Reply: Neutral tumor evolution?

Timon Heide[1,*], Luis Zapata[1,*], Marc J. Williams[2-4,*], Benjamin Werner[1,*], Giulio Caravagna[1], Chris P. Barnes[3,5], Trevor A. Graham[2,^] & Andrea Sottoriva[1,^]

[1] Evolutionary Genomics & Modelling Lab, Centre for Evolution and Cancer, The Institute of Cancer Research, London, UK.
[2] Evolution and Cancer Laboratory, Barts Cancer Institute, Queen Marry University of London, London, UK.
[3] Department of Cell and Developmental Biology, University College London, London, UK.
[4] Centre for Mathematics and Physics in the Life Sciences and Experimental Biology (CoMPLEX), University College London, London, UK.
[5] Department of Genetics, Evolution and Environment, University College London, London, UK.

* These authors contributed equally to this work
^ For correspondence: t.graham@qmul.ac.uk and andrea.sottoriva@icr.ac.uk

We thank Tarabichi and colleagues for the constructive criticism of our Williams *et al.* 2016[1] work. Their critique has four main points that we address in this document using a simulation approach and reanalysis of public datasets.

1. Impact of clonal copy number alterations

In Williams *et al.* 2016[1], we assessed the cumulative VAF distribution M(f) over the frequency range of [0.12,0.24] in order to restrict our analysis to subclonal variants within a range that would be applicable to the diverse datasets we considered. Tarabichi and colleagues note that tumours with a tetraploid genome will have a 'peak' of clonal mutations at f~0.25 (mutations in a single allele, Supplementary Figure 1A), thus causing an 'artificial deviation from the linear fit' and incorrect rejection of neutrality. The integration range we chose was based on a triploid tumour with read depth of 100X, giving an upper threshold of 0.26 (see Supplementary Methods).Although this is suitable for the majority of cases, we agree that a tetraploid tumour could have a peak within our integration range, and therefore potentially more tumours could be consistent with neutral evolution. In Supplementary Figure 1B we show that the additional clonal peak causes a false rejection of neutrality, whereas identifying the tumour correctly as tetraploid and selecting the correct integration range fixes the problem (Supplementary Figure 1C).

We do acknowledge that the 1/f integration method, while representing the correct analytical derivation, is sensitive to the choice of integration range and is most accurate when applied to the whole VAF spectrum of subclonal mutations only. We have recently developed better metrics, as well as a Bayesian model selection framework that directly compares the neutral model against models with selection using the entire VAF distribution[2]. We also contributed to the development of additional classification methods for neutrality that exploit multi-region sequencing[3]. We care to stress however, that the majority of cancers we analysed were *not neutral* and showed signs of subclonal selection.

2. Interpretation of the 1/f statistical test

Tarabachi and colleagues correctly note that failing to reject the null is not necessarily evidence for the null. This is absolutely true, but evolutionary analysis of cancer genomic data requires a hypothesis-driven approach based on a sensible null. Analysing data without knowing what to expect in the simplest scenarios may lead to wrong conclusions, as we have recently highlighted[3]. The fundamental message in our original manuscript is that neutrality, the null model of molecular evolution[4], was often a sufficient explanation of the available data. This logic is a valid frequentist perspective on hypothesis testing[5]. The test we applied quantifies the deviation from the null distribution in terms of a change in model parameter ($s = 0$ vs $s > 0$). This structure arises from a frequentist approach and arguments for setting up the test in any other way are arbitrary. There are an infinite number of models of selection, some produce vanishing deviations from neutrality that are not measurable (e.g. weak selection), while others are biologically unrealistic (e.g. each

mutation is a driver, constant population size in cancer). This is why in molecular evolution, neutrality is the null[4]. In our view, selection remains arguably the most important force in cancer, but maintaining a sensible null model is required to avoid over-interpreting the data.

Tarabichi and colleagues state that the deterministic solution we report in our manuscript (Eq.7) relies on the strong assumption of synchronous cell divisions. That is not the case, Eq.7 is the convergent solution of a continuous-time stochastic branching process[6]. They also state that simulating stochastic processes is more realistic. This is indeed what we did in our original manuscript using stochastic simulations (Fig S9-S11), demonstrating the convergence to the deterministic solution. A comprehensive analysis of the underlying stochastic Luria-Delbrück model shows that the scaling behaviour remains unchanged even in the explicit presence of stochastic cell death[7]. In Tarabichi's letter Figure 1b, the claim that a stochastic neutral model does not imply 1/f is therefore incorrect, as also demonstrated by others before us[6-8].

