Functional versus non-functional intratumor heterogeneity in cancer.

Williams M.J.[1], Werner B.[2], Graham T.A.[1], Sottoriva A. [2]

[1]Evolution and Cancer Laboratory, Barts Cancer Institute, Queen Mary University of London, London, UK
[2]Centre for Evolution and Cancer, The Institute of Cancer Research, London, UK

Abstract
Next generation sequencing data from human cancers are often difficult to interpret within the context of tumour evolution. We developed a mathematical model describing the accumulation of mutations under neutral evolutionary dynamics and showed that 323/904 cancers (~30%) from multiple types were consistent with the neutral model of tumour evolution.

Deep sequencing of tumour samples reveals the fraction of cells that harbour individual mutations. These mutations are the result of replication errors during cell division and their frequency distribution is a consequence of each cancers evolutionary history. In our latest work [1] we utilised the distribution of mutant allele frequencies of tumors to infer how individual cancers grew.

The shape of the mutant allele frequency distribution is determined by numerous factors, such as the growth characteristics of a tumor, the mutation rate per division and potentially, the presence of differentially growing sub-clones. In particular, the occurrence of multiple sub-clones can introduce arbitrary complexity in the clonal composition of the tumor, making the interpretation of the mutant allele frequency distribution extremely difficult.
We started from the simplest possible scenario where all cells in the tumour grow at the same rate, a situation where the tumour follows 'neutral evolution', which is the null model for molecular evolution[2,3]. Mathematically, this leads to a distribution of mutations that follows a power law. Given a mutation rate μ, a probability β that cells produce two surviving offspring and the frequency of mutations f, our model predicts that the cumulative number of mutations with a frequency greater than f, $M(f)$, is linearly proportional to $1/f$,

$$M(f) \sim \frac{\mu}{\beta} \frac{1}{f}. \qquad [1]$$

Equation [1] has been described previously in the population genetics literature[2,3], but has not to our knowledge been applied to cancer genomes. We tested the prediction of our model on next-generation sequencing data from 904 tumors of many different types. Remarkably, we found that more than 1/3 of all cases were consistent with this simple model.

Under this neutral model of tumour evolution, the frequency of a mutation in the tumor is a proxy for the time it occurred, as early events are at a high frequency, while later events remain at low frequencies. Mutational timelines can therefore be resolved, as can measurements of the mutation rate in individual cancers. Our mutation rate measurements yield rates of the order

10^{-6}-10^{-7} bp/division, and were higher than previous estimates (10^{-9} bp/division [4]). This is because what we measure is the 'effective mutation rate', μ/β. If there is high cell death in the tumor (low β) the true mutation rate may be much lower. Moreover, we note that this was one of the first attempts to measure mutation rates *in vivo* in human malignancies, whereas previous measurements were performed *in vitro* or were speculative.

Deviations from the predictions of equation [1] indicate more complex growth patterns such as the presence of functionally distinct subclones. On-going clonal selection in cancer has been the traditional view of tumour evolution[5]. However, recent evidence might change this presumption. Multi-region profiling studies in colorectal cancer have found neutral evolution to be consistent with genomic data: using the diversity of methylation patterns as a measure of the age of a tumour, Siegmund *et al*[6] observed "flat" clonal expansions where the age of samples from opposite parts of the tumour appeared the same, consistent with neutral evolution. More recently we proposed and validated a "Big Bang" model of colon cancer, where malignancies were characterised by numerous intermixed subclones and lack of stringent selection[7]. Our results extended these first observations of neutral evolution, demonstrating that neutral evolution can also explain the intra-tumour heterogeneity seen in other cancer types. Interestingly, some cancer types seem more prone to neutral evolution than others, indicating that the cellular architecture and local microenvironmental conditions are likely important factors in determining the rules of tumour evolution.

Our measurement of neutrality has the benefit that it only requires a single sample to perform, and can therefore be applied to a large amount of existing data. However, our approach only reveals the evolutionary dynamics within this sample. A 'global' inference of neutrality across an entire tumour requires multiple samples from different locations of the same tumour. In a recent study, Ling *et al*[8] considered a large number of samples from a single hepatocellular carcinoma. They also found the dynamics to conform with a neutrally expanding tumor based on an estimate of the total number of mutations in the tumour. Future work will need to address the basis of 'local' versus 'global' neutrality.

Overall, our model predicts that neutral evolution leaves a characteristic signature in the allelic frequency distribution. Identifying this signature in cancer genomes enabled us to vastly simplify the interpretation of what appears inherently complex and noisy data. Consequently, our work demonstrates how integrating physically informed mathematical models and cancer genomics data can provide new and perhaps unexpected insight into these complex data.

Recently, physicists were able to demonstrate the existence of gravitational waves by detecting an incredibly weak signal amidst a noisy dataset[9]. This was feasible because scientists knew *a priori* what the signal of a gravitational wave would look like thanks to Einstein's field equations, formulated a century earlier. Cancer genomics at present suffers from the reverse problem: large quantities of data can already be generated but the underlying theory remains

underdeveloped. What is lacking is the connection between theoretical and experimental frameworks that would allow us to fully interpret and evaluate the wealth of genomic data. Our latest work is, we hope, a first step in this direction.

Figure 1
Mutations and their frequency encode the history of individual tumours. **(A)** Mutations label distinct lineages and as a tumour grows, the size of these lineages becomes progressively smaller. Here, with one mutation per division and all cells growing at the same rate the 2 mutations occurring during the first division are each present in 50% of the population, mutations occurring during the final division are present in 12.5%. **(B)** When sequencing tumour biopsies we measure the frequency of mutations in the population. Neutral tumor evolution imprints a characteristic 1/f signature in the distribution of subclonal mutant allele frequencies.

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. Durrett, R. Population genetics of neutral mutations in exponentially growing cancer cell populations. *The Annals of Applied Probability* **23,** 230–250 (2013).
3. Griffiths, R. C. & Tavaré, S. The age of a mutation in a general coalescent tree. *Communications in Statistics. Stochastic Models* **14,** 273–295 (1998).
4. Jones, S. *et al.* Comparative lesion sequencing provides insights into tumor evolution. *Proc. Natl. Acad. Sci. U.S.A.* **105,** 4283–4288 (2008).
5. Fearon, E. R. & Vogelstein, B. A genetic model for colorectal tumorigenesis. *Cell* **61,** 759–767 (1990).
6. Siegmund, K. D., Marjoram, P., Woo, Y.-J., Tavaré, S. & Shibata, D. Inferring clonal expansion and cancer stem cell dynamics from DNA methylation patterns in colorectal cancers. *Proc. Natl. Acad. Sci. U.S.A.* **106,** 4828–4833 (2009).
7. Sottoriva, A. *et al.* A Big Bang model of human colorectal tumor growth. *Nature Genetics* **47,** 209–216 (2015).
8. Ling, S. *et al.* Extremely high genetic diversity in a single tumor points to prevalence of non-Darwinian cell evolution. *Proc. Natl. Acad. Sci. U.S.A.* **112,** E6496–505 (2015).
9. Abbott, B. P. *et al.* Observation of Gravitational Waves from a Binary Black Hole Merger. *Phys. Rev. Lett.* **116,** 061102 (2016).

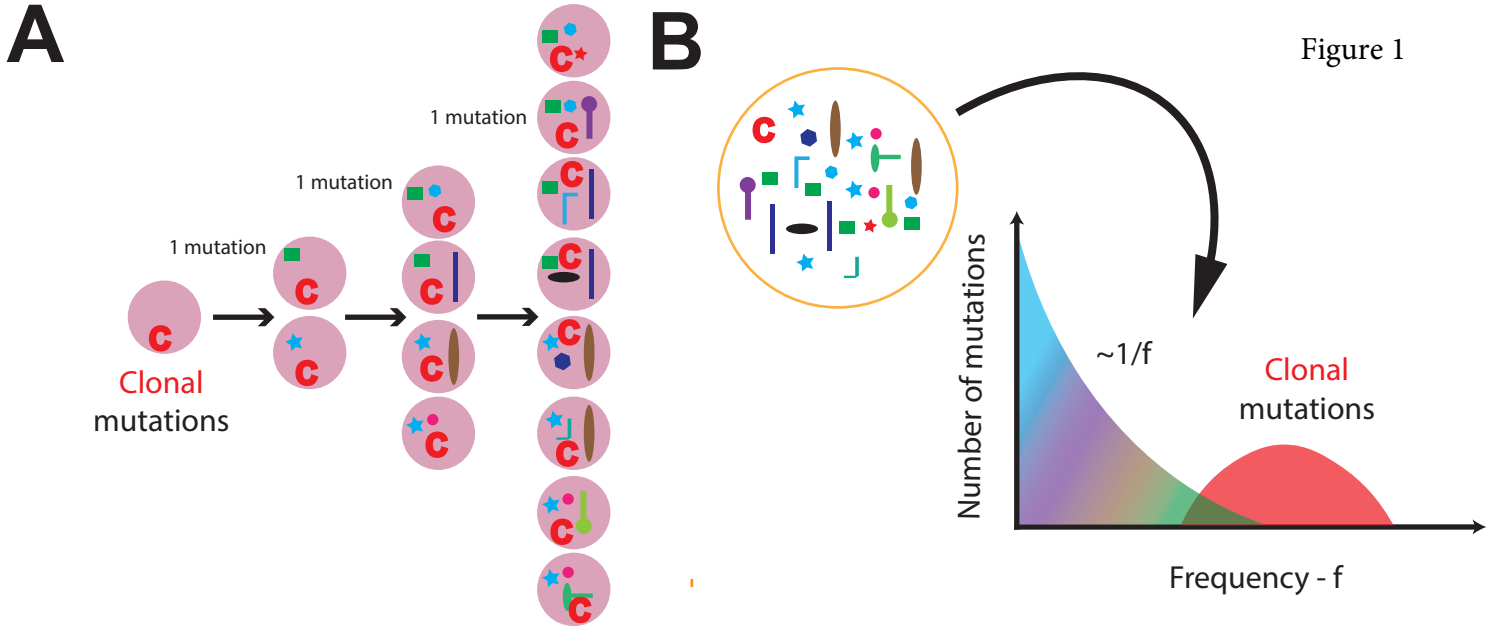
A
1 mutation
1 mutation
1 mutation
Clonal
mutations
B
Number of mutations
~1/f
Clonal
mutations
Frequency - f

Figure 1