Somatic evolution of cancer genes in sex chromosomes

Amarachi C. Akachukwu[1†], Jasmine S. Ratcliff[1,2†], Antonio Marco[1,*]

[1]School of Life Sciences and [2]Institute for Social and Economic Research, University of Essex, Wivenhoe Park, Colchester CO43SQ, United Kingdom.

† These authors contributed equally to the work.

*To whom correspondence should be addressed:

amarco@essex.ac.uk

ABSTRACT

Genes on sex chromosomes have higher evolutionary rates than those on autosomes. However, this does not necessarily apply to somatic evolution in cancer. Many dominant mutations have been described in the so-called proto-oncogenes (OGs), while recessive mutations are typically described in tumor-suppressor genes (TSGs). Evidence indicates that mutations in X-chromosome TSGs are more likely to contribute to cancer than those in autosomal TSGs. Here, we formalize this in several dynamic models and predict, as expected, that mutations spread faster in TSGs located on the X chromosome than on autosomes (faster-X effect). Conversely, mutations in OGs spread faster on autosomes than on the X chromosome, but under high selective pressure, this difference is negligible. Published genomic screenings of cancer samples show evidence of the faster-X effect in TSGs. This pattern is observed in both sexes, suggesting that the maintenance of X-chromosome inactivation during cancer progression plays an important role in the evolution of TSGs. Strikingly, the relative mutation incidence in X-linked TSGs among females across individual studies is bimodal, with one group of studies showing a faster-X effect and another group showing similar incidences for X-linked and autosomal TSGs. This differentiation between cancer samples is not associated with the specific type of cancer or the tissue of origin. This may indicate that X-chromosome inactivation plays a differential role in the involvement of X-linked TSGs across individual cancers.

INTRODUCTION

The rate at which beneficial mutations are fixed in a population has been a central topic in population genetics. If beneficial mutations are fixed in one gene more often than in another gene, we say that the first gene evolves faster than the second. One classic result is that, overall, genes in haploid organisms evolve comparatively faster than those in diploid organisms (Crow and Kimura 1970; Hartl and Clark 2007). This is because a majority of newly introduced mutations are recessive with respect to the wild-type allele in diploid populations, and the beneficial effects of the new mutant are not manifested in heterozygosis. This result can be extended to the relative evolutionary rate of genes in autosomes and sex chromosomes. A gene located on an X chromosome in mammals, when in males, is effectively in haploidy (hemizygosis). Theoretical models have shown that, in the long term, genes located on the X chromosome should evolve faster than genes located on autosomes: the faster-X effect (Charlesworth et al. 1987). However, some contradictory data related to the faster-X effect exists (e.g. Avila et al. 2014), which may indicate that the underlying assumptions for the faster-X model should be carefully checked for each individual case and its assumptions.

Cancer results from the abnormal growth of populations of cells that escape multiple cellular control mechanisms (Weinberg and Weinberg 2006; Ruddon 2007). Cancer is a multistage process and typically requires one or more genes to be mutated at the somatic level (Vogelstein and Kinzler 2002; Greaves and Maley 2012; Martincorena and Campbell 2015). A somatic mutation in a ‘cancer gene’, which is detrimental for the host, is actually beneficial to the somatic tumor as it may help it spread in the organism. Although this picture is much more complex, traditionally there have been two mayor types of cancer genes identified, depending on the type of mutation they require to be beneficial for the progressing tumor: oncogenes and tumor-suppressor genes. Oncogenes (OGs), or proto-oncogenes when they are normally functioning in the host, are usually activated by gain-of-function mutations, typically dominant, as only one of the two copies needs to be mutated to have an impact. On the other hand, tumor-suppressor genes (TSGs) usually require both copies to be inactive (by either mutation, deletion, or silencing) to have an effect in cancer, and therefore these

mutations follow a traditional recessive inheritance pattern. This is also known as the Knudson two-hit hypothesis, which states that a TSG involved in cancer development should have both of its alleles inactivated, either by mutation or by another mechanism (Knudson 1971).

The main complication in developing genetic models of cancer evolution is that these are models of somatic evolution, for which many of the assumptions made in standard diploid germline models do not hold (mainly gamete sorting). Likely for this reason, somatic evolution models are not as well developed as germline genetic models. Some early models comparing haploid and diploid systems were developed by Crow and Kimura (1965) based on earlier work in theoretical population genetics (Fisher 1930; Muller 1932). These models compare the evolution of genes in haploid and diploid (sexual) systems, concluding that sexual reproduction results in faster genes evolution. This conclusion was challenged by Maynard Smith, which triggered some debate (Maynard Smith 1968; Crow and Kimura 1969). In any case, these works were not specifically about somatic tissues (as they allowed recombination in diploids), yet Crow and Kimura (1965) proposed that diploidy may confer some protection against damaging mutations in somatic tissues. This was explicitly modelled by Orr (1995), who concluded that when deleterious mutations are recessive and relatively common, diploidy is advantageous compared to haploidy in somatic tissues.

Cancer somatic evolutionary models are often based on the Luria-Delbrück model or its derivations (reviewed in Frank (2018). These models, on one hand, ignore the haplotype of the genes involved, but on the other hand, implement more realistic parameters to account for cancer complexity, such as explaining the different rates of cancer spread under different conditions. For instance, more complex models can implement multistage processes (Frank 2010; Nunney 1999), compartmentalization (Frank and Nowak 2004), the joint impact of pro- and anticancer mutations (Michor et al. 2003), or even epistatic interactions (Alfaro-Murillo and Townsend, 2023). (See Cairns (1998) for a historical overview on the development of cancer evolution models.) However, our goal is not to evaluate fixation rates of cancer or its associated mutations, but the expected differences in spread between cancer genes located on

autosomes and the X chromosome. In this context, we aim to investigate whether there is a faster-X effect in either TSGs and/or OGs and the relative role of selective pressure and somatic mutation rates in their evolutionary rates during cancer progression. To do so, we here develop models of cancer gene evolution in somatic tissues, considering haploidy as a proxy for chromosome location, and evaluating our predictions by analyzing whole genome mutation screens from multiple cancer samples.

