

**Editorial Note:** This manuscript has been previously reviewed at another journal that is not operating a transparent peer review scheme. This document only contains reviewer comments and rebuttal letters for versions considered at Nature Communications. Mentions of the previous journal have been redacted as indicated.

Reviewers' comments:

Reviewer #1 (Remarks to the Author): This reviewer was Reviewer #1 for [redacted]; evolution of resistance and trade-offs

The authors have failed to convince me that the comments made on the [redacted] version of this manuscript are not relevant. I fail to see the fundamental difference between this manuscript and the earlier barcoding manuscripts. There are no new biological insights gained from this study.

Reviewer #4 (Remarks to the Author): Recruited to replace Reviewer #3; modelling and clonal evolution

Acar and colleagues describe in this manuscript an experimental set up to study growth dynamics of multi-clonal cancer cells under treatment. The hyperflask setting is certainly intriguing as it allows to grow unprecedentedly large cell populations and, thus, to test shifts of clone fractions under treatment in reproducible experiments.

The authors provide here an interesting proof of principle of this approach using a NSCLC cell line with a previously reported subclone harboring a MET amplification, which leads to resistance to the EGFR inhibitor Gefitinib. I'm not sure whether the CDKN2A deletion-subclone was also previously reported, but the data shown in this paper strongly support it as pre-existing (although it is not conclusive).

Although I have the feeling that more could have been done to prove the power of this approach (the results supporting the value of this approach to discover effective sequential treatments are limited, e.g. validation experiments of the drug screening would have been nice) I understand this manuscript has already been reviewed and therefore I do not intend to ask for substantial new experiments or analyses. Overall, I have no strong objections against its publication in Nature Communications.

I'd like to offer either way a few suggestions and minor comments that I'd ask the authors to address:

1) As mentioned above, the main limitation of this study is that it remains a proof of principle with no new results, except for the approach itself. Given that, it would be helpful to provide/discuss more extensively guidelines on when and how this approach is meaningful and should be used. For example, what happen if a cell line is clonal, can this approach still be used to study the emergence of new mechanisms of resistance in a faster way, given the large sample size? Is it a particularly suited system to be combined with high-throughput screening like the one performed in Figure 6? Do the author envision it could be expanded to grow immortalized cells from primary tumors (which could display higher heterogeneity than cell lines) or the condition are too difficult to grow such models?

2) Figure 1, 2, and 3 are almost exclusively representative diagrams / toy images of the approaches here taken, we only get to actual data in Figure 4. This seems a bit odd. I suggest condensing this information into 1 figure and move most of the panels to the supplemental material (e.g. Figure 2). It is my opinion that 4 figures in total are here more than enough.

3) The authors talk about mathematical modeling in the abstract and on page 8, the latter in reference to the estimated growth rates. The only mathematical model that I could find described in this manuscript is the one used to estimate the time of emergence of resistant mutations in Figure 1D. Growth rates are estimated from measured barcode frequencies using a simple log fold-change expression (in the methods the authors indeed talk about "bioinformatics analysis"). Unless I missed something (my apologies if so), I would remove the term 'mathematical modeling' from the abstract and the text (except maybe for the data shown in Figure 1D).

4) "Moreover, cell plasticity and drug tolerance, instead of Darwinian adaptation, often occurs in current model systems, leading to resistance that is non-heritable, potentially reversible, and that does not represent what happens in the clinic. Non-heritable drug resistance can arise through epithelial-mesenchymal transition<sup>34</sup> or upregulation of drug-efflux pumps<sup>35</sup>. Although these are very important cellular mechanisms of resistance, they do not pertain to clonal evolution, which drives persistent resistance in human cancers over long timescales"

This is clearly an overstatement and a misleading/self-contradicting sentence. Saying that cell plasticity does not represent what happens in the clinic and then mention EMT as an example is just wrong. Cell plasticity is relevant in the clinic, as is EMT, it is heritable (in a cancer pertaining sense, i.e. transmissible through cell generations), these mechanisms do pertain to clonal evolution as demonstrated by the authors themselves in Figure 6B and the argument on 'persistent resistance' although reasonable is not proven. The authors should just remove this paragraph (which is unnecessary) or substantially correct it.

5) It is surprising to see in Figure 6F that DMSO and POT do not respond to Gefinitib, given this should correspond to the initial population that is sensitive to this drug. Is this dose related? Can the authors comment on that?

Given these resistant clones are pre-existing, one is left to wonder whether a combination of MET and EGFR inhibitors (at different concentrations) would have been effective from the beginning or whether the sequential treatment would have still been better. This could be a simple and interesting experiment to do and could reinforce the relevance of the approach.

5bis) Related to this: A deep targeted sequencing of the cell line here used would have led to the discovery of MET and CDKN2A alterations? (both MET and CDKN2A are in many targeted panels)

6) Minor typos:

- Page 8: "As this group comprises of barcodes" should be "As this group comprises barcodes" or "As this group is composed of barcodes"
- Page 11: "CDK4/5 inhibitor" should be "CDK4/6 inhibitor"
- Fig 6G: legend is missing the label, what do these colors/numbers represent?