3. Insights from simulated tumours

Tarabichi and colleagues use a stochastic branching process, virtually identical to the one[1], to generate synthetic genomic data and test our method. In their Figure 1, Tarabichi *et al.* present a synthetic analysis of the 1/f test using the analytical deterministic solution (Figure 1a) and stochastic simulations (Figure 1b). In both analyses, a new subclone is introduced at a certain fixed time point. First, we noted that the stochastic simulation result does not appear to converge to the deterministic one (i.e. Tarabichi's Figure 1a is different from 1b), which is at odds with previous literature[1,6-9].

We note that the choice of simulated parameters in their synthetic test is unrealistic. The driver event in the newly selected subclone does not just induce a selective advantage, modelled as an increase net-growth rate ($adv_{subclone}$), but *at the same time*, also a change in mutation rate. Curiously, the new subclone can have lower mutation rate (up to 8 times lower) or higher mutation rate (up to 100 times higher). A mutation rate of 1024 mutations per cell division (see Tarabichi's Figure 1a, x-axis, $\mu_{subclone}=2^{10}$) is found only in a very small set of colorectal or uterine cancers with mutations in the proof-reading domain of POLE or POLD. A POLE subclone arising within a POLE wild-type background appears to be a very rare event[10]. Thus, we urge caution when considering the implications of the parameters at the extremities of the range considered by Tarabichi *et al*.

Nonetheless, we have reproduced Tarabichi's Figure 1b using our stochastic branching process simulation with precisely the same parameters (Figure 1A in this document, see Supplementary Methods). We found that when selection generated a detectable subclonal cluster with $f_{subclone}\geq 10\%$, this was correctly identified by the 1/f test in the majority of cases and neutrality was rejected (top left quadrant of Figure 1A, example in Figure 1B, 1/f tail of new subclone in green). For the majority of cases where the 1/f test failed, this was due to the new subclone being very small because of weak selection ($adv_{subclone}<0.5$, bottom half of Figure 1A, example in Figure 1C). Figure 1D illustrates the relationship between selective advantage and the subclone cell fraction in the final tumour, highlighting the issue of the limit of detectability (LOD) even assuming 100x depth of coverage. We have specifically quantified this effect[2], identifying a 'wedge of selection' that describes the detectability problem in cancer genomic data at current resolution. In general, the fact that selection is inefficient in expanding populations is well known in population genetics[11] and we have demonstrated this effect in colorectal cancer[12]. If subclonal selection does not significantly change the clonal composition of the tumour, the signature of neutral growth ('1/f tail') dominates the detectable VAF spectrum (Figure 1C, bottom part of Figure 1A).

Notably, on the right-hand side of Figure 1A, where the subclone had both a selective advantage and was also hypermutant ($\mu_{subclone}>=64$), the analysis showed that a hypermutant subclone generates a massive 1/f tail containing thousands of the subclone's own private mutations that dominate the entire VAF distribution, obscuring the underlying subclonal structure (which was generated with the old mutation rate, and hence with many less mutations). In the VAF distribution

of these tumours it is very difficult to identify the subclonal cluster because it is tiny with respect to the subclone's enormous 1/f tail, generated with up to 1000 mutations per cell division (top right quadrant of Figure 1A, example in Figure 1E). It is not surprising that our test, or any other test, would struggle to detect any subclonal cluster or deviation from 1/f in these VAF distributions. Nevertheless, in this potentially unrealistic scenario Tarabichi and colleagues have clearly demonstrated how 1/f tails can be the dominant signal in cancer genomic data, to the extent where they can dominate the entire subclonal structure. Curiously, for moderate values of selection ($adv_{subclone}$~0.5), a change in mutation rate from normal to hypermutant could be detected, leading to rejection of neutrality (mid-right area in Figure 1A and example in Figure 1F). An analogous deviation from 1/f caused by an increase in mutation rate was discussed in Williams *et al.* 2016, Figure S11H (see Supplementary Figure 2). For weak selection and a hypermutator subclone, the new subclone did not reach a detectable size and therefore neutrality could not be rejected (as in Figure 1C).

Importantly, we note that the lack of discriminatory power under these peculiar scenarios is not dependent on our test, but rather on the intrinsic limitations and confounding factors in the data. To demonstrate this we compared the performance of our 1/f test using an extended VAF integration range of subclonal mutations (Figure 2G) with DPclust, an established subclonal reconstruction approach based on Dirichlet clustering[13] developed by some of the authors of the letter (Figure 2H). Application of these two methods to the same set of synthetic data in Figure 2A (see Supplementary Methods), demonstrated that severe identifiability issues affect clustering methods as well. Even under optimal circumstances of high selection, the sensitivity of a DPclust remained low in the vast majority of cases (Figure 2H).

Importantly, we are pleased that in their letter, Tarabichi and colleagues confirm that both in the presence of a purely neutral process and in the case of subclonal selection, 1/f neutral tails are predicted to be pervasive in cancer data. In some of their simulations, they are so pervasive that they dominate the entire VAF distribution. This is because 1/f tails are a simple consequence of clonal growth, with each individual clone generating its own neutral tail during the expansion[1,2].

4. Analysis of subclonal selection using dN/dS ratios

Using a test inspired by the classical dN/dS method, the authors claim to find evidence of subclonal selection in cancers classified as neutral with our 1/f test. Specifically, they first classify tumours as neutral/non-neutral using our 1/f method, and then pool the subclonal mutations in 192 known cancer genes from different tumours to calculate a dN/dS value for neutral vs non-neutral groups. Subclonal mutations in the neutral group should not contain evidence of selection (dN/dS should not be significantly higher than 1). Conversely, subclonal mutations in non-neutral cancers, as well as clonal mutations in all cancers are expected to contain selected genes (in neutrally growing tumours selection was present during tumorigenesis), thus leading to dN/dS>1.

In the attempt to address this criticism, we have reproduced Tarabichi *et al*'s analysis using the same dN/dS method[14]. We first measured dN/dS values in the colorectal and gastric cancers analysed in our original manuscript using the 369 cancer genes from Martincorena *et al*[14]. We could not fully reproduce Tarabichi *et al* pan-cancer TCGA analysis because the CAVEMAN somatic calls the authors used are not publicly available. We therefore reanalysed the pan-cancer TCGA dataset using the variant calls publicly available from the GDC data portal (see Supplementary Methods).

We found that in all three cohorts, subclonal missense mutations in neutral-classified tumours, precisely as predicted by neutrality, had dN/dS vales that were not higher than 1, thus confirming our findings (Figure 2A-C, missense mutations on the left, green bars). This also recapitulated Tarabichi *et al*'s analysis for missense mutations. Moreover, in the pan-cancer cohort, missense mutations in non-neutral cancers were significantly higher than 1 (Figure 2C).

We then inspected nonsense mutations and found a very small group of neutrally classified patients with 3 or more subclonal nonsense mutations in cancer driver genes in the gastric cohort

(1/57 patients, 1.7%) and in the pan-cancer cohort (11/278, 3.9%). Another 7/278 pan-cancer cases (2.5%) had 2 subclonal nonsense mutations.

It is important to note that dN/dS is a cohort-level analysis (all mutations from all patients are pooled together) whereas our neutrality test is a patient-specific analysis. dN/dS statistics are extremely sensitive to nonsense mutations because they are rare, and therefore even a single misclassified patient carrying several nonsense mutations in cancer driver genes would significantly boost the dN/dS value due to mutations in the same patient 'counted' multiple times. This results from the fact that, although dN/dS can reveal an excess or depletion of mutations in a cohort, it doesn't differentiate whether this excess is coming from one or from multiple patients because the statistic is not normalised for the number of mutations per case. This means that a set with >90% neutral tumours could still show dN/dS>1, and the signal could be driven by a few outliers in the cohort.

We manually explored the pan-cancer patients with multiple subclonal nonsense variant (see Supplementary Figure 3) and found that these were due to clonal mutations 'bleeding' into the subclonal range, as well as misclassifications caused by ploidy errors, and possibly the presence of selected subclonal clusters hidden underneath 1/f tails. Moreover, we note that these TCGA tumours were classified with our original 1/f test and limited integration range, and we have now developed better classification methods.

