes, we would like to highlight that all of our CRISPR-A targeting vectors, which utilize a 14bp dead-guide RNA, achieve a maximum of 10-fold induction of the target mRNA. This is true for all four genes and sgRNAs we have reported in this study including *Plgrkt*, *Mrps18b*, *Esr1* and *Erbb2* (Please see **Extended Data Fig. 4f and g and Extended Data Fig. 7 f and g**). Importantly this holds true across multiple other gRNAs we have profiled in the lab as well, where we achieve on average a 3-10 fold induction of targets.



Extended Data Fig. 4f and g (left): *RT-qPCR analysis of 2409 cells transduced with dgRNAs (n=3 biological replicates) targeting Erbb2 (**f**) or Esr1 (**g**). Statistical analysis was performed using an ordinary one-way ANOVA with Tukey's correction for multiple testing compared to dgNTC controls. Data are mean ± s.e.m.*
Extended Data Fig. 7f and g (right): *qRT-PCR from Pten$^{-/-}$;Trp53$^{-/-}$ tumors transduced with either dgPlgrkt (**f**) or dgMrps18b (**g**) (n=4 tumors per group). Statistical analysis was performed using a two-tailed unpaired t-test.*

In fact, we purposefully spent a lot of time and effort to create and titrate a system to achieve these 'relatively' low levels of transactivation compared to many other CRISPR-A systems (including Zhou *et al*. *Nature Neuroscience* 2018, Gemberling *et al*. *Nature Methods* 2021, Deng *et al*. *Nature Communications* 2022, or Wu *et al*. *Molecular Cell* 2023), which generate on average 100-5000-fold inductions of target genes. These more conventional CRISPR-A systems are widely regarded to achieve non-physiological levels of activation and are known to induce cellular stress and cytotoxicity (Johnston *et al*. *The CRISPR Journal* 2022, Wu *et al*. *Molecular Cell* 2023, and Liang *et al*. *Nature Communications* 2022). Importantly, our levels of mRNA induction (4x for *Mrsp18b*, 6x for *Erbb2*, and 10x for *Plgrkt*) suffice to significantly accelerate and increase tumorigenesis. As such, we actually think our system harbors many advantages and consider it physiologically relevant or at least reasonable, and likely one of the best systems established to date to model lower-level copy number gain seen in human tumors. Of note, all these constructs are in the process of being made available through Addgene for the broad science community.

Probably most importantly, the protein expression level observed in our mouse tumors closely matches that of human basal-like breast cancer as shown for tumors that have high PLGRKT expression. We carefully quantified the PLGRKT expression level in human breast cancer TMAs and while all normal tumor-adjacent epithelium showed little to no expression, 17/35 and 18/35 BLBC tumors showed low and high PLGRKT expression, respectively. Importantly, this closely mimicked the expression level and pattern see in our mouse tumors. Moreover, this staining was performed with a PLGRKT antibody that was validated with KO tissue (hence we can be certain that it is specific) and the staining of mouse and human tumors were performed, imaged, and quantified in parallel. As such, we feel confident that we are modelling physiologically relevant protein levels in our mouse model. This data is shown as the new **Fig. 5d**.

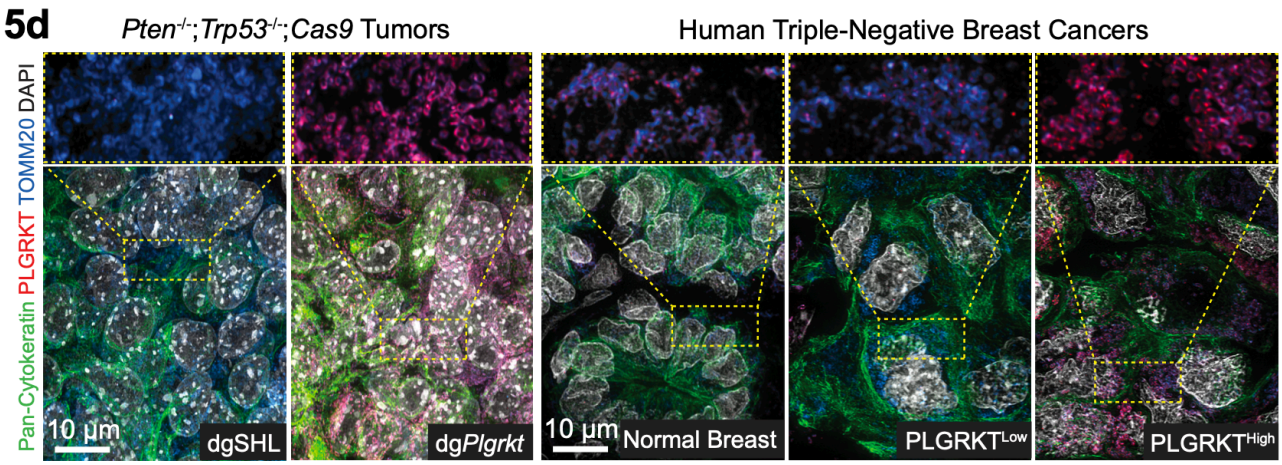


Fig. 5d: *Immunofluorescent staining of autochthonous mouse mammary tumors (n=4 tumors per genotype) and human breast cancer samples (n=35 basal-like breast cancer tumors) and normal adjacent tissue (n=14) showing localization of PLGRKT within the mitochondria. Scale bar, 10 μm.*

In addition, our extensive transcriptional cross-species analysis of *FBXW7*, *CSNK1A1*, *METTL14* and *PLGRKT* altered human and mouse tumors were significantly concordant. This indicates to us that we are indeed faithfully modeling the driving genetic events – naturally of only one gene rather than the whole chromosome arm, using our CRISPR-KOALA system. Please see the new **Fig. 3f-i** for the cross-species transcriptional analysis**.**

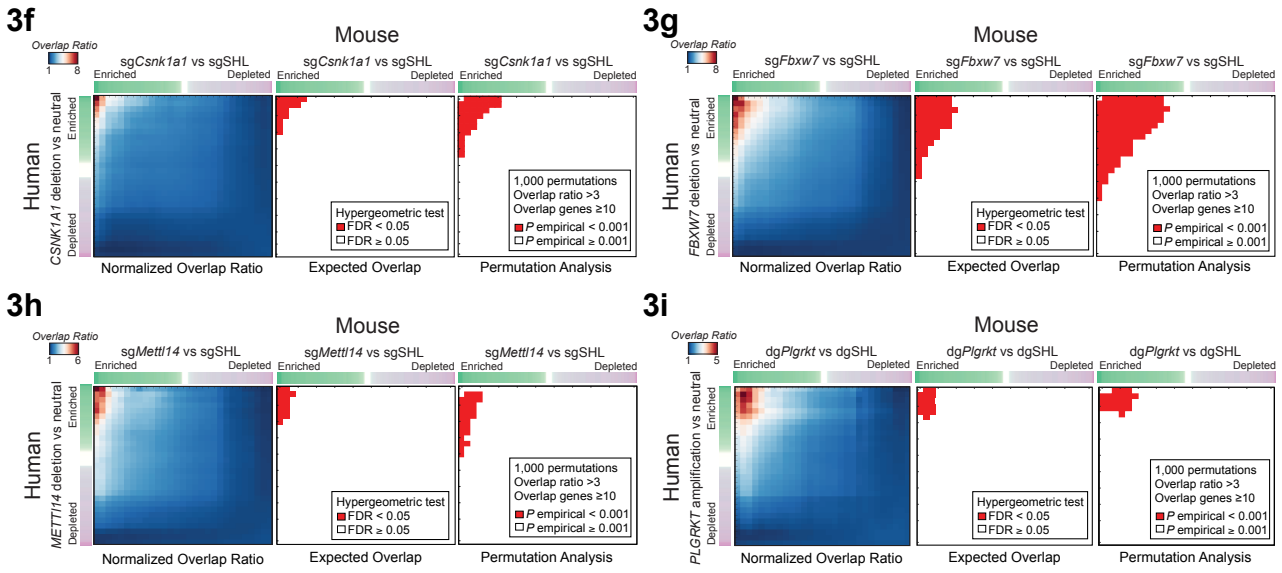


Fig. 3f-i: *One-sided hypergeometric tests were performed across thresholds (every 100 genes up to 2,000; every 500 between 2,500–5,000; and every 2,000 thereafter) for concordant upregulated genes comparing mouse CNA driver gene tumors and their corresponding human basal-like breast cancer sample harboring the given aneuploidy (CSNK1A1 loss; **f**, FBXW7 loss; **g**, METTl14 loss; **h**, and PLGRKT gain; **i**). Enrichment was quantified as $-\log_{10}$ FDR-adjusted p values and normalized overlap ratios (left), with $-\log_{10}$ p values FDR-adjusted normalized to the expected overlap at each threshold (middle). Alternatively, overlap significance was evaluated by permutation testing, where the mouse gene ranking was randomly permuted 1,000 times to derive a null distribution of overlap scores, and empirical p values were computed as the proportion of permutations yielding overlaps ≥ the observed one. Only overlaps satisfying ratio > 3 and ≥ 10 overlapping genes were retained as significant (right).*

However, we also agree with the reviewer that we cannot experimentally ascertain this dose-dependency holds true for all of our 90+ hits. Thus, we have now explicitly acknowledged this limitation in the discussion on Lines 491-501: "*A limitation of the CRISPR-KOALA methodology, inherent to all high-throughput formats, is the difficulty to ascertain whether the induced expression levels always precisely model the low-level CNAs typical of human tumors. This challenge is compounded by the fact that the precise protein levels resulting from a given CNA in patients are often unclear. However, we have provided substantial evidence for the physiological relevance of our approach. For gain-of-function hits, we demonstrated that our system induces protein*

*expression for the proto-oncogene PLGRKT at levels that closely recapitulate the dosage seen in human tumors (**Fig 5d**) and* similarly induces *MRPS18B* four-fold, which is sufficient to drive mammary tumor development*. Second, our screen identified several established haploinsufficient tumor suppressors, including FBXW7, CSNK1A1, and ADAM10, whose dosage-sensitive functions have been independently confirmed in conditional mouse models by us[34] and others[69,86]. An important future direction will be to determine whether these dosage sensitivities apply to the other 85 hits from our screen.*"

Minor Points:

1. Manuscript could be improved in terms of language and description; there is no proper flow.

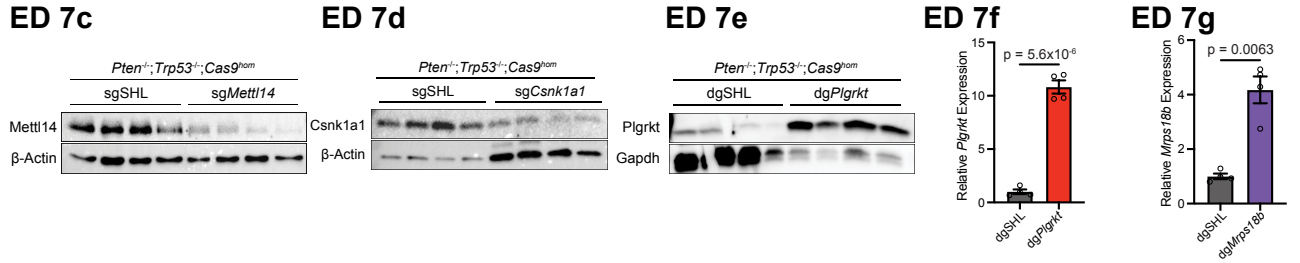
Reply: Thank you for this comment – we have now improved our language and description and think that the new structure of the manuscript from (1) performing the CRISPR-KOALA screen, (2) the validation of all screen hits in the mouse, (3) assessing the relevance of our mouse screening data in patient datasets and the human MCF10A system, (4) interrogating the effect of our cancer driver genes on the accumulation of aneuploidy and finally (5) the extended results section on PLGRKT which has significantly improved the flow of the paper.

Importantly, we added several new subheadings to the manuscript to improve readability and flow:

Establishing a multiplexed in vivo CRISPR-knockout and activation platform
In vivo CRISPR-KOALA screens identify novel BLBC CNA cancer genes
Molecular characterization tumors driven by CNA cancer genes
Modeling CNA cancer genes bypasses the need for aneuploidy
PLGRKT drives breast and skin cancer independent of its canonical plasminogen-activating function
PLGRKT regulates mitochondrial dynamics
PLGRKT regulates detoxification of reactive oxygen species

2. Can they check by qPCR or WB the extent of KO or overexpression for their candidates in figure 2?

Reply: While we had initially included in the manuscript direct evidence of the INDEL formation for all of the driver gene tumors by direct sequencing of tumor DNA around the sgRNA target sequence (now **Extended Data Fig. 7a and b**), we did only show qRT-PCR and/or Western Blot data for sg*Mettl14* and dg*Plgrkt* (**now Extended Data Fig. 7c and e and f**). We have now included Western Blot confirmation of the initial DNA-sequencing on isolated tumor cells from the remaining cancer driver genes that we have extensively validated such as *Csnk1a1* (please see **new Extended Data Fig. 7d)** and included qRT-PCR of the newly included hit Mrpsb18b; **Extended Data Fig. 7g**).



***Extended Data Fig. c-e,** Immunoblot for Mettl14 (**c**), Csnk1a1 (**d**), and Plgrkt (**e**) in Pten$^{-/-}$;Trp53$^{-/-}$ tumors transduced with the indicated gRNA (n=4 tumors per group). **f,g,** qRT-PCR from Pten$^{-/-}$;Trp53$^{-/-}$ tumors transduced with either dgPlgrkt (**f**) or dgMrps18b (**g**) (n=4 tumors per group). Statistical analysis was performed using a two-tailed unpaired t-test.*

3. Figure 4a: is it skin cancer or mammary cancer? legend is not correct.

Reply: We apologize for the confusion with this figure legend. This experiment was performed through injection of mid-gestation embryos, which results in infection in both the skin and the mammary gland. We noted this point in the manuscript on Lines 325-333 "*we performed ultrasound-guided in utero injection to deliver LV-Cre-dgRNAs into the surface ectoderm of mid-gestation embryos, creating long-term clones in the skin and/or mammary gland".* In the figure itself, we quantified tumor-free survival in these mice as any skin or mammary tumor. We understand the confusion as all of the other experiments were performed with intraductal injections directly into the mammary gland and have added to the figure legend the following for clarity: "*Overexpression of Plgrkt through in utero injection of the surface ectoderm targeting both the skin and mammary gland,*

accelerates both cutaneous squamous cell carcinoma (cSCC) and mammary tumor development in mice (n=8 per group).”

4. Did the authors estimate “MOI” in vivo?

Reply: Yes, we have previously optimized and carefully determined the titre and MOI of our lentiviral systems *in vivo* such that we can achieve a transduction efficiency of 30% within the luminal mammary epithelial population, 5% within the basal epithelial compartment and less than 0.5% within the stroma in Langille *et al*. *Cancer Discovery* 2022, Supplementary Fig. S1: “We optimized the parameters for an *in vivo* CRISPR screen by using a mixture of lentiviruses expressing GFP or RFP to determine the viral titer that transduces the mammary epithelium at clonal density (multiplicity of infection <1). Higher viral titers were associated with double infections, whereas a 15% overall transduction level minimized double infections while generating sufficient clones to screen (Supplementary Fig. S1F–S1I). Flow cytometry revealed that the third and fourth mammary glands each contain >3.5x10^5 epithelial cells, and that EPCAM^{hi}/CD49f^{mid} luminal cells showed a higher infectivity (~30%) compared with EPCAM^{mid}/CD49f^{hi} basal cells (~5%; Fig. 1B; Supplementary Fig. S1H and S1I).”

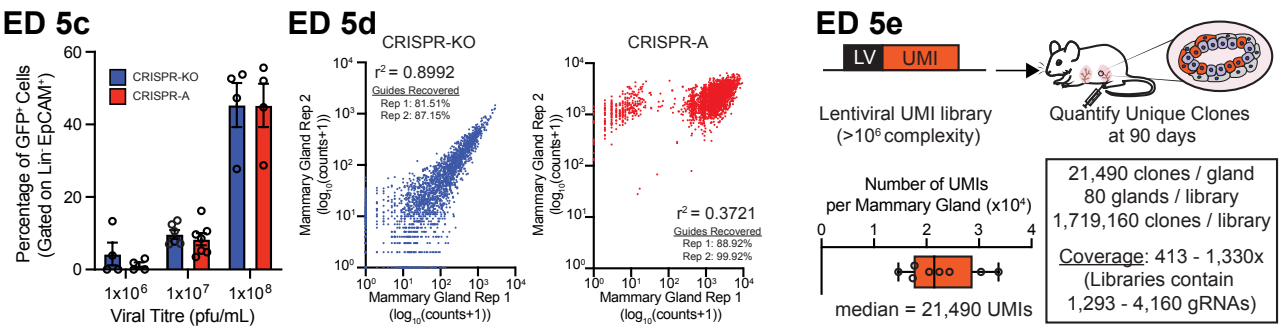
FIGURE REDACTED

TEXT REDACTED

We used the same parameters and titers to calculate MOI and screening coverage for CRISPR-KOALA screen. We have also experimentally validated that these parameters apply to the bigger CRISPR-A construct by doing a side-by-side comparison of KO and A vectors at 1e8, 1e7, and 1e6 pfu/ml, which confirmed that they have the same infectivity in the mammary gland. We have also done replicates of an NTC library showing that we can recover all guides from the mammary glands of independent mice and that replicate mice have a strong R-squared value for both KO and A libraries, all of which we have now specified in more detail in the Methods section.

To be thorough and calculate the true multiplicity of clones achieved upon intraductal injections into the mammary gland, we have now also used a lentiviral library with unique molecular identifiers at a complexity of >1M unique barcodes. We injected this library into eight independent mammary glands and quantified clones with unique barcodes 3 months later by next-generation sequencing. This complex library confirmed that we obtain 20,000 independent UMI clones per mammary gland at the given 15% transduction rate, which translates to a coverage of between 400 and 1,330 clones for any given sgRNA within every library tested. We have added this information to the Methods section.

For completeness, ~~all this data has now be~~en included as new **Extended Data Fig. 5c-e**. We have also updated the text to refer to these important MOI optimizations as follows: “*We titrated the libraries to a 15% transduction level to minimize double infections while ensuring sufficient generation of clones for high content screening (**Extended Data Fig. 5c-e**)[25,34]. We then generated cohorts of 20 mice per library and genotype (440 total mice), achieving an overall coverage of between 413 and 1330 clones per gRNA (**Extended Data Fig. 5e**), resulting in a functional genomics CNA screen across a fifth of the genome.*” on Lines 149-154.



Extended Data 5c (left): *Representative in vivo titering of CRISPR-KO and CRISPR-A viruses directly in the mouse mammary gland (n=4-6 mammary glands per condition). Data are mean ± s.e.m.*

Extended Data 5d (middle): *gRNA representation correlations between mammary glands transduced with CRISPR-KO (left) or CRISPR-A (right) libraries.*

Extended Data 5e (right): *Transduction of mouse mammary glands (n=8) with a barcoded lentiviral library followed by next generation sequence to quantify the number of individually infected epithelial clones per injection. A summary of the coverage calculations is provided in the box. Data are mean with the box extending from the 25th to 75th percentile with whiskers from the minimum to maximum datapoint.*

Reply to Reviewer #2:

In this manuscript, Al-Zahrani et al. describe the development and broad application of CRISPR-KOALA, a method that allows for the inactivation and activation of libraries of candidate genes in autochthonous mouse models of cancer using CRISPR-KO and CRISPR-A. They use this technique to study the impact of over 3000 genes that are found in common chromosomal arm deletions and amplifications in BLBC. This is an important system that allows the analysis of a large number of genes in parallel in an immunocompetent setting. The authors rightly point out that many of their hits are not found in cell line studies, underscoring the unique value of these types of in vivo (i.e. non-cell line) approaches.

In their parallel inactivation and activation study they start with two models that recapitulate genetic alterations that are common in human breast cancer, namely P53/PTEN and P53/RB1. Through their screen they identify 90 cancer driver genes, most of which do not have previously known roles in BLBC progression. They convincingly show that 5 hits from their screen act as tumor suppressors or oncogenes in vivo in single-gene validation experiments. Perhaps interestingly, alteration of these genes did not drive the tumors to converge on some aggressive cell state but rather each genotype generated unique effects. They also show (somewhat unexpectedly) that genetic modulation of any of these 5 genes reduces the selection for aneuploidy. They delve into the mechanism of tumorigenesis of one novel oncogene (Plgrkt) and show that Plgrkt upregulation is associated with mitochondrial reprogramming, specifically an inhibition of oxidative phosphorylation, in **in vitro models**. Their approach of inactivating and activating libraries of genes in a high-throughput and physiological system is novel and could be applied to the study of driver genes in other cancer types.

Reply: We thank the reviewer for their thoughtful and positive evaluation of our manuscript and for highlighting the novelty of our findings, including the strengths of our new CRISPR-KOALA platform to evaluate gene function in a physiological settings and the number of novel and potent cancer driver genes identified with the surprising divergence of the induced cell states/pathway while uniformly overcoming the need for aneuploidy and the importance of our work extending previous studies to an *in vivo*, immune competent setting. We have addressed all of the reviewers' comments below in full, which has significantly strengthened our manuscript.

In particular (and along with the editors major concerns), we have completely (1) re-analyzed our screening data as well as performed re-screens in both human and mouse models which have all confirmed our initial findings as well as providing greater insight into lower penetrance driver genes on aneuploid regions, (2) performed validation and follow-up studies in the *Rb1*;*Trp53* model for all of our top driver genes and (3) a completely reworked section providing new data demonstrating and confirming the oncogenic role of PLGRKT as a regulator of several mitochondrial functions in both human and mouse models.

Major Points:

1. It was not well described what they consider a tumor. Are these individual masses that form? I presume that they are as they only inject 20 mice per library = 80 mammary glands. Let's image that is not quite the case and some tumors are not clearly clonal masses. From the methods, they only count the dominant sgRNA and any sgRNA that has at least 90% as many reads. This seems like a peculiar decision as an sgRNA that is within 90% likely means that the clonal tumor arose from an initial cell that got transduced with two lentiviral vectors (this also make sense that those might gain the most advantage). Meanwhile a second clonal tumors in that mass which has 50% or 10% as many reads to ignored. While their hits will likely be hits regardless of what arbitrary cutoffs they use, I would expect that more information could be glean with a more thoughtful analysis approach. They also entire lacks statistical in this analysis. The sgRNAs have >100x differences in representation in the pool (from their Sup Fig), I would have expected this representation to factor into a null expectation for sgRNA representation in tumors. It would have likely been helpful if they had sg/sgNTC within all of their pools, rather than separate (as most people do for in vitro CRISPR screens).

Reply: The reviewer raises an excellent point. We agree that further careful analysis of our screening data is important, particularly for understanding lower penetrance drivers and the potential poly-clonality in tumors. We are really thankful for the suggestion.

We have now taken a two-pronged approach: **1) a new set of experimental screens** designed to unmask weaker hits and **2) a comprehensive bioinformatic re-analysis** of our entire dataset, using a model-based approach to complement our initial ranked-based analysis. These new analyses have not only strengthened our original conclusions and provided statistical power to our large *in vivo* screening dataset but have also

yielded significant new biological insights. To clarify the approach, we took for these analyses, we have provided a schematic as the new **Fig. 2a**.

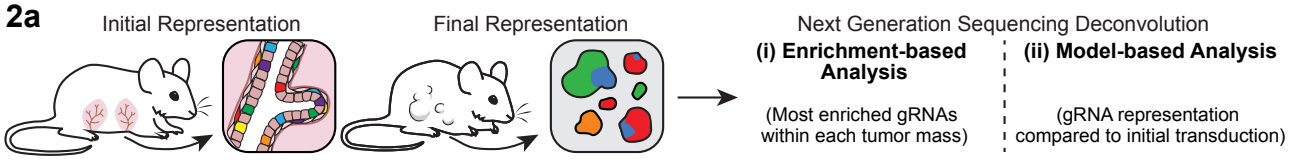
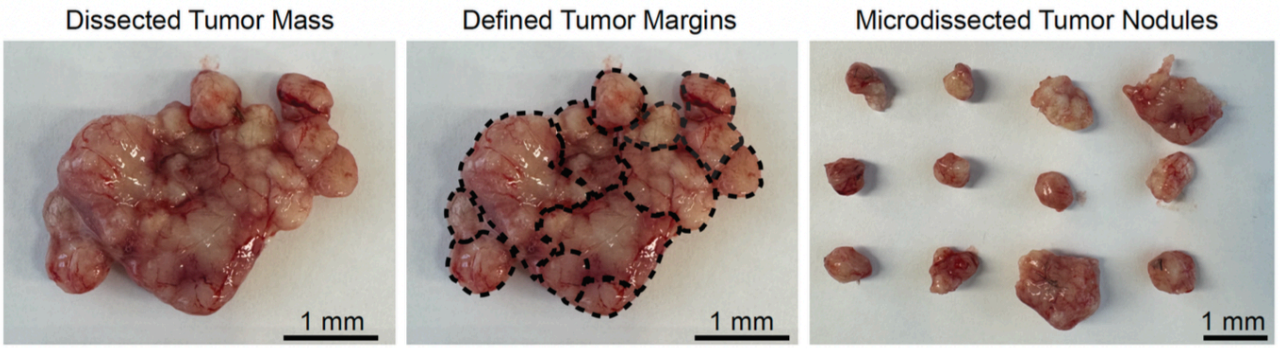


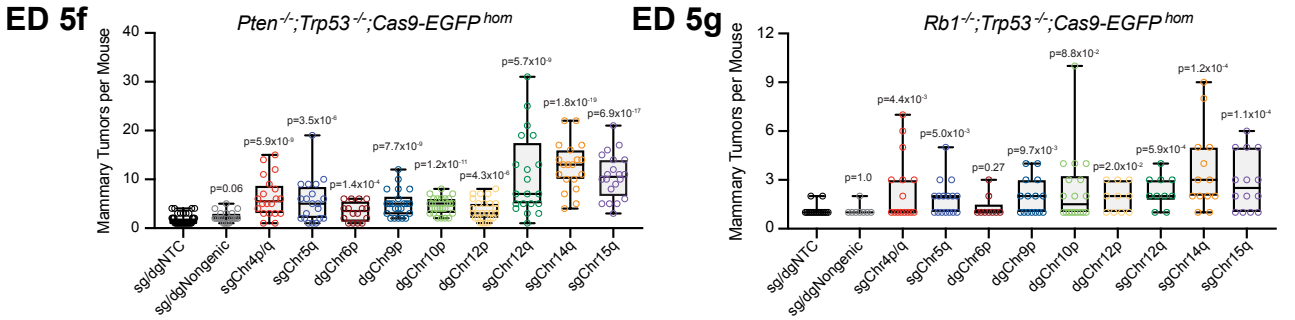
Fig. 2a: *Schematic representation of the in vivo CRISPR-KOALA screen, library deconvolution and analysis pipelines.*

Importantly, the reviewer's assessment of our tumors is correct. In the original submission, we defined a tumor as a palpable mass of at least 1 mm³ in size (though most were substantially larger and confirmed histologically). Each mass was microdissected along obvious margins to separate distinct tumors followed by deep sequencing and library deconvolution. To provide an example of these microdissections, we have updated the methods section to read "*At humane tumor endpoint mice were euthanized and tumors were immediately dissected and placed on ice.*" on Line 756. Additionally, we have provided an example of a representative tumor mass from one of our mammary glands at endpoint, its individual tumor margins, and the resulting individual tumor nodules that were easily dissected and subject to next-generation sequencing as the new **Supplementary Figure 3.**



Supplementary Figure 3: *Representative tumor mass at humane endpoint (left) from a chromosome-arm level library. Tumor margins along which individual tumors were microdissected are indicated with the dashed lines (middle) and the resulting tumor nodules are shown (right).*

Additionally, we enumerated tumors by biweekly palpation and observed distinct differences in tumor burden across our models: Control *Pten/Trp53* mice developed on average 1-4 tumors, whereas experimental *Pten/Trp53* mice transduced with chromosome arm sg/dgRNA libraries developed 3-15 and up to 25 tumors for some of the more aggressive libraries. The *Rb1/Trp53* model typically developed fewer tumors on average, between 1-2 tumors for controls, though some experimental libraries increase this to 10 tumors/mouse (please see the new **Extended Data Fig. 5f and g** which we have now represented as box plots to better represent the data).