RESULTS

Oncogene somatic evolution in a chromosomal context

Let us consider a population of cells in an individual in which a driver mutation in an oncogene confers selective advantage: an increase in the relative growth with respect to the non-mutated cells. While this is an oversimplification of real cancer evolution, maintaining these conditions and assumptions across all modelsallows us to isolate the relative impact of chromosomal context in cancer evolution (*ceteris paribus*). In this model, c represents the proportion of cancer cells in a tissue, u is the mutation rate, and s is the selective advantage of cells with the driver mutation in the oncogene. Assuming that driver mutations in oncogenes are gain-of-function and dominant, we derived an approximate formula for the ratio of the growth of cancer cells for oncogenes (R_{OG}) in diploidy (ΔC_{dd}) relative to haploidy (ΔC_h) (see Methods for derivation):

$$R_{OG} = \frac{\Delta C_{dd}}{\Delta C_h} \approx \frac{2(1-c)u + (1-c^2)s}{(1-c)u + (1-c^2)s} \qquad (1)$$

From equation (1), it can be easily seen that R_{OG} is greater than 1.Therefore, for two OGs under the same conditions (mutation rate and selective pressure) but in different chromosomal contexts, the mutations in the autosome (diploid) will spread faster in the new cancer cell population than those on theX chromosome (assuming one copy in male cells).

However, if the selective pressure is high enough, the dominant terms in both numerator and denominator will be $(1-c^2)s$ and the ratio will tend to 1. (This can be seen also by dividing the numerator and denominator by *s*, redefining *r* as the ratio *u/s* and expanding the equation about *r=0*, resulting in a value of 1.). In conclusion, our model suggests that OGs evolve faster in an autosome than on an X chromosome in somatic tissues, unless mutation rates are particularly high.

We also ran our models by iterating generations without solving them analytically, monitoring the change in cancer cells frequencies for different selection and mutation rates. We observed that in oncogenes, under a moderate selective pressure (s=0.01), cancer cells spread faster when the driver mutation occurs in an autosome compared to a haploid X chromosome (Figure 1A). However, as we increase selective pressure, and as predicted by equation (1), the fixation in both chromosomes becomes similar (Figures 1B and 1C). We then considered the time (in generations) it takes in our model to reach a frequency of cancer cells that compromises one half of the population: $t_{1/2}$. This approach allows us to explore a wide range of selection pressures in a single plot. In Figure 2A, we compared the $t_1/_2$ across all selective pressure values from 0 to 1, in both diploid and haploid configurations, confirming that as selective pressure gets stronger, the evolutionary rate of oncogenes in both types of chromosomes becomes similar. For higher mutation rates, the pattern remains comparable, but the differences between X chromosomes and autosomes become even smaller (Figures 2B and 2C).

In conclusion, the models of somatic evolution predict that if an oncogene is located on an autosome, driver mutations will lead to a faster spread of cancer cells compared to oncogenes located on the X chromosome in males. However, these differences are likely to be negligible in a realistic cancer context with relatively high mutation and selective pressures. In other words, when selection is strong, ploidy does not have a significant impact on gain-of-function mutations in proto-oncogenes leading to cancer.

Faster-X somatic evolution of tumour-suppressor genes

To evaluate the evolutionary rate of TSGs, we can, as in the OG case, determine the relative rate of change in the diploid recessive (ΔC_{dr}) model with respect to the rate in the haploid (ΔC_h) model (see Methods for derivation):

$$R_{TSG} = \frac{\Delta C_{dr}}{\Delta C_h} \approx \frac{c(1-c)s}{(1-c)u + (1-c^2)s} \qquad (2)$$

In this equation, (1-c)u is always positive, and $(1-c^2)s > c(1-c)s$ except for the trivial state c=1. Hence, for any selection or mutation rate, the evolution of TSGs is faster on an X chromosome (again, assuming haploidy) than on the autosomes.

When we ran the model iteratively, we found that, regardless the level of selective pressure, the fixation of cancer alleles is faster on the X chromosome in males than on autosomes (Figures 3A to C). This is also observed for $t_{1/2}$ in Figure 4A. Similar to the case of oncogenes, we can use $t_{1/2}$ to simultaneously evaluate various selective and mutational pressures. Although evolution is always faster on the X chromosome for all evaluated values (Figures 4A to C), for high mutation rates, the differences between the X chromosome and autosomes become smaller, as in the case of oncogenes.

In summary, for any level of selective pressure, the computer models show that driver mutations in tumour-suppressor genes spread faster if they are located on the X chromosome in males compared to autosomes. This faster-X effect persists even under very high mutation rates, although the effect is less pronounced.

Empirical evaluation of the models

The predictions from the models developed in the previous sections can be compared with the mutation profiles observed in cancer genomic screens. Here, we downloaded and processed (see Methods) the information available in the COSMIC database, computing the number of observed mutations in genomic screens per nucleotide as a proxy for the evolutionary rate.

First, we combined multiple studies and calculated the evolutionary rate for each cancer gene. When comparing the evolutionary rate of oncogenes between the X chromosome and autosomes (including studies on both males and females for the X chromosome), we observed that the evolutionary rate is higher in autosomes compared to the X chromosome (Figure 5A). This observation aligns with the predictions from the model. Conversely, for TSGs, the pattern is reversed, showing a faster-X effect (Figures 5B), as also predicted by the models. However, the differences were not statistically sound (Mann-Whitney test, $p= 0.475$ and $p=0.346$ for OGs and TSGs respectively). This may be due to the merging of many heterogeneous studies and the lack of consideration for sex-specific information.

Thus, we analyzed each study separately, calculating the overall evolutionary rate of pooled TSGs and OGs on the X chromosome or autosomes for each study. In this analysis, TSGs appear to evolve faster on the X chromosome than on autosomes ($p<2.2E-6$, Wilcoxon test; Figure 5C) while oncogenes tend to evolve slightly faster in an autosomal context, although the statistical support in this case is weak ($p=0.0411$).

Next, we split the studies into those from male and those from female cancer samples. As expected, TSGs evolve faster on the X chromosome than on autosomes in males ($p=0.005$; Figure 6A), yet oncogenes evolve faster on autosomes than on the X chromosome ($p=0.18$). Also as expected, oncogenes show a comparable evolutionary pattern in both types of chromosomes in females, where the X is in diploidy ($p=0.566$; Figure 6B). Strikingly, TSGs in female cancers also evolve faster on the X chromosome compared to autosomes ($p<2.2E-16$). This may be explained by pervasive X-chromosome silencing or, in some cases, loss of

heterozygosity (see Discussion). When we consider only nonsense mutations (gain of a stop codon) we observe that in males, both OG and TSGs have mutations more frequently on the X chromosome than on autosomes Figure 6 C and D). This may be indicative of known oncogenes that can be affected by loss-of-function mutations. In the case of females (Figure 6D), for TSGs, there are two populations (bimodal distribution), one of which has a higher mutation rate on the X chromosome, potentially corresponding to samples with an X-chromosome silenced (or lost). Lastly, when we consider only synonymous mutations, the distribution of mutations for both TSGs and OGs in males (Figure 6E) and females (Figure 6F) indicates a comparable evolutionary rate between autosomes and the X chromosome, as expected, since these mutations are not likely to have an important functional impact on cancer development.

Lastly, we considered specific cancers to evaluate our predictions. The advantage of this approach is that it helps control biases arising from comparing very different samples and pooling gene information. The disadvantage is, however, a smaller sample size. In this case, we explored the evolutionary rate of TSGs and OGs in prostate cancer (an unambiguously male cancer) and breast cancer (mostly a female cancer, with all samples considered derived from females). From these analyses, we observed that the faster-X effect described above is present in both breast (Figure 7B) and prostate (Figure 7D) cancers, although the difference was not statistically significant ($p=0.120$ and $p=0.533$ in breast and prostate respectively), most likely due to the reduced sample size. Conversely, a slightly faster evolution of oncogenes in autosomes compared with X chromosomes was observed in both breast (Figure 7A) and prostate cancers (Figure 7C). However, these differences, in addition to not being statistically supported ($p=0.765$ and $p=0.996$) are not biologically meaningful as the effect size is negligible. In summary, the slower-X effect in oncogenes is negligible in cancer, but the faster-X effect in tumor-suppressor genes is pervasive in both male and female cancer samples.

DISCUSSION

Here, we developed a series of deterministic models to compare the expected evolutionary ratio of cancer genes in diploidy versus haploidy, serving as a proxy for male X chromosome-linked genes versus autosomal genes in somatic evolution. The first caveat is that these models are not intended to evaluate overall evolutionary rates in realistic situations. More complex stochastic models have been developed for this purpose (see Introduction). Nonetheless, our deterministic models are useful for comparing the evolutionary rates of the same gene as if they were in two different contexts (haploidy or diploidy). Importantly, stochastic models also support our findings for TSGs: if two hits are needed to reach a cancerous cell status, the probability of fixation at a given time is smaller than if only one hit is needed (see equation [3] in Michor et al. (2004). Moreover, the time to fixation of a cancer cell type requiring two hits is longer than that for cells requiring only one hit (Iwasa et al. 2005).

The observation that TSGs and OGs exhibit different evolutionary dynamics could potentially be used to identify novel cancer genes, particularly TSGs, based on the differences in their mutation profiles between the X chromosome and autosomes. Indeed, the somatic evolutionary features of cancer genes have been used to classify them as either TSGs or OGs (Chandrashekar et al. 2020). However, TSGs and OGs also have different germline evolution histories. For instance, TSGs on the X chromosome are, on average, younger than those on autosomes, suggesting that selection against TSGs on the X chromosome has driven a movement of genes out of the X (Wang et al. 2022). This type of selection-driven movement of genes across chromosomes has been extensively studied, such as in the demasculinization of the X chromosome (Sturgill et al. 2007). In summary, the distinct functional roles of TSGs and OGs in cancer are mirrored in their differing evolutionary dynamics in both germline and somatic contexts.

One caveat of our TSG model for the X chromosome is that for very high mutation rates, the approximations may not hold well (Supplementary Figure 3). Additionally, the numerical evaluation of half-times indicates very small differences for high selection rates in both OGs

(Figure 2) and TSGs (Figure 4). This might suggest that for cancers with a mutator phenotype (Loeb 2010), which results in very high mutation rates, our predictions may not be accurate. However, this is not necessarily the case. First, although mutator phenotypes have been described, they are not necessary for cancer development (Tomlinson and Bodmer 1999). Second, issues with mutation rates arise only when approaching a rate of 10%, which is unrealistically high. In non-recombining systems (such as cancer somatic evolution), very high mutation rates can result in a mutational meltdown, where the benefits of faster evolution are outweighed by the detrimental effects of unwanted mutations (Lynch et al. 1993). This phenomenon has also been proposed to occur in cancer (Frank and Nowak 2004).

Our results from the analysis of genomics screens also indicated that X-chromosome inactivation is critical in cancer development, as it affects the expression of TSGs in females. Indeed, TSGs that escape X inactivation and therefore require two-hits to reach cancer status are associated with a reduction of cancer incidence in females compared to males in certain cancers (Dunford et al. 2017).

In summary, TSGs tend to evolve faster on the X chromosome than on autosomes in cancer tissuess, either due to hemizygosity in males or the prevalence of X chromosome inactivation in females.

MATERIALS AND METHODS

Dynamic models of somatic cancer gene evolution

To compare the somatic evolutionary dynamics of cancer genes in different chromosomal contexts, we need three distinct dynamic models: one for haploid systems (modelling both TSG and OG evolution in X chromosomes in male-derived cells), one for diploid recessive

systems (modelling TSG evolution in autosomes) and one for diploid dominant systems (for OGs in autosomes).

Haploid

We first define a gene with two possible alleles: A and a, where A represents the wild-type allele and a confers a selective advantage (1+s) in a cancer/pre-cancer state. The mutation A to a models the loss-of-function of a TSG or the gain-of-function of an OG. To avoid confusion with classic germline evolutionary models, we use n and c instead of p and q to represent the proportions of cells in a tissue with the normal (non-cancerous) allele and the cancer (or cancer-predisposition) allele, respectively. Thus, as in germline models, n+c=1. Assuming a non-reversible mutation from A to a at a rate of u, the proportions of n and c after mutation are given by:

$$n' = n(1-u)$$

$$c' = c + (1-c)u$$

and after selection, if the relative selection in a is increased in *s* compared to that of A, will be given by:

$$n'' = n'/\bar{w}$$

$$c'' = c'(1+s)/\bar{w}$$

where $\bar{w}$ is the average selection given by:

$$\bar{w} = 1 + [c + (1-c)u](1+s)$$

This is analogous to the classic germline evolutionary model for haploids, with the difference that the assumptions made for further simplifications of the equation do not necessarily hold. From these equations, one can derive the expected change in the frequency of cancer cells after one 'generation' (a mutation plus a selection event):

$$\Delta c_h = c'' - c = \frac{(1-c)u + [1 - c^2 - c(1-c)u]s}{1 + [c + (1-c)u]s} \qquad (3)$$

Although in a cancer somatic evolution context we cannot assume that *s* and/or *u* are very small, these are moderately low and the product *su* can be assumed to be very small compared to *s* or *u*. Hence, we can approximate this equation as:

$$\Delta c_h \approx \frac{(1-c)u+(1-c^2)s}{1+cs} \quad (4)$$

The accuracy of this approximation is evaluated for a wide range of s and u values (both from 0.001 to 0.1) and the approximation closely followed the exact results (Supplementary Figure 1).

Diploid dominant

In this model, a cell with alleles AA has a mutation rate of *2u* towards Aa. If the cell is Aa, there is a rate *u* for it to change to aa, assuming no reverse mutation. However, since the selective advantage of Aa can be comparable to that of aa (in a completely dominant model), the model becomes similar to that of the haploid model, with the difference that the mutation rate is now *2u* instead of *u*. The chance in the frequency of c is therefore:

$$\Delta c_{dd} = c'' - c = \frac{2(1-c)u+[1-c^2-2c(1-c)u]s}{1+[c+2(1-c)u]s} \quad (5)$$

As in the haploid model, assuming that the product *su* is very small, we can simplify this to:

$$\Delta c_{dd} \approx \frac{2(1-c)u+(1-c^2)s}{1+cs} \quad (6)$$

The accuracy of this approximation is evaluated for a wide range of s and u values (both from 0.001 to 0.1) and the approximation closely followed the exact results (Supplementary Figure 2).

Diploid recessive

Since this is a somatic evolution model, there is no gamete formation; therefore, mutations occur at the diploid level. This results in a two-step mutation process, where the probability

of mutation from AA to Aa is *2u* and the probability of mutation from AA to aa is *u*. The relative selective advantage of the aa genotype is given by 1+s. As in the previous case, we are interested not in the actual genotype but in its contribution to cancer development. For this reason, we can redefine n as the proportion of cells with either AA and Aa genotypes, and *c* as the proportion of cells with the aa genotype. As a first approximation, we can assume that the mutation from AA/Aa to aa results from two simultaneous independent mutation events with rates *2u* and *u.* Therefore, the overall mutation rate will be $2u^2$. Taking all of this into account, the change in frequencies after mutation will be:

$$n' = n\left(1-2u^2\right)$$

$$c' = c + 2u^2\left(1-c\right)$$

and after selection:

$$n'' = n'/\bar{w}$$

$$c'' = c'\left(1+s\right)/\bar{w}$$

where:

$$\bar{w} = 1 + \left[c + 2u^2\left(1-c\right)\right]s$$

Giving the equation of change in cancer cells frequency:

$$\Delta c_{dr} = c'' - c = \frac{2u^2\left(1-c\right) + \left(c - c^2 + 2u^2 - 4cu^2 + 2c^2u^2\right)s}{1 + \left[c + 2u^2\left(1-c\right)\right]s} \qquad (7)$$

And if we assume that the term u^2 is significantly smaller than the other terms we can find an approximate form:

$$\Delta c_{dr} = c'' - c \approx \frac{c\left(1-c\right)s}{1+cs} \qquad (8)$$

As in the previous analysis, we must evaluate the accuracy of the approximated solution given in equation (8). Although the approximation was sufficiently accurate for realistic values of s and u, it becomes less accurate when u is very large (~0.1) and/or u>s (Supplementary Figure 3). However, these scenarios are unrealistic , and we expect mutation rates to be way below 0.10.

Analysis of somatic mutations in cancer

We retrieved information on cancer genes from the COSMIC database, version 99 (Forbes et al. 2015). We analyzed 492 census genes: 254 OGs and 238 TSGs, after discarding genes classified under both categories. We identified mutations described in multiple genome screens cataloged in COSMIC, tracking the individual studies and the sex of the donor. To measure the impact of mutation and selection on those genes, we computed the average number of missense mutations per nucleotide. All the code, with annotations, is available on GitHub at https://github.com/antoniomarco/CancerSomaticEvolutionX.

REFERENCES

Alfaro-Murillo JA, Townsend JP. 2023. Pairwise and higher-order epistatic effects among somatic cancer mutations across oncogenesis. *Mathematical Biosciences* **366**: 109091.

Avila V, de Procé SM, Campos JL, Borthwick H, Charlesworth B, Betancourt AJ. 2014. Faster-X effects in two Drosophila lineages. *Genome Biol Evol*.

Cairns J. 1998. Mutation and cancer: the antecedents to our studies of adaptive mutation. *Genetics* **148**: 1433–1440.

Chandrashekar P, Ahmadinejad N, Wang J, Sekulic A, Egan JB, Asmann YW, Kumar S, Maley C, Liu L. 2020. Somatic selection distinguishes oncogenes and tumor suppressor genes. *Bioinformatics* **36**: 1712–1717.

Charlesworth B, Coyne JA, Barton NH. 1987. The Relative Rates of Evolution of Sex Chromosomes and Autosomes. *The American Naturalist* **130**: 113–146.

Crow JF, Kimura M. 1970. *An Introduction to Population Genetics Theory*. Harper & Row.

Crow JF, Kimura M. 1965. Evolution in Sexual and Asexual Populations. *The American Naturalist* **99**: 439–450.

Crow JF, Kimura M. 1969. Evolution in Sexual and Asexual Populations: A Reply. *The American Naturalist* **103**: 89–91.

Dunford A, Weinstock DM, Savova V, Schumacher SE, Cleary JP, Yoda A, Sullivan TJ, Hess JM, Gimelbrant AA, Beroukhim R, et al. 2017. Tumor-suppressor genes that escape from X-inactivation contribute to cancer sex bias. *Nat Genet* **49**: 10–16.

Fisher RA. 1930. *The genetical theory of natural selection*. Oxford Clarendon Press http://archive.org/details/geneticaltheoryo00fishuoft (Accessed June 20, 2024).

Forbes SA, Beare D, Gunasekaran P, Leung K, Bindal N, Boutselakis H, Ding M, Bamford S, Cole C, Ward S, et al. 2015. COSMIC: exploring the world's knowledge of somatic mutations in human cancer. *Nucl Acids Res* **43**: D805–D811.

Frank SA. 2018. *Dynamics of Cancer: Incidence, Inheritance, and Evolution*. Princeton University Press.

Frank SA. 2010. Somatic evolutionary genomics: Mutations during development cause highly variable genetic mosaicism with risk of cancer and neurodegeneration. *Proceedings of the National Academy of Sciences* **107**: 1725–1730.

Frank SA, Nowak MA. 2004. Problems of somatic mutation and cancer. *Bioessays* **26**: 291–299.

Greaves M. 2018. Nothing in cancer makes sense except…. *BMC Biology* **16**: 22.

Greaves M, Maley CC. 2012. Clonal evolution in cancer. *Nature* **481**: 306–313.

Hartl DL, Clark AG. 2007. *Principles of Population Genetics*. Sinauer Associates.

Iwasa Y, Michor F, Komarova NL, Nowak MA. 2005. Population genetics of tumor suppressor genes. *Journal of Theoretical Biology* **233**: 15–23.

Knudson AG. 1971. Mutation and cancer: statistical study of retinoblastoma. *Proc Natl Acad Sci USA* **68**: 820–823.

Loeb LA. 2010. Mutator phenotype in cancer: origin and consequences. *Semin Cancer Biol* **20**: 279–280.

Lynch M, Bürger R, Butcher D, Gabriel W. 1993. The mutational meltdown in asexual populations. *J Hered* **84**: 339–344.

Martincorena I, Campbell PJ. 2015. Somatic mutation in cancer and normal cells. *Science* **349**: 1483–1489.

Maynard Smith J. 1968. Evolution in Sexual and Asexual Populations. *The American Naturalist* **102**: 469–473.

Michor F, Frank SA, May RM, Iwasa Y, Nowak MA. 2003. Somatic selection for and against cancer. *J Theor Biol* **225**: 377–382.

Michor F, Iwasa Y, Nowak MA. 2004. Dynamics of cancer progression. *Nat Rev Cancer* **4**: 197–205.

Muller HJ. 1932. Some Genetic Aspects of Sex. *The American Naturalist* **66**: 118–138.

Nunney L. 1999. Lineage selection and the evolution of multistage carcinogenesis. *Proc Biol Sci* **266**: 493–498.

Orr HA. 1995. Somatic mutation favors the evolution of diploidy. *Genetics* **139**: 1441–1447.

Ruddon RW. 2007. *Cancer Biology*. Oxford University Press, USA.

Sturgill D, Zhang Y, Parisi M, Oliver B. 2007. Demasculinization of X chromosomes in the Drosophila genus. *Nature* **450**: 238–241.

Tomlinson I, Bodmer W. 1999. Selection, the mutation rate and cancer: ensuring that the tail does not wag the dog. *Nat Med* **5**: 11–12.

Vogelstein B, Kinzler KW. 2002. *The Genetic Basis of Human Cancer*. 2nd edition. McGraw-Hill Professional, New York.

Wang X, Hu W, Li X, Huang D, Li Q, Chan H, Zeng J, Xie C, Chen H, Liu X, et al. 2022. Single-Hit Inactivation Drove Tumor Suppressor Genes Out of the X Chromosome during Evolution. *Cancer Research* **82**: 1482–1491.

Weinberg RA, Weinberg RA. 2006. *The Biology of Cancer*. 1st edition. Garland Science, New York.

FIGURE LEGENDS

Figure 1. Simulation of somatic evolution of oncogenes. The simulations run for 250 generations (x axis) and the frequency of cancerous (mutated) plotted (y axis). The simulations were performed for three different selective pressures: 0.01 (A), 0.05 (B) and 0.1 (C). Solid line represents diploid cells (autosomes) and dashed lines haploid cells (X chromosome in males).

Figure 2. Simulation of somatic evolution of oncogenes comparing selection and mutation levels. The simulations run for 250 generations and the time that took to reach 50% of cancer cells is shown in the y axis with respect to different selective pressures in the x axis. The analysis is done for three mutation levels: 0.01 (A), 0.05 (B) and 0.1 (C). Solid line represents diploid cells (autosomes) and dashed lines haploid cells (X chromosome in males).

Figure 3. Simulation of somatic evolution of tumour-suppressor genes. The simulations run for 250 generations (x axis) and the frequency of cancerous (mutated) plotted (y axis). The simulations were performed for three different selective pressures: 0.01 (A), 0.05 (B) and 0.1 (C). Solid line represents diploid cells (autosomes) and dashed lines haploid cells (X chromosome in males).

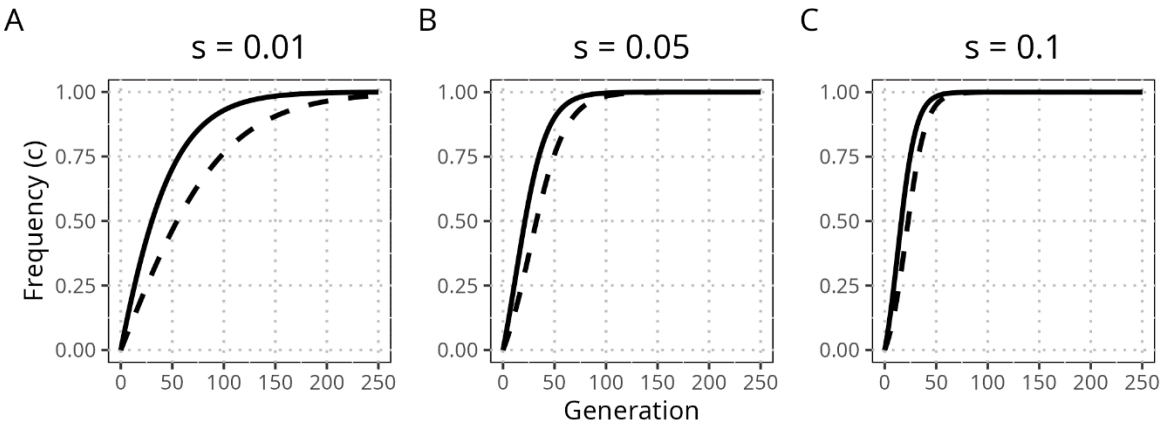
Figure 4. Simulation of somatic evolution of tumour-suppressor genes comparing selection and mutation levels. The simulations run for 250 generations and the time that took to reach 50% of cancer cells is shown in the y axis with respect to different selective pressures in the x axis. The analysis is done for three mutation levels: 0.01 (A), 0.05 (B) and 0.1 (C). Solid line represents diploid cells (autosomes) and dashed lines haploid cells (X chromosome in males).

Figure 5. Somatic mutations in cancer genes from cancer samples. The number of missense mutations per nucleotide is measures for each cancer gene using all available genome screens samples from COSMIC. Panel A shows the distribution for oncogenes, for both X chromosomes and autosomes. Panel B shows the equivalent data for tumour-suppressor genes. In panel C the distribution of log2 fold ratio of the number of mutations between autosomes and X chromosomes for tumour-suppressor genes and oncogenes pulled together for each genomic scan study.

Figure 6. Relative somatic evolutionary rate between autosomes and X chromosomes in cancer samples. The distribution of the log2 fold ratio between the number of mutations in autosomes and X chromosomes in cancer samples for tumour-suppressor genes and oncogenes combined for each genomic scan. The distributions are shown for males (A, C and E) and females (B, D, F) for missense mutations (A and B), nonsense mutations (C and D) and synonymous mutations (E and F).

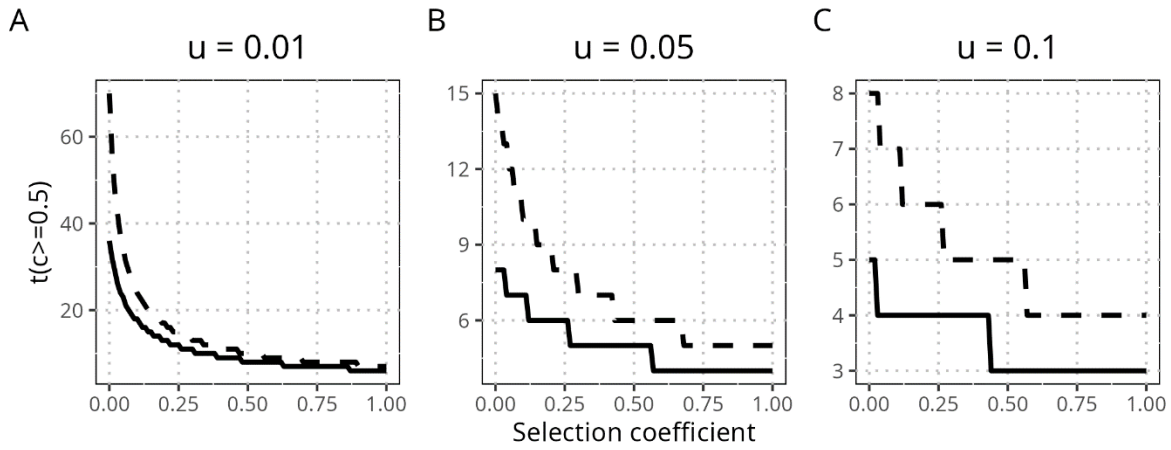
Figure 7. Mutations per nucleotide for breast and prostate cancer. This figure compares the number of missense mutations in breast cancer (A and B) and prostate cancer (C and D) for oncogenes (A and C) and tumour-suppressor genes (B and D) located either on the X chromosome or in a autosome.

1 **Figure 1**


A
s = 0.01
B
s = 0.05
C
s = 0.1
Frequency (c)
1.00
0.75
0.50
0.25
0.00
0
50
100
150
200
250
Generation

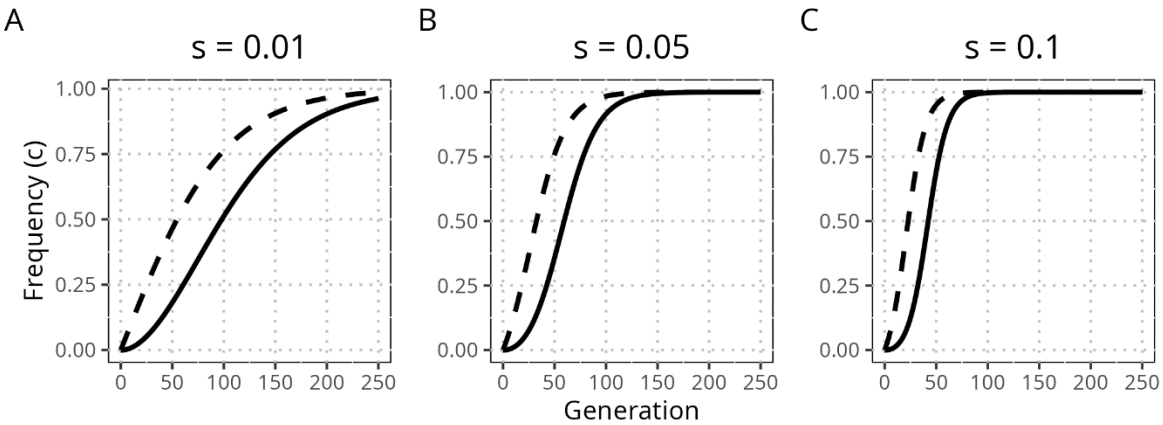
2

1 **Figure 2**


A
u = 0.01
B
u = 0.05
C
u = 0.1
t(c>=0.5)
60
40
20
15
12
9
6
8
7
6
5
4
3
0.00
0.25
0.50
0.75
1.00
Selection coefficient

2

1 **Figure 3**

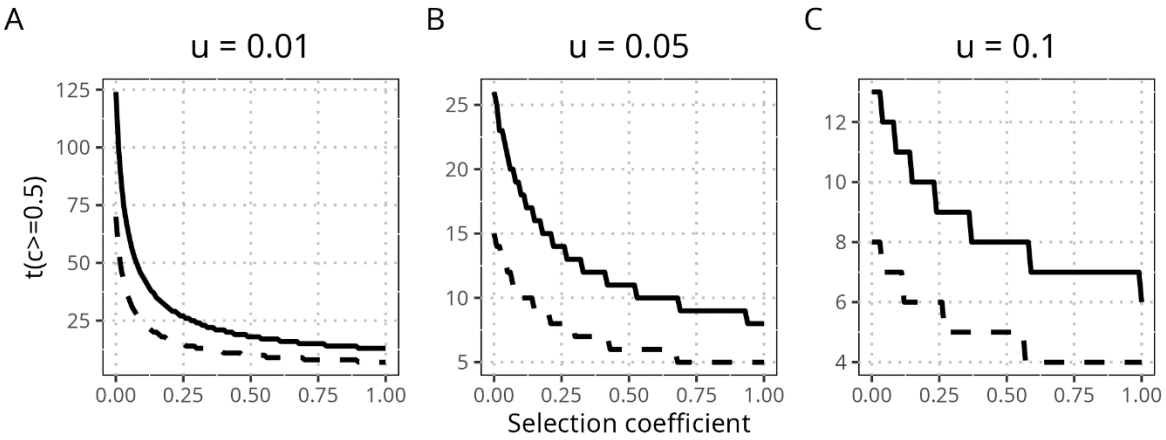


2

3

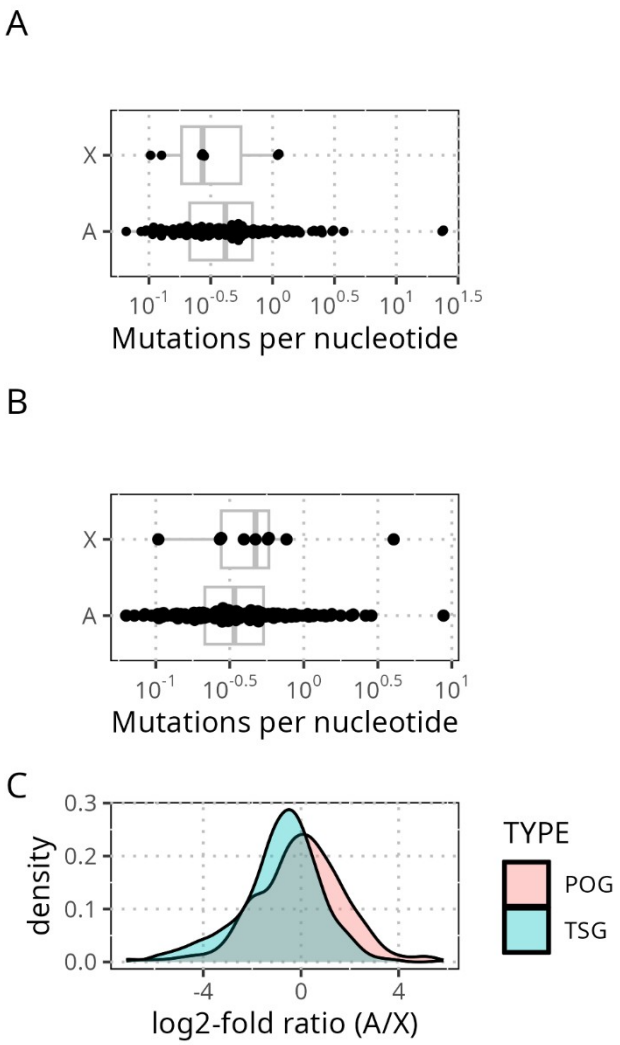
4

1 **Figure 4**


A
u = 0.01
B
u = 0.05
C
u = 0.1
t(c>=0.5)
125
100
75
50
25
25
20
15
10
5
12
10
8
6
4
0.00
0.25
0.50
0.75
1.00
Selection coefficient

2

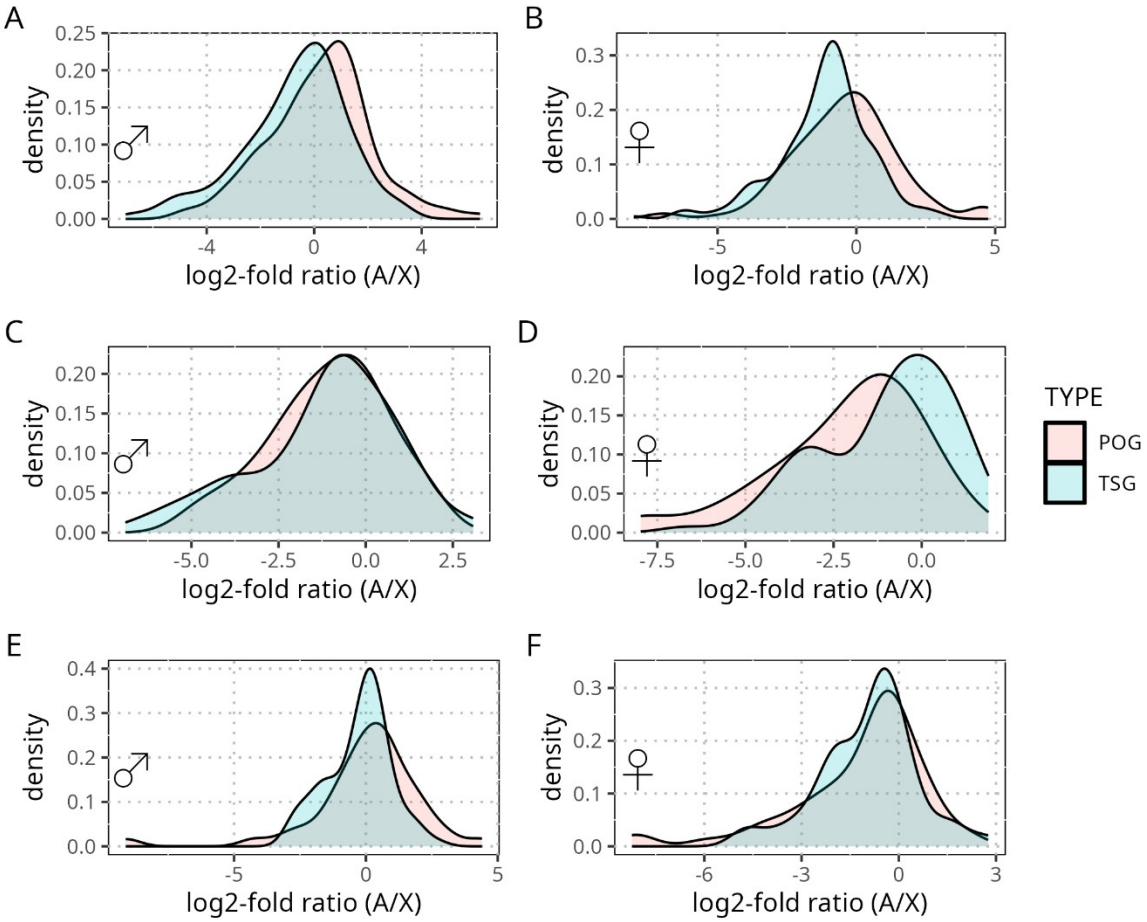
1 **Figure 5**


A
X
A
10^{-1} $10^{-0.5}$ 10^{0} $10^{0.5}$ 10^{1} $10^{1.5}$
Mutations per nucleotide
B
X
A
10^{-1} $10^{-0.5}$ 10^{0} $10^{0.5}$ 10^{1}
Mutations per nucleotide
C
density
0.3
0.2
0.1
0.0
-4
0
4
log2-fold ratio (A/X)
TYPE
POG
TSG

2

3

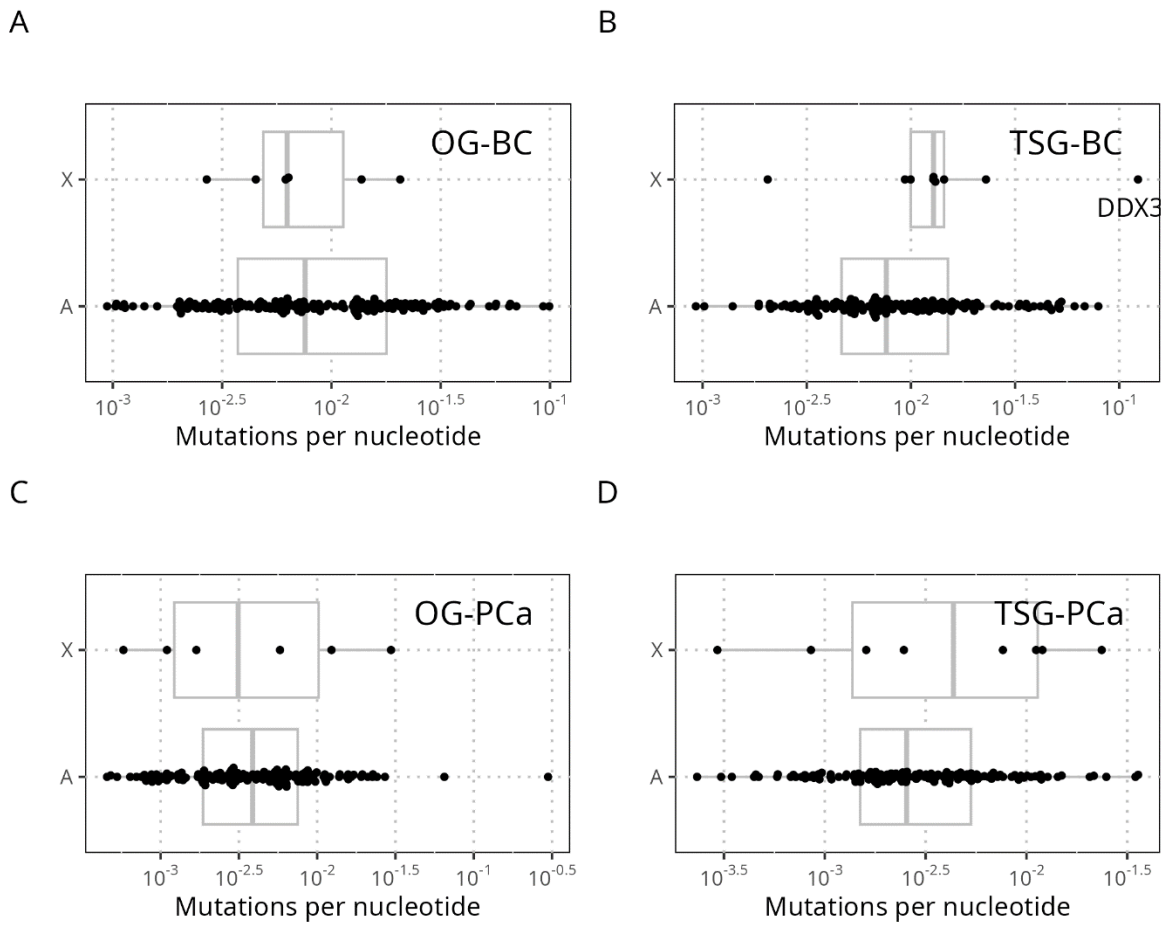
1 **Figure 6**



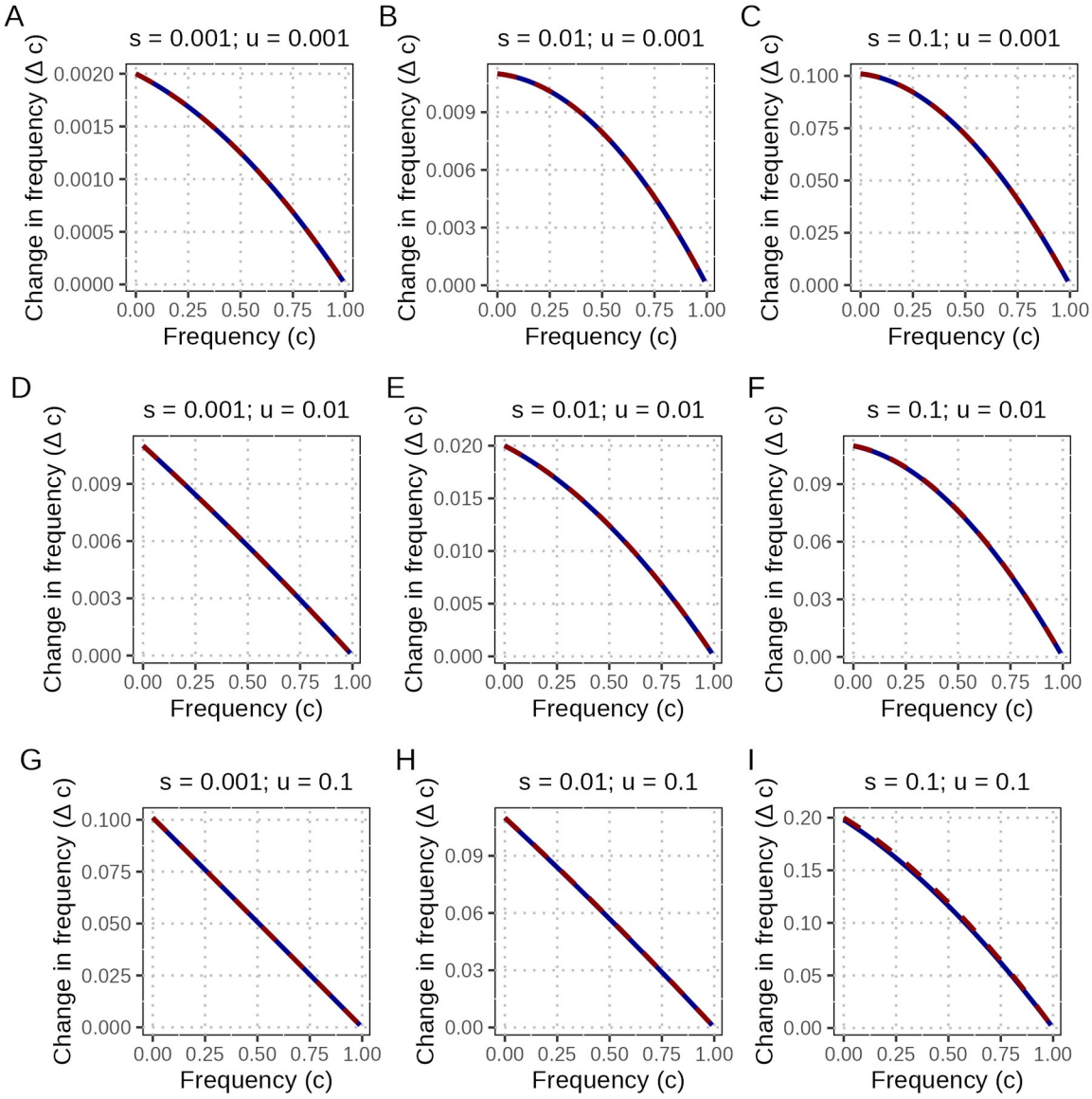
2

3