Reviewer #5 (Remarks to the Author): Recruited to replace Reviewer #2; lung cancer and EGFR signalling

The authors have submitted a revised manuscript and response letter. This reviewer was asked to assess the comments of reviewer #2 (original). The authors have done a good job in addressing the comments. The main issue this reviewer sees with the revised manuscript is the rather incremental advance of the study compared to the prior literature - as commented upon by both reviewer #1 and #3.

Reviewers' comments:

We thank all reviewers again for their assessment of our manuscript and constructive criticisms. Below is the point by point response to reviewer #4 comments.

Reviewer #4 (Remarks to the Author): Recruited to replace Reviewer #3; modelling and clonal evolution

Acar and colleagues describe in this manuscript an experimental set up to study growth dynamics of multi-clonal cancer cells under treatment. The hyperflask setting is certainly intriguing as it allows to grow unprecedentedly large cell populations and, thus, to test shifts of clone fractions under treatment in reproducible experiments. The authors provide here an interesting proof of principle of this approach using a NSCLC cell line with a previously reported subclone harboring a MET amplification, which leads to resistance to the EGFR inhibitor Gefitinib. I'm not sure whether the CDKN2A deletion-subclone was also previously reported, but the data shown in this paper strongly support it as pre-existing (although it is not conclusive).

We thank the reviewer for appreciating the novelty of the approach presented in this study. We do confirm that to the best of our knowledge, CDKN2A loss has never been associated to trametinib resistance. We highlight this more clearly in our revised version of the manuscript (page 6, text in red).

Although I have the feeling that more could have been done to prove the power of this approach (the results supporting the value of this approach to discover effective sequential treatments are limited, e.g. validation experiments of the drug screening would have been nice) I understand this manuscript has already been reviewed and therefore I do not intend to ask for substantial new experiments or analyses. Overall, I have no strong objections against its publication in Nature Communications. I'd like to offer either way a few suggestions and minor comments that I'd ask the authors to address:

1) As mentioned above, the main limitation of this study is that it remains a proof of principle with no new results, except for the approach itself. Given that, it would be helpful to provide/discuss more extensively guidelines on when and how this approach is meaningful and should be used. For example, what happens if a cell line is clonal, can this approach still be used to study the emergence of new mechanisms of resistance in a faster way, given the large sample size? Is it a particularly suited system to be combined with high-throughput screening like the one performed in Figure 6? Do the authors envision it could be expanded to grow immortalized cells from primary tumors (which could display higher heterogeneity than cell lines) or the conditions are too difficult to grow such models?

This is a very interesting point. Indeed, in a clonal cell line with no pre-existing resistance one would expect that all resistance dynamics are driven by drug tolerance followed by de novo resistance. These dynamics are often described by bet-hedging phenomenon (Nichol et al. 2016 – Genetics). This could be definitely studied with the presented platform, although one would expect different barcodes in different replicas, making the evolution highly stochastic. On the other hand, the floating barcodes allow to determine the waiting time of a de novo mutant with great precision, and hence measure the temporal dynamics in the context of bet-hedging, which are key to understand mutation rates and dynamics of resistance. We discuss this aspect further in the revised version of the manuscript (page 13, paragraph in red). The point of whether this approach can be extended to patient-derived lines is a great point and we discuss this potential application in the revised version of the manuscript (page 13, paragraph in red).

2) Figure 1, 2, and 3 are almost exclusively representative diagrams / toy images of the approaches here taken, we only get to actual data in Figure 4. This seems a bit odd. I suggest condensing this information into 1 figure and move most of the panels to the supplemental material (e.g. Figure 2). It is my opinion that 4 figures in total are here more than enough.

We agree with this reviewer. We have combined Figure 1 and 2 into a single figure. Given the complexity of the experimental assay and for clarity, we have left old Figure 3 on its own however.

3) The authors talk about mathematical modeling in the abstract and on page 8, the latter in reference to the estimated growth rates. The only mathematical model that I could find described

in this manuscript is the one used to estimate the time of emergence of resistant mutations in Figure 1D. Growth rates are estimated from measured barcode frequencies using a simple log fold-change expression (in the methods the authors indeed talk about "bioinformatics analysis"). Unless I missed something (my apologies if so), I would remove the term 'mathematical modeling' from the abstract and the text (except maybe for the data shown in Figure 1D).

We do agree that we do not sufficiently expand on the importance of mathematical modelling in the manuscript. However, we maintain it is key to do simulations with stochastic branching processes to perform experimental design, hence our introductory work in Figure 1, Material and Methods (first section) and Supplementary Note. Following from that it's the case that we simplified the maths to simply calculate growth rates, since the time-course barcodes give enough information to fit a simple model. In light of this comment and others, we have extended the part on mathematical modelling of the experimental design and barcodes (see Material and Methods, page 16, text in red and new Figure S2). Given the mathematical modelling embedded at the core of the design of the whole experiment (see Figure 1), we kept the term "mathematical modelling" in the abstract.

