A lottery simulation model for predicting cancer risk based on the strength and abundance of individual mutations

Authors:

Alexander J Wolfe and Adam Giangreco

Affiliation:

Lungs for Living Research Centre, University College London, 5 University Street WC1E 6JF

Corresponding author:

Adam Giangreco

a.giangreco@ucl.ac.uk

Abstract

Models for predicting an individual's risk of developing cancer are complicated by a lack of well-conserved mutations present in most types of disease. To address this challenge we developed a simple lottery simulation of cancer development based solely on the abundance and relative survival advantage of mutations present within individual stem or progenitor cells. In simulations requiring few mutations or that involve mutations exhibiting a strong survival advantage cancers develop in our simulation in a progressive manner amenable to oncogene-based risk prediction. In contrast, in simulations involving numerous mutations that lack a strong survival advantage, a stochastic process of neutral drift and punctuated equilibrium determines eventual tumour formation. In these situations, the development of cancer is largely attributable to the rate of mutational variability amongst progenitor cells rather than the identity and relative survival advantage of the mutations present at any single time point. These results suggest that measuring rates of stem cell mutational variability represents an important component of predicting an individual's risk of developing cancer.

INTRODUCTION

The theory that cancers are the end-result of accumulated oncogenic changes was first posited by Nordling, Armitage and Doll based on their observations regarding carcinogen dose responsiveness, tumour latency, and age-dependent mortality (Nordling 1953, Armitage and Doll 1954). Given enough time, individual progenitor or stem cells are thought to accumulate sufficient numbers of mutations in a stepwise manner. These progressively accumulated mutations provide cells their capacity to proliferate indefinitely, evade immunosurveillance, and invade the local tissue, thereby resulting in cancer formation (Stratton, Campbell et al. 2009). Implicit in this model is the understanding that the risk of developing most cancers depends largely on the pro-tumorigenic properties of specific oncogenic mutations and the relative survival advantage conferred by each of these changes.

In recent years, high throughput sequencing has raised the possibility of comprehensively cataloguing all mutations and chromosomal aberrations present within cancers (Stratton, Campbell et al. 2009, Pe'er and Hacohen 2011). The aim of these studies has been to identify conserved genetic mutations suitable for predicting cancer risk and facilitating the early clinical diagnosis of disease. Surprisingly, this work has uncovered a staggering diversity and lack of conservation amongst inter-individual and intra-individual mutations that precludes the identification of oncogenes suitable for cancer risk prediction (Wood, Parsons et al. 2007, Pleasance, Cheetham et al. 2010, Pleasance, Stephens et al. 2010). Studies have also established that the number of mutations required for cancer development varies widely depending on tissue type, and that the mutational burden within physiologically normal stem and progenitor cells is often similar to that of cancers (Ding, Getz et al. 2008, Brody 2011, Martincorena, Roshan et al. 2015). Thus, the relationship between the identity and abundance of mutations within specific tissue progenitor cells and overall cancer risk appears to be more complex than originally anticipated (Alexandrov, Nik-Zainal et al. 2013).

To address this complexity, we developed a lottery-based simulation of cancer development based on the observation that a random process of stem cell turnover governs the homeostasis of most tissues (Simons and Clevers 2011). We defined a series of randomly generated numbers, or ‘events’ as representing individual stem or progenitor cell mutations. Each of these ‘events’ confers variable degrees of selective advantage, similar to the biological variability amongst individual oncogenic mutations. To complete each simulation (equivalent to winning a lottery jackpot; here representing the development of cancer), all events must be selected. We designed our simulation to vary both the number of events required for cancer development as well as their individual survival advantage. Our simulated results provide new insights regarding the relative importance of oncogene abundance, strength, and mutation rate in human tumorigenesis.

RESULTS

In standard lottery simulations, each winning number (event) must be chosen simultaneously and the correct sequence of these events changes after each cycle of play. In this simulation, the correct or 'winning' sequence of events remains a fixed constant. Additional modifications include the ability to vary the odds of a single event being chosen (representing the mutational frequency of any given gene), the ability to vary the total number of events required (representing the number of oncogenes required for cancer development), and the introduction of a 'dwell time', or number of plays for which each correctly chosen event persists. Dividing this dwell time over the odds of an event being chosen establishes the relative survival advantage of individual events. By varying both the total number of events and their relative survival advantage, four possible scenarios for cancer development emerge (Figure 1A).

To test our model, we examined whether changes to the number of events influenced the number of 'plays' necessary to complete each simulation. This was done under conditions in which correctly selected events were maintained for a dwell time equal to the odds of each event occurring (100% survival advantage; scenarios A and B, Figure 1A). Using these parameters, we observed a positive correlation between the number of events and the number of plays required (Figure 1B). Each simulation was run 1000 times, and we plotted the number of plays required to complete each simulation against the proportion of total simulations (Figure 1C). These results resembled an approximation of the age-dependent incidence of the 4 most common adult human cancers (breast, prostate, lung, colorectal) as reported by the SEER database (http://seer.cancer.gov). Thus, our simulation adequately reproduces the natural kinetics of human tumour formation.

We next studied whether varying the survival advantage of individual events influenced the number of plays required to complete each simulation. We examined simulations requiring either 6 or 12 total events and varied the survival advantage of individual events between

10% - 100% (6 events, scenarios A versus C, Figure 1A) or 40%-100% (12 events, scenarios B versus D, Figure 1A). We also introduced a 'powerball' event that, once present, persisted for the entirety of each simulation. This was done to mimic the occurrence of fundamental driver mutations (such as p53) that are present within the majority of human cancers and thought to influence subsequent oncogene accumulation within affected cells. Consistent with our previous results, we observed an inverse correlation between the relative survival advantage of individual events and the number of plays required to complete each simulation (Figure 2A, B). Interestingly, although varying the relative survival advantage of individual events increased the number of required plays, it had no appreciable effect on the kinetics of each simulation (Figure 2C, D). Thus, we were unable to distinguish which of these simulated scenarios might best reflect natural human tumorigenesis. There was also no increase whatsoever in the rate of cancer development following the introduction of a 'powerball' event, consistent with the observation that p53 and other driver mutations are frequently observed at high abundance within otherwise physiologically normal tissues (Martincorena, Roshan et al. 2015).

In simulations requiring either 6 or 12 events, we examined the rate and pattern of individual event accumulation. In simulations where events exhibited a 100% survival advantage, we invariably observed a linear, direct relationship between event accumulation and the number of plays required to complete each simulation (Figure 3A, B). Similarly, we found that under conditions of moderate (40-60%) survival advantage scenarios requiring 6 total events exhibited a progressive, linear event accumulation (Figure 3C). In contrast, under conditions of moderate survivial advantage the accumulation of events in scenarios requiring 12 total events occurred in a random, stochastic manner (Figure 3D). A similar pattern in simulations involving 6 total events was observed under conditions of low (<20%) survival advantage (data not shown).

DISCUSSION

The present study uses a simple lottery simulation to provide new theoretical insights regarding the role of oncogene strength, abundance, and mutation rates in determining the probability of developing cancer. An important component of this simulation is that it does not require *a priori* knowledge of the identity of specific oncogenes associated with individual cancers. Rather, the only information needed is an approximation of the survival advantage conferred by individual progenitor cell mutations and the number of mutations required for initiating tumour formation. In simulations that require few mutations or that involve mutations exhibiting a strong survival advantage (scenarios A-C, Figure 1A) we found that cancers develop in a progressive manner amenable to oncogene-based risk prediction. In contrast, we observed that in simulations involving numerous oncogenic mutations that lack a strong survival advantage (scenario D, Figure 1A), a stochastic process of mutational variability is likely to precede tumour formation. Using these data, we establish that in addition to the strength of individual oncogenic mutations, the rate of mutational variability within individual progenitor cells is an important, previously unrecognised predictor of cancer risk. Our data also suggest that, while undoubtedly important, the presence of single driver mutations such as p53 is unlikely to influence the kinetics of tumour development. Furthermore, our results support the recently described ‘big bang’ scenario of tumorigenesis, in which a process of punctuated equilibrium is thought to facilitate the accumulation of significant mutational burdens within cells over a short space of time.

In order to determine the clinical utility of our simulation it will be necessary to establish the identity of mutations present within longitudinally collected precancerous biopsy samples. The goal of such a study will be to identify whether oncogenic mutations accumulate in a progressive, stepwise fashion amenable to gene-based risk prediction (scenarios A-C, Figure 4) or whether mutations occur in a random, unpredictable fashion with a high degree of temporal variability (scenario D, Figure 4). Scenarios A through C, in which oncogenic mutations accumulate linearly with time, would facilitate oncogene based risk stratification and early patient diagnosis. This is likely to be the case for cancers with highly conserved oncogene profiles such as childhood leukemias. In contrast, should scenario 'D' be observed, this would suggest that these tumours are driven by a stochastic process of punctuated equilibrium resulting in the progressive accumulation and subsequent disappearance of numerous ontogenic mutations. In this case, an assessment of an individual's cancer risk based on individual oncogenes is unlikely to be effective. Rather, predicting cancer risk in these patients may only be achieved via an analysis of the rate of mutational variability. This possibility is consistent with the recent observation that multiple oncogenic mutations are present within hundreds of mutated clones throughout physiologically normal skin and oesophagus tissue (Martincorena, Roshan et al. 2015). This scenario is also in agreement with the surprisingly high number of mutations and lack of oncogenic conservation present within most types of cancer (Wood, Parsons et al. 2007, Pleasance, Stephens et al. 2010, Pe'er and Hacohen 2011).

Overall, our simulation predicts that, given adequate time, all proliferative tissues exhibiting imperfect DNA replication will eventually develop cancers. Our model also predicts that the mechanisms by which cancers arise depend primarily on the strength and number of oncogenic mutations. These findings extend the work of Tomasetti and Vogelstein regarding stem cell division and cancer risk and support their suggestion that the most promising avenue for reducing cancer risk is through prevention (Tomasetti and Vogelstein 2015). Specifically, our results indicate that strategies specifically aimed at reducing rates of mutational accumulation and variability over time will reduce overall cancer incidence. These strategies will inevitably involve reducing cellular turnover, toxicant and mutation exposure via dietary and lifestyle modification. Thus, contrary to the rules of a normal lottery game, in our simulation 'the only way to win (i.e., prevent cancer development) is not to play'.

Acknowledgements

We gratefully acknowledge Professors Sam Janes, Benjamin Simons, and Peter Campbell for assistance in the conceptualization and interpretation of these results. We acknowledge Jeff Miller and Jason Seba for help with designing the stochastic lottery simulation model, and members of the Lungs for Living Research Centre for helpful comments regarding data interpretation. Dr Adam Giangreco is supported by a European Research Council Starting Investigator Grant (#260290) and the UCL Comprehensive Biomedical Research Centre.

Figure legends

Figure 1. A modified lottery simulation for cancer development. (A) Varying event abundance and survival advantage produces four possible scenarios for cancer development. (B) Results of simulations requiring 6, 8, 10, or 12 total events, each of which exhibits 100% survival advantage. (C) Probability curves of cancer development versus the number of total plays for simulations requiring 6, 8, 10, or 12 events exhibiting 100% survival advantage. Error bars (B) represent the standard error of the mean; each simulation was run 1000 times to minimize sampling variability.

Figure 2. Reducing the survival advantage of individual oncogenes slows the rate of cancer development. (A, B) Results of simulations requiring 6 (A) or 12 (B) total events, each of which exhibit between 10-100% survival advantage as indicated. (C, D) Probability curves of cancer development plotted against the number of plays requiring 6 (C) or 12 (D) total events with varying survival advantage. Error bars (A, B) represent the standard error of the mean; each simulation was run 1000 times.

Figure 3. The pattern of oncogene accumulation varies based on the abundance and survival advantage of individual mutations required for cancer development. (A, B) Plots of individual simulations requiring 6 (A) or 12 (B) events, each of which exhibits 100% survival advantage. (C, D) Plots of simulations requiring 6 (C) or 12 (D) total events, each of which exhibits 40% survival advantage. Total event abundance is represented by a red line; 'powerball' events by a blue line (A-D).

Figure 4. Depictions of theoretical progressive versus stochastic mutation accumulation in precancerous biopsy specimens. Scenarios A-C depict the accumulation of mutations in a progressive, stepwise fashion amenable to oncogene-based risk prediction. Scenario D depicts the random, unpredictable fashion with which mutations are expected to occur in cancers requiring numerous oncogenes that lack a significant survival advantage.

References

Alexandrov, L. B., S. Nik-Zainal, D. C. Wedge, S. A. Aparicio, S. Behjati, A. V. Biankin, G. R. Bignell, N. Bolli, A. Borg, A. L. Borresen-Dale, S. Boyault, B. Burkhardt, A. P. Butler, C. Caldas, H. R. Davies, C. Desmedt, R. Eils, J. E. Eyfjord, J. A. Foekens, M. Greaves, F. Hosoda, B. Hutter, T. Ilicic, S. Imbeaud, M. Imielinski, N. Jager, D. T. Jones, D. Jones, S. Knappskog, M. Kool, S. R. Lakhani, C. Lopez-Otin, S. Martin, N. C. Munshi, H. Nakamura, P. A. Northcott, M. Pajic, E. Papaemmanuil, A. Paradiso, J. V. Pearson, X. S. Puente, K. Raine, M. Ramakrishna, A. L. Richardson, J. Richter, P. Rosenstiel, M. Schlesner, T. N. Schumacher, P. N. Span, J. W. Teague, Y. Totoki, A. N. Tutt, R. Valdes-Mas, M. M. van Buuren, L. van 't Veer, A. Vincent-Salomon, N. Waddell, L. R. Yates, I. Australian Pancreatic Cancer Genome, I. B. C. Consortium, I. M.-S. Consortium, I. PedBrain, J. Zucman-Rossi, P. A. Futreal, U. McDermott, P. Lichter, M. Meyerson, S. M. Grimmond, R. Siebert, E. Campo, T. Shibata, S. M. Pfister, P. J. Campbell and M. R. Stratton (2013). "Signatures of mutational processes in human cancer." Nature **500**(7463): 415-421.

Armitage, P. and R. Doll (1954). "The age distribution of cancer and a multi-stage theory of carcinogenesis." Br J Cancer **8**(1): 1-12.

Brody, J. P. (2011). "Age-specific incidence data indicate four mutations are required for human testicular cancers." PLoS One **6**(10): e25978.

Dewhurst, S. M., N. McGranahan, R. A. Burrell, A. J. Rowan, E. Gronroos, D. Endesfelder, T. Joshi, D. Mouradov, P. Gibbs, R. L. Ward, N. J. Hawkins, Z. Szallasi, O. M. Sieber and C. Swanton (2014). "Tolerance of whole-genome doubling propagates chromosomal instability and accelerates cancer genome evolution." Cancer Discov **4**(2): 175-185.

Ding, L., G. Getz, D. A. Wheeler, E. R. Mardis, M. D. McLellan, K. Cibulskis, C. Sougnez, H. Greulich, D. M. Muzny, M. B. Morgan, L. Fulton, R. S. Fulton, Q. Zhang, M. C. Wendl, M. S. Lawrence, D. E. Larson, K. Chen, D. J. Dooling, A. Sabo, A. C. Hawes, H. Shen, S. N. Jhangiani, L. R. Lewis, O. Hall, Y. Zhu, T. Mathew, Y. Ren, J. Yao, S. E. Scherer, K. Clerc, G. A. Metcalf, B. Ng, A. Milosavljevic, M. L. Gonzalez-Garay, J. R. Osborne, R. Meyer, X. Shi, Y. Tang, D. C. Koboldt, L. Lin, R. Abbott, T. L. Miner, C. Pohl, G. Fewell, C. Haipek, H. Schmidt, B. H. Dunford-Shore, A. Kraja, S. D. Crosby, C. S. Sawyer, T. Vickery, S. Sander, J. Robinson, W. Winckler, J. Baldwin, L. R. Chirieac, A. Dutt, T. Fennell, M. Hanna, B. E. Johnson, R. C. Onofrio, R. K. Thomas, G. Tonon, B. A. Weir, X. Zhao, L. Ziaugra, M. C. Zody, T. Giordano, M. B. Orringer, J. A. Roth, M. R. Spitz, Wistuba, II, B. Ozenberger, P. J. Good, A. C. Chang, D. G. Beer, M. A. Watson, M. Ladanyi, S. Broderick, A. Yoshizawa, W. D. Travis, W. Pao, M. A. Province, G. M. Weinstock, H. E. Varmus, S. B. Gabriel, E. S. Lander, R. A. Gibbs, M. Meyerson and R. K. Wilson (2008). "Somatic mutations affect key pathways in lung adenocarcinoma." Nature **455**(7216): 1069-1075.

Gerlinger, M., A. J. Rowan, S. Horswell, J. Larkin, D. Endesfelder, E. Gronroos, P. Martinez, N. Matthews, A. Stewart, P. Tarpey, I. Varela, B. Phillimore, S. Begum, N. Q. McDonald, A. Butler, D. Jones, K. Raine, C. Latimer, C. R. Santos, M. Nohadani, A. C. Eklund, B. Spencer-Dene, G. Clark, L. Pickering, G. Stamp, M. Gore, Z. Szallasi, J. Downward, P. A. Futreal and C. Swanton (2012). "Intratumor heterogeneity and branched evolution revealed by multiregion sequencing." N Engl J Med **366**(10): 883-892.

Martincorena, I., A. Roshan, M. Gerstung, P. Ellis, P. Van Loo, S. McLaren, D. C. Wedge, A. Fullam, L. B. Alexandrov, J. M. Tubio, L. Stebbings, A. Menzies, S. Widaa, M. R. Stratton, P. H. Jones and P. J. Campbell (2015). "Tumor evolution. High burden and pervasive positive selection of somatic mutations in normal human skin." Science **348**(6237): 880-886.

Nordling, C. O. (1953). "A new theory on cancer-inducing mechanism." Br J Cancer **7**(1): 68-72.
Pe'er, D. and N. Hacohen (2011). "Principles and strategies for developing network models in cancer." Cell **144**(6): 864-873.

Pleasance, E. D., R. K. Cheetham, P. J. Stephens, D. J. McBride, S. J. Humphray, C. D. Greenman, I. Varela, M. L. Lin, G. R. Ordonez, G. R. Bignell, K. Ye, J. Alipaz, M. J. Bauer, D. Beare, A. Butler, R. J. Carter, L. Chen, A. J. Cox, S. Edkins, P. I. Kokko-Gonzales, N. A. Gormley, R. J. Grocock, C. D. Haudenschild, M. M. Hims, T. James, M. Jia, Z. Kingsbury, C. Leroy, J. Marshall, A. Menzies, L. J. Mudie, Z. Ning, T. Royce, O. B. Schulz-Trieglaff, A. Spiridou, L. A. Stebbings, L. Szajkowski, J. Teague, D. Williamson, L. Chin, M. T. Ross, P. J. Campbell, D. R. Bentley, P. A. Futreal and M. R. Stratton (2010). "A comprehensive catalogue of somatic mutations from a human cancer genome." Nature **463**(7278): 191-196.

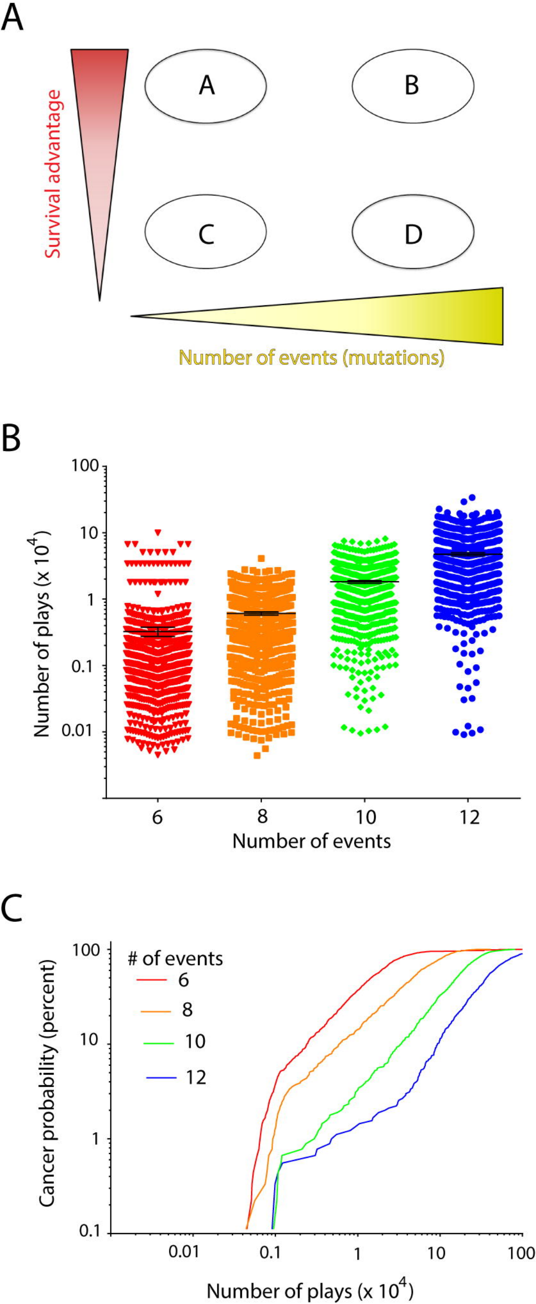
Pleasance, E. D., P. J. Stephens, S. O'Meara, D. J. McBride, A. Meynert, D. Jones, M. L. Lin, D. Beare, K. W. Lau, C. Greenman, I. Varela, S. Nik-Zainal, H. R. Davies, G. R. Ordonez, L. J. Mudie, C. Latimer, S. Edkins, L. Stebbings, L. Chen, M. Jia, C. Leroy, J. Marshall, A. Menzies, A. Butler, J. W. Teague, J. Mangion, Y. A. Sun, S. F. McLaughlin, H. E. Peckham, E. F. Tsung, G. L. Costa, C. C. Lee, J. D. Minna, A. Gazdar, E. Birney, M. D. Rhodes, K. J. McKernan, M. R. Stratton, P. A. Futreal and P. J. Campbell (2010). "A small-cell lung cancer genome with complex signatures of tobacco exposure." Nature **463**(7278): 184-190.

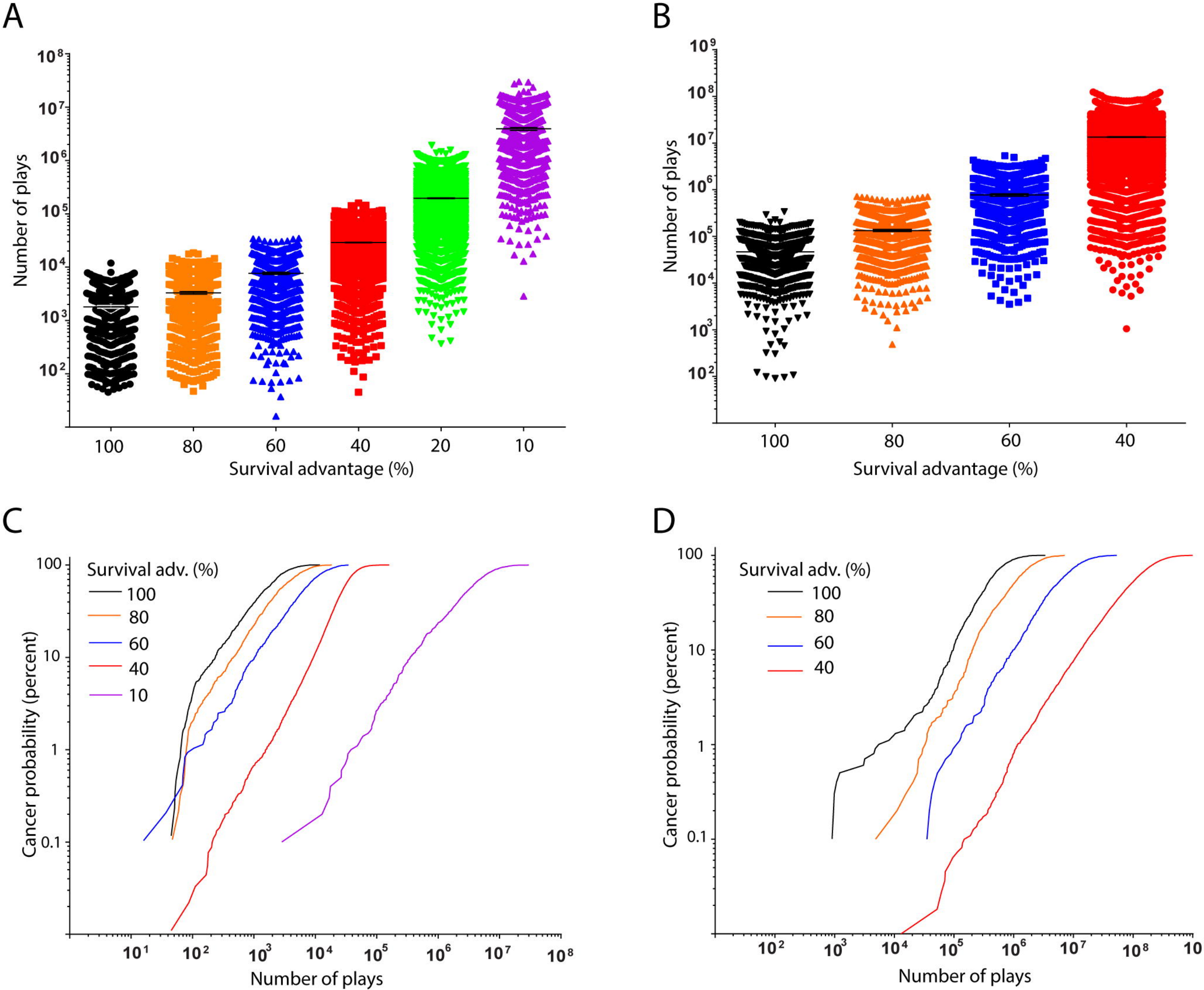
Simons, B. D. and H. Clevers (2011). "Strategies for homeostatic stem cell self-renewal in adult tissues." Cell **145**(6): 851-862.

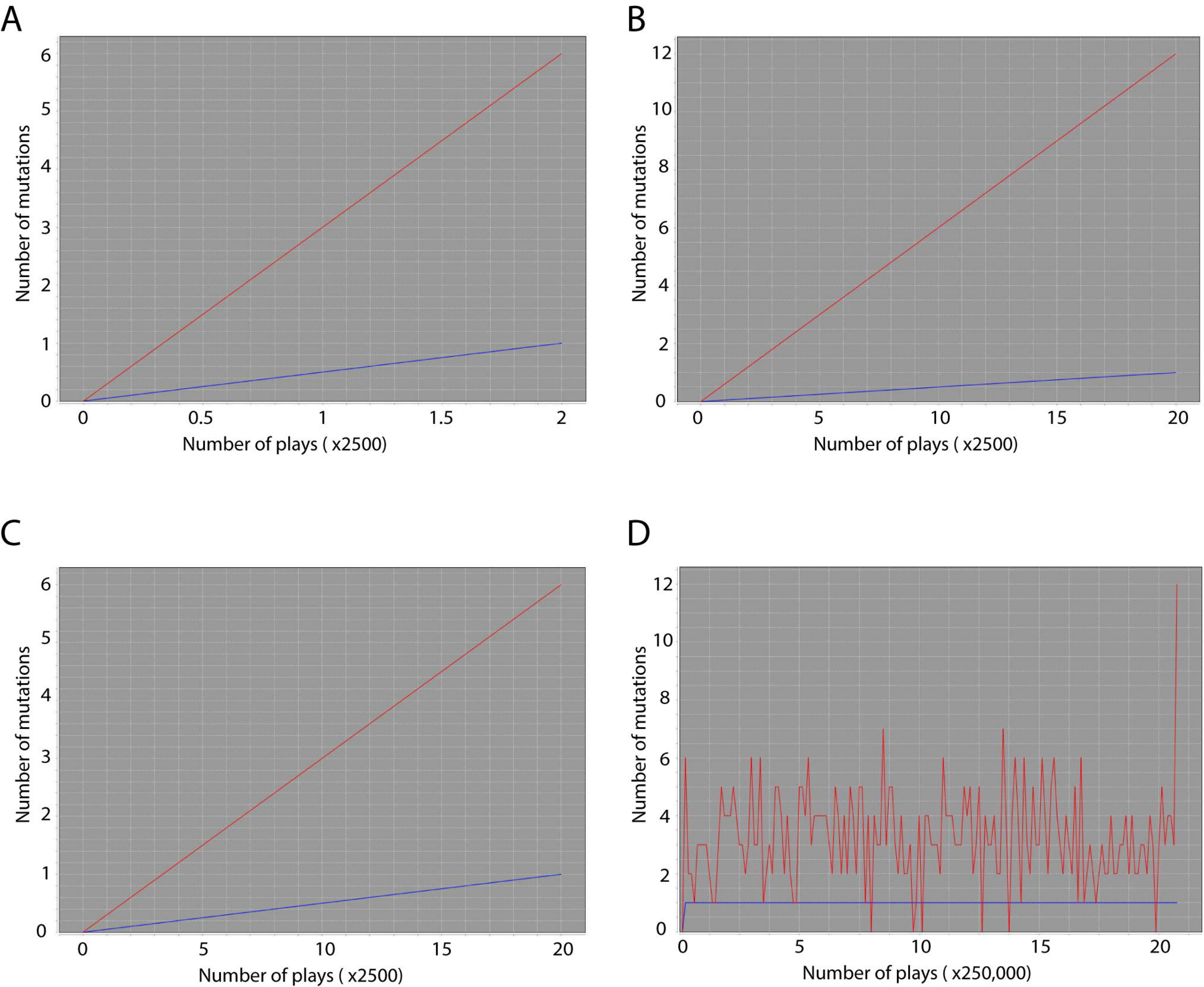
Stratton, M. R., P. J. Campbell and P. A. Futreal (2009). "The cancer genome." Nature **458**(7239): 719-724.