Importantly, removing these few misclassified patients lowered the dN/dS value of the whole cohort to non-significant (Figure 2C, nonsense, green bars). After removing just 3.9% of patients with 3 or more nonsense mutations from the pan-cancer cohort (leaving 96.1% of putatively neutral cases), the dN/dS value for nonsense became not significant with respect to the background (Figure 3D, p=0.19). Even lower dN/dS values were reached by removing the additional 7/278 patients with 2 or more subclonal nonsense mutations (p=0.36). We demonstrate this clearly by generating dN/dS values for 'control sets' of passenger genes using bootstrapping of 1,000 random sets of 198 non-driver genes as well as neutral genes (Figure 3D). We noted that there was a systematic positive bias for the estimation of dN/dS that were consistently above 1. This could be due to the fact that public GDC calls are depleted of synonymous somatic mutations that were removed because present in the dbSNP database, thus skewing the dN/dS values, as mentioned in Martincorena *et al* 2017. Importantly, for tumours classified as non-neutral, dN/dS values remained higher than background for both missense and nonsense, in line with our classification.

Although we do acknowledge that the original 1/f test was not optimal, removing a few misclassified patients and applying the dN/dS orthogonal method confirmed neutrality in >93% of tumours classified as neutral by our method proposed in Williams *et al.* 2016. Our analysis also highlights the fact that dN/dS values for nonsense mutations are not a representative summary statistic for a cohort because a few outliers can drive all the signal. For a fair comparison between these two methods, a per-patient dN/dS analysis is required.

Despite some disagreements, Tarabachi and colleagues provided some valid constructive criticism of our original manuscript. In our assessment of this critique however, our original conclusion remains valid: that neutral evolution provides an entirely adequate description of the pattern of intra-tumour heterogeneity that has been observed to date across many tumours. Importantly, we thank Tarabichi and co-authors because they led us to the finding that VAF distribution analyses applied to single patients, like our neutrality test, can be carefully combined with cohort-level statistics like dN/dS to increase the power to discriminate between neutral dynamics and selection in cancer.

Contributions

TH performed simulation and bioinformatic analyses. LZ performed dN/dS analysis. MW performed additional simulation analysis. BW performed mathematical analysis. GC provided expertise on clustering analysis. All authors addressed the criticism and wrote the manuscript.

Acknowledgements
A.S. is supported by The Chris Rokos Fellowship in Evolution and Cancer and by Cancer Research UK (A22909). T.A.G. is supported by Cancer Research UK (A19771). C.P.B. is supported by the Wellcome Trust (097319/Z/11/Z). B.W. is supported by the Geoffrey W. Lewis Post-Doctoral Training fellowship. A.S. and T.A.G. are jointly supported by the Wellcome Trust (202778/B/16/Z and 202778/Z/16/Z respectively). M.J.W is supported by a Medical Research Council student scholarship.

Competing financial interests
We declare no competing financial interests.

Figure Legends

Figure 1. Insights from stochastic simulations of cancer growth. (A) Heatmap recapitulating Tarabichi's Figure 1b with same parameter set and showing proportion of simulations where neutrality was rejected. **(B)** Example VAF distribution with a detectable subclonal cluster (dashed line indicates subclone frequency) in which the 1/f test rejects neutrality in favour of selection (R^2 reported). **(C)** Example VAF distribution with a weakly-selected subclone that remains below the limit of detection in the data (100X depth). **(D)** Subclone cell fraction in the final tumour as a function of fitness advantage, for $adv_{subclone}$<0.5 the subclone rarely reaches the detectable size of ~10% cell fraction, the approximate limit of detectability (LOD) of 100x depth of sequencing. **(E)** Example VAF distribution for a subclone with selective advantage and, at the same time, high mutation rate. **(F)** Example VAF distribution for a selected and extreme mutator subclone. **(G)** Sensitivity of the 1/f test applied to subclonal mutations in the extended range of VAF=[0.025, 0.45] from the simulations in panel A, numbers report proportion of cases where neutrality was rejected (R^2<0.98). **(H)** Sensitivity of subclone detection of DPclust, a Dirichlet subclonal clustering method, when applied to the same simulated data in panel A. Numbers report the proportion of cases (20 cases per parameter combination) where the correct subclone has been identified (allowing for a 5% CCF error with respect to the true position).

Figure 2. Detecting subclonal selection with dN/dS analysis. dN/dS analysis using Martincorena *et al.* 2017 method applied to the colorectal cancers **(A)**, gastric cancers from ref[15] analysed in Williams *et al.* 2016 **(B)**, and TCGA pan-cancer analysis using newly available GDC calls to reproduce Tarabichi's dN/dS analysis **(C)**. In each type of cancers, the cancers were classified as neutral or non-neutral using the 1/f test, and the dN/dS values of clonal and subclonal variants assessed using the Martincorena method for the pooled variants in each group. **(D)** Analysis of systematic bias in the dN/dS estimates calculated from neutral and non-neutral subclonal mutations from the pan-cancer cohort for three sets of genes unlikely to contain drivers: neutral genes from Zapata *et al.*[16], non-driver genes (all but the 198 cancer genes), and neutral genes from Martincorena *et al.*{Martincorena:2017bla}.

References

1. Williams, M. J., Werner, B., Barnes, C. P., Graham, T. A. & Sottoriva, A. Identification of neutral tumor evolution across cancer types. *Nature Genetics* **48,** 238–244 (2016).
2. Williams, M. J. *et al.* Quantification of subclonal selection in cancer from bulk sequencing data. *Nature Genetics* **50,** 895–903 (2018).
3. Sun, R. *et al.* Between-region genetic divergence reflects the mode and tempo of tumor evolution. *Nature Genetics* **49,** 1015–1024 (2017).
4. Wu, C.-I., Wang, H.-Y., Ling, S. & Lu, X. The Ecology and Evolution of Cancer: The Ultra-Microevolutionary Process. *Annual Review of Genetics* (2016). doi:10.1146/annurev-genet-112414-054842
5. Cox, D. R. *Principles of Statistical Inference*. (Cambridge University Press, 2009). doi:10.1017/CBO9780511813559
6. Durrett, R. Population genetics of neutral mutations in exponentially growing cancer cell populations. *The Annals of Applied Probability* **23,** 230–250 (2013).
7. Kessler, D. A. & Levine, H. Large population solution of the stochastic Luria-Delbruck evolution model. *Proc. Natl. Acad. Sci. U.S.A.* **110,** 11682–11687 (2013).
8. Kessler, D. A. & Levine, H. Scaling solution in the large population limit of the general asymmetric stochastic Luria-Delbrück evolution process. *J Stat Phys* **158,** 783–805 (2015).

9. Bozic, I., Gerold, J. M. & Nowak, M. A. Quantifying Clonal and Subclonal Passenger Mutations in Cancer Evolution. *PLoS Comput. Biol.* **12,** e1004731 (2016).
10. Temko, D. *et al.* Somatic POLE exonuclease domain mutations are early events in sporadic endometrial and colorectal carcinogenesis, determining driver mutational landscape, clonal neoantigen burden and immune response. *J. Pathol.* **245,** 283–296 (2018).
11. McFarland, C. D., Korolev, K. S., Kryukov, G. V., Sunyaev, S. R. & Mirny, L. A. Impact of deleterious passenger mutations on cancer progression. *PNAS* **110,** 2910–2915 (2013).
12. Sottoriva, A. *et al.* A Big Bang model of human colorectal tumor growth. *Nature Genetics* **47,** 209–216 (2015).
13. Nik-Zainal, S. *et al.* The life history of 21 breast cancers. *Cell* **149,** 994–1007 (2012).
14. Martincorena, I. *et al.* Universal Patterns of Selection in Cancer and Somatic Tissues. *Cell* **0,** (2017).
15. Wang, K. *et al.* Whole-genome sequencing and comprehensive molecular profiling identify new driver mutations in gastric cancer. *Nature Genetics* **46,** 573–582 (2014).
16. Zapata, L. *et al.* Negative selection in tumor genome evolution acts on essential cellular functions and the immunopeptidome. *Genome Biol.* **19,** 924 (2018).