Extended Data Fig. 5f and g: *Quantification of the number of mammary tumors per mouse for each arm-level CRISPR-KOALA library in both Pten-/-;Trp53-/- (**f**) or Rb1-/-;Trp53-/- (**g**) models. Each dot represents an individual mouse. Statistical analysis was performed using two-tailed unpaired t-tests compared to sg/dgNTC controls. Data are mean with the box extending from the 25th to 75th percentile with whiskers from the minimum to maximum datapoint.*

The reviewer is also absolutely correct that the physical constraint of the mammary gland means that some tumors could be polyclonal, with multiple clones growing into one other. Our original ranked-based analysis, which focused on the most dominant sgRNAs, was designed to identify the strongest hits in a competitive *in vivo* environment. We believe that this is reasonable as we were analyzing 80 independently injected glands for a given library and 1,478 tumors across the entire screen. While some tumors might be polyclonal in a given mammary gland, it would be expected that strong hits would emerge as monoclonal tumor outgrowths across other independent glands in the screen. This is largely why we defined screen hits as: "*Genes targeted by at least 2 different gRNAs and present in at least 1% of the total tumors within a library*" (on Lines 161-162).

However, we completely agree with the reviewer's point that this could overlook weaker drivers or subclones. Our new experimental and bioinformatic analyses described below were designed to address these possibilities directly and are provided on Lines 179-189: "*We also analysed our screen using the statistical mode MAGeCK I[37], which compares the gRNA across all tumors to the initial viral library (**Fig. 2a**). While this approach confirmed our high-penetrance hits, its increased sensitivity to subclonal gRNA representation and potential cooperative effects between gRNAs identified 78 candidate driver genes with lower penetrance (**Fig. 2d and Extended Data Fig. 6c and Supplementary Table 3**). A key advantage of this approach is the detection of significantly depleted gRNAs, which showed the expected depletion of core essential genes as defined by DepMap[38] (**Fig. 2d and Extended Data Fig. 6d and Supplementary Table 3**). Underscoring the importance of performing these screens in vivo, this analysis also revealed that several genes classified as core essential in cell culture, such as Csnk1a1, Ino80, Mettl14, and Ube2n, instead function as potent tumor suppressors in the mammary gland (**Fig. 2d and Extended Data Fig. 6d and Supplementary Table 3**).*"

This approach accounts for the differences in gRNA representation within the original library and provides statistics for every gene across all tumors without relying on a "dominant hit" cutoff (also addressed experimentally in response to **Major Comment #3**). This new analysis has added a rich new layer of biological insights to our study, all of which we have added to the manuscript (Lines 179-189, as quoted immediately above). It also allowed us to run rigorous statistics, and we have now added the p-value for every gene into **Fig. 2b** and to calculate a LFC for each gRNA, which is now presented in the **Supplemental Table 3.**

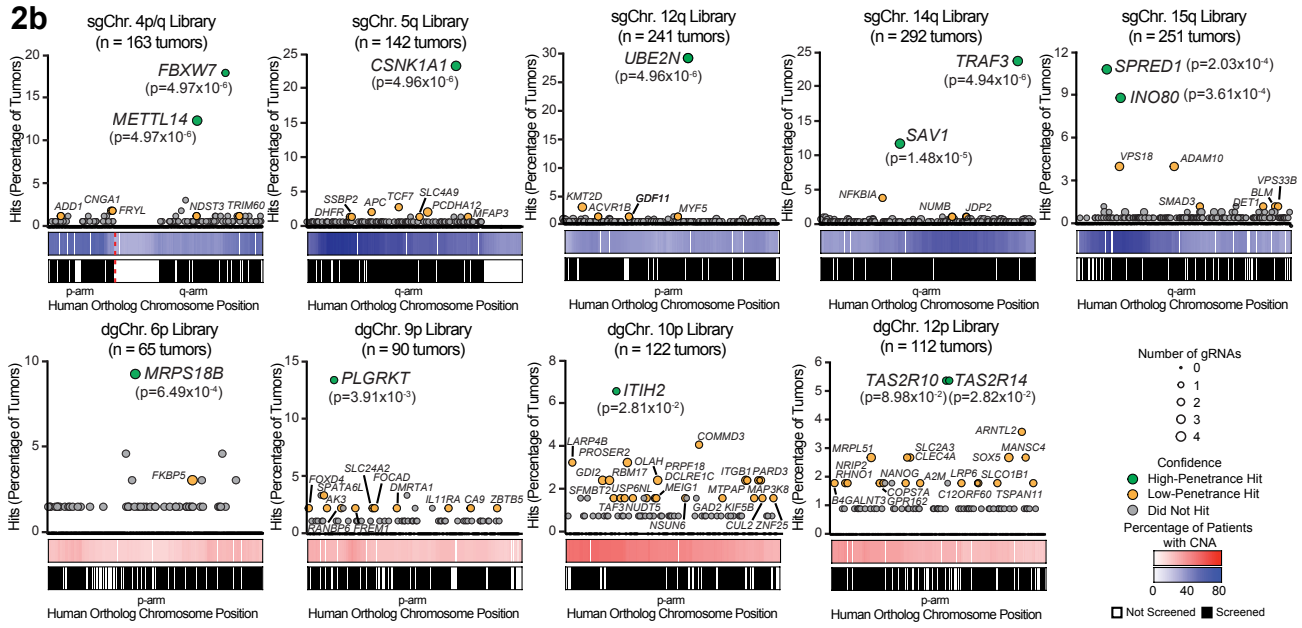


Fig. 2b: *gRNA representation from the CRISPR-KOALA screen of the top ten specific and recurrent basal-like breast cancer copy number alterations (n=1,478 tumors). Data is divided by chromosome library and the percentage of tumors within each library enriched for a particular gRNA is plotted on the y-axis with the x-axis representing the human chromosome coordinate from the distal end of the p-arm to the distal end of the q-arm. High penetrance, low penetrance and genes which did not hit are shown in green, orange and grey, respectively. The size of each dot represents the number of gRNAs which scored for each gene. Below each plot is the proportion of patients from METABRIC dataset harboring CNAs at each specific gene, with red and blue representing amplifications and deletions, respectively. Black bars represent human genes for which the mouse ortholog was screened. Statistical analysis was performed using a permutation test where gRNAs are randomly assigned to random genes and performed using MAGeCK.*

We also present the results of the MAGeCK analysis as new main Figure panel **Fig 2d**:

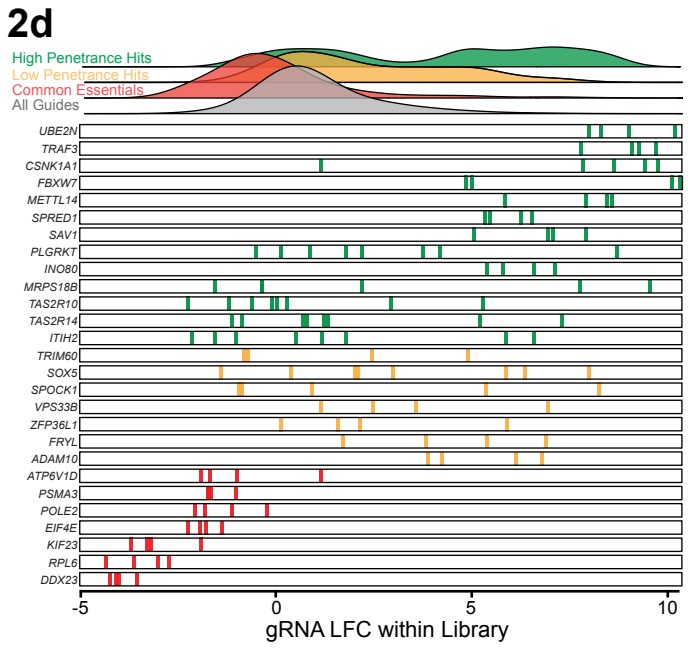


Fig. 2d: *Histogram distribution of the log fold change for each gRNA targeting high penetrance, low penetrance and common essential genes along with the overall distribution of all genes within the library (top). The log fold change of each gRNA targeting representative genes for each category is shown (bottom).*

The reviewer also raises the question as to why we have not included non-targeting and/or non-genic controls within our libraries, as is common with *in* vitro screens. Importantly, we routinely choose to have the control non-targeting gRNAs within a separate library, as this allows us to (1) monitor tumor multiplicity and latency separately of the experimental library and (2) measure the tumorigenic potential of all genes within a library without diluting any effects of bystander controls. Of note, we have now repeated the screen with a new control library containing sgRNA and dgRNAs target non-genic loci (2,600 loci; 5,200 sg/dgRNAs) and followed survival and tumor multiplicity in both the *Pten;*Trp53 and *Rb1;Trp53* models. As expected, we obtained statistically identical results to our initial non-targeting control libraries. The update survival curves are provided in **Fig. 1c**.

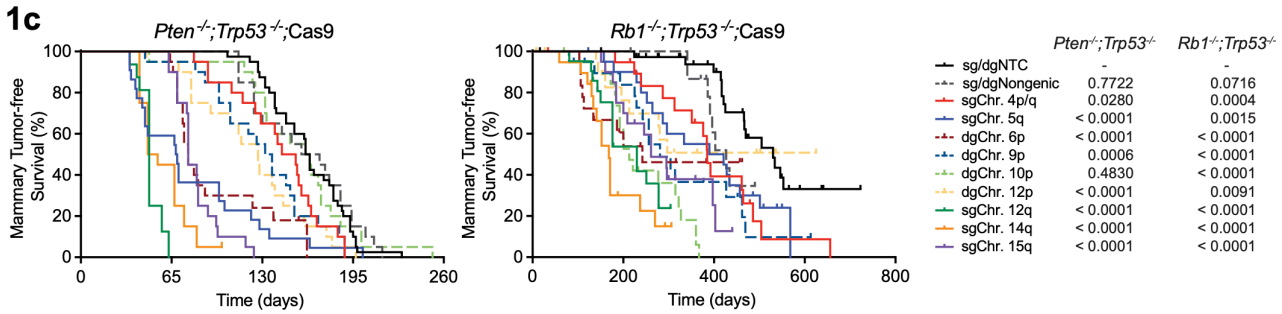
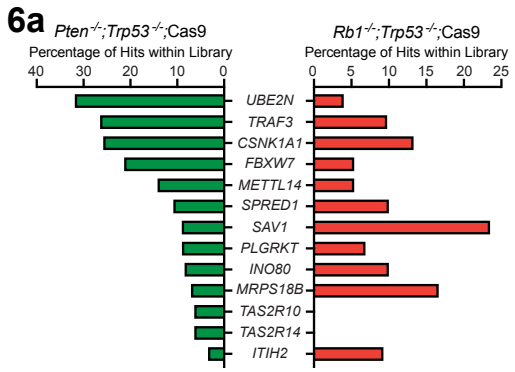


Fig. 1c: *Intraductal injections of the CRISPR-KOALA libraries targeting the mouse orthologs of the top ten specific basal-like breast cancer copy number alterations significantly accelerate tumor onset in both Pten$^{fl/fl}$;Trp53$^{fl/fl}$;Cas9 and Rb1$^{fl/fl}$;Trp53$^{fl/fl}$;Cas9 mice (n=20 mice/80 mammary glands per background), compared to control libraries (n=40 mice/160 mammary glands per background for sg/dgNTC libraries and n=20 mice/80 mammary glands in the Pten$^{fl/fl}$;Trp53$^{fl/fl}$;Cas9 or n=15 mice/60 mammary glands in the Rb1$^{fl/fl}$;Trp53$^{fl/fl}$;Cas9 background for sg/dgNongenic libraries) with the exception of the dgChr10p library in the Pten$^{fl/fl}$;Trp53$^{fl/fl}$;Cas9 mice. Statistical analysis was performed using the Gehan-Breslow-Wilcoxon test for each individual chromosome library compared to the sg/dgNTC control library within a genotype.*

2. They only present merged data from their Pten/p53 and Rb1/p53 screens. These results should also be shown separately. Perhaps their data has insufficient resolution to compare across the Pten/p53 and Rb1/p53 screens but the data should still be presented.

Reply: We would like to highlight that the individual data across both backgrounds was already presented in the original **Supplementary Table 3**, but we completely agree that it should be shown more prominently and have now added it as a **new Extended Data Fig. 6a** to enhance visibility. Importantly, we believe our data has

sufficient resolution across both models, with the only discordance occurring for the Chr12p hits. This may simply reflect that certain hits, such as *TAS2R10* and *TAS2R14* may function in an *Rb1*-dependent pathway.



Extended Data Fig. 6a: *Distribution of high penetrance hits across both Pten$^{-/-}$;Trp53$^{-/-}$ and Rb1$^{-/-}$;Trp53$^{-/-}$ screening cohorts.*

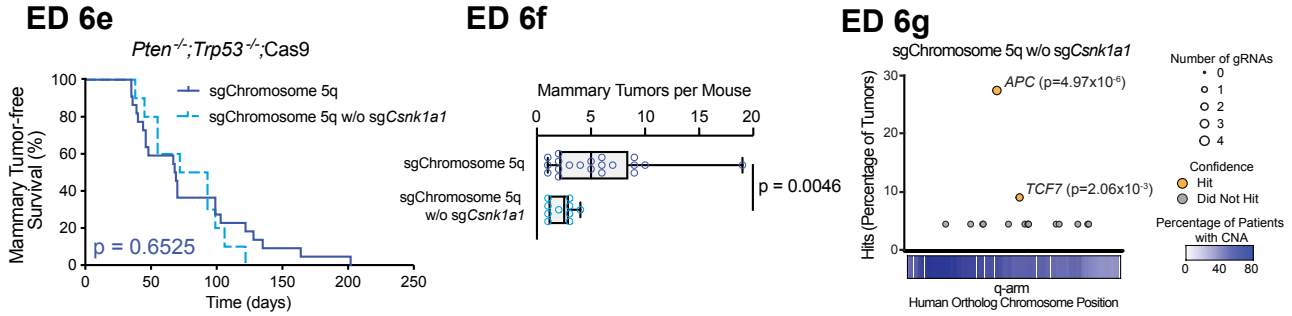
3. Their approach suffers from only identifying the very largest driver genes as these greatly increase tumor growth and their sequencing-based analysis only focus on the largest clonal growth. There are likely many other genes in these regions that have more moderate effects that they are not uncovering. This might be something that they should acknowledge along with any other limitation, along with potential solutions. For example, would repeating these library screens without sgRNA/dgRNAs with the largest effects could help identify others with more moderate effects? dgRNA induced genes are not necessarily induced at the correct levels, many could be too low and that could lead to false negative in their screen etc.

Reply: We agree with the reviewer, and as with any large-scale screen, a key consideration is whether the screen has reached saturation or if false negatives were missed due to technical issues like poor guide efficiency.

To address this, we performed the experiment that was suggested by the reviewer and re-screened one of our most aggressive libraries after removing the top hit. We chose to perform this re-screen with the knockout of library of Chromosome 5q as (1) Chromosome 5q was the top-most recurrent and specific copy number alteration that we identified across all basal-like breast cancer patients and (2) the initial screen presented with an incredibly dominant, high-penetrant driver gene (*Csnk1a1*) that we identified as the top hit in approximately 25% of the tumors. These conditions made this an ideal library to address whether *Csnk1a1* within the library was masking the emergence of lower penetrance hits.

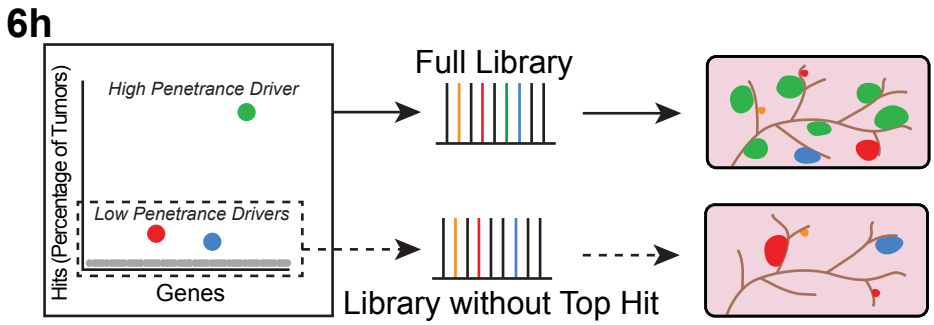
Interestingly, while removing guides that target *Csnk1a1* drastically reduced tumor penetrance, the latency of tumors that developed was surprisingly unchanged (**Extended Data Fig. 6e and f**). Next-generation sequence of the resulting tumors in the '*hit-depleted*' library failed to identify any new hits outside of the lower penetrance drivers, *Apc* and *Tcf7*, that were also already recovered in the primary screen. Overall, the lower penetrant tumors still form mammary tumors with similar latency as high penetrance ones, but the reduced overall tumor multiplicity observed supports our notion that the false negative rate within our screens is likely very low.

We have described this approach and the results in the manuscript Lines 190-198: "*To rule out the possibility that the high penetrance hits masked the emergence of additional lower penetrance ones, we cloned and re-screened the chr5q library excluding the major hit Csnk1a1. We chose the knockout of library of chr5q as (1) chr5q is the top-most recurrent and specific CNA in BLBC patients and (2) the initial screen revealed Csnk1a1 as the sole, highly penetrant driver gene. While tumor penetrance was reduced, the latency of tumors that developed in this chr5q library without Csnk1a1 was surprisingly unchanged (**Extended Data Fig. 6e and f**). In the absence of Csnk1a1, we uncovered Apc and Tcf7 as the dominant driver genes without identifying any other new hits (**Extended Data Fig. 6g**). This result suggests that our primary screens were at or near saturation and confirms the bona fide identity of the lower-penetrant tumor suppressors (**Extended Data Fig. 6h**)."*



Extended Data Fig. 6e (left): *Re-screening a Chromosome 5q library, depleted of sgRNAs targeting Csnk1a1 (n=10 mice) does not alter mammary tumor-free survival compared to the initial library. Note that the full Chromosome 5q library is the same mice that are presented in Fig. 1c. Statistical analysis was performed using the Gehan-Breslow-Wilcoxon test.*
Extended Data Fig. 6f (middle): *A significant reduction in tumor multiplicity was observed at endpoint in mice transduced with the Chromosome 5q library without sgCsnk1a1 compared to the full library. Statistical analysis was performed using a two-tailed unpaired t-test. Data are mean with the box extending from the 25th to 75th percentile with whiskers from the minimum to maximum datapoint.*
Extended Data Fig. 6g (right): *gRNA representation from the sgChromosome 5q library without Csnk1a1. The percentage of tumors within the library enriched for a particular gRNA is plotted on the y-axis with the x-axis representing the human chromosome coordinate from the distal end of the p-arm to the distal end of the q-arm. The size of each dot represents the number of gRNAs which scored for each gene. Below each plot is the proportion of patients from METABRIC dataset harboring CNAs at each specific gene.*

Additionally, for clarity and ease of reading, we have outlined this experimental strategy and conclusions in a schematic shown in new **Extended Fig 6h**:



Extended Data Fig. 6h: *Schematic depicting the effect of removing the top penetrant hit from a given chromosome arm library. The lower penetrant tumors still form mammary tumors with similar latency, but much fewer tumors form overall in the absence of the top driver gene.*

Another consideration in large-scale screens is the false positive rate. In the original submission, we had already individually validated four high-penetrance hits (*Fbxw7, Mettl14, Csnk1a1, Plgrkt*) in our *Pten*;*Trp53* mouse model. We have now expanded that to include another high penetrance CRSIPR-A hit, *Mrps18b* but also tested all five of these of these high penetrance genes in the *Rb1*;*Trp53* model. Importantly, across both models, each high-penetrant driver gene significantly accelerated mammary tumor onset compared to safe-harbor targeting controls. Please see the new **Fig. 3a**:

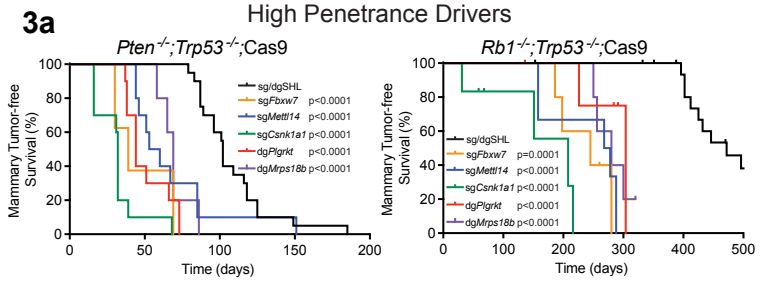


Fig. 3a*: Modulation of high penetrance BLBC driver genes significantly accelerates mammary tumor onset in Pten$^{fl/fl}$;Trp53$^{fl/fl}$;Cas9 mice (left; n=20 mice for sg/dgSHL controls and n=10 mice [5 for each of two independent gRNAs] for each driver gene) and Rb1$^{fl/fl}$;Trp53$^{fl/fl}$;Cas9 mice (right; n=10 mice for sg/dgSHL controls and n=5 mice [2-3 for each of two independent gRNAs] for each gene). Statistical analysis was performed using the Gehan-Breslow-Wilcoxon test for each driver gene compared to the sg/dgSHL controls.*

In addition, to further address the potential of having identified a high rate of false positives, we have now tested the tumorigenic potential of five lower penetrance driver genes (*Fryl*, *Trim60*, *Adam10*, and *Sox5*) in our *Pten*;*Trp53* mouse model. Interestingly, all but one driver (*Fryl*) significantly accelerated tumor development when injected individually into sensitized *Pten*;*Trp53* mice, showing the robustness of our screening platform and its findings. Please see the new **Fig. 3b**:

3b

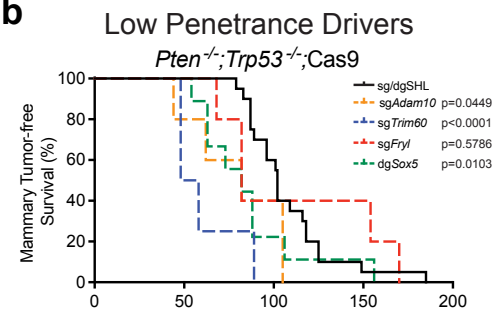


Fig. 3b: *Modulation of most low penetrance BLBC driver genes significantly accelerates mammary tumor onset in Pten$^{fl/fl}$;Trp53$^{fl/fl}$;Cas9 mice (n=20 mice for sg/dgSHL controls and n=5 mice [2-3 for each of two independent gRNAs] for each driver gene). Note that the sg/dgSHL curves for the Pten$^{fl/fl}$;Trp53$^{fl/fl}$;Cas9 mice provided in **a** and **b** are from the same mice. Statistical analysis was performed using the Gehan-Breslow-Wilcoxon test for each driver gene compared to the sg/dgSHL controls.*

A third potential concern of any mouse screen is its relevance to the human setting. To address this, we have now performed a parallel *in vitro* and *in vivo* validation screen using a human breast cancer cell line model. These experiments critically confirmed that (1) the most enriched hits in human cell line xenografts overlap strongly with our high penetrance mouse screen hits and (2) gratifyingly support the strength of our direct *in vivo* approaches as several of our bona fide tumor suppressor genes (*Csnk1a1*, *Ino80*, *Mettl14* and *Ube2n*) which are common essential genes in cell culture strongly deplete in any conventional cell line based screen.

This validation screen and results are described on Line 199-214: "*To functionally validate our primary findings in human breast cancer, we performed parallel in vitro and in vivo CRISPR-KOALA screens in human MCF10A mammary epithelial cells, which were sensitized towards transformation with PIK3CA^{H1047R} and MYC[39] (**Extended Data Fig. 6i**). We designed a 'hit-containing' validation library with gRNAs that target all 90 enrichment-based hits but also contained gRNAs targeting the flanking genes of every hit. Additionally, we cloned an identical 'hit-depleted' library which did not contain any of the hits from our primary screen (**Extended Data Fig. 6i and Supplementary Table 3**). Orthoptic transplantation of cells transduced with the 'hit-containing' library significantly accelerated tumor onset and outgrowth in the mammary gland of immune-compromised NOD scid gamma (NSG) compared to the 'hit-depleted' library (**Fig. 2e**). Next-generation sequencing of the MCF10A 'hit-containing' tumors and cell lines revealed strong enrichment of several hits including FBXW7, APC, SPRED1 and PLGRKT whose oncogenic potential only manifest in vivo (**Fig. 2f and Extended Data Fig. 6j-l and Supplementary Table 3**). Interestingly, several common essential genes, including CSNK1A1, INO80, METTL14 and UBE2N were significantly depleted in MCF10A cell lines and/or tumors (**Fig. 2f and Extended Data Fig. 6j-l and Supplementary Table 3**), but score as strong tumor suppressors in our autochthonous mouse models in vivo (**Fig. 2b-d and Extended Data Fig. 6**), confirming the context-dependency of our screens.*"

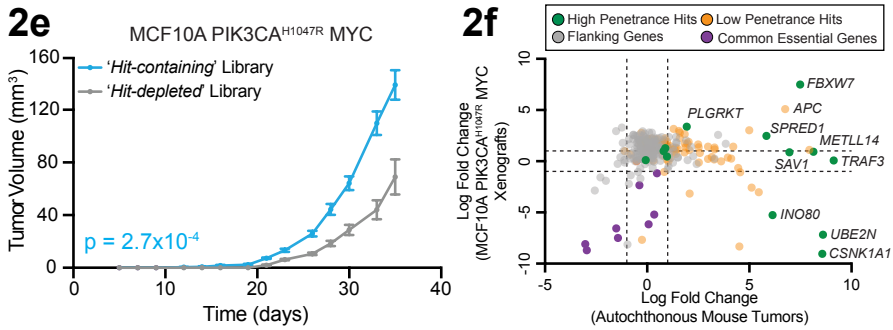
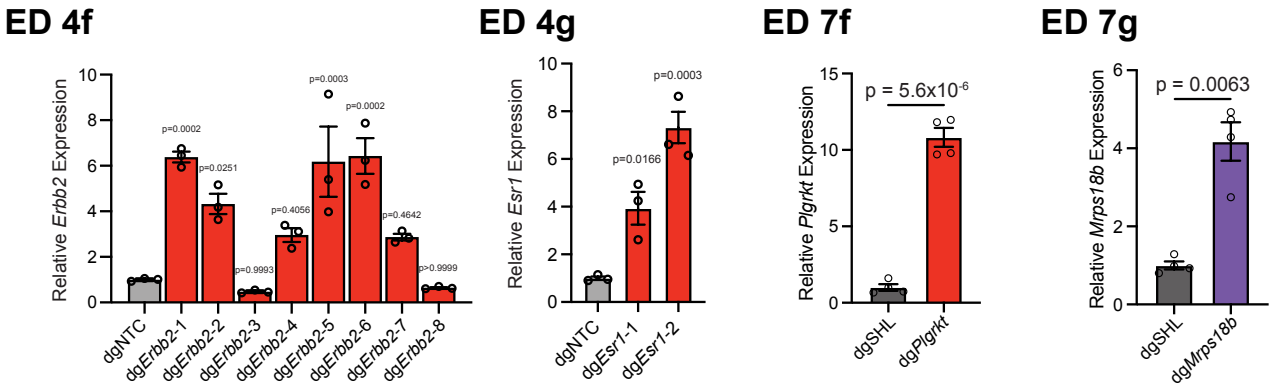


Fig. 2e (left): *Transduction of a library containing all 90 human CNA driver genes and each flanking gene ('Hit-containing') significantly accelerates tumor growth compared to the same library depleted of these driver genes ('Hit-depleted'; n=5 mice and 10 tumors per library). Statistical analysis was performed using a two-tailed unpaired t-test at tumor endpoint. Data are presented as mean ± s.e.m.*
Fig. 2f (right): *Comparison of gene-level log fold changes between the 'Hit-containing' human MCF10A PIK3CA^{H1047R} MYC xenograft screen and autochthonous chromosome-arm mouse screens.*

Lastly, the reviewer raises the potential that our dgRNAs may be inducing genes to incorrect levels as a potential confound of our screen. Firstly, to address gene dosage for activation of genes we would like to highlight that all of our CRISPR-A targeting vectors, which utilize a 14bp dead-guide RNA, achieve a maximum of 10-fold induction in of the target mRNA. This is true for all four genes we have reported in this study including *Plgrkt*, *Mrps18b*, *Esr1* and *Erbb2* (Please see **Extended Data Fig. 4f and g and Extended Data Fig. 7 f and**

g). Importantly this holds true across multiple other gRNAs we have profiled in the lab as well, where we achieve on average a 3-10 fold induction of targets.



Extended Data Fig. 4f and g (left): *RT-qPCR analysis of 2409 cells transduced with dgRNAs (n=3 biological replicates) targeting Erbb2 (**f**) or Esr1 (**g**). Statistical analysis was performed using an ordinary one-way ANOVA with Tukey's correction for multiple testing compared to dgNTC controls. Data are mean ± s.e.m.*
Extended Data Fig. 7f and g (right): *qRT-PCR from Pten$^{-/-}$;Trp53$^{-/-}$ tumors transduced with either dgPlgrkt (**f**) or dgMrps18b (**g**) (n=4 tumors per group). Statistical analysis was performed using a two-tailed unpaired t-test.*

In fact, we purposefully spent a lot of time and effort to create and titrate a system to achieve these 'relatively' low levels of transactivation compared to many other CRISPR-A systems (including Zhou *et al*. *Nature Neuroscience* 2018, Gemberling *et al*. *Nature Methods* 2021, Deng *et al*. *Nature Communications* 2022, or Wu *et al*. *Molecular Cell* 2023), which generate on average 100-5000-fold inductions of target genes. These more conventional CRISPR-A systems are widely regarded to achieve non-physiological levels of activation and are known to induce cellular stress and cytotoxicity (Johnston *et al*. *The CRISPR Journal* 2022, Wu *et al*. *Molecular Cell* 2023, and Liang *et al*. *Nature Communications* 2022). Importantly, our levels of mRNA induction (4x for *Mrsp18b*, 6x for *Erbb2*, and 10x for *Plgrkt*) suffice to significantly accelerate and increase tumorigenesis. As such, we are actually very proud of our system and consider it physiologically relevant or at least reasonable, and the best system established to date to model lower-level copy number gain seen in human tumors. Of note, all these constructs are in the process of being made available through Addgene for the broad science community.

Probably most importantly, the protein expression level observed in our mouse tumors closely matches that of human basal-like breast cancer as shown for tumors that have high PLGRKT expression. We carefully quantified the PLGRKT expression level in human breast cancer TMAs and while all normal tumor-adjacent epithelium showed little to no expression, 17/35 and 18/35 BLBC tumors showed low and high PLGRKT expression, respectively. Importantly, this closely mimicked the expression level and pattern see in our mouse tumors. Moreover, this staining was performed with a PLGRKT antibody that was validated with KO tissue (hence we can be certain that it is specific) and the staining of mouse and human tumors were performed, imaged, and quantified in parallel. As such, we feel confident that we are modelling physiologically relevant protein levels in our mouse model. This data is shown as the new **Fig. 5d**:

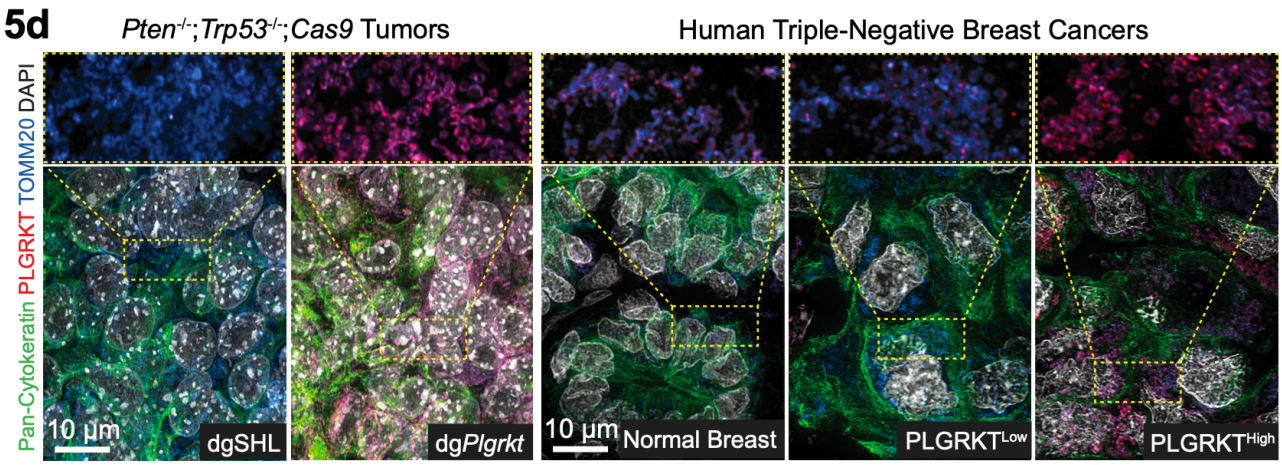


Fig. 5d: *Immunofluorescent staining of autochthonous mouse mammary tumors (n=4 tumors per genotype) and human breast cancer samples (n=35 basal-like breast cancer tumors) and normal adjacent tissue (n=14) showing localization of PLGRKT within the mitochondria. Scale bar, 10 μm.*

Together, re-screening of a chromosome arm depleted of the top hit, follow-up validation of additional high and low penetrance driver genes in multiple mouse BLBC models, cross-species validation screens in human MCF10A cell lines and careful titering of our CRISPR-KOALA platform provides strong evidence that our screen was robust, with low rates of false positives and negatives. We are hopeful that the reviewer will agree that we have sufficiently addressed these points.

4. Their phrasing of how inactivation of their new cancer genes (i.e. tumor suppressor and oncogene hits) leads to an overall reduction in aneuploidy is a little imprecise. They suggest on line 237-239 that alteration of a gene of interest would alleviate the pressure to accumulate that CNA. That make perfect sense and has been shown for before Pten in a mouse SCLC model (McFadden et al, Cell.). But then my interpretation, which might also be theirs, is that when they generate tumors with one of these genotypes of interest that tumor is fit enough on its own that cells that arise with aneuploidy have insufficiency selective advantage to emerge. Otherwise, it is difficult to image how many different single genes “suppress” aneuploidy.

Reply: We thank the reviewer for this excellent and insightful comment. Their interpretation is exactly what we intended to convey, and we apologize if our original phrasing was imprecise. We have now revised the manuscript to be much clearer on this important point.
To be explicit, our model is not that these driver genes mechanistically suppress aneuploidy. Rather, it is a matter of selective pressure. By engineering the key driver gene alteration from the outset, we are providing an immediate fitness advantage, making the incipient tumor cell already "fit enough" to proliferate efficiently.
We think this *a priori* fitness advantage has three major consequences:

1. It obviates the need for the tumor to undergo the time-consuming process of acquiring and selecting for a specific CNA to gain that same advantage.

2. It allows the tumor to bypass the significant fitness cost associated with aneuploidy imposed by the widespread gene dosage imbalance of a large-scale CNA. This is likely true during tumor initiation (which explains the increased tumor multiplicity observed in our engineered models) as well as during tumor growth (which explains the incredible speed of tumor outgrowth in our engineered models).

3. It creates a fitness landscape where cells that subsequently acquire random aneuploidies have no significant selective advantage and are outcompeted. Therefore, as the reviewer correctly surmised, any cells that subsequently acquire aneuploidy in these tumors lack a sufficient selective advantage to emerge and dominate the population. This would explain why none of the tested tumors with engineered drivers eventually develop aneuploidies even when serially transplanted. The aneuploidy cost still outweighs any potential benefit of further aneuploidies.

While we acknowledge that human tumorigenesis is undoubtedly more complex - requiring more genetic alterations and involving pressures from therapy and long-term immune-editing as well as exposures to environmental factors - the fundamental principles our study has uncovered remain relevant. Our extensive

cross-species validation, now included in the manuscript, provides direct evidence that the key driver genes we identified in mice also function in human cells.

We have now dedicated a whole paragraph about this to the discussion and believe this clarification greatly improves the manuscript, and we are grateful for the reviewer's input. Of note, we have now also added the McFadden *et al. Cell* 2014 reference, which we have missed in the original submission, as well as several other references showing that the loss of a whole chromosome arms can be driven by the need to inactivate a single tumor suppressor gene (please see **Lines 447-463**):

"By engineering these key driver gene alterations from the outset, we provide the incipient tumor cell with an immediate and substantial fitness advantage. This allows the tumor to bypass two major hurdles of spontaneous tumorigenesis: the time-consuming process of stochastically acquiring a specific CNA and the significant fitness cost imposed by the resulting gene dosage imbalance. This directly addresses the 'aneuploidy paradox'[2,8], wherein aneuploidy, despite being detrimental to normal cells, is well-tolerated and even selected for in cancer. Our data, consistent with previous findings[67], indicate that engineering these driver genes bypasses this fitness cost while capitalizing on the selective advantage of the altered gene dosage of the driver gene, explaining both the increased tumor multiplicity and the accelerated growth in our models.

Consequently, this initial advantage creates a fitness landscape whereby cells that subsequently acquire random aneuploidies have no significant selective edge and are outcompeted. This explains why our engineered tumors do not develop aneuploidy, even when serially transplanted - the inherent cost of any new chromosomal alteration simply outweighs any potential benefit.

*Building on studies showing that chromosome arm loss can be driven by the selective pressure to inactivate a single canonical tumor suppressor, such as Pten[68] **(Note this is the MacFadden Reference)**, Fbxw7[69] or Smad4[70], our data expands this 'single-driver' model in three key ways: (1) we show it applies to numerous novel tumor suppressor genes identified in this study; (2) is also applies to oncogenes and (3) crucially, that a single events obviates the need for aneuploidies altogether."*

I hope that the reviewer agrees that we have sufficiently addressed this point, but we would be eager to receive further input and/or tips for enhanced clarity.

5. In Figure 4, the authors convincingly disprove the hypothesis that Plgrkt drives tumor initiation through plasminogen activation and hypothesize that Plgrkt increases tumorigenesis through mitochondrial reprogramming. The authors show separately that Plgrkt increases mammary tumor development in vivo and that 4T1 cells in vitro exhibit mitochondrial reprogramming but it would be more convincing if they could provide a link between these two observations (i.e MT disfunction in cancer cells from the autochthonous tumors and/or increased in vivo growth of 4T1 cells with Plgrkt over expression).

Reply: Thank you for this comment, and we agree and have now done a large body of work connecting the *in vitro* observations with the *in vivo* observations.

In the original submission, we had already shown that PLGRKT and its mutants which cannot activate plasminogen and remodel the ECM, are still oncogenic in our autochthonous mouse model (now **Fig. 5c**). As the reviewer suggested, we have now also demonstrated that PLGRKT overexpression significantly increases tumor growth of 4T1 cells as well as human MCF10A PIK3CA^{H1047R} cells *in vivo*. Please see **new Figure 4 a and b**).

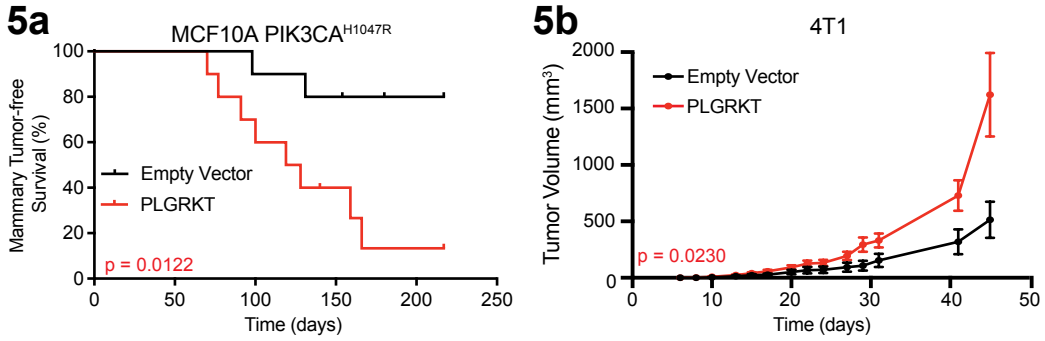
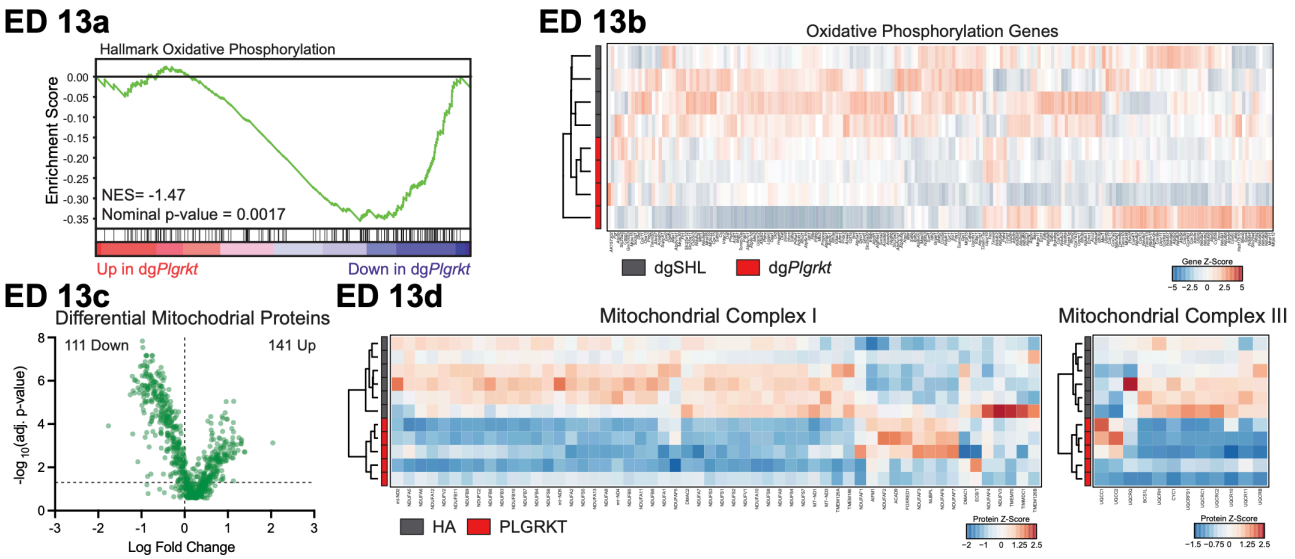


Fig. 5a and b: *Orthotopic transplantation of MCF10A PIK3CA^{H1047R} (**a**) or 4T1 (**b**) cells overexpressing PLGRKT significantly accelerates mammary tumor outgrowth (n=10 tumors per group). Statistical analysis was performed using the Gehan-Breslow-Wilcoxon test (**a**) or two-tailed unpaired t-tests (**b**) compared to empty vector controls. Data are mean ± s.e.m (**b**).*

To link the mitochondrial dysfunctions and reduced oxygen consumption that we observed *in vitro* to our *in vivo* models, we transcriptionally profiled our autochthonous dg*Plgrkt* overexpressing tumors and performed mitochondrial fractionation followed by mass spectrometry using our 4T1 allograft tumors. These experiments showed that *PLGRKT* overexpressing tumors show a pronounced and highly significant downregulation of transcripts and proteins associated with the electron transport chain and oxidative phosphorylation, and in particular Complexes I and III. Please see new **Extended Figure 13a-d**:



Extended Data Fig. 13a (top left): *Gene set enrichment plot of the Hallmark Oxidative Phosphorylation pathway from transcriptional profiling of dgSHL or dgPlgrkt mammary tumors (n=4 tumors per group).*
Extended Data Fig. 13b (top right): *Unsupervised hierarchical clustering of dgSHL and dgPlgrkt mammary tumors based on all genes within the Hallmark Oxidative Phosphorylation gene set. Note that clustering based on these genes stratifies Plgrkt overexpressing tumors from controls.*
Extended Data Fig. 13c (bottom left): *Volcano plot of differential proteins identified from isolated mitochondrial of 4T1 xenografts overexpressing PLGRKT (n=5) to empty vector controls (n=6).*
Extended Data Fig. 13d (bottom right): *Unsupervised hierarchical clustering of mitochondrial proteins from empty vector and PLGRKT 4T1 xenografts mammary tumors based on all detected proteins belonging to Mitochondrial Complexes I (left) and III (right).*

Together, we hope this reviewer agrees that this data provides the requested link between mitochondrial dysfunction in cancer cells from the autochthonous tumors and increased in vivo growth of 4T1 cells with PLGRKT overexpression. We would like to also direct the reviewer to their next comment (**Major Comment #6**), where our additional experiments to deepen our mechanistic understanding of how PLGRKT regulates mitochondrial dynamics has several sections that were performed in our autochthonous mouse and human breast cancer samples *in vivo*.

Lastly, we also found that human BLBC tumors with high PLGRKT expression show significant enrichment of gene sets associated with oxidative damage and stress responses, which now extends our findings to the human setting, which we think is critically important. Please see new **Fig. 5n**:

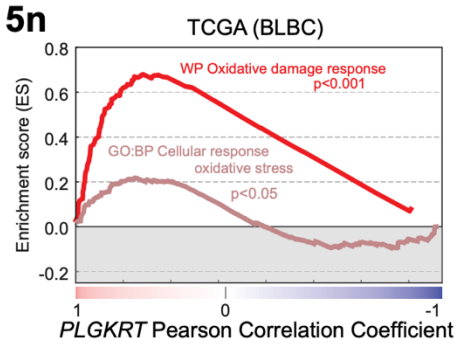


Fig. 5n: *Gene-set enrichment analysis of oxidative response pathways in human BLBC tumors. Genes were ranked by their Pearson correlation coefficient with PLGRKT expression.*

6. While the screen is elegant, it would be of value if the authors could provide some additional mechanistic insights into how Plgrkt changes mitochondrial function and/or additional evidence that this gene is an oncogene in human tumors/cell lines.

We completely agree with this reviewer and have spent a considerable amount of effort to deepen our mechanistic understanding of PLGRKT's oncogenic function.

Importantly, we have now demonstrated that endogenous PLGRKT localizes to the mitochondria within autochthonous mouse tumors, human basal-like breast cancers and mouse and human breast cancer cell lines, using super-resolution microscopy showing that PLGRKT co-localizes with the inner mitochondrial membrane marker TIMM23 and is in close proximity to the outer mitochondrial membrane marker TOMM20 (using a tightly validated PLGRKT antibody). In addition, we found that PLGRKT localizes to the mitochondria, that overexpression of PLGRKT increases it mitochondrial localization using immunofluorescence, super-resolution microscopy and Stimulated Emission Depletion (STED) microscopy in 4T1, MCF10A, mouse tumors and patient breast cancer tissue arrays (please see Lines 367-376 and the new **Fig. 5d and e and Extended Data Fig. 12 g-n**).

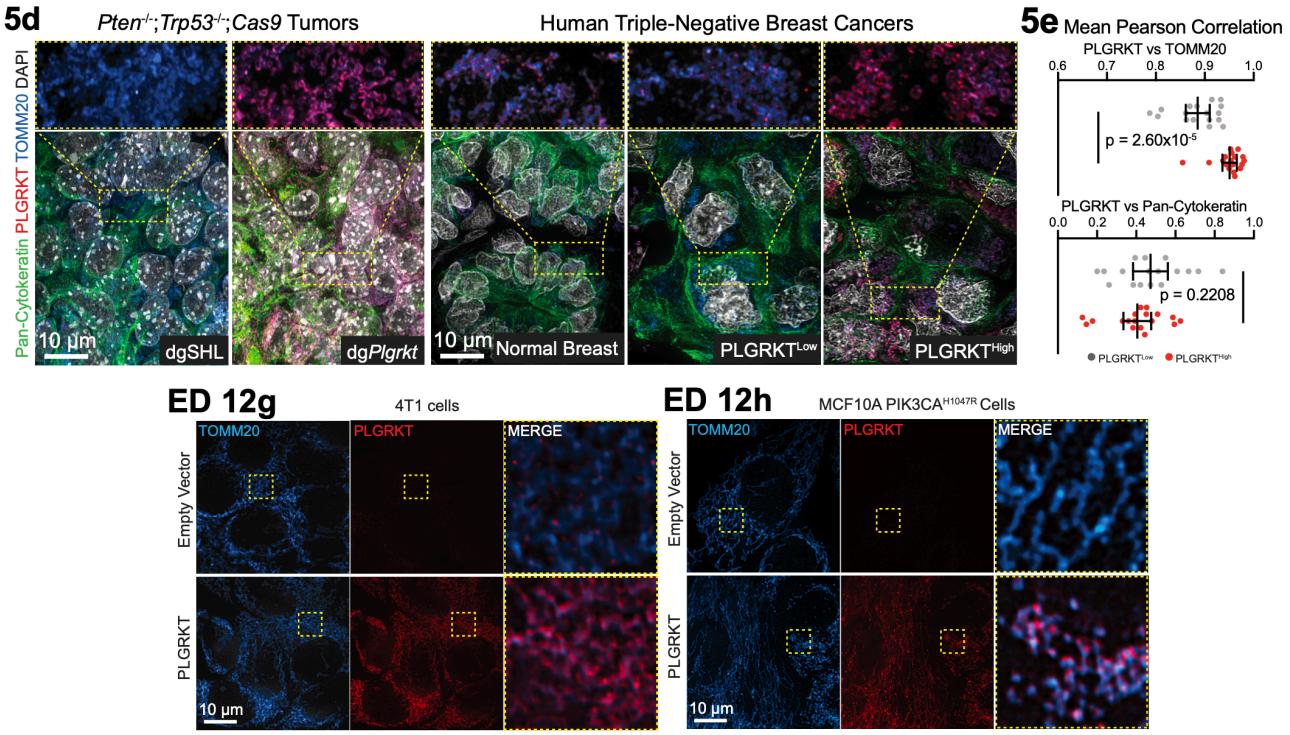
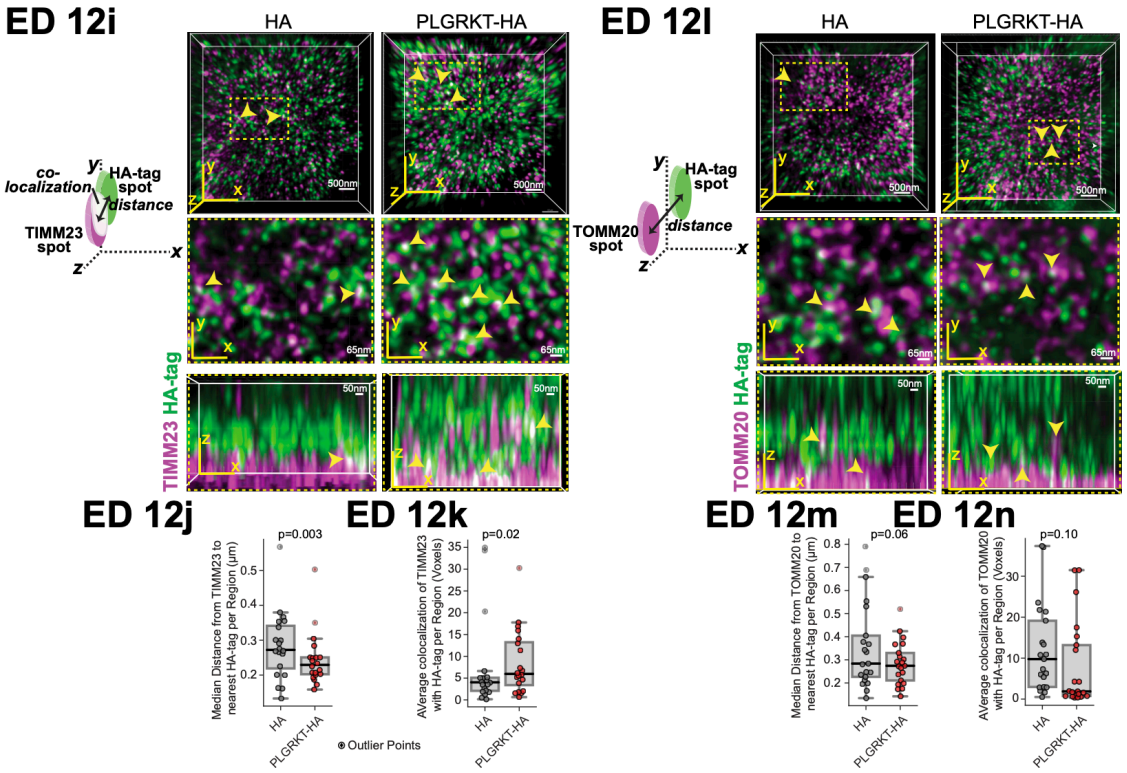


Fig. 5d (top left): *Immunofluorescent staining of autochthonous mouse mammary tumors (n=4 tumors per genotype) and human breast cancer samples (n=35 basal-like breast cancer tumors) and normal adjacent tissue (n=14) showing localization of PLGRKT within the mitochondria. Scale bar, 10 μm.*
Fig. 5e (top right): *Quantification of co-localization of PLGRKT with TOMM20 and Pan-Cytokeratin from the human breast cancer samples in (**d**) stratified by low (grey; n=17) and high (red; n=18) PLGRKT expression. Statistical analysis was performed using two-tailed unpaired t-tests. Data are mean ± s.e.m.*
Extended Data Fig. 12g and h (bottom): *Representative immunofluorescence of 4T1 (**g**) and MCF10A PIK3CA^{H1047R} (**h**)cell lines overexpressing empty vector or PLGRKT (n=3 replicates) showing localization of PLGRKT within the mitochondria. Scale bar, 10 μm.*



Extended Data Fig. 12i-n: *STED microscopy of 4T1 cells overexpressing HA-tagged empty vector or PLGRKT-HA vector, co-stained for inner mitochondrial membrane protein TIMM23 (**i-k**) or outer mitochondrial membrane protein TOMM20 (**l-n**). For **i** and **l**, representative 3D reconstructions of STED data are shown, with axis planes as indicated, showing close proximity and overlap (white signal) between PLGRKT-HA and TIMM23 specifically. For **j** and **m**, quantification of nearest neighbour distance from mitochondrial membrane marker to HA-tag (n=20 regions across 10 cells per condition) is shown. For **k** and **n**, quantification of co-localized voxels between mitochondrial membrane marker and HA-tag (n=20 regions across 10 cells per condition) is shown. Statistical analysis was performed using two-tailed unpaired t-tests. Data are mean with the box extending from the 25th to 75th percentile with whiskers from the minimum to maximum datapoint.*

In addition, we found that PLGRKT overexpression enhances mitochondrial motility in stressed mouse and human cells through Enhanced Super Radial Fluctuation (eSRRF) and live cell imaging. Please see Lines 377-386, the new **Fig. 5f-i and Supplementary Videos 1-10**.