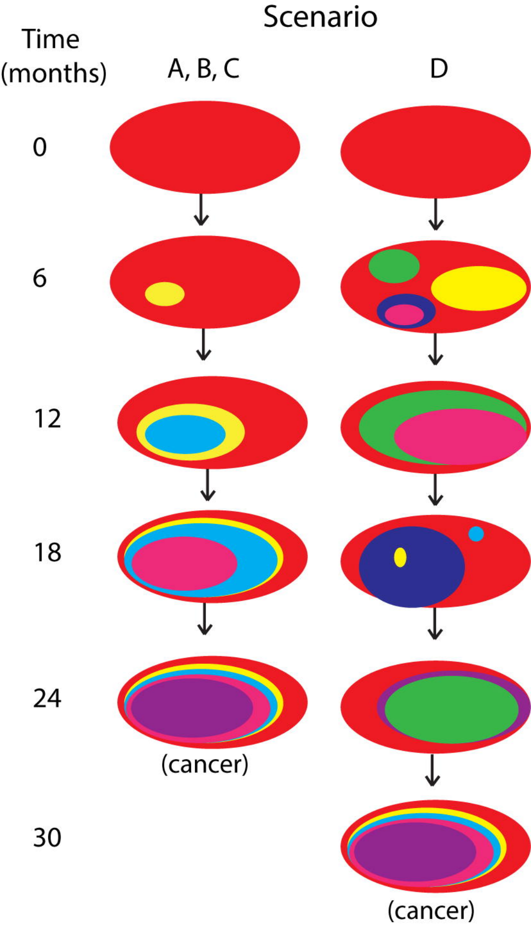
Tomasetti, C. and B. Vogelstein (2015). "Cancer etiology. Variation in cancer risk among tissues can be explained by the number of stem cell divisions." Science **347**(6217): 78-81.

Wood, L. D., D. W. Parsons, S. Jones, J. Lin, T. Sjoblom, R. J. Leary, D. Shen, S. M. Boca, T. Barber, J. Ptak, N. Silliman, S. Szabo, Z. Dezso, V. Ustyanksky, T. Nikolskaya, Y. Nikolsky, R. Karchin, P. A. Wilson, J. S. Kaminker, Z. Zhang, R. Croshaw, J. Willis, D. Dawson, M. Shipitsin, J. K. Willson, S. Sukumar, K. Polyak, B. H. Park, C. L. Pethiyagoda, P. V. Pant, D. G. Ballinger, A. B. Sparks, J. Hartigan, D. R. Smith, E. Suh, N. Papadopoulos, P. Buckhaults, S. D. Markowitz, G. Parmigiani, K. W. Kinzler, V. E. Velculescu and B. Vogelstein (2007). "The genomic landscapes of human breast and colorectal cancers." Science **318**(5853): 1108-1113.


A
Survival advantage
A
B
C
D
Number of events (mutations)
B
Number of plays (x 10^4)
100
10
1
0.1
0.01
6
8
10
12
Number of events
C
Cancer probability (percent)
100
10
1
0.1
# of events
6
8
10
12
0.01
0.1
1
10
100
Number of plays (x 10^4)


A
Number of plays
10^8
10^7
10^6
10^5
10^4
10^3
10^2
100
80
60
40
20
10
Survival advantage (%)
B
Number of plays
10^9
10^8
10^7
10^6
10^5
10^4
10^3
10^2
100
80
60
40
Survival advantage (%)
C
Cancer probability (percent)
100
10
1
0.1
Survival adv. (%)
100
80
60
40
10
10^1
10^2
10^3
10^4
10^5
10^6
10^7
10^8
Number of plays
D
Cancer probability (percent)
100
10
1
0.1
Survival adv. (%)
100
80
60
40
10^2
10^3
10^4
10^5
10^6
10^7
10^8
10^9
Number of plays


A
Number of mutations
0
1
2
3
4
5
6
Number of plays (x2500)
0
0.5
1
1.5
2
B
Number of mutations
0
2
4
6
8
10
12
Number of plays (x2500)
0
5
10
15
20
C
Number of mutations
0
1
2
3
4
5
6
Number of plays (x2500)
0
5
10
15
20
D
Number of mutations
0
2
4
6
8
10
12
Number of plays (x250,000)
0
5
10
15
20


Scenario
Time (months)
A, B, C
D
0
6
12
18
24
30
(cancer)
(cancer)