4) "Moreover, cell plasticity and drug tolerance, instead of Darwinian adaptation, often occurs in current model systems, leading to resistance that is non-heritable, potentially reversible, and that does not represent what happens in the clinic. Non-heritable drug resistance can arise through epithelial-mesenchymal transition<sup>34</sup> or upregulation of drug-efflux pumps<sup>35</sup>. Although these are very important cellular mechanisms of resistance, they do not pertain to clonal evolution, which drives persistent resistance in human cancers over long timescales"

This is clearly an overstatement and a misleading/self-contradicting sentence. Saying that cell plasticity does not represent what happens in the clinic and then mention EMT as an example is just wrong. Cell plasticity is relevant in the clinic, as is EMT, it is heritable (in a cancer pertaining sense, i.e. transmissible through cell generations), these mechanisms do pertain to clonal evolution as demonstrated by the authors themselves in Figure 6B and the argument on 'persistent resistance' although reasonable is not proven. The authors should just remove this paragraph (which is unnecessary) or substantially correct it.

Indeed, we fully agree with this reviewer and we apologise for the confusion. We have now changed the introduction and the conclusions to highlight the importance of cell plasticity and heritable non-genetic changes such as EMT (page 2, text in red).

5) It is surprising to see in Figure 6F that DMSO and POT do not respond to Gefitinib, given this should correspond to the initial population that is sensitive to this drug. Is this dose related? Can the authors comment on that?

Given these resistant clones are pre-existing, one is left to wonder whether a combination of MET and EGFR inhibitors (at different concentrations) would have been effective from the beginning or whether the sequential treatment would have still been better. This could be a simple and interesting experiment to do and could reinforce the relevance of the approach.

5bis) Related to this: A deep targeted sequencing of the cell line here used would have led to the discovery of MET and CDKN2A alterations? (both MET and CDKN2A are in many targeted panels)

This is a good point that concerned us in the original manuscript. In Supplementary Figure 4 we performed digital droplet PCR, which is ~10 times more sensitive than targeted sequencing, to assess whether MET amplification and CDKN2A loss was detectable in the original POT population. The issue is that contrary to point mutations, copy number alterations are read as an average of the copies per cell in a bulk sample, hence ddPCR nor sequencing can pick up low frequency copy number changes in those samples. However, we do note that in the scRNA-seq experiment, a small subset of single cells with MET amplification and CDKN2A loss are found in the original POT (see Supplementary Figure 12).

It is indeed surprising that combination of gefitinib and captamatinib shows decreased sensitivity in POT with respect to gefitinib alone. We speculate that given the consistent sensitive profiles for gefitinib in this line, this is possibly due to drug antagonism. We report this in the revised manuscript.

6) Minor typos:

- Page 8: "As this group comprises of barcodes" should be "As this group comprises barcodes" or "As this group is composed of barcodes"

Sorry, corrected.

- Page 11: "CDK4/5 inhibitor" should be "CDK4/6 inhibitor"

Sorry, corrected.

- Fig 6G: legend is missing the label, what do these colors/numbers represent?

Sorry, corrected.

## **Description of Additional Supplementary Files**

File Name: Supplementary Data 1

Description: Copy number estimates from whole-exome sequencing.

File Name: Supplementary Data 2

Description: Barcode frequencies.

File Name: Supplementary Data 3

Description: Coverage values of whole-exome sequencing per sample.

File Name: Supplementary Data 4

Description: Single nucleotide variant calls.

File Name: Supplementary Data 5

Description: List of compounds tested in the high-throughput drug screening.

## Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

### Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided  
*Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g.  $F$ ,  $t$ ,  $r$ ) with confidence intervals, effect sizes, degrees of freedom and  $P$  value noted  
*Give  $P$  values as exact values whenever suitable.*
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's  $d$ , Pearson's  $r$ ), indicating how they were calculated

*Our web collection on [statistics for biologists](#) contains articles on many of the points above.*

### Software and code

Policy information about [availability of computer code](#)

Data collection All described in the Methods section

Data analysis All described in the Methods section and Supplementary Information

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

### Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Sequence data have been deposited at the European Genome-phenome Archive (EGA), which is hosted by the EBI and the CRG, under accession number EGAS00001003200. Further information about EGA can be found on <https://ega-archive.org>.

### Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- Life sciences  Behavioural & social sciences  Ecological, evolutionary & environmental sciences

## Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	<i>Describe how sample size was determined, 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 sample sizes are sufficient.</i>
Data exclusions	<i>Describe any data exclusions. If no data were excluded from the analyses, state so OR if data were excluded, describe the exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.</i>
Replication	<i>Describe the measures taken to verify the reproducibility of the experimental findings. If all attempts at replication were successful, confirm this OR if there are any findings that were not replicated or cannot be reproduced, note this and describe why.</i>
Randomization	<i>Describe how samples/organisms/participants were allocated into experimental groups. If allocation was not random, describe how covariates were controlled OR if this is not relevant to your study, explain why.</i>
Blinding	<i>Describe whether the investigators were blinded to group allocation during data collection and/or analysis. If blinding was not possible, describe why OR explain why blinding was not relevant to your study.</i>

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

### Materials & experimental systems

n/a	Included in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data

### Methods

n/a	Included in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging