Reply to Noorbakhsh and Chuang

Marc J. Williams[1-3], Benjamin Werner[4], Chris P. Barnes[2,5], Trevor A. Graham[1,*] & Andrea Sottoriva[4,*]

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

[*] Correspondence should be addressed to T.A.G. (t.graham@qmul.ac.uk) or A.S. (andrea.sottoriva@icr.ac.uk)

Noorbakhsh and Chuang question the resolvability of evolutionary dynamics from cancer sequencing data. In Williams et al., 2016[1] we showed that the cumulative frequency distribution of sub-clonal mutations, *M(f)*, expected in a neutrally-evolving cancer is:

$$M(f) \sim \frac{1}{f}$$

and found that this model well-approximated the frequency distribution in more than 35% of the 900 cancers of 14 different types that we analysed. Neutral evolution describes the situation where there is no clonal selection within a population: all mutations that accrue are passengers and all drivers were already present in the first transformed cell. Neutrality is the null model of molecular evolution[2], and our analysis showed that for a considerable proportion of tumors it was not possible to reject the null model. One important consequence is that intra-tumor heterogeneity is not necessarily a read-out of clonal selection as had been previously argued, but can resemble a purely neutral process.

For clarity we summarize briefly the biological basis of our analysis. Selection fundamentally acts on phenotypes and their relation to the microenvironmental context, not on genotypes. Many different biological mechanisms could influence the evolutionary dynamics, including tissue architecture, the genotype-phenotype map and accessibility of genotype space, and the constancy of the microenvironment. However, a tumor is composed by asexually reproducing cancer cells and any expanding clone will naturally carry all other heritable properties of the clone to higher frequency in the tumor. This effect is called ‘hitchhiking’ and well known in the absence of recombination. Population genetics approaches like the one we employed exploit genetic hitchhiking of passenger mutations to discern the effects of selection[3], even when the driver lesion itself is unidentified or immeasurable. Selection leads to the selected clone becoming over-represented in the tumor, and importantly this is the case irrespective of what actually drives a clonal expansion (whether it is a copy number driver, an SNV, an epigenetic event or even a non-cell autonomous effect). Moreover, because of hitchhiking, any subset of mutations from a tumor will inform precisely on the same clonal structure, whether they are

mutations from only diploid regions, coding mutations, or noncoding mutations. We demonstrated this property in our analysis of gastric whole genome sequencing data (Figure 2C in the original manuscript; the same dynamics were recovered irrespective of the type of SNVs used in the analysis).

Noorbakhsh and Chuang argue that models of clonal selection better explain the available data than neutral (null model). They suggest that the family of distributions:

$$M(f) \sim \frac{1}{f^{\alpha}}$$

represent the effect of clonal selection where α>1 represents disruptive (diversifying) selection, and α<1 negative (stabilising) selection. However, these models are not phenomenological in their derivation, and hence do not represent the underlying biology. Furthermore, negative selection is compatible with neutrality[4] and in the case of positive selection, Noorbakhsh and Chuang discuss a special case of selection (disrupting/diversifying selection) which increases variation, as opposed to positive directional selection that in cancer drives clonal expansions, thus reducing variation[5,6].

To our knowledge, there is no known analytical solution for the subclonal dynamics of expanding populations under selection and in fact this is a longstanding open problem in population genetics. Various complications that have hindered a general model of selection include the plethora of potential phenotypic changes that might alter fitness in different ways, the importance of context (including both cell-intrinsic (epistasis) and cell-extrinsic (microenvironment) effects), and the timing of mutation in an expanding population. Although no derivation is provided, Noorbakhsh and Chuang's model appears to assume multiple very commonly mutated alleles that all experience some selection, suggesting the existence of a large number of driver alterations. The paucity of recurrent driver mutations that can be detected in any given cancer questions the realism of this assumption[7]. The alternative model, considered in Williams et al. is that driver mutations are rare in the genome (and thus infrequently acquired) but exhibit large effects. When a driver mutation is acquired, neutral mutations continue to be acquired within the driver clone, and these passengers are carried to higher frequency by the clonal expansion of the driver. Consequently the frequency distribution shows a characteristic peak centered at the frequency of the expanding sub-clone (Williams Fig S11) – a distribution that is very different to that predicted by Noorbakhsh and Chuang.