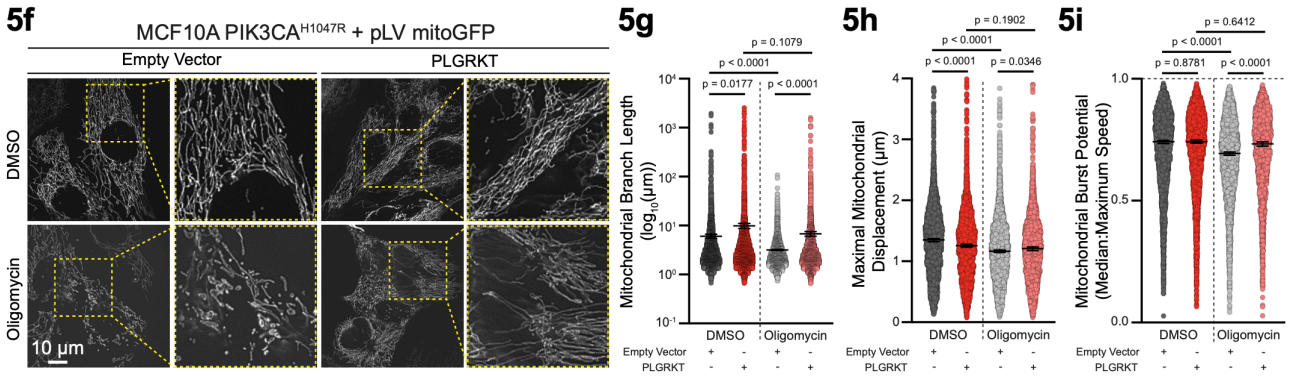
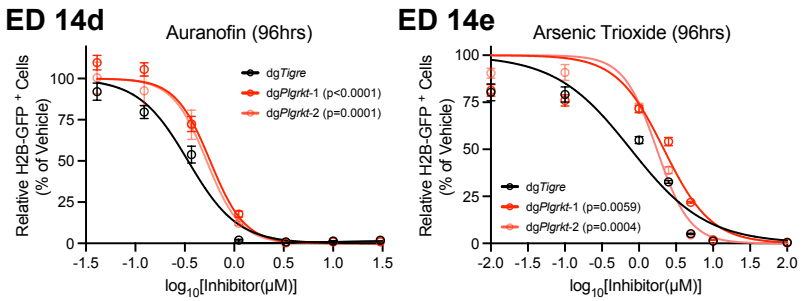


Fig. 5f (left): *Representative images of mitoGFP from enhanced super resolution-radial fluctuation (eSRRF) imaging of MCF10A PIK3CA^{H1047R} cells overexpressing empty vector or PLGRKT and treated with DMSO or oligomycin. Scale bar, 10 µm.*
Fig. 5g-i (right): *Quantification of mitochondrial branch length from eSRRF imaging (**g**), and maximal mitochondrial displacement (**h**) or mitochondrial burst potential (**i**) from mitochondrial live imaging of MCF10A PIK3CA^{H1047R} cells overexpressing empty vector or PLGRKT and treated with DMSO or oligomycin (n=24 fields of view per condition collected over three independent biological replicates). Statistical analysis was performed using or an ordinary one-way ANOVA with Tukey's correction for multiple testing. Data are median ± 95% confidence interval.*

Next, we sought to connect the mitochondrial dysfunction to a cellular mechanism and found that PLGRKT not only conferred a significant growth advantage to oligomycin, but to several other ROS-inducing chemotherapies including auranofin and arsenic trioxide. Please see the new **Extended Data Fig. 14d and e**).



DataExtended Data Fig. 14d and e: *Relative viability of 4T1 cells expressing dgTigre, dgPlgrkt-1 or dgPlgrkt-2 treated with Auranofin (**d**) or Arsenic Trioxide (**e**) for 96 hours. Statistical analysis was performed by fitting normalized response curve with a variable slope and comparing differences in fits. Data are mean ± s.e.m*

This led us to speculate that PLGRKT overexpression regulates resistance and promotes cell survival when challenged with elevated levels of ROS. We now show that PLGRKT itself regulates mitochondrial superoxide and total cellular ROS detoxification. Please see the new **Fig. 5l and m and Extended Data Fig 14 a-c.**

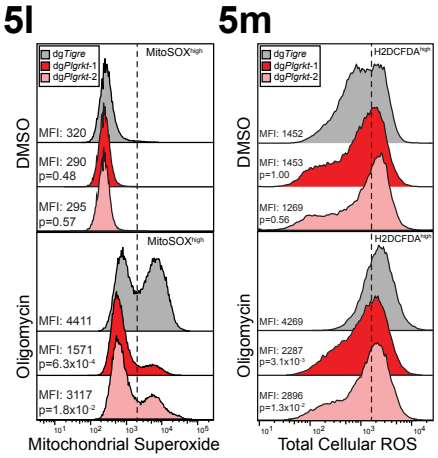
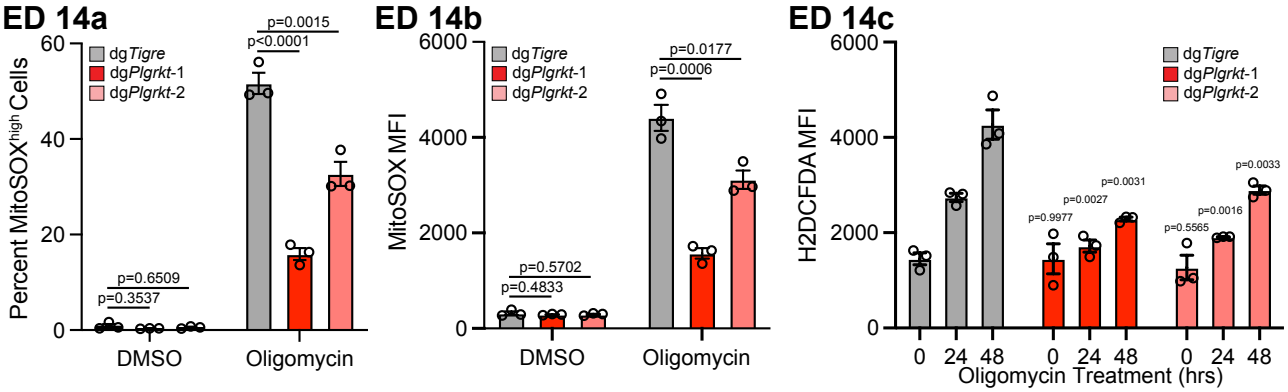


Fig. 5l and m: *Representative histograms of the mean fluorescence intensity (MFI) for mitochondrial superoxide (mitoSOX; **l**) or total cellular ROS (H2DCFDA; **m**) as measured by flow cytometry in 4T1 cells expressing dgTigre, dgPlgrkt-1 or dgPlgrkt-2 and treated with DMSO or oligomycin (n=3 replicates per condition). Statistical analysis was performed two-tailed unpaired t-tests compared to empty vector controls.*



Extended Data Fig. 14a-c: *Quantification of the percentage of MitoSOX^{High} (**a**), MitoSOX Mean Fluorescence Intensity (MFI; **b**), and H2DCFDA MFI (**c**) in 4T1 cells expressing dgSHL, dgPlgrkt-1 or dgPlgrkt-2 and treated for 48 hours with DMSO or Oligomycin (**a** and **b**) or across a 48 hour time course (**c**). Statistical analysis was performed using a two-tailed unpaired t-test Data are mean ± s.e.m.*

Lastly, treatment of PLGRKT overexpressing cell lines with oligomycin or hydrogen peroxide to induce ROS resulted in a significant elevated expression of antioxidant genes, which we confirmed by qRT-PCR. Please see the new **Fig. 5o and Extended Data Fig. 14g**.

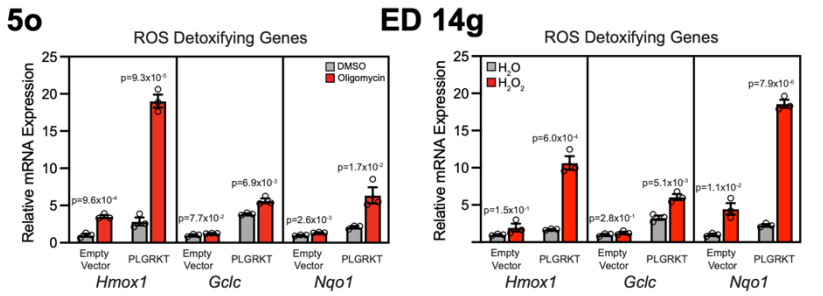


***Fig. 5o and Extended Data Fig. 14g**, qRT-PCR analysis for several antioxidant target genes from 4T1 cells overexpressing empty vector or PLGRKT at baseline or treated with oligomycin (**5o, left**) or H_2O_2 (**ED14g, right**) for 12 hours (n=3 samples per group). Statistical analysis was performed using a two-tailed unpaired t-test Data are mean ± s.e.m.*

Overall, we believe that we have significant enhanced our molecular understanding of how PLGRKT functions in breast cancer independent of its known role as a plasminogen binding protein. Together, our new data suggests that PLGRKT is a potent proto-oncogene on chr9p, that drives tumorigenesis through its increased localization to the mitochondria, dysregulation of the electron transport chain, elevated ROS detoxification, and promoting healthier, stress-resistant mitochondria. Please see the updated summary schematic of the functions of PLGRKT as the new **Fig. 5p**:

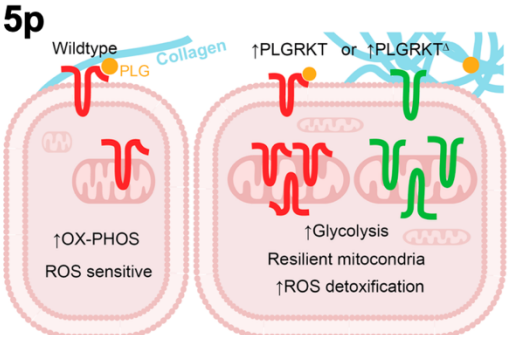


***Fig. 5p:** Schematic of the canonical and non-canonical functions for PLGRKT in promoting tumorigenesis.*

7. While the authors do their screens in both the Pten/p53 and Rb1/p53 models they do not do any validation or follow-up studies in the Rb1/p53 model. Would this change any of their ultimate conclusions, I am not sure. It would seem reasonable to look at at least their top hit in the Rb1/p53 model and its effect on aneuploidy.

As highlighted in response to **Major Comment #3**, we have now performed extensive validation of all hits originally presented in the first submission in the *Rb1*;*Trp53* model and extended our analysis it to another high penetrance CRISPR-A hit, Mrps18b. Please see new **Fig. 3a**:

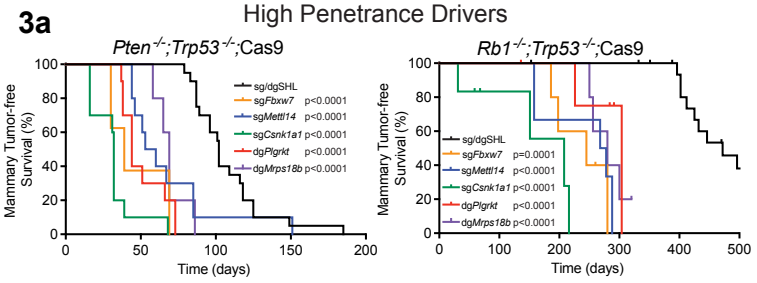


***Fig. 3a**: Modulation of high penetrance BLBC driver genes significantly accelerates mammary tumor onset in Pten$^{fl/fl}$;Trp53$^{fl/fl}$;Cas9 mice (left; n=20 mice for sg/dgSHL controls and n=10 mice [5 for each of two independent gRNAs] for each driver gene) and Rb1$^{fl/fl}$;Trp53$^{fl/fl}$;Cas9 mice (right; n=10 mice for sg/dgSHL controls and n=5 mice [2-3 for each of two independent gRNAs] for each gene). Statistical analysis was performed using the Gehan-Breslow-Wilcoxon test for each driver gene compared to the sg/dgSHL controls.*

To be thorough, we have also investigated how *Rb1*;*Trp53* driver gene tumors impacted aneuploidy as suggested by the reviewer. To start, we performed whole genome sequencing (WGS) on 23 new tumors, including four new *Pten*;*Trp53* tumors overexpressing *Mrps18b* (a new hit we have added during revisions) and a larger cohort of 19 *Rb1*;*Trp53* tumors with various driver genes (*Fbxw7* (n=2), *Mettl14* (n=1), *Csnk1a1* (n=3), *Plgrkt* (n=1) and *Mrps18b* (n=2)) and controls (sg/dgSHL; n=10), for a total of 89 WGS profiles.

Importantly, the additional data from the *Pten*;*Trp53*;dg*Mrps18b* and the entire *Rb1*;*Trp53* cohort confirms our central hypothesis that manipulating these driver genes overcomes the need to accumulate aneuploidy. In fact, *Rb1*/*Trp53* tumors actually show significantly higher baseline aneuploidy ($p = 2.09 \times 10^{-5}$) compared to *Pten*;*Trp53* tumors, likely due to the increased genetic instability upon loss of Rb1 and the significantly longer latency of this model. Still, *Rb1*/*Trp53* tumors with our driver hits have significantly less aneuploidy and a stochastic pattern of gains and losses, providing powerful evidence of our paper's main conclusion. Please see new **Fig 4d** and **Extended Data Fig. 11e and h**):

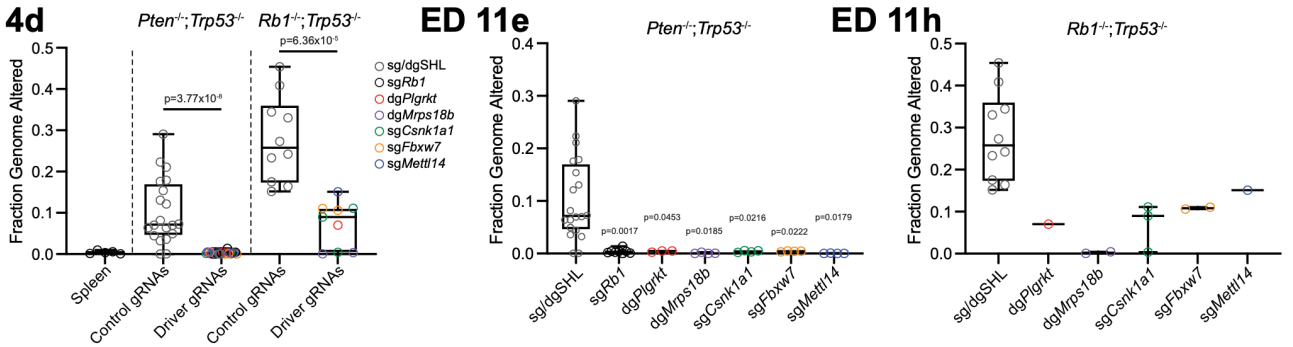


Fig. 4e (left): *Fraction genome altered from spleens (n=5), Pten$^{-/-}$;Trp53$^{-/-}$ tumors transduced with Control gRNAs (sg/dgSHL; n=20) or CNA driver gene gRNAs (n=27), and Rb1$^{-/-}$;Trp53$^{-/-}$ tumors transduced with Control gRNAs (sg/dgSHL; n=9) or CNA driver gene gRNAs (n=8) Individual driver genes are represented by the coloured circles within the boxplot. Statistical analysis was performed using a two-tailed unpaired t-tests. Data are mean with the box extending from the 25th to 75th percentile with whiskers from the minimum to maximum datapoint pooled from across four independent whole genome sequencing runs.*

Extended Data 11e (middle): *Fraction genome altered from Pten$^{-/-}$;Trp53$^{-/-}$ tumors transduced with sg/dgSHL (n=20), sgRb1 (n=8), sgFbxw7 (n=4), sgMettl14 (n=4), sgCsnk1a1 (n=4), dgMrps18b (n=4) or dgPlgrkt (n=3). Statistical analysis was performed using a two-tailed unpaired t-tests compared to sg/dgSHL. Data are mean with the box extending from the 25th to 75th percentile with whiskers from the minimum to maximum datapoint pooled from across four independent whole genome sequencing runs.*

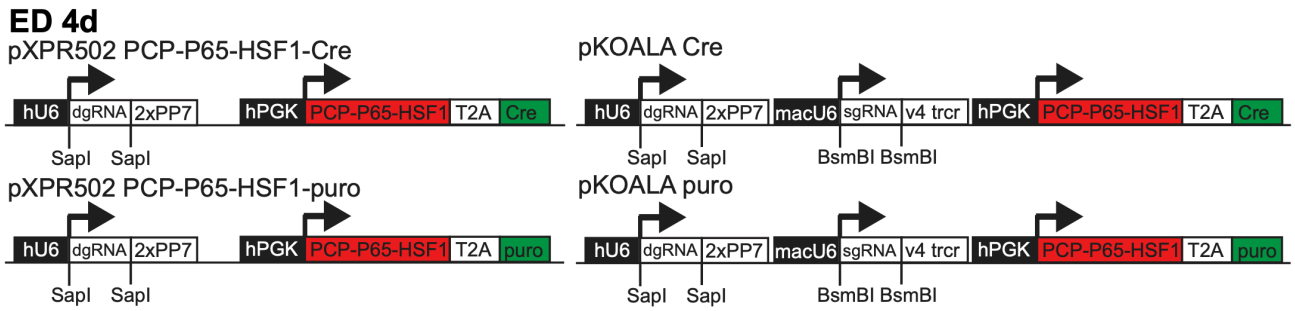
Extended Data 11h (right): *Fraction genome altered from Rb1$^{-/-}$;Trp53$^{-/-}$ tumors transduced with sg/dgSHL (n=10), sgFbxw7 (n=2), sgMettl14 (n=1), sgCsnk1a1 (n=3), dgMrps18b (n=2) or dgPlgrkt (n=1). Statistical analysis was performed using a two-tailed unpaired t-tests compared to sg/dgSHL. Data are mean with the box extending from the 25th to 75th percentile with whiskers from the minimum to maximum datapoint pooled from across four independent whole genome sequencing runs.*

Minor Points:

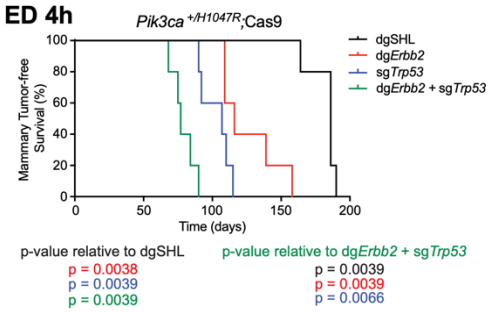
1. Their screening platform allows gene inactivation and gene activation, however they don't really employ this as a "bidirectional" screen. For me that would mean that they inactivation and activate the same genes, and get reciprocal effects. Line 134 claims "simultaneous"(line 134) which is not exactly correct. They weren't in the same mice.

Reply: We thank the reviewer for pointing out our overstatement at the beginning of the results section and we do agree that the screen for the top 10 chromosome arm-level alterations was not explicitly performed in a bidirectional manner where the same gene is modulated in both directions. To clarify our overstatement, we have removed the words bidirectional and simultaneous from the following sentences to read: "*resulting in a bidirectional functional genomics CNA screen across a fifth of the genome*." (now on Line 153) and "*enabling us to screen for oncogenic or tumor suppressive cancer driver genes within CNAs.*" (now on Lines 115-116).

We did however want to highlight the power of our new CRISPR-KOALA technology for allowing the multiplexed combination of CRISPR-knockout and activation within the same cell line both *in vitro* and *in vivo*, as we strongly believe this is one of the strengths of our new approach for in functional genomics. To this end we have utilized developed and added to the manuscript a version of the CRISPR-KOALA backbone which allows for the expression of both a 20 bp and 14 bp guide RNAs simultaneously which allows for knockout and activation of two different genes simultaneously in the same cell. As a proof of concept for this system we simultaneously overexpressed *Erbb2* and knocked out *Trp53* in sensitized *Pik3ca*^{H1047R};*Cas9* mice, significantly accelerating mammary tumor onset compared to either perturbation alone. To show this we have added the following to the manuscript "*Furthermore, CRISPR-KOALA may be used to simultaneously activate and knockout multiple genes within the same cell, with combined amplification of Erbb2 and knockout of Trp53 significantly accelerating tumor onset compared to either perturbation alone (**Extended Data Fig. 4d and h**).*" *on lines 135-138.*



Extended Data Fig. 4d: *Schematic of the CRISPR-A and dual lentiviral constructs used in the CRISPR-KOALA platform.*



Extended Figure 4h: CRISPR-A mediated activation of *Erbb2*, CRISPR-KO mediated knockout of *Trp53*, and CRISPR-KOALA mediated combined perturbation accelerates mammary tumor onset in *Pik3ca*$^{+/H1047R}$;*Cas9*^{het} mice (n=5 mice per group). Statistical analysis was performed using the Gehan-Breslow-Wilcoxon test with the indicated comparisons provided under the graph.

We have also performed a simultaneous CRSIPR-KOALA experiment in our human MCF10A PIK3CA MYC cells, *in vitro* and *in vivo*, which validated our mouse screens. Please see new **Fig. 2e and f and Extended Data 6i-l**.

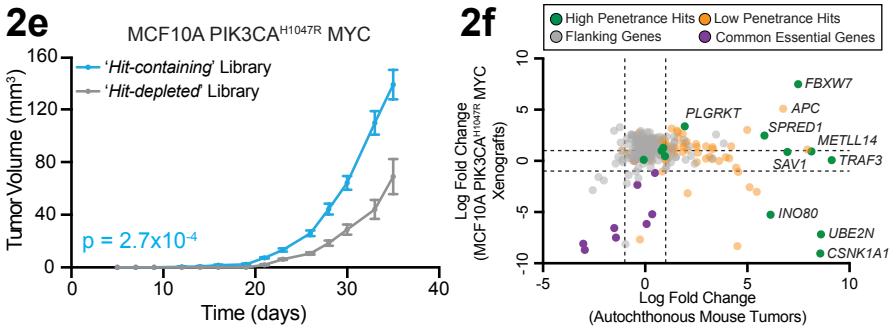


Fig. 2e (left): Transduction of a library containing all 90 human CNA driver genes and each flanking gene ('Hit-containing') significantly accelerates tumor growth compared to the same library depleted of these driver genes ('Hit-depleted'; n=5 mice and 10 tumors per library). Statistical analysis was performed using a two-tailed unpaired t-test at tumor endpoint. Data are presented as mean ± s.e.m.
Fig. 2f (right): Comparison of gene-level log fold changes between the 'Hit-containing' human MCF10A PIK3CA^{H1047R} MYC xenograft screen and autochthonous chromosome-arm mouse screens.

Additionally, we have prepared all of the plasmids required for *in vitro* and *in vivo* CRISPR-KOALA screening and have begun the process of depositing them at Addgene to serve as a public resource. In addition, we are in the process of writing a detailed methods paper describing library design, screening parameters, NGS sequencing protocols and analysis pipelines for publication in JOVE, that will allow all our colleagues to perform these simultaneous KOALA screens in vitro and in vivo in transplant and autochthonous models.

2. Line 250 I don't think that "discrepancy" is the correct word here.

Reply: We thank the reviewer for pointing out the awkward use of the word "discrepancy", which does not fit in the sentence or convey our point properly. To clarify this section, we have replaced this section (now on Lines 296) with the following: "*We postulate that this is due to the genetic ablation of Trp53 within our mouse models, which eliminates the need to select for CNA events that inactivate the p53 pathway.*"

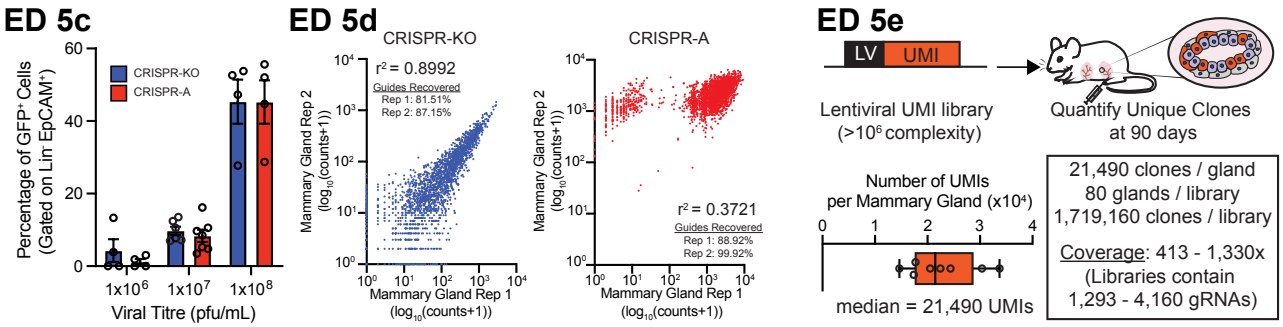
3. I think that they should be more careful claiming 1000x coverage. Most cells have no ability to become a tumor at all ever. So it seems disingenuous to count those.

Reply: We agree with the reviewer and that not all cells have the potential to form tumors, but once they have lost p53 and PTEN, they should be sensitized to do so and upon another gene alteration, they might be able

to form a tumor, which is actually the very experiment we are performing. As such, we feel it is important to calculate the coverage as transduced cells in the beginning of the experiment. Having said that, we agree that only long-term surviving cells/tissue stem cells would have a real chance to do so. As such, we have now performed several experiments to ascertain the long-term survival of infected cells.

To be thorough and calculate the true multiplicity of clones achieved upon intraductal injections into the mammary gland, we have now also used a lentiviral library with unique molecular identifiers at a complexity of >1M unique barcodes. We injected this library into eight independent mammary glands and quantified clones with unique barcodes 3 months later by next-generation sequencing. This complex library confirmed that we obtain 20,000 independent UMI clones per mammary gland at the given 15% transduction rate, which translates to a coverage of between 400 and 1330 clones for any given sgRNA within every library tested. We have added this information to the Methods section.

For completeness, all this data has now been included as new **Extended Data Fig. 5c-e**. We have also updated the text to refer to these important MOI optimizations as follows: "*We titrated the libraries to a 15% transduction level to minimize double infections while ensuring sufficient generation of clones for high content screening (**Extended Data Fig. 5c-e**)[25,34]. We then generated cohorts of 20 mice per library and genotype (440 total mice), achieving an overall coverage of between 413 and 1330 clones per gRNA (**Extended Data Fig. 5e**), resulting in a functional genomics CNA screen across a fifth of the genome.*" on Lines 149-154.



Extended Data 5c (left): *Representative in vivo titering of CRISPR-KO and CRISPR-A viruses directly in the mouse mammary gland (n=4-6 mammary glands per condition). Data are mean ± s.e.m.*
Extended Data 5d (middle): *gRNA representation correlations between mammary glands transduced with CRISPR-KO (left) or CRISPR-A (right) libraries.*
Extended Data 5e (right): *Transduction of mouse mammary glands (n=8) with a barcoded lentiviral library followed by next generation sequence to quantify the number of individually infected epithelial clones per injection. A summary of the coverage calculations is provided in the box. Data are mean with the box extending from the 25th to 75th percentile with whiskers from the minimum to maximum datapoint.*

4. Line 179, what do they mean by tumor penetrance, it seems that essentially all mice get tumors at some point.

Reply: To clarify this point, nearly all mice eventually develop mammary tumors in the *Pten*$^{-/-}$;*Trp53*$^{-/-}$ background due to a tumor onset across all models of approximately 150 days. However, mice develop tumors with a median onset of 400 days in the *Rb1*$^{-/-}$;*Trp53*$^{-/-}$ model around which point many of the injected mice reach humane endpoint, established by our Animal Care Committee for age-related issues including skin ulcerations, distended abdomens and general poor body condition. This is reflected in **Fig. 1c** where for several libraries mice were censored from the Kaplan-Meier plot, resulting in less than 100% of mice developing tumors prior to end point, which we defined as tumor latency. To make this clear with regards to how we generated the survival curves we have now added the following to the Materials and Methods section: "*All injected animals were included in the survival analyses and mice which reached humane endpoint for non-related concerns were censored from the study.*" on Lines 751-753.

5. I am not sure what Figure 4f is showing at all. Is there a fractionation control in there? Are the mutants behaving as expected or something?

Reply: The intention of this panel was to confirm that both wildtype Plgrkt and its corresponding mutants were in fact localized to the mitochondria. While the Western blot confirms this observation, we agree with the reviewer that it is not the most convincing way to represent this data and is not suitable for publication. We

have now performed super resolution immunofluorescence and Super Transmission Electron Microscopy in both mouse and human breast cancer cell lines and tumor sections confirming that PLGRKT is localized to the mitochondria (**Fig. 5d and Extended Data 12g-n**). Additionally, we have performed subcellular mitochondrial fractionations followed by mass spectrometry and identified that indeed PLGRKT and its plasminogen-deficient binding mutants are localized to the mitochondria (**Extended Data Fig. 12f**). We believe that these new results much more clearly convey this important observation, and we have removed the initial Western blot from the manuscript.

6. While not the direct point of their study, the results of whole arm aneuploidies (or other large CNAs) is the change in expression of many genes and it's the summation of those that needs to lead to an overall change in fitness. If a loss contains a TSG but also many essential genes then that might not be selected for. Not sure if they can better articulate that their finding represent some of the major drivers but that they overall effect on fitness will be influences by many other genes in the same region. Perhaps altering combination of genes is a future direction for the field.

Reply: We thank the reviewer for this point, which we have now discussed in detail in the discussion on Lines 447-463:

"By engineering these key driver gene alterations from the outset, we provide the incipient tumor cell with an immediate and substantial fitness advantage. This allows the tumor to bypass two major hurdles of spontaneous tumorigenesis: the time-consuming process of stochastically acquiring a specific CNA and the significant fitness cost imposed by the resulting gene dosage imbalance. This directly addresses the 'aneuploidy paradox'[2,8], wherein aneuploidy, despite being detrimental to normal cells, is well-tolerated and even selected for in cancer. Our data, consistent with previous findings[67], indicate that engineering these driver genes bypasses this fitness cost while capitalizing on the selective advantage of the altered gene dosage of the driver gene, explaining both the increased tumor multiplicity and the accelerated growth in our models.

Consequently, this initial advantage creates a fitness landscape whereby cells that subsequently acquire random aneuploidies have no significant selective edge and are outcompeted. This explains why our engineered tumors do not develop aneuploidy, even when serially transplanted - the inherent cost of any new chromosomal alteration simply outweighs any potential benefit.

Building on studies showing that chromosome arm loss can be driven by the selective pressure to inactivate a single canonical tumor suppressor, such as Pten[68], Fbxw7[69] or Smad4[70], our data expands this 'single-driver' model in three key ways: (1) we show it applies to numerous novel tumor suppressor genes identified in this study; (2) is also applies to oncogenes and (3) crucially, that a single events obviates the need for aneuploidies altogether.

Our data also complement and extend previous work[7,9,12,71-77], supporting the existence of cancer gene islands - genomic regions containing clustered cancer driver genes. Our findings suggest a slightly revised cancer gene island hypothesis: typically, there are only one to two strong cancer driver genes among less penetrant ones located within the gained/lost chromosome arms. The extent to which these cancer driver genes cooperate and lead to cumulative effects within a given chromosome arm remains to be determined[72]."

7. Line 358, "extent" not "extend"

Reply: We apologize for this typo and have corrected the mistake which is now located on Line 467.

Reply to Reviewer #3:

Genome instability and aneuploidy are important hallmarks of cancer. Specific arm level copy number alterations across tumor types and focal amplifications driving classic oncogenes have long suggested that these events play important roles in cancer formation and increasing fitness. Until recently however, deconvolution of target genes on arm-level CNA has been challenging due to the large number of affected genes.

This manuscript functionally characterizes oncogenic and tumor suppressor genes in the context of basal-like breast cancer (BLBC). The authors focus on 10 arm-level CNA events that are both specific to and prevalent in BLBC. They deploy their previously described CRISPR KOALA technique that multiplexes guides allowing for gene activation and inactivation using a single CRISPR Cas9 system by intraductal mammary injection in two genetic backgrounds (TP53/PTEN and RB/TP53). The sg/dgRNA libraries targeting 3,752 genes yielded 90 hits across 1,478 mouse tumors. The authors functionally characterize one of these hits, Plgrkt, in the murine ecoderm of PIKC3CA-H1047R animals and describe induced squamous cell carcinomas and changes of mitochondrial reprogramming.

Overall, this research paper is a solid body of work describing an approach to interrogate genes affected by copy number alterations in vivo and has several preliminary, yet interesting conclusions. However, considering that the technology (KOALA) and approach (in vivo positive selection functional genomic screens to validate genes targeted by copy number events) have been previously reported, the paper's ultimate impact is predicated on the depth of discovery and mechanistic follow-up, generalizability and therapeutic relevance.

These elements are more preliminary in the current manuscript, tempering enthusiasm.

Reply: We sincerely thank the reviewer for their insightful and constructive comments, which have been invaluable in strengthening our manuscript. We have addressed all of the points in full in the revised version.

We would also like to take this opportunity to clarify a key aspect of our work. We appreciate the reviewer's contextualization of our study; however, we wish to respectfully note that the CRISPR-KOALA platform is a new technology that we introduce for the first time in this manuscript.

We have previously performed *in vivo* CRSIPR-KO studies (e.g. Loganathan *et al*., Science 2020; Langille *et al*., Cancer Discovery 2022, Yanchus *et al*., Science 2022; Martinez *et al*., Nature Communications 2024), but never on this scale and never with CRISPR-A and never combined. We apologize if this was not made sufficiently clear. Specifically, we believe the ability to perform multiplexed, simultaneous gene inactivation and CRISPR-mediated activation using a single Cas9 system is a significant technological advance. To our knowledge, this is the first time a comprehensive CRISPR activation screen has been deployed to functionally identify driver oncogenes from large copy-number altered regions directly *in vivo,* which is also where we see the novelty (in addition to the identification and characterization of dozens of novel cancer driver genes). We believe this novel application represents a powerful new paradigm for deconvoluting the complex genetics of cancer.

Major Points:

1. Specificity of the hits and models to individual CN events: for each of the 90 oncogenic/tumor suppressor hits, it would be helpful to see the enrichment of each targeting guide across each of the individual tumors and genetic backgrounds analyzed.
Each individual tumor and genetic background harbors different arm level copy number gains and we cannot fully evaluate the specificity of the hits in regards to the central hypothesis (mutual exclusivity for copy number change), beyond the one example in Figure 3 (MDM4). It seems somewhat strange that tumors without the oncogene (e.g. METTL14) are highly aneuploid while those with the oncogene are stable. This does not appear to fully validate the hypothesis that the target specifically removes the selective pressure for one specific chromosomal event.
In a reciprocal manner, evaluating the arm-level CNAs at baseline across many individual tumors in each murine genetic background would help us understand the relevance of each murine model to the human arm-level data (e.g. a version of extended data 2b could be shown in main figure for both human and murine data).

Reply: We sincerely thank the reviewer for this insightful comment and valuable suggestions. We have broken down our replies to this comment to facilitate ease of reading:

1.1 Specificity of the hits and models to individual CN events: for each of the 90 oncogenic/tumor suppressor hits, it would be helpful to see the enrichment of each targeting guide across each of the individual tumors and genetic backgrounds analyzed.

We have now provided all the enrichment of each targeting across each of the individual tumors and genetic background (Pten/p53 and Rb1/p53) in **Supplemental table 3**.

We have also performed **a comprehensive bioinformatic re-analysis** of our entire dataset, using a model-based approach to complement our initial rank-based analysis. These new analyses have not only strengthened our original conclusions and provided statistical power to our large *in vivo* screening dataset but have also yielded significant new biological insights. To clarify the approach, we took for these analyses, we have provided a schematic as the new **Fig. 2a**.

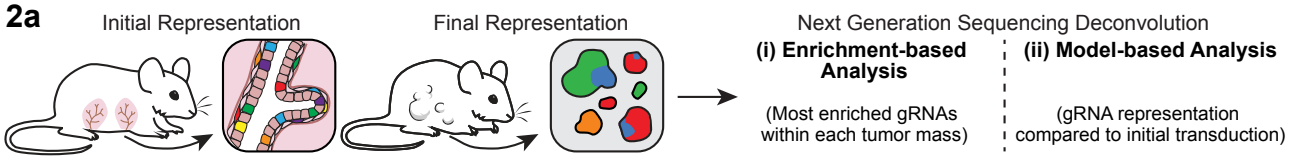


Fig. 2a: *Schematic representation of the in vivo CRISPR-KOALA screen, library deconvolution and analysis pipelines.*

Our new bioinformatic analyses described below were designed to address the enrichment of each targeting guide:

*'We also analysed our screen using a MAGeCK-based statistical model[37], which compares the gRNA across all tumors to the initial viral library (**Fig. 2a**). While this approach confirmed our high-penetrance hits, its increased sensitivity to subclonal gRNA representation and potential cooperative effects between gRNAs identified 78 candidate lower penetrance driver genes (**Fig. 2d and Extended Data Fig. 6c and Supplementary Table 3**). A key advantage of this approach is the detection of significantly depleted gRNAs, which showed the expected depletion of core essential genes as defined by DepMap[38] (**Fig. 2d and Extended Data Fig. 6d and Supplementary Table 3**). Underscoring the importance of performing these screens in vivo, this analysis also revealed that several genes classified as core essential in cell culture, such as Ube2n, Csnk1a1, Mettl14, and Ino80, instead function as potent tumor suppressors in the mammary gland (**Fig. 2d and Extended Data Fig. 6d and Supplementary Table 3**).'*

Importantly this approach accounts for the difference in sgRNA representation within the original library and provides statistics for every gene across all tumors and also allowed us to run rigors statistics and we have now added the p-value for every hit into the Figure panel 2b and the Supplemental Table 3. As requested by this reviewer, it also allowed us to calculate a LFC for each sgRNA, which is now presented in **Supplemental Table 3** for all genes and for the top hits in the **new Fig. 2d**.

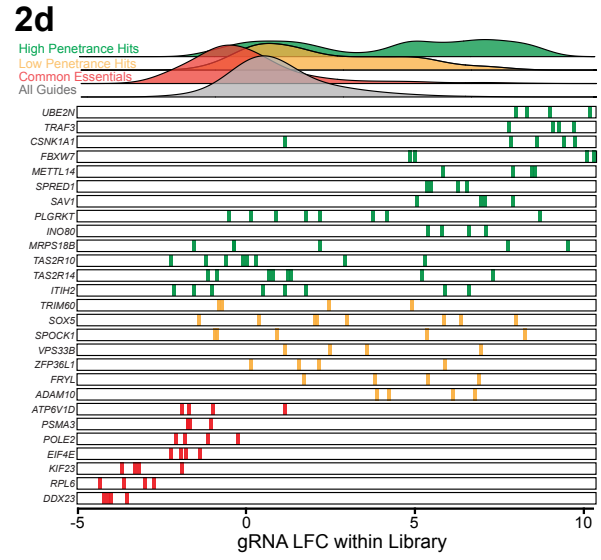


Fig. 2d: *Histogram distribution of the log fold change for each gRNA targeting high penetrance, low penetrance and common essential genes along with the overall distribution of all genes within the library (top). The log fold change of each gRNA targeting representative genes for each category is shown (bottom).*

We hope the reviewer agrees that now a reader can easily see the difference between high- and lower-penetrance hits, the distribution of the whole library and the distribution of common essential genes, that drop out, as expected.

1.2 Each individual tumor and genetic background harbors different arm level copy number gains and we cannot fully evaluate the specificity of the hits in regards to the central hypothesis (mutual exclusivity for copy number change), beyond the one example in Figure 3 (MDM4). It seems somewhat strange that tumors without the oncogene (e.g. METTL14) are highly aneuploid while those with the oncogene are stable. This does not appear to fully validate the hypothesis that the target specifically removes the selective pressure for one specific chromosomal event.

Just to clarify, we have never performed WGS on tumors from the screen. We have only sequenced baseline tumors from Pten/p53-mutant mice, as well as tumors that were initiated by transducing these Pten/p53-mutant mice with individual guide RNAs targeting hits from the screen. While the baseline Pten/p53-mutant tumors showed widespread aneuploidy, the Pten/p53-mutant tumors with engineered gain or loss of hits were euploid. Figure 3 shows that baseline tumors do not show gains of the mouse homologous chr1q arm, which was at first unexpected, as chr1p amplification are very frequent in human tumors. We then found that a likely explanation for this is the fact that Jason Sheltzer's group identified MDM4 as the driving oncogene on chr1q, which inactivates p53. As our tumors are already p53-deficient, we hypothesized that this could be a very likely explanation.

We agree with the reviewer that the finding that tumors without additional driving oncogene/tumor suppressor genes are aneuploid, but those with additional drivers are euploid was - at least at first - very surprising. We thus have repeated the whole WGS analysis in the Rb1/p53-mutant mouse model, using again baseline Rb1/p53-mutant tumors as well as Rb1/p53-mutant tumors with engineered additional hits. Importantly, we see a very similar pattern in that Rb1/p53-mutant with additional driver alterations exhibits a much less complex tumor genome.

We have now performed whole genome sequencing (WGS) on 23 new tumors, including four new Pten;Trp53 tumors overexpressing *Mrps18b* (a new hit we have added during revisions) and a large cohort of 19 Rb1;Trp53 tumors with various driver genes (*Fbxw7* (n=2), *Mettl14* (n=1), *Csnk1a1* (n=3), *Plgrkt* (n=1) and *Mrps18b* (n=2)) and controls (sg/dgSHL; n=10), for a total of 89 WGS profiles. Importantly, the data from *Mrps18b* and the whole Rb1;Trp53 cohort confirms our central hypothesis that manipulating these driver genes overcomes the need to accumulate aneuploidy. In fact, Rb1/Trp53 tumors actually show significantly higher baseline aneuploidy ($p = 2.09x10^{-5}$) compared to Pten;Trp53 tumors, likely due to the increased genetic instability upon loss of Rb1 and the significantly longer latency of this model. Still, Rb1/Trp53 tumors with our driver hits have significantly less aneuploidy and a stochastic pattern of gains and losses, providing powerful evidence of our paper's main conclusion. Please see new **Fig 4d** and **Extended Data Fig. 11e and h:**

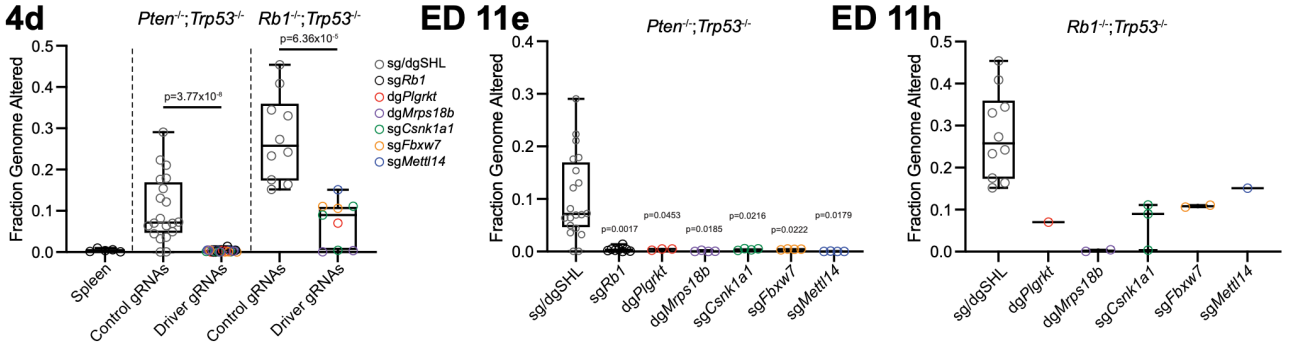


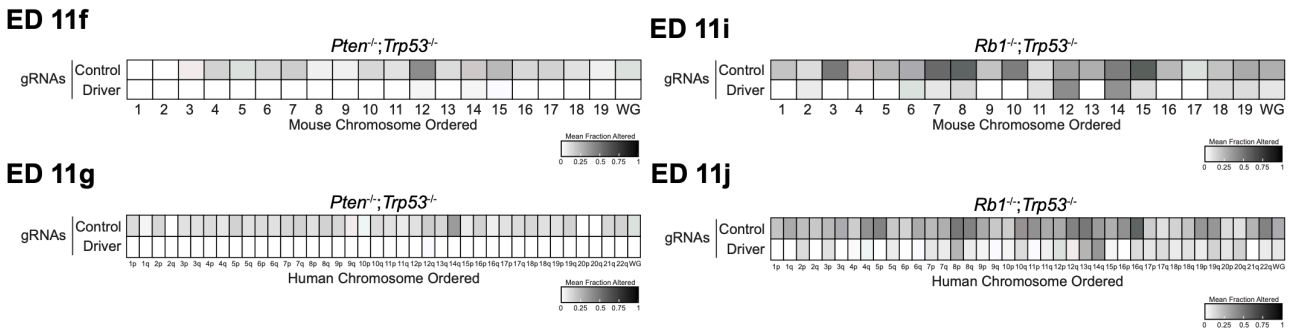
Fig. 4e (left): *Fraction genome altered from spleens (n=5), Pten$^{-/-}$;Trp53$^{-/-}$ tumors transduced with Control gRNAs (sg/dgSHL; n=20) or CNA driver gene gRNAs (n=27), and Rb1$^{-/-}$;Trp53$^{-/-}$ tumors transduced with Control gRNAs (sg/dgSHL; n=9) or CNA driver gene gRNAs (n=8) Individual driver genes are represented by the coloured circles within the boxplot. Statistical analysis was performed using a two-tailed unpaired t-tests. Data are mean with the box extending from the 25th to 75th percentile with whiskers from the minimum to maximum datapoint pooled from across four independent whole genome sequencing runs.*

Extended Data 11e (middle): *Fraction genome altered from Pten$^{-/-}$;Trp53$^{-/-}$ tumors transduced with sg/dgSHL (n=20), sgRb1 (n=8), sgFbxw7 (n=4), sgMettl14 (n=4), sgCsnk1a1 (n=4), dgMrps18b (n=4) or dgPlgrkt (n=3). Statistical analysis was performed using a two-tailed unpaired t-tests compared to sg/dgSHL. Data are mean with the box extending from the 25th to 75th percentile with whiskers from the minimum to maximum datapoint pooled from across four independent whole genome sequencing runs.*

Extended Data 11h (right): *Fraction genome altered from Rb1$^{-/-}$;Trp53$^{-/-}$ tumors transduced with sg/dgSHL (n=10), sgFbxw7 (n=2), sgMettl14 (n=1), sgCsnk1a1 (n=3), dgMrps18b (n=2) or dgPlgrkt (n=1). Statistical analysis was performed using a two-tailed unpaired t-*

tests compared to sg/dgSHL. Data are mean with the box extending from the 25th to 75th percentile with whiskers from the minimum to maximum datapoint pooled from across four independent whole genome sequencing runs.

We also calculate the chromosome-level or arm-level aneuploidy as new the new **Extended Fig.11f, g, I and, j** in human or mouse chromosome order:



Extended Data Fig. 11f and g (left): *Individual chromosome or chromosome-arm level alterations from Pten$^{-/-}$;Trp53$^{-/-}$ in mouse (**f**) or human (**g**) chromosome ordering with samples stratified as control (n=20) or driver (n=27) gRNAs. WG represents the whole genome.*
Extended Data Fig. 11i and j (right): *Individual chromosome or chromosome-arm level alterations from Rb1$^{-/-}$;Trp53$^{-/-}$ in mouse (**f**) or human (**g**) chromosome ordering with samples stratified as control (n=20) or driver (n=27) gRNAs. WG represents the whole genome.*

Additionally, the raw data for all 89 individual WGS profiles are provided in the Sequencing Reads Archive under accession PRJNA1146341 (available for revision at the following secure link: https://dataview.ncbi.nlm.nih.gov/object/PRJNA1146341?reviewer=5r6mdq8n4j59gqq4a2km2v7rju) and the copy number profiles of each individual tumor and control spleens have now been included and are available for review as new **Supplementary Figure 7**), as suggested by the reviewer. A reference to this has been added to the Methods sections as follows "*Individual plots for each sequenced mammary tumor are provided in Supplementary Figure 7*)." on Lines 1121-1122.

While at first sight somewhat counterintuitive, the data paints a clear picture, and we have now dedicated a whole paragraph to explain this point in the discussion and are grateful for the reviewer's input (please **Line 447-463**):

'By engineering these key driver gene alterations from the outset, we provide the incipient tumor cell with an immediate and substantial fitness advantage. This allows the tumor to bypass two major hurdles of spontaneous tumorigenesis: the time-consuming process of stochastically acquiring a specific CNA and the significant fitness cost imposed by the resulting gene dosage imbalance. This directly addresses the 'aneuploidy paradox'[2,8], wherein aneuploidy, despite being detrimental to normal cells, is well-tolerated and even selected for in cancer. Our data, consistent with previous findings[67], indicate that engineering these driver genes bypasses this fitness cost while capitalizing on the selective advantage of the altered gene dosage of the driver gene, explaining both the increased tumor multiplicity and the accelerated growth in our models.

Consequently, this initial advantage creates a fitness landscape whereby cells that subsequently acquire random aneuploidies have no significant selective edge and are outcompeted. This explains why our engineered tumors do not develop aneuploidy, even when serially transplanted - the inherent cost of any new chromosomal alteration simply outweighs any potential benefit.

Building on studies showing that chromosome arm loss can be driven by the selective pressure to inactivate a single canonical tumor suppressor, such as Pten[68] (=McFadden, 2014), Fbxw7[69] or Smad4[70], our data expands this 'single-driver' model in three key ways: (1) we showed that it applies to numerous novel tumor suppressor genes identified in this study; (2) it also applies to oncogenes; and (3) crucially, that a single event obviates the need for aneuploidies altogether.

I hope that this reviewer agrees that we have sufficiently addressed this point but we would be eager to receive further input and/or tips for enhanced clarity.

1.3 In a reciprocal manner, evaluating the arm-level CNAs at baseline across many individual tumors in each murine genetic background would help us understand the relevance of each murine model to the human arm-level data (e.g. a version of extended data 2b could be shown in main figure for both human and murine data).

As suggested by the reviewer, we now evaluated the arm-level CNAs at baseline across many individual tumors (n=30) in each murine genetic background to assess the relevance of the murine models to the human arm-level data. We have assembled WGS data for 20 Pten/p53- and 10 Rb1/p53-mutant baseline tumors and assessed the similarity of CNA compared to human tumors (please see the revised version of the figure, **Fig. 4b and c**), which showed a surprising concordance between species:

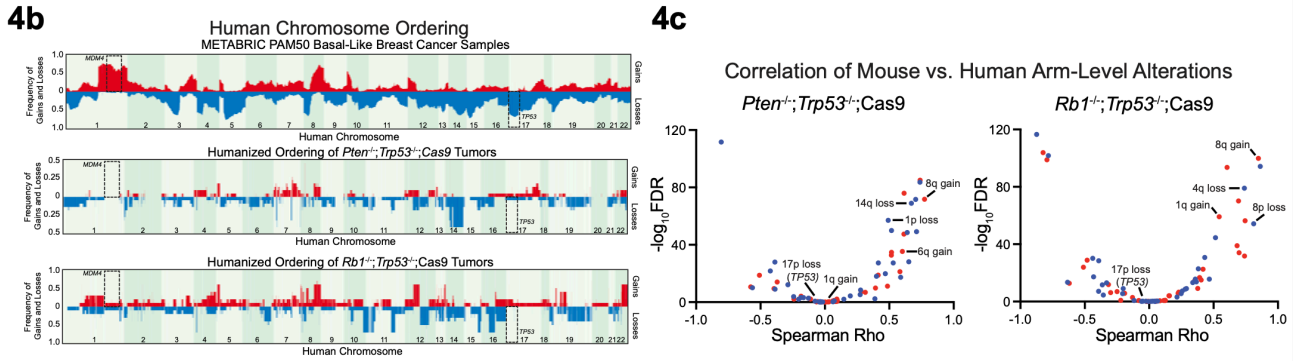
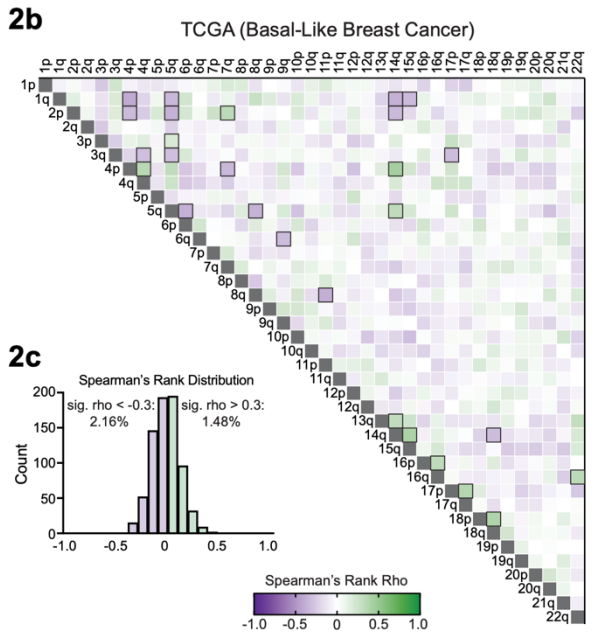


Fig. 4b (an adapted version; left): *Humanized ordering of whole genome sequencing data. The frequency of gains and losses is ordered by human chromosome position for METABRIC BLBC patients, Pten$^{-/-}$;Trp53$^{-/-}$ tumours (n=20), Pten$^{-/-}$;Trp53$^{-/-}$ tumours with CNA driver gene alterations (n=27), Rb1$^{-/-}$;Trp53$^{-/-}$ tumors (n=10), or Rb1$^{-/-}$;Trp53$^{-/-}$ tumors with CNA driver gene alterations (n=9), from top to bottom. Dashed insets show the location of MDM4 and TP53 loci.*

Fig. 4c: *Correlations of the frequency of each human chromosome arm-level gain (red) or loss (blue) compared to their corresponding humanized chromosome frequencies in Pten$^{-/-}$;Trp53$^{-/-}$ or Rb1$^{-/-}$;Trp53$^{-/-}$ tumors. Select arm-level alterations are labelled.*

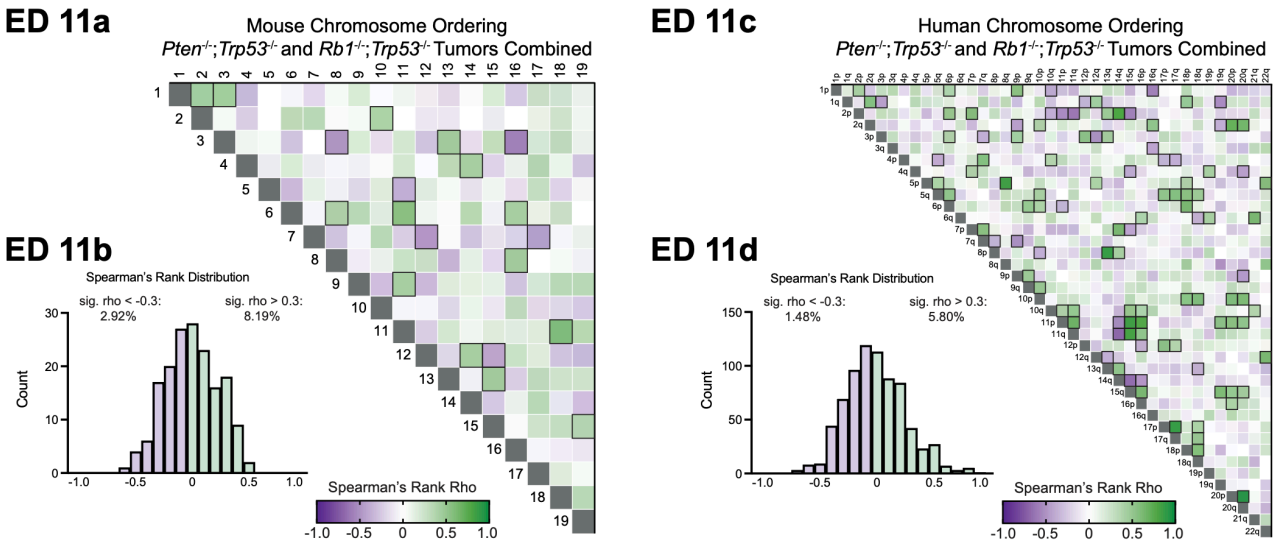
To further extend this analysis to CNA cooccurrences, we have now performed the correlation analysis of extended data 2b, but for mouse tumors, as suggested by the reviewer. This showed a very similar random distribution of chromosome and chromosome-arm level alterations that were observed in human basal-like breast cancers. As suggested by the reviewer, we have now generated a version of **Extended Data Fig. 2b** in both mouse- and human-chromosome ordering from our murine data as **Extended Data Fig. 11a-d.**

For reference, here is the original **Extended Data Fig. 2b and c** (human):



Extended Data Fig. 2 b and c: *Spearman correlation of each pairwise arm-level copy number alterations in basal-like breast cancer patients from TCGA USCS Xena database. Black boxes represent significant Spearman correlations |0.3|.* ***c****, Distribution of correlated and anti-correlated chromosome arms for basal-like breast cancer from TCGA USCS Xena database.*

And the new mouse data is provided as the new **Extended Fig 11a-d (mouse)**:



Extended Data Fig. 11 a-d: *Spearman correlation of each chromosome-level copy number alterations in autochthonous mouse Pten$^{-/-}$;Trp53$^{-/-}$ and Rb1$^{-/-}$;Trp53$^{-/-}$ (n=30) tumor cells. Black boxes represent significant Spearman correlations |0.3|.* ***b****, Distribution of correlated and anti-correlated chromosomes for Pten$^{-/-}$;Trp53$^{-/-}$ and Rb1$^{-/-}$;Trp53$^{-/-}$tumor cells.* ***c,*** *Spearman correlation of each pairwise human arm-level copy number alterations in in autochthonous mouse Pten$^{-/-}$;Trp53$^{-/-}$ and Rb1$^{-/-}$;Trp53$^{-/-}$ (n=30) tumor cells when genes are ordered according to their human chromosome position. Black boxes represent significant Spearman correlations |0.3|.* ***d****, Distribution of correlated and anti-correlated chromosome arms for chromosome arms for Pten$^{-/-}$;Trp53$^{-/-}$ and Rb1$^{-/-}$;Trp53$^{-/-}$tumor cells.*

2. Relevance of hits to human BLBC: Even though the authors provide a nice comparison between CNA in human and mouse BLBC samples, a more in-depth and side by side association to human BLBC samples (e.g. TCGA) would increase the relevance of this work and evaluation of hits. For instance, comparing transcriptional signatures derived in this paper to those of patients with similar gains and losses of human homologs could be helpful.

Reply: We thank the reviewer for this comment. In addition to the cross-species comparisons mentioned in response to comment #1, we have now also compared the transcriptional profiles of our mouse tumors to human BLBC on a global level as well as focused on a particular driver gene.

To further ascertain that these mouse tumors recapitulate the biology seen in human BLBC, we have now performed a cross-species analysis, which shows that the base model (sg/dgSHL) as well as all genetically engineered models that were transcriptionally profiled (*Fbxw7*, *Mettl14*, *Csnk1a1*, and *Plgrkt*) closely cluster with human BLBC and not Luminal or HER2+ breast cancers. Please see new **Fig. 3d and e**.

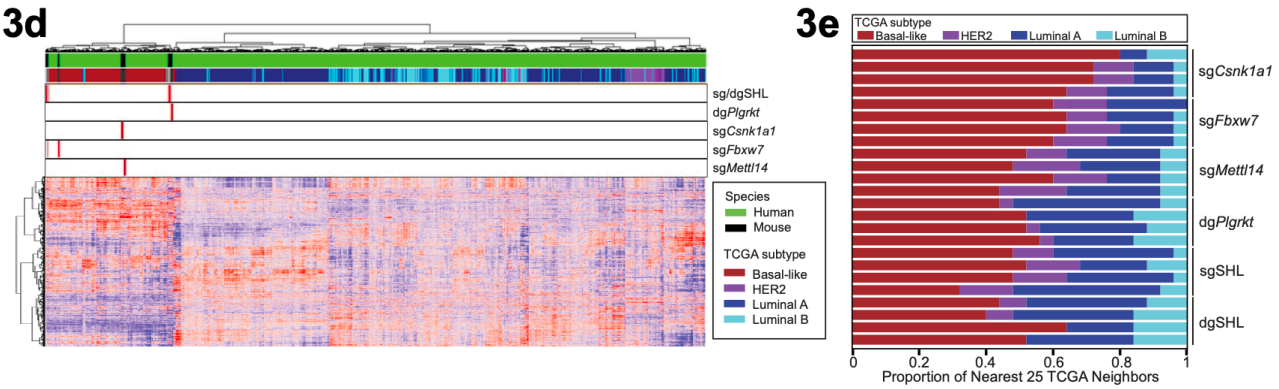


Fig. 3d (left): *Integration of human breast cancer and mouse CNA driver gene transcriptomes followed by unsupervised hierarchical clustering reveals that all mouse tumors cluster within the human basal-like breast subtype.*
Fig. 3e (right): *Nearest neighbor analysis for each mouse CNA driver gene tumor (n=4 tumors per gene) showing that the majority of the top 25 nearest human breast cancer samples belong to the human basal-like breast subtype.*

In addition, further analysis revealed a highly significant, cross-species conservation of transcriptional programs driven by all four driver genes further supporting the clinical relevance of our findings and indicating that these

CNA driver genes help shape the transcriptional state of human tumors harboring the respective alteration. Please see new **Fig. 3f-i and line 256-267**:

'*To establish the clinical relevance of our mouse models, we integrated their transcriptional profiles with data from breast cancer samples in TCGA. We first confirmed that our engineered mouse tumors molecularly resemble human BLBC, clustering exclusively within this subtype and away from others (**Fig. 3d and e**). We then tested whether our identified driver genes exert the same transcriptional effects in human tumors. To do this, we stratified human BLBC tumors from TCGA by copy-number alterations - losses of CSNK1A1, METTL14 and FBXW7, or gains of PLGRKT - and compared their transcriptional profile to diploid samples. We then ranked the differential genes from the TCGA data and compared them to the corresponding mouse transcriptomes using a Rank-Rank Hypergeometric Overlap (RRHO)[40]. This revealed a highly significant, cross-species conservation of transcriptional programs driven by all four driver genes (**Fig. 3f-i and Extended Data Fig. 10a-d**). This conservation supports the clinical relevance of our findings and indicates that these CNA driver genes help shape the transcriptional state of human tumors harboring the respective alteration.'*

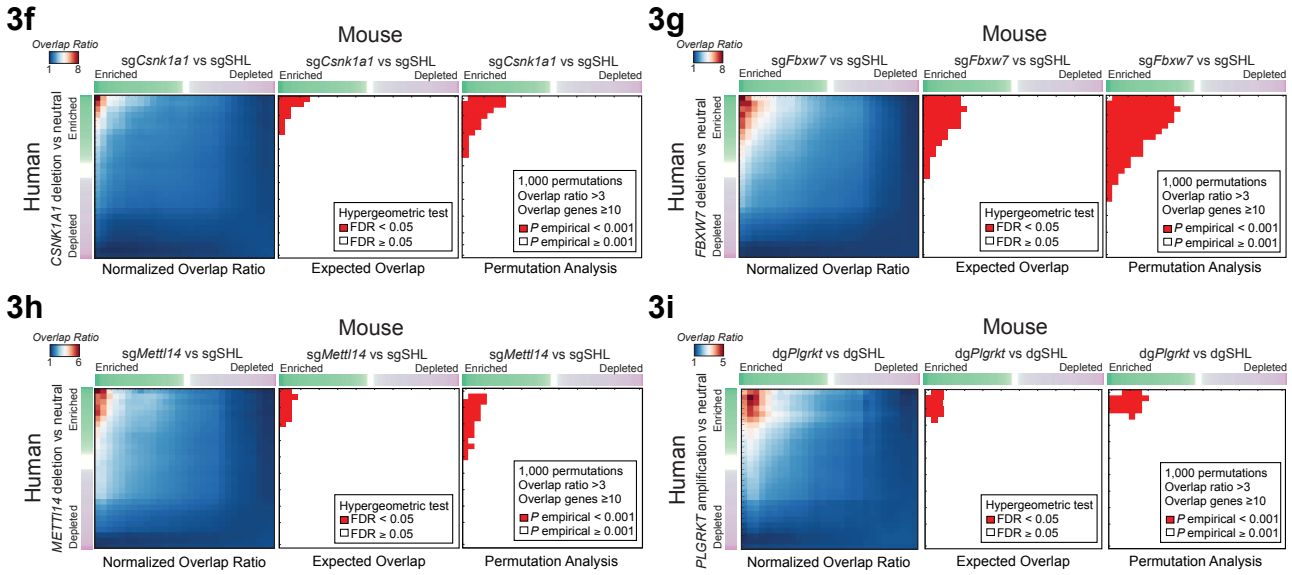


Fig. 3f-i: *One-sided hypergeometric tests were performed across thresholds (every 100 genes up to 2,000; every 500 between 2,500–5,000; and every 2,000 thereafter) for concordant upregulated genes comparing mouse CNA driver gene tumors and their corresponding human basal-like breast cancer sample harboring the given aneuploidy (CSNK1A1 loss; **f**, FBXW7 loss; **g**, METTl14 loss; **h**, and PLGRKT gain; **i**). Enrichment was quantified as $-\log_{10}$ FDR-adjusted p values and normalized overlap ratios (left), with $-\log_{10}$ p values FDR-adjusted normalized to the expected overlap at each threshold (middle). Alternatively, overlap significance was evaluated by permutation testing, where the mouse gene ranking was randomly permuted 1,000 times to derive a null distribution of overlap scores, and empirical p values were computed as the proportion of permutations yielding overlaps ≥ the observed one. Only overlaps satisfying ratio > 3 and ≥ 10 overlapping genes were retained as significant (right).*

We hope the reviewer agrees that these suggested extensive transcriptional cross-species analysis of *FBXW7*, *CSNK1A1*, *METTL14* and *PLGRKT* altered tumors showed a highly significant correlation, which indicates to us that we are indeed faithfully modeling the genetic events – naturally only of the one gene rather than the whole chromosome arm.

3. Translational relevance of hits: the authors convincingly demonstrate that this approach can identify novel genes that when perturbed can accelerate tumorigenesis. It could be powerful to demonstrate that modulation of these hits once a tumor is established can have the opposite effect on growth dynamics.

Reply: We thank the reviewer for this excellent and important point regarding the translational relevance of our findings.

To address the translational relevance, we successfully extended our primary findings from the mouse screen to human cells. We have now performed a targeted, multiplexed screen of all 90 hits in non-transformed human mammary epithelial cells (MCF10A) engineered to express a common breast cancer driver ($PIK3CA^{H1047R}$) and Cas9. As controls, we included non-targeting control s/dgRNAs as well as s/dgRNAs targeting the neighboring genes for each hit. While these MCF10A $PIK3CA^{H1047R}$ cells are typically non-tumorigenic, introducing the library of sg/dgRNAs targeting our 90 hits was sufficient to drive robust tumor formation in mice. In a more

aggressive model co-expressing MYC, our library significantly accelerated tumor growth. Deconvolution of these human-cell-derived tumors by next-generation sequencing confirmed the enrichment of the same driver genes identified in our original mouse screen, strongly supporting their relevance in a human context. Please see new **Fig. 2e and f**:

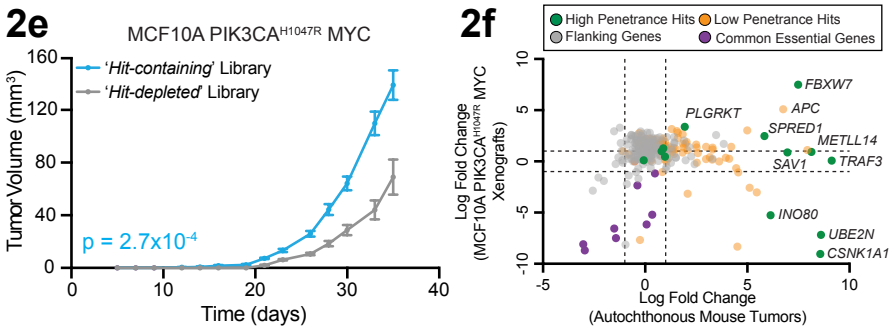


Fig. 2e (left): *Transduction of a library containing all 90 human CNA driver genes and each flanking gene ('Hit-containing') significantly accelerates tumor growth compared to the same library depleted of these driver genes ('Hit-depleted'; n=5 mice and 10 tumors per library). Statistical analysis was performed using a two-tailed unpaired t-test at tumor endpoint. Data are presented as mean ± s.e.m.*
Fig. 2f (right): *Comparison of gene-level log fold changes between the 'Hit-containing' human MCF10A PIK3CA^{H1047R} MYC xenograft screen and autochthonous chromosome-arm mouse screens.*

Furthermore, to show clinical relevance of PLGRKT in a human setting, we have extensively validated PLGRKT and could show that it triggers tumorigenesis of human MCF10A cells (sensitized toward transformation using a PIK3CA^{H1047R} mutation). Please see new **Fig 5a**.

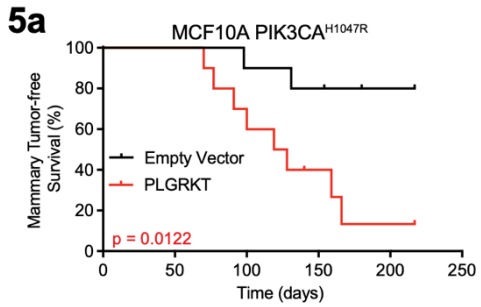


Fig. 5a: *Orthotopic transplantation of MCF10A PIK3CA^{H1047R} cells overexpressing PLGRKT significantly accelerates mammary tumor outgrowth (n=10 tumors per group). Statistical analysis was performed using the Gehan-Breslow-Wilcoxon test compared to empty vector.*

We agree that testing whether modulating these hits in established tumors affects growth dynamics is a critical next step. We have initiated studies to test for such oncogene addiction using a standard Dox-inducible system for our top hit, PLGRKT, using mouse 4T1 and human PIK3CA^{H1047R} MCF10A cells. However, the explosive tumor growth induced by PLGRKT expression in this model presents a significant technical challenge, as the kinetics of drug washout and protein depletion are likely too slow to observe a therapeutic effect before tumors reach endpoint (which is why we are reluctant to include this data into the manuscript).

Recognizing the importance of this question, we are actively developing more sophisticated models better suited to definitively test for therapeutic dependency. These include using an auxin-inducible degron (AID) system for rapid protein degradation and establishing Dox-inducible systems in the slower-growing, autochthonous mouse models. However, the AID system still requires optimization, and we could not yet perform the experiment in the autochthonous model, as we had to generate and use >400 *Pten*$^{fl/fl}$;*Trp53*$^{fl/fl}$;Cas9 to perform all other revision experiments.

Therefore, while interrogating oncogene addiction is a key future direction for this work, we believe a comprehensive therapeutic study of all 90 hits represents a substantial undertaking that builds upon the foundational discoveries presented here. The primary scope of this manuscript is the development of a novel platform and its successful application to identify a large cohort of previously unknown cancer drivers. We hope that the reviewer agrees and we have clarified this in the discussion, highlighting that testing the therapeutic potential of these genes is an intriguing and critical next step as highlighted in our discussion (please see line 486-490):

'While gene losses can underpin exceptional therapeutic responses[85], identifying recurrently amplified driver genes may offer more promising therapeutic opportunities, akin to ERBB2/HER2. This is particularly important

for CNA-driven diseases with limited single nucleotide variants and few known driver genes, including BLBC, glioma, and ovarian cancer. ***Testing the therapeutic potential of our newly identified driver genes will be an intriguing next step.'***

4. Depth of mechanistic follow-up. The follow-up data on the Plgrkt gene is performed in a distinct system than the screen and produces SCC tumors (not BLBCs). The data is interesting but preliminary for a paper in the intended journal.

Reply: We agree with this reviewer and have spent a considerable amount of time to increase the mechanistic insights into PLGRKT's oncogenic function.

Firstly, just to iterate our response to **Major Comment #2**, we have shown that PLGRKT, results in increased outgrowth of human MCF10A cells (sensitized toward transformation using a PIK3CA^{H1047R} mutation). Please see **Fig. 5a**.

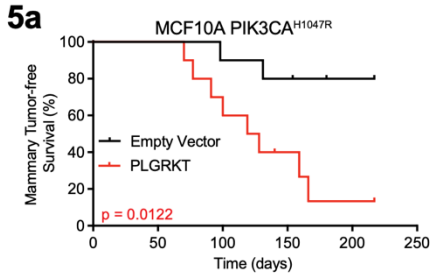


Fig. 5a: *Orthotopic transplantation of MCF10A PIK3CA^{H1047R} cells overexpressing PLGRKT significantly accelerates mammary tumor outgrowth (n=10 tumors per group). Statistical analysis was performed using the Gehan-Breslow-Wilcoxon test compared to empty vector.*

We have now added a large body of work exploring PLGRKT's function. We have now demonstrated that endogenous PLGRKT localizes to the mitochondria within autochthonous mouse tumors, human basal-like breast cancers and mouse and human breast cancer cell lines, using super-resolution microscopy showing that PLGRKT co-localizes with the inner mitochondrial membrane marker TIMM23 and is in close proximity to the outer mitochondrial membrane marker TOMM20 (using a tightly validated PLGRKT antibody). In addition, we found that PLGRKT localizes to the mitochondria, that overexpression of PLGRKT increases it mitochondrial localization using immunofluorescence, super-resolution microscopy and Stimulated Emission Depletion (STED) microscopy in 4T1, MCF10A, mouse tumors and patient breast cancer tissue arrays (please see Lines 367-376 and the new **Fig. 5d and e and Extended Data Fig. 12 g-n**).

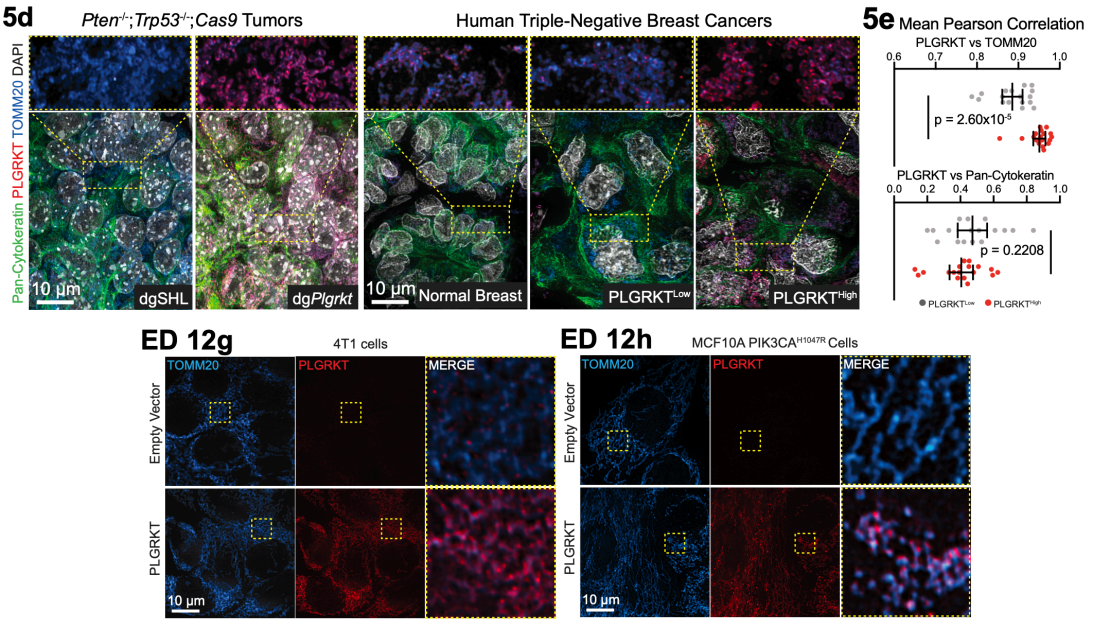
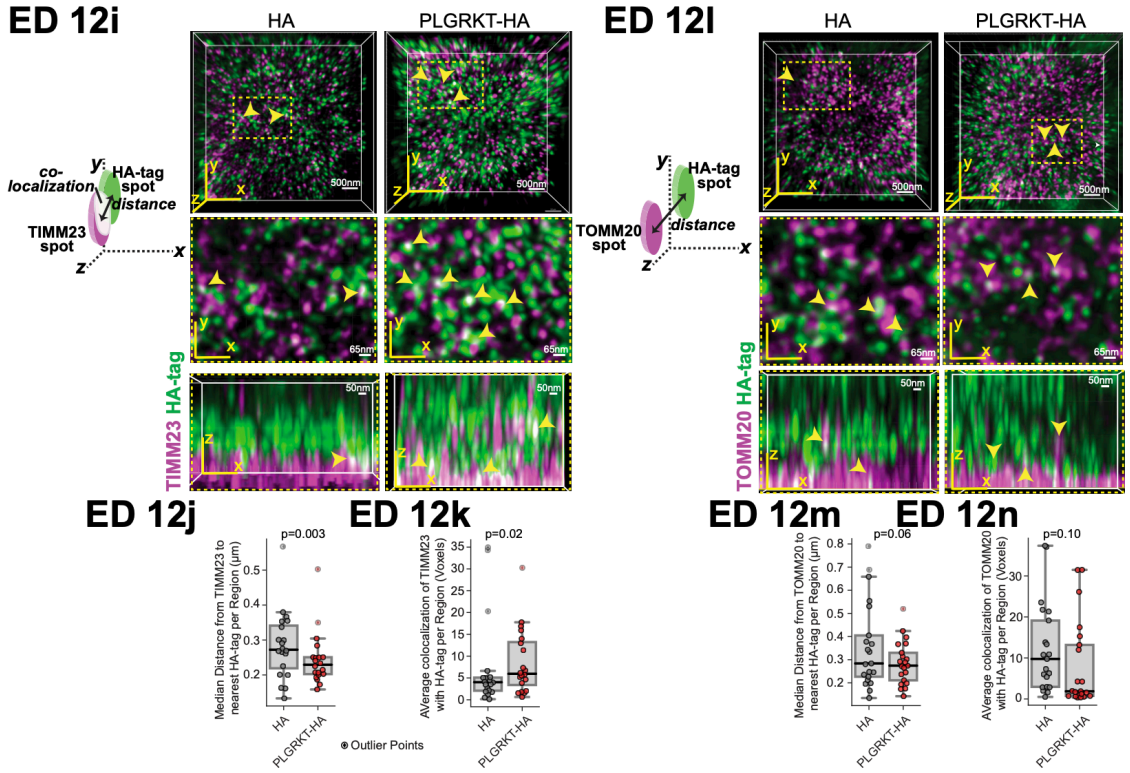


Fig. 5d (top left): *Immunofluorescent staining of autochthonous mouse mammary tumors (n=4 tumors per genotype) and human breast cancer samples (n=35 basal-like breast cancer tumors) and normal adjacent tissue (n=14) showing localization of PLGRKT within the mitochondria. Scale bar, 10 μm.*

Fig. 5e (top right): *Quantification of co-localization of PLGRKT with TOMM20 and Pan-Cytokeratin from the human breast cancer samples in (**d**) stratified by low (grey; n=17) and high (red; n=18) PLGRKT expression. Statistical analysis was performed using two-tailed unpaired t-tests. Data are mean ± s.e.m.*

Extended Data Fig. 12g and h (bottom): *Representative immunofluorescence of 4T1 (**g**) and MCF10A PIK3CA^{H1047R} (**h**)cell lines overexpressing empty vector or PLGRKT (n=3 replicates) showing localization of PLGRKT within the mitochondria. Scale bar, 10 μm.*



Extended Data Fig. 12i-n: *STED microscopy of 4T1 cells overexpressing HA-tagged empty vector or PLGRKT-HA vector, co-stained for inner mitochondrial membrane protein TIMM23 (**i-k**) or outer mitochondrial membrane protein TOMM20 (**l-n**). For **i** and **l**, representative 3D reconstructions of STED data are shown, with axis planes as indicated, showing close proximity and overlap (white signal) between PLGRKT-HA and TIMM23 specifically. For **j** and **m**, quantification of nearest neighbour distance from mitochondrial membrane marker to HA-tag (n=20 regions across 10 cells per condition) is shown. For **k** and **n**, quantification of co-localized voxels between mitochondrial membrane marker and HA-tag (n=20 regions across 10 cells per condition) is shown. Statistical analysis was performed using two-tailed unpaired t-tests. Data are mean with the box extending from the 25th to 75th percentile with whiskers from the minimum to maximum datapoint.*

In addition, we found that PLGRKT overexpression enhances mitochondrial motility in stressed mouse and human cells through Enhanced Super Radial Fluctuation (eSRRF) and live cell imaging. Please see Lines 377-386, the new **Fig. 5f-i and Supplementary Videos 1-10**.

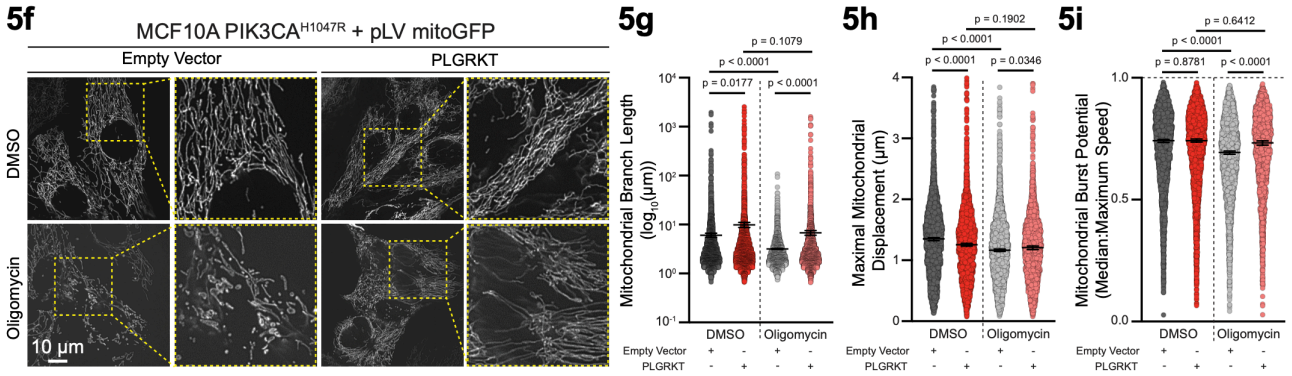
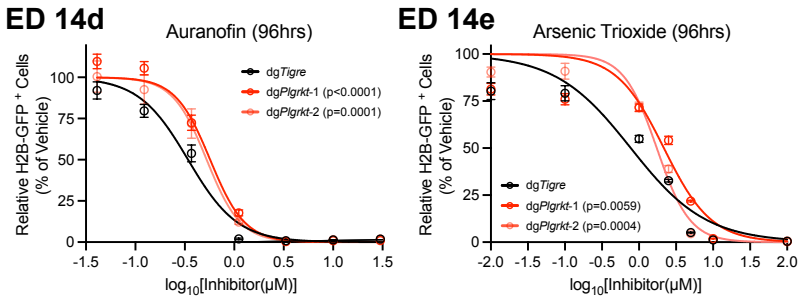


Fig. 5f (left): *Representative images of mitoGFP from enhanced super resolution-radial fluctuation (eSRRF) imaging of MCF10A PIK3CA^{H1047R} cells overexpressing empty vector or PLGRKT and treated with DMSO or oligomycin. Scale bar, 10 μm.*

Fig. 5g-i (right): *Quantification of mitochondrial branch length from eSRRF imaging (**g**), and maximal mitochondrial displacement (**h**) or mitochondrial burst potential (**i**) from mitochondrial live imaging of MCF10A PIK3CA^{H1047R} cells overexpressing empty vector or PLGRKT and treated with DMSO or oligomycin (n=24 fields of view per condition collected over three independent biological replicates). Statistical*

analysis was performed using or an ordinary one-way ANOVA with Tukey's correction for multiple testing. Data are median ± 95% confidence interval.

Next, we sought to connect the mitochondrial dysfunction to a cellular mechanism and found that PLGRKT not only conferred a significant growth advantage to oligomycin, but to several other ROS-inducing chemotherapies including auranofin and arsenic trioxide. Please see the new **Extended Data Fig. 14d and e**).



Extended Data Fig. 14d and e: *Relative viability of 4T1 cells expressing dgTigre, dgPlgrkt-1 or dgPlgrkt-2 treated with Auranofin (**d**) or Arsenic Trioxide (**e**) for 96 hours. Statistical analysis was performed by fitting normalized response curve with a variable slope and comparing differences in fits. Data are mean ± s.e.m*

This led us to speculate that PLGRKT overexpression regulates resistance and promotes cell survival when challenged with elevated levels of ROS. We now show that PLGRKT itself regulates mitochondrial superoxide and total cellular ROS detoxification. Please see the new **Fig. 5l and m and Extended Data Fig 14 a-c.**

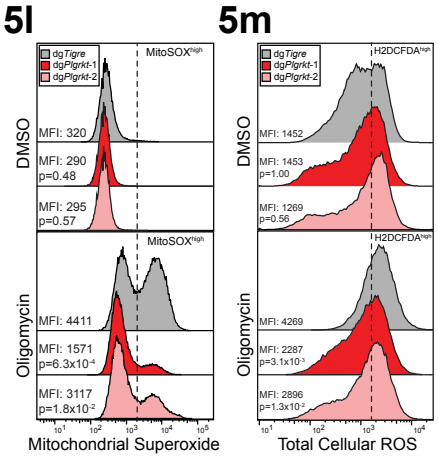
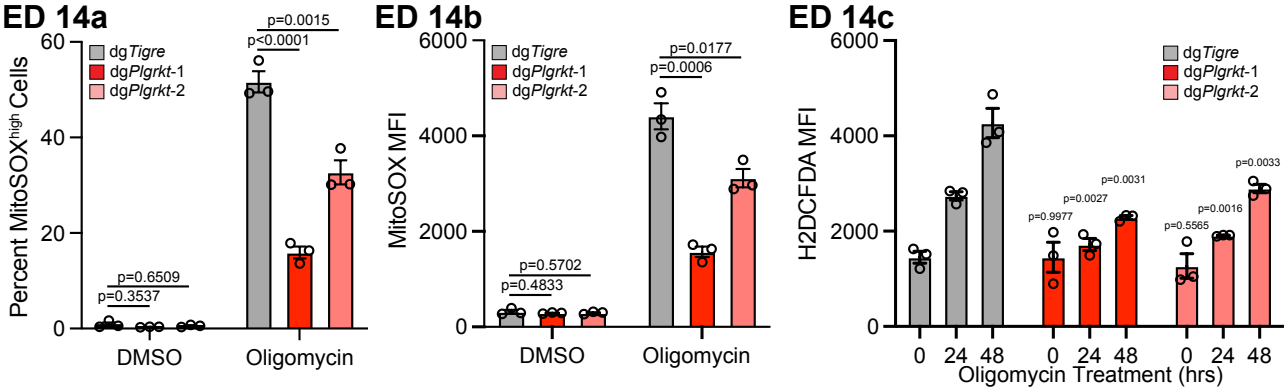


Fig. 5l and m: *Representative histograms of the mean fluorescence intensity (MFI) for mitochondrial superoxide (mitoSOX; **l**) or total cellular ROS (H2DCFDA; **m**) as measured by flow cytometry in 4T1 cells expressing dgTigre, dgPlgrkt-1 or dgPlgrkt-2 and treated with DMSO or oligomycin (n=3 replicates per condition). Statistical analysis was performed two-tailed unpaired t-tests compared to empty vector controls.*



Extended Data Fig. 14a-c: *Quantification of the percentage of MitoSOX^{+high} (**a**), MitoSOX Mean Fluorescence Intensity (MFI; **b**), and H2DCFDA MFI (**c**) in 4T1 cells expressing dgSHL, dgPlgrkt-1 or dgPlgrkt-2 and treated for 48 hours with DMSO or Oligomycin (**a** and **b**) or across a 48 hour time course (**c**). Statistical analysis was performed using a two-tailed unpaired t-test Data are mean ± s.e.m.*

Lastly, treatment of PLGRKT overexpressing cell lines with oligomycin or hydrogen peroxide to induce ROS resulted in a significant elevated expression of antioxidant genes, which we confirmed by qRT-PCR. Please see the new **Fig. 5o and Extended Data Fig. 14g**.

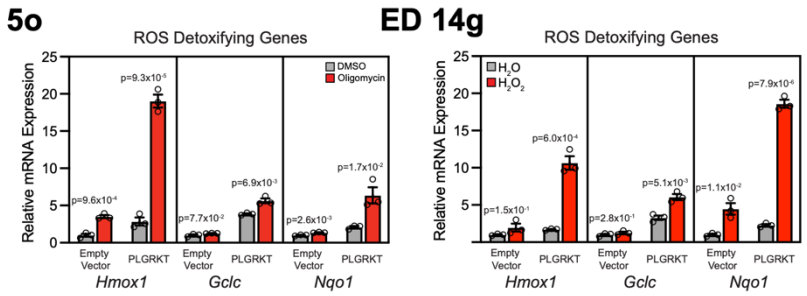


***Fig. 5o and Extended Data Fig. 14g**, qRT-PCR analysis for several antioxidant target genes from 4T1 cells overexpressing empty vector or PLGRKT at baseline or treated with oligomycin (**5o, left**) or H_2O_2 (**ED14g, right**) for 12 hours (n=3 samples per group). Statistical analysis was performed using a two-tailed unpaired t-test Data are mean ± s.e.m.*

Overall, we believe that we have significant enhanced our molecular understanding of how PLGRKT functions in breast cancer independent of its known role as a plasminogen binding protein. Together, our new data suggests that PLGRKT is a potent proto-oncogene on chr9p, that drives tumorigenesis through its increased localization to the mitochondria, dysregulation of the electron transport chain, elevated ROS detoxification, and promoting healthier, stress-resistant mitochondria. Please see the updated summary schematic of the functions of PLGRKT as the new **Fig. 5p**.

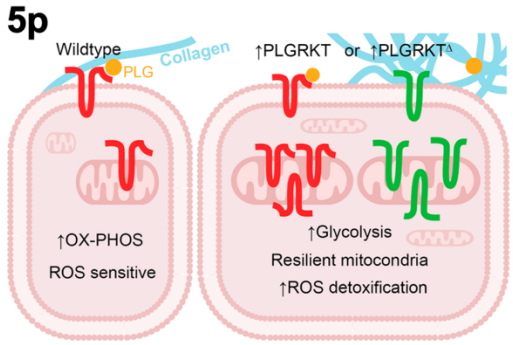


***Fig. 5p:** Schematic of the canonical and non-canonical functions for PLGRKT in promoting tumorigenesis.*

5. A comprehensive evaluation of Plgrkt oncogenesis and chr9p gain across cancer types, in human and murine data, and relating these changes to mitochondrial phenotypes is needed.

Reply: We want to thank this reviewer for this comment and agree that elucidating PLGRKT's function in other cancers will be very important.

To this end, further analysis revealed a highly significant, cross-species conservation of transcriptional programs driven by PLGRKT in human BLBC further supporting the clinical relevance of our findings. Please see new lines 268-279 and new **Fig. 3i and Extended Data Fig. 10d.**

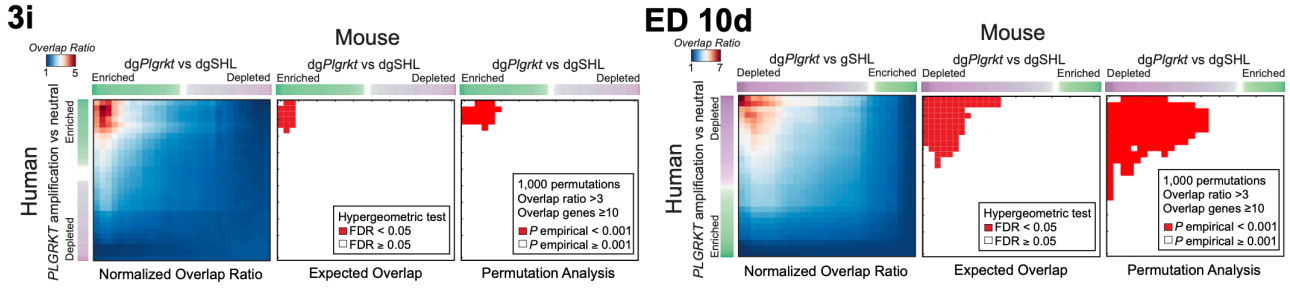


***Fig. 3i and Extended Data Fig. 10d:** One-sided hypergeometric tests were performed across thresholds (every 100 genes up to 2,000; every 500 between 2,500-5,000; and every 2,000 thereafter) for concordant upregulated (**3i**) or downregulated (**ED 10d**) genes comparing mouse CNA driver gene tumors and their corresponding human basal-like breast cancer samples harboring PLGRKT gains. Enrichment was quantified as $-\log_{10}$ FDR-adjusted p values and normalized overlap ratios (left), with $-\log_{10}$ p values FDR-adjusted normalized to the expected overlap at each threshold (middle). Alternatively, overlap significance was evaluated by permutation testing, where the mouse gene ranking was randomly permuted 1,000 times to derive a null distribution of overlap scores, and empirical p values were computed as the proportion of permutations yielding overlaps ≥ the observed one. Only overlaps satisfying ratio > 3 and ≥ 10 overlapping genes were retained as significant (right).*

In addition, PLGRKT-high human BLBC tumors also showed a significant enrichment of gene sets associated with oxidative damage and stress responses. Please see the new **Fig. 5n**.

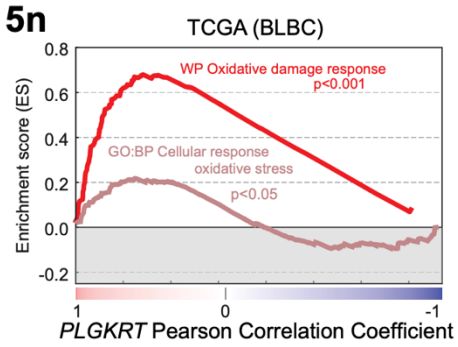
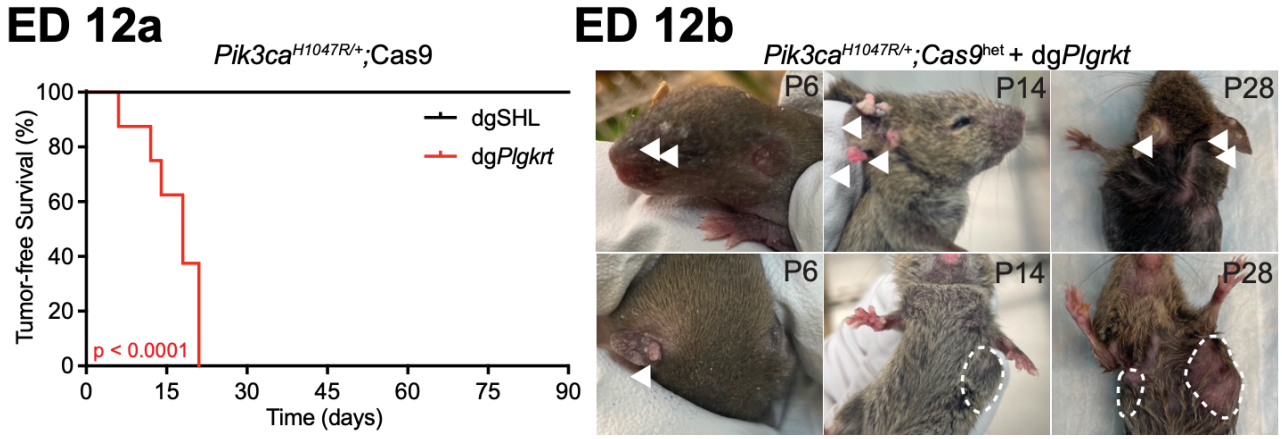


Fig. 5n: *Gene-set enrichment analysis of oxidative response pathways in human BLBC tumors. Genes were ranked by their Pearson correlation coefficient with PLGRKT expression.*

In addition, we have also already clearly shown that PLGRKT can dramatically accelerate cancer development in a mouse model of PIK3CA^{H1047R}-driven cutaneous and Head&Neck squamous tumorigenesis (please see **Extended Data 12a and b**).

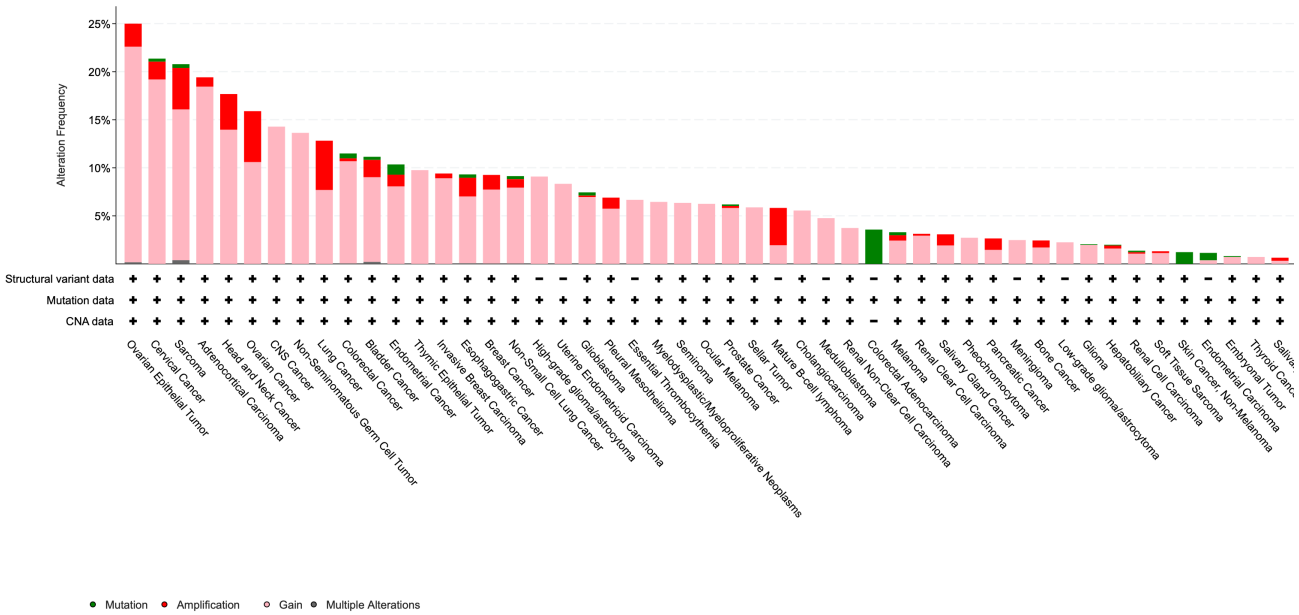


Extended Data Fig. 12a (left): *Overexpression of Plgrkt significantly accelerates cutaneous squamous cell carcinoma (cSCC) and mammary tumor development in mice (n=8 mice per group). Statistical analysis was performed using the Gehan-Breslow-Wilcoxon test*
Extended Data Fig. 12b (right): *Representative photographs of mice following in utero injection of dgPlgrkt. White arrows and dashed areas indicate cSCC and mammary tumors, respectively. Post-natal age of the mice is labelled within the photograph.*

While it was encouraging to see that PLGRKT can also drive another cancer type, we can speculate that the CNA drivers are highly cancer-specific and that it is impossible to bioinformatically weed out if this is an oncogene in other cancer types. **TEXT REDACTED**

FIGURE REDACTED

TEXT REDACTED Overall, the amplification or gain of PLGRKT are quite broad (please see bar graph below) and its effect in these other cancer types remains to be analysed, which is an important next step, but beyond the scope of this current manuscript.



For more detailed analysis, please see: https://bit.ly/4hF7rZK

TEXT REDACTED

\

FIGURE REDACTED

TEXT REDACTED

Reply to Reviewer #4:

Nature is committed to facilitate training in peer-review and to ensure that everyone involved in our peer review process is appropriately recognised. This reviewer co-reviewed one of the listed reports.

Reply: We would like to directly thank the reviewer for taking the time to contribute to the peer-review process and providing feedback that has undoubtedly strengthened our manuscript. We hope that we have adequately addressed your comments, concerns and suggestions.

Point-by-Point Response to Referees' Comments (2024-08-17371)

We sincerely thank the four reviewers for their thorough analyses and valuable suggestions, which have greatly guided us in improving our study.

Reviewer 1: The authors have thoroughly addressed all of my previous concerns. The addition of non-genic control libraries, comprehensive WGS profiling with arm-level aneuploidy calculations, and the substantially expanded mechanistic characterization of PLGRKT considerably strengthen the manuscript. The cross-species transcriptional analyses and demonstration that PLGRKT protein levels in mouse tumors match those in human BLBC are particularly compelling. I am satisfied with the revised title and the acknowledgment of limitations regarding dosage sensitivity.

The authors should confirm that CRISPR-KOALA plasmids have been deposited at Addgene, the WGS data will be released upon publication and all raw sequencing data from the screens are deposited in an appropriate repository.

We thank Reviewer 1 for the incredibly positive feedback on our revised manuscript. In our substantially revised manuscript, we have addressed all Reviewer 1's major and minor concerns and have added a wealth of new data that further support our original findings. We greatly appreciate how these additional experiments have strengthened both the manuscript and our conclusions. Reviewer 1 shares this view and is satisfied that our conclusions are now fully supported by our new data.

Lastly, we can confirm that our plasmids are in the process of being deposited on Addgene, the WGS data will be released upon publication, and our raw RNA-sequencing data will be deposited and released upon publication.

Reviewer 2: The manuscript it greatly improved, both in scope and precision (of data and language). They provide new analysis and data from new pools, both of which were request quite specifically by the Reviewers. The provide additional single gene validation. They have provided sufficient additional discussion of the caveats of the approach. On one hand, they have done a huge amount of work; on the other hand, some experiments (e.g. having Safe targeting control sgRNAs) should have been included in the first place without the reviewers asking. Overall, this paper has employed novel and interesting in vivo approaches to uncover new cancer drivers.

We thank the Reviewer 2 for their thoughtful and positive evaluation of our manuscript and for highlighting the new data and analysis as well as single gene validation. As with reviewer 1, this reviewers' comments have also significantly strengthened our manuscript.

Reviewer 3: We would like to thank the authors for their thorough and thoughtful response to our, and other reviewers' questions, generating an impressive and substantial body of work. Demonstrating the technical feasibility of screening many hundreds of genes *in vivo* is certainly an important advance and the amount of new data, results and figures is quite vast.

The authors provided data and analyses to begin to address the five major points we raised, which we feel were partially addressed in the manuscript revisions. Additional analyses using MAGeCK, expanded data generation in multiple genetic background and whole-genome sequencing data sufficiently addressed concerns regarding the specificity and penetrance of the identified hits. Moreover, the penetrance of PLGRKT oncogene activity was also demonstrated in additional models and backgrounds.

The rebuttal clarifies the authors' belief that the technical approach described in this paper is a fundamentally novel technical advance beyond previous work. While we do not dispute this argument, we believe that CRISPR-KOALA is best categorized as an important iteration on existing methods. Of course, there are other (both experimental and computational) methods for deconvoluting driver genes in copy number regions as well and it is unclear what other approaches might have discovered.

We think the paper is stronger than before. It is a tour-de-force that should be published but we remain unsure of its suitability for the current journal.

We thank Reviewer 3 for their careful review of our revised manuscript. We appreciate that they recognize the extensive new work performed during revision, and their recommendation for publication. We are grateful for the constructive feedback provided throughout this process, which has undoubtedly strengthened the manuscript. Regarding the remaining concerns, we believe these offer a valuable opportunity to further clarify our conceptual framework and the specific scope of our findings.

We have carefully addressed these final points through targeted textual refinements to ensure our conclusions are communicated with maximum clarity. While some of the suggested directions represent fascinating avenues for future research, we believe they fall beyond the intended scope of this foundational study. We hope the reviewer finds that these revisions successfully resolve the outstanding queries and demonstrate the suitability of this work for *Nature*.

Overall comment: Novelty and value of CRISPR-KOALA

Regarding the classification of CRISPR-KOALA, we agree with the reviewer's perspective that the system is best described as an important iteration on existing methodologies and our primary objective was to refine these approaches to enable simultaneous, bidirectional manipulation in a single, high-throughput setting.

We have revised the Results section to explicitly frame CRISPR-KOALA as an **important iteration** that "**builds upon established technology**" to facilitate these specific screening capabilities (please see line 111-113). We believe this framing accurately reflects the lineage of the technique while highlighting the unique biological insights it enables. Ultimately, we believe the most significant contribution of this work lies in the discovery of previously unrecognized driver genes and their non-canonical roles in tumorigenesis, an advance enabled by the utility of this system.

Reviewer 1 and 2 also highlighted the novelty and broad applicability of CRISPR-KOALA, describing it as a "*novel, simple and valuable technique that could be used across several fields*" and a "*broadly applicable, high-throughput and physiological system, which is novel and could be applied to the study of driver genes in other cancers*". Together, we hope the reviewer finds that these textual refinements successfully resolve the outstanding queries.

Point-by-Point Response to remaining comments:

1. Reducing selective pressure for CN events. The revised data nicely demonstrate that the latency of tumor formation is an important factor in determining whether sufficient time has elapsed for subclones to become clonal (and therefore detectable via bulk sequencing). Given that each of the genes induces different biology (Ext Fig 7i) we would expect, for example, that the presence of PLGRKT reduces selective pressure on Chr19 events, but has no impact on other CN events (unless all possible events are functionally the same). Based on the updated Figure ED11a, certain CNAs do co-occur—for example, Chr19 and Chr15—with Chr19 hosting PLGRKT – so this would suggest there is likely to be complementary biology here. We simply remain unsure whether the observations here are best explained by (1) highly tumorigenic cells produce polyclonal tumors requiring less time for selection or (2) there is unique biology that obviates CN selection in meaningful ways. The analysis remains somewhat premature to distinguish.

We agree with Reviewer 3 that distinct driver genes should reduce selection only for specific CN events rather than broadly bypassing aneuploidy, which was our initial expectation as well. This hypothesis indeed applies to genetically homogeneous cancers, as demonstrated recently by Scott Lowe's lab in pancreatic cancer models (Baslan *et al.* Nature 2022), for example, where *Smad4* knockdown alleviates selective pressure to lose chromosome 9. Pancreatic cancer is genetically very uniform, with most tumors harboring KRas mutations, followed by inactivation of p53 and/or the Smad4/TGFBR pathway. As such, tumors typically follow a deterministic genetic trajectory, and when this trajectory is experimentally provided, there is no need to acquire the corresponding CNAs.

However, our data strongly indicate that basal-like breast cancer, which is genetically highly heterogeneous, with multiple alternative evolutionary genetic trajectories leading to tumorigenesis. This is reflected by the extensive CNAs affecting, on average, 80% of the genome (Watkins *et al.* Nature 2020). When one such trajectory is experimentally provided through a defined driver perturbation, tumors no longer need to search for alternative fitness-conferring CNAs and remain CNA-silent. Although individual drivers act through diverse

biological processes, these converge on a common endpoint: **enhanced cellular fitness**. Once this fitness advantage is provided, aneuploidy no longer confers additional selective benefit.

We believe this conceptual framework is VERY important and a distinct finding of our study, which was also previously raised by Reviewer 2. We have now further streamlined text and refined **Fig. 4g** to make this conceptual model even more explicit (biological heterogeneity -> convergence on increased cellular fitness -> no requirement for aneuploidy). We have also rephrased the Discussion (please see lines 422-430) to explicitly describe this point:

'BLBC is genetically highly heterogeneous, characterized by multiple alternative evolutionary trajectories and extensive CNAs affecting most of the genome[1]. Our data indicate that when a specific trajectory is experimentally provided through defined genetic perturbations, tumors no longer need to search for alternative, fitness-conferring CNAs and remain CNA-silent. This directly addresses the 'aneuploidy paradox'[2,3], wherein aneuploidy, despite being detrimental to normal cells, is selected for in cancer to determine a fitness advantage. Once this advantage is provided by a specific driver, aneuploidy no longer confers a selective benefit. Our data thus broaden the 'single-driver' model, which posits that the selective pressure to inactivate a single canonical tumor suppressor can drive the loss of an entire chromosome arm as seen with Fbxw7[4], Pten[5] or Smad4[6]. We demonstrate that this principle extends to newly identified tumor suppressors and, notably, oncogenes and show that in BLBC a single driver event can obviate the need for aneuploidies altogether.'

Lastly, we agree with this reviewer that an alternative explanation would be that these highly tumorigenic cells simply did not have enough time to accumulate CNAs. We have experimentally tested this possibility as describe in the result section:
'To rule out the possibility that the CNA cancer gene tumors were genomically stable due to insufficient time to accumulate alterations, we assessed the CNA profile of Mettl14-deficient tumors after serial transplantation. Importantly, the secondary transplants also remained CNA-free (Fig. 4f).'

Together, we hope that the reviewer agrees that these results show that it is indeed the unique biology of BLBC and the CNA driver genes that obviates CN selection.

2. Relevance to human BLBC: We appreciate and recognize the authors' efforts to demonstrate that the GEMMs used in this study closely resemble the transcriptional profiles of human BLBC in TCGA data. Additional cross-species data demonstrating *in vivo* transformation of high-penetrance hits in MCF10A cells further strengthens their findings. However, we find the genomics data in 3d and 3e preliminary.

We first would like to thank Reviewer 3 for their positive comments regarding our cross-species analysis and our human MCF10A CRISPR-KOALA screen, noting that these "*further strengthen [our] findings*".

To further clarify the clinical relevance to human BLBC, we would like to highlight that we also directly demonstrated that PLGRKT overexpression increases penetrance and accelerates orthotopic tumor growth of non-transformed human MCF10A PIK3CA^{H1047R} (**Fig. 5a**). This provides a clear functional link to human tumorigenesis (see also next point). We also show significant associations between PLGRKT expression and mitochondrial dysfunction in human BLBC TCGA data (**Fig. 5n**), demonstrating strong concordance between mouse and human driver-associated expression programs – not only for PLGRKT, but for all of our major hits (**Fig. 3. f-i)**. As such, it is important to note that the genomics data in **Fig. 3d** and **3e** is further substantiated by **Fig. 3. f-i** and **Extended Data Fig. 10a-g**.

We have endeavored to provide a comprehensive body of validation within the traditional scope and timeframe of a revision. While we recognize that further specialized studies will be essential in the future to fully explore the downstream therapeutic potential of these drivers, we believe our current integration of large-scale *in vivo* screening, cross-species genomics, and human functional validation provides a robust foundation for these findings.

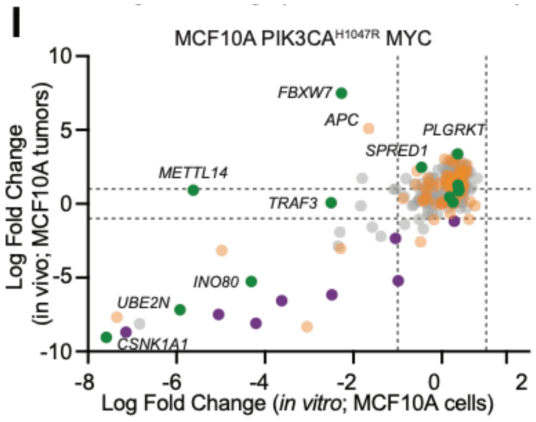
Given that PLGRKT (as well as the other hits) can (i) transform non-transformed human cells (in both *in vitro* and *orthograft* assays), (ii) is significantly altered in cancer patients, and (iii) triggers rapid tumorigenesis in GEMMs, we consider this multifaceted evidence the gold standard for establishing the functional relevance of a novel cancer

driver. Should this reviewer or the Editorial team desire any specific further analysis, we would be grateful for guidance on how to best address those priorities.

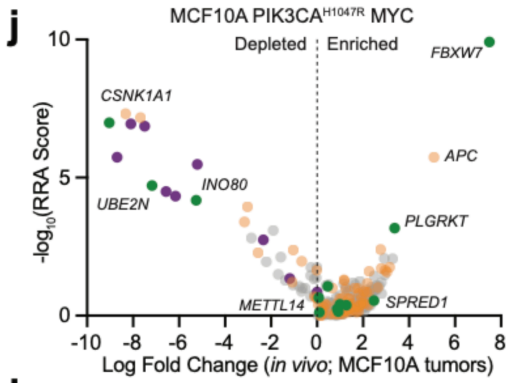
3. Tumor maintenance. While it is true that a systematic validation of all driver genes as targets might be challenging, we believe it is possible to perform experiments to ascertain PLGRKT as a cancer target. Extended Data Fig 6L emphasizes that the magnitude of the PLGRKT effect may be somewhat modest in human cells, compared to other hits. This further emphasizes the importance of being very sure this is a robust human-relevant result.

Reviewer 3 suggests "*that the magnitude of the PLGRKT effect may be somewhat modest in human cells, compared to other hits*" based on **Extended Data Fig. 6l**. We appreciate the opportunity to clarify the effect size of PLGRKT.

Extended Data Fig. 6l presents data from a screening assay, within which PLGRKT in fact exhibits one of the strongest effect sizes, both within the screen itself and compared to other published screens. In addition, PLGRKT exhibits a huge effect size when tested individually within human cells, which might have been overlooked by the reviewer. We thus respectfully disagree with this interpretation, for reasons outlined below:



In **Extended Data Fig. 6l**, the y-axis displays the results of the ***in vivo*** orthotopic allograft screening experiment, while the x-axis shows the results from the parallel ***in vitro*** cell culture screen. On the y-axis, *PLGRKT* overexpression is in fact the **third strongest hit** in the human re-screen, with an *in vivo* **log fold change of 3.38** relative to controls (i.e., a 10.5-fold increase). By contrast, the x-axis reflects *in vitro* proliferation, where PLGRKT is minimally affected, which is a key conclusion of our study, namely that many drivers manifest their phenotypes only *in vivo*. Consistent with this, our data further show that *PLGRKT* overexpression produces a measurable phenotype *in vitro* only under oxidative stress conditions, while under baseline growth conditions, no proliferative advantage is observed (**Fig. 5k**).



The effect of *PLGRKT* overexpression in our human re-screen is perhaps easier to discern in **Extended Fig. 6j** (left), where once again one can observe that *PLGRKT* is the third strongest hit with a **significant log fold change of 3.38** (=10.5-fold increase).

While our screen identified exceptionally strong drivers such as Fbxw7 and Apc - which effectively serve as robust positive controls - the presence of these established benchmarks does not categorize PLGRKT as a modest hit. On the contrary, we observe a 10.5-fold increase in tumorigenicity upon PLGRKT expression, an effect size that is highly significant for a previously uncharacterized oncogene.

To place this in a broader context, an effect of this magnitude is comparable to, or greater than, those reported for canonical drivers in seminal *in vivo* screening studies. For instance, primary data from the foundational work by Martin *et al*. (2021) demonstrate that loss of *Pten* or *Trp53* yields similar fold-change ranges in certain *in*

vivo settings. We believe that since these genes are universally recognized as potent drivers, the 10.5-fold acceleration driven by PLGRKT establishes it as a major contributor to mammary tumorigenesis.

TEXT REDACTED **FIGURE REDACTED**

Lastly, it is important to note that this screen was performed in MCF10A/PIK3CA^{H1047R}/MYC overexpressing cells, which are fully tumorigenic. When we tested the effect of PLGRKT overexpression outside the screening context, in weakly tumorigenic MCF10A/PIK3CA^{H1047R} lacking MYC overexpression, PLGRKT alone converted these cells into highly penetrant tumors *in vivo* (**Fig. 5a**), increasing tumor incidence from 20% to 90% (which is, in fact, similar to the effect of overexpressing MYC):

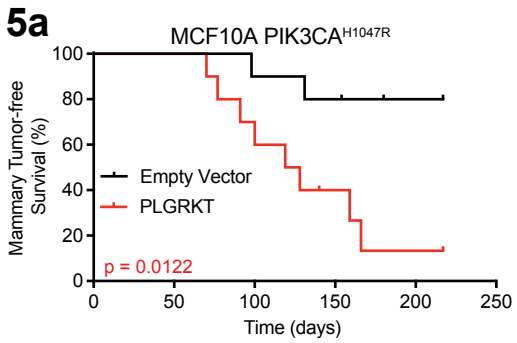


Fig. 5a: *Orthotopic transplantation of MCF10A PIK3CA^{H1047R} cells overexpressing PLGRKT significantly accelerates mammary tumor outgrowth (n=10 tumors per group). Statistical analysis was performed using the Gehan-Breslow-Wilcoxon test (compared to empty vector control).*

We therefore respectfully disagree that PLGRKT exhibits a "*modest*" effect in human models and believe these data represent robust human-relevant oncogenic activity, consistent with the Reviewer 3's request for functional validation in human systems.

To validate PLGRKT's potential as cancer target will require extensive additional studies and time. For example, we have now sourced Plgrkt$^{fl/fl}$ mice to enable conditional ablation in mammary tumors and are currently optimizing inducible degradation systems (e.g., degron tags) in our human MCF10A models. However, ensuring that these tags do not interfere with the non-canonical oncogenic functions and optimizing degradation efficiency *in vivo* in orthotopic tumors are technically demanding processes that require significant time (i.e. many months based on the KM blot in **Fig. 5a**) to validate and replicate. As such, while these experiments are most certainly vital next steps, we hope this reviewer agrees that they are outside of the scope of the current study, which focuses

on identifying driver genes of CNA and CNA biology rather than validating each driver gene's potential oncogene addiction and druggability.

We hope the reviewer agrees that our demonstrations of PLGRKT's ability to transform human cells and its strong association with clinical outcomes in patient data constitute a robust, human-relevant discovery that stands as a significant advance in its own right. We have added a section to the Discussion highlighting these ongoing efforts and the importance of future therapeutic validation.

4. Depth of mechanistic follow up: An expanded figure has been added to further deconvolute the mechanism by which PLGRKT enhances tumor formation, as described in earlier figures. However, this work stops short of elucidating the precise mechanism by which mitochondrial activity is increased; instead, it documents additional models in which this phenotype is observed, the co-localization of PLGRKT to mitochondria, and several assays used to measure mitochondrial activity.

We agree with Reviewer 3 that we do not fully "*elucidate the exact mechanism by which PLGRKT regulates mitochondrial activity*". However, the primary goal of this work is to identify and validate CNA driver genes through large-scale *in vivo* functional genomics - an area with limited prior knowledge. Within this scope, we provide in-depth, carefully evaluated mechanistic evidence demonstrating that PLGRKT is a *bona fide* oncogene, acting not through its canonical role in regulating plasminogen, but via a COMPLETELY novel function acting through regulation of mitochondrial homeostasis and ROS, supported by multiple orthogonal assays in mouse and human systems. Full structural or biophysical characterization would require extensive future work, and we hope that this reviewer and the editors agree that it is beyond the intended scope of the present study. Notably, seminal oncogene discovery papers historically did not include complete mechanistic resolution at the time of initial discovery.

TEXT REDACTED

FIGURE REDACTED

TEXT REDACTED

5.Systematic analysis of PLGRKT in human cancer. The authors address oncogene penetrance in additional tumor types, noting that fully resolving this question would require substantial additional data generation and may fall outside the scope of this manuscript. We agree with this assessment and appreciate the inclusion of PLGRKT amplification data across additional tumor types, noting that several exhibit significantly higher gain/amplification

frequencies than BLBC. The recurrent mutation data could be interpreted as somewhat counterintuitive, with a recurring frameshift mutation most likely resulting in loss of function of PLGRKT.

Reviewer 3 acknowledges that systematic analysis across tumor types "*would require additional data generation and may fall outside the scope of this manuscript*". We agree and in line with Reviewer 1 have focused the revised manuscript on breast cancer, including the revised title, which now reads "*Aneuploidy selects for the acquisition of driver genes in breast cancer*".

We have provided the pan-cancer analysis in the point-by-point response (but not the manuscript) and would be happy to include it, if the Editorial team believes it would strengthen the paper. Reviewer 3 did "*appreciate the inclusion of PLGRKT amplification data across additional tumor types, noting that several exhibit significantly higher gain/amplification frequencies than BLBC*". Lastly, we definitely agree with the Reviewer that "*the recurrent mutation data could be interpreted as somewhat counter-intuitive, with a recurring frameshift mutation most likely resulting in loss of function of PLGRKT*". However, this would obviously require empirical testing, and since this mutant is not observed in breast cancer, we feel it is beyond the scope of our current study.

Concluding Remarks

The core contributions of our work comprise:

1. **Methodological Innovation:** Establishment of **CRISPR-KOALA** as a broadly applicable *in vivo* screening platform.
2. **Functional Mapping:** Execution of a large-scale, bidirectional *in vivo* screen uncovering previously unrecognized CNA driver genes.
3. **Conceptual Resolution:** Demonstration that diverse drivers provide a fitness advantage that obviates the need for aneuploidy, explaining the heterogeneous landscape of BLBC.
4. **Cross-Species Validation:** Validation of hits in human model systems and TCGA datasets, ensuring clinical relevance.
5. **Mechanistic Discovery:** Identification of a novel, non-canonical mitochondrial function for PLGRKT in metabolic reprogramming and ROS detoxification.

TEXT REDACTED

As such, we are convinced that this manuscript represents the beginning of a new and exciting field in CNA driver gene biology, with substantial clinical relevance and strong interest from both academia and industry as these drivers, especially the oncogenes, are potential future drug targets and biomarkers. Furthermore, we show that the heterogeneous transcriptional landscape of BLBC drivers converges on a common outcome: providing incipient tumor cells with a strong fitness advantage that obviates the need for aneuploidy — a principle that appears to hold across cancer types.

References:

1 Watkins, T. B. K. *et al.* Pervasive chromosomal instability and karyotype order in tumour evolution. *Nature* **587**, 126-132 (2020). https://doi.org:10.1038/s41586-020-2698-6

2 Sheltzer, J. M. & Amon, A. The aneuploidy paradox: costs and benefits of an incorrect karyotype. *Trends Genet* **27**, 446-453 (2011). https://doi.org:10.1016/j.tig.2011.07.003

3 Ben-David, U. & Amon, A. Context is everything: aneuploidy in cancer. *Nat Rev Genet* **21**, 44-62 (2020). https://doi.org:10.1038/s41576-019-0171-x

4 Mao, J. H. *et al.* Fbxw7/Cdc4 is a p53-dependent, haploinsufficient tumour suppressor gene. *Nature* **432**, 775-779 (2004). https://doi.org:10.1038/nature03155

5 McFadden, D. G. *et al.* Genetic and clonal dissection of murine small cell lung carcinoma progression by genome sequencing. *Cell* **156**, 1298-1311 (2014). https://doi.org:10.1016/j.cell.2014.02.031

6 Baslan, T. *et al.* Ordered and deterministic cancer genome evolution after p53 loss. *Nature* **608**, 795-802 (2022). https://doi.org:10.1038/s41586-022-05082-5

Point-by-Point Response to Reviewer's Comments (*2024-08-17371*)

All of the editorial changes addressed below have been tracked within the submitted manuscript in green font.

Acknowledgment of Limitations

We have now included a section in the Discussion section that directly addresses several of the limitations associated with our study, including the consideration of potential off-target effects when performing CRISPR-KOALA, the potential for missing haploinsufficient tumor suppressive effects in our screens, and the inability of our current screening approaches to model the tumorigenic potential of gene-gene interactions:

"*A limitation of the CRISPR-KOALA methodology, inherent to all high-throughput formats, is the difficulty to ascertain whether the induced expression levels always precisely model the low-level CNAs typical of human tumors. This challenge is compounded by the fact that the precise protein levels resulting from a given CNA in patients are often unclear. However, we have provided substantial evidence for the physiological relevance of our approach. For gain-of-function hits, we demonstrated that our system induces protein expression for the proto-oncogene PLGRKT at levels that closely recapitulate the dosage seen in human tumors and similarly induces MRPS18B four-fold, which is sufficient to drive mammary tumor development. Second, our screen identified several established haploinsufficient tumor suppressors, including FBXW7, CSNK1A1, and ADAM10, whose dosage-sensitive functions have been independently confirmed. Despite these successes, potential off-target effects remain a challenge inherent to CRISPR formats, and our current approach cannot yet model the complex gene-gene interactions or epistatic synergies often seen in cancer progression. Future efforts will be required to determine whether these dosage sensitivities apply to the remaining 85 hits and to explore the cooperative landscape of these drivers through combinatorial screening.*"

Method and Statistical Reporting Checklists

The following clarifications were specifically requested as in the Reporting Summary by the editorial team. We have summarized the comments below and provided our response to each here.

1. Please see the extended comments document for a list of figures and figure legends that require additional information.

Reply: We have now updated all the requested clarifications to the statistical tests, exact replicate numbers, error bars, exact p-values, molecular weight markers, and data deposition in all the figure legends and figures that were highlighted as requiring attention. A detailed point-by-point description for each of these changes is included in the word document that was provided as an accompanying article file: **489435_2_attach_68_24802_revised**.

2. Please ensure the following software/packages/tools/algorithms are mentioned in the manuscript as well, since they have been listed in the reporting summary: BD FACS diva, etc.

Reply: We have included the software, tools, packages, and algorithms in the main text and in the reporting summary.

3. Please provide the version numbers of the following software in the reporting summary: BD FACS diva, etc.

Reply: We have included the version numbers for those that were missing in the reporting summary.

4. Please ensure that information on datasets used/generated is provided under the 'data' section of the reporting summary, as this information is not relevant to this field.

Reply: We have removed all the relevant dataset information out of the 'Software and code' section of the reporting summary and moved it to the 'Data' section, as requested.

5. Please note that the following GEO and SRA accession numbers are currently not available/private: GSE274219 and PRJNA1146341. Please ensure that this data is publicly released by the time you resubmit your final manuscript; we will not be able to accept your manuscript if the data is not publicly available.

Reply: Both datasets that we have generated in this study, GSE274219 and PRJNA1146341, have been made publicly available as of April 16, 2026.

6. Please mention all databases/datasets used in the study along with appropriately accessible links/accession-codes in the manuscript under the “Data availability” section as well as in this reporting summary: OncoKB database, etc.

Reply: We have amended the Data Availability Statement in the manuscript to include the accession numbers of the RNA-seq and WGS data that we generated, in addition to including the information regarding the other public datasets that were used in our analysis (METABRIC, TCGA, OncoKB, and various single cell RNA-sequencing data sets). We have also added the gene-level copy number calls for all of the mouse whole genome sequencing data as a resource through Figshare (through the manuscript submission portal), and have added the DOI and link as indicated into the Data Availability Statement.

7. Please determine how sample size was determined for experiments other than those mentioned here, detailing any statistical methods used to predetermine sample size OR if no sample size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these are sufficient.

Reply: We have provided additional information in the reporting summary for describing the specific sample size for the remaining mouse experiments (in addition to the screening data that was initially provided) and for the cell culture-based experiments. We would also like to highlight here that during the revision, we failed to include the information on data exclusions and have updated this to provide the two instances used in our manuscript.

8. Please state how often experiments other than CRISPR hits were replicated or performed independently and also confirm if all attempts at replication were successful.

Reply: We have included that all additional experiments to the CRISPR validations were performed in at least three independent replicates and that all attempts at replication were successful in the reporting summary.

9. For experiments other than those involving mice, please describe how samples were allocated into experimental groups. If allocation was not random, describe how covariates were controlled OR if this is not relevant to those experiments, explain why.

Reply: We have now described in the reporting summary how all of the cell culture experiments were pre-assigned randomly into experimental groups prior to initiating the experiment.

10. For remaining experiments, please provide a rationale as to why no blinding was used.

Reply: We have clarified this section of the reporting summary to much greater detail than was initially reported. We have included now a list of all experiments/methods where reviewers were or were not blinded at the data collection and/or analysis stage, and that in instances where no blinding was used it was due to unbiased sample collection on an instrument which pre-informed the researcher of the experimental groups.

11. Please provide the clone names for the following monoclonal antibodies in the reporting summary: anti-Pan-Cytokeratin, etc.

Reply: We have now included the clone names for all the monoclonal antibodies used in the reporting summary (Pan-Cytokeratin, HA, TOMM20, TOMM20-AF647, ERBB2, beta-Actin, and GAPDH).

12. Please specify the dilution used for the following antibody in the methods section of the manuscript/supplementary table 11: Anti-Rabbit IgG (Vector Laboratories).

Reply: The Anti-Rabbit IgG (Vector Laboratories) secondary is used without dilution. We had previously written “neat” in Supplementary Table 11 but have now changed this to “Undiluted” for increased clarity.

13. Please describe the validation of all primary antibodies for the species and application, noting any validation statements on the manufacturer’s website and in relevant citations. Please include this information directly in the reporting summary.
Reply: In the reporting summary we have now included that all primary antibodies were validated using knockout cell lines and/or recombinant proteins for the respective applications. We have also stated that the validation information for each individual antibody is available directly from the manufacturer’s website. We have also updated the “Histology” paragraph of the Methods section to reflect this as well.

14. For MCF10A cell lines sensitized towards oncogenic transformation with PIK3CAH1047R and MYC, please describe the authentication procedures for these cell lines used OR declare which of the cell lines used were not authenticated.

Reply: We have clarified that the MCF10A cell lines sensitized towards transformation with PIK3CAH1047R and/or MYC were not independent authenticated.

15. Please provide information on housing conditions for the mice, describing dark/light cycle, ambient temperature, and humidity in the manuscript.

Reply: We have now included the information on the dark/light cycle, ambient temperature, access to food and water, and humidity within the animal facility to the reporting summary and the Methods section “Animal husbandry”.