Furthermore, our derivations show that Noorbakhsh and Chuang's proposed model of selection also describes neutral evolution in a population that is following boundary driven growth (e.g. the growth follows a power-law such that $N(t) \sim t^{\gamma}$). Replacing the exponential growth function with power-law growth and following the derivation in Williams *et al* yields the cumulative distribution of mutations in the form:

$$M(f) \sim \frac{1}{f^{\frac{\gamma}{\gamma-1}}}$$

and setting γ=2 leads to $M(f)=1/f^2$. Noorbakhsh and Chuang's analysis shows that this quadratic model well approximates the data from many cancers. We would suggest this may indicate more the change in tumour size over time, than the evolutionary dynamics of tumour subclones. We wholly agree with the authors' comment that their analysis "does not clearly show a lack of neutrality".

Noorbakhsh and Chuang also question our use of the R^2 statistic to fit our model to the data. We must agree that R^2 values are not the optimal statistic to fit the 1/f distribution and in follow up work we are now developing improved tools. However, we were aware of potential shortcomings of the R^2 value, which is why we used a very high $R^2 = 0.98$ value as our cut off, in order to avoid a misclassification and thus an over-interpretation of the data. Our analysis as it stands shows when the most parsimonious 'null' model of the evolutionary dynamics is an adequate description of the data. In addition, we note that we provided extensive statistical analysis in our original manuscript to test the robustness of our fit method and our ability to recover the correct evolutionary dynamics (Supplementary Figure 9 and 10).

Nevertheless, we agree that our analysis, as any bioinformatics analysis, relies on the quality of the underlying genomic data. The authors discuss the problem of 'noise' in measured variant allele frequency (VAF) in the TCGA data due to limited sequencing depth. Depth clearly constrains the accuracy of VAF measurement. However, we note that when we looked at an independent dataset of gastric cancers, whole-genome sequenced at very high depth (100x)[8] (representing an exemplar dataset), we found that our neutral model fitted the data very frequently (60 of 78 cases were consistent with neutral growth). Moreover, we had previously computationally investigated the effect of limited sequencing depth on the ability of our model to correctly recover neutral dynamics, and observed asymptotic improvements in the fit as depth increased, that essentially saturate at depths greater than 75X (Williams Fig S10). Hence, sequencing depth and VAF determination is only a confounding issue in low depth sequencing. We suggest that due to limited depth in some samples and our choice of a very conservative threshold, potentially more cases may actually follow neutral evolution than currently identified.

We fully acknowledge the need for better, multi-region and deeper sequenced data, which will inform on the way in which tumours grow with higher accuracy.

Furthermore, we note that slight deviations from neutrality may beg the question of the clinical relevance of weak selection effects that do not have the power to significantly change the clonal composition of a tumour.

Taken together, we are glad that our paper is fuelling scientific discussion around such an important topic of interpretation of cancer genomic data, and we also appreciate the suggestions of the authors to potentially investigate other measures of fit for our model other than R^2. Nevertheless, we maintain that the models of selection they proposed would require further theoretical understanding before they can become applicable to cancer genomic data.

Acknowledgments

A.S. is supported by the Chris Rokos Fellowship in Evolution and Cancer. B.W. is supported by the Geoffrey W. Lewis Post-Doctoral Training fellowship. This work was supported by the Wellcome Trust (105104/Z/14/Z). C.P.B. acknowledges funding from the Wellcome Trust through a Research Career Developmental Fellowship (097319/Z/11/Z). This work was supported by a Cancer Research UK Career Developmental Award to T.A.G. M.J.W. is supported by a UK Medical Research Council student fellowship.

Author Contributions

All authors contributed equally to this response letter.

Competing Financial Interests

The authors declare no competing financial interests.

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* (2016). doi:10.1038/ng.3489
2. P Donnelly, A. & Tavaré, S. Coalescents and Genealogical Structure Under Neutrality. *Annual Review of Genetics* **29,** 401–421 (2003).
3. Gillespie, J. H. Genetic drift in an infinite population. The pseudohitchhiking model. *Genetics* **155,** 909–919 (2000).
4. Kimura, M. *The Neutral Theory of Molecular Evolution*. (Cambridge University Press, 1984).
5. 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
6. Greaves, M. & Maley, C. C. Clonal evolution in cancer. *Nature* **481,** 306–313 (2012).
7. Tomasetti, C., Marchionni, L., Nowak, M. A., Parmigiani, G. & Vogelstein, B. Only three driver gene mutations are required for the development of lung and colorectal cancers. *PNAS* **112,** 118–123 (2015).
8. Wang, K. *et al.* Whole-genome sequencing and comprehensive molecular profiling identify new driver mutations in gastric cancer. *Nature Genetics* **46,** 573–582 (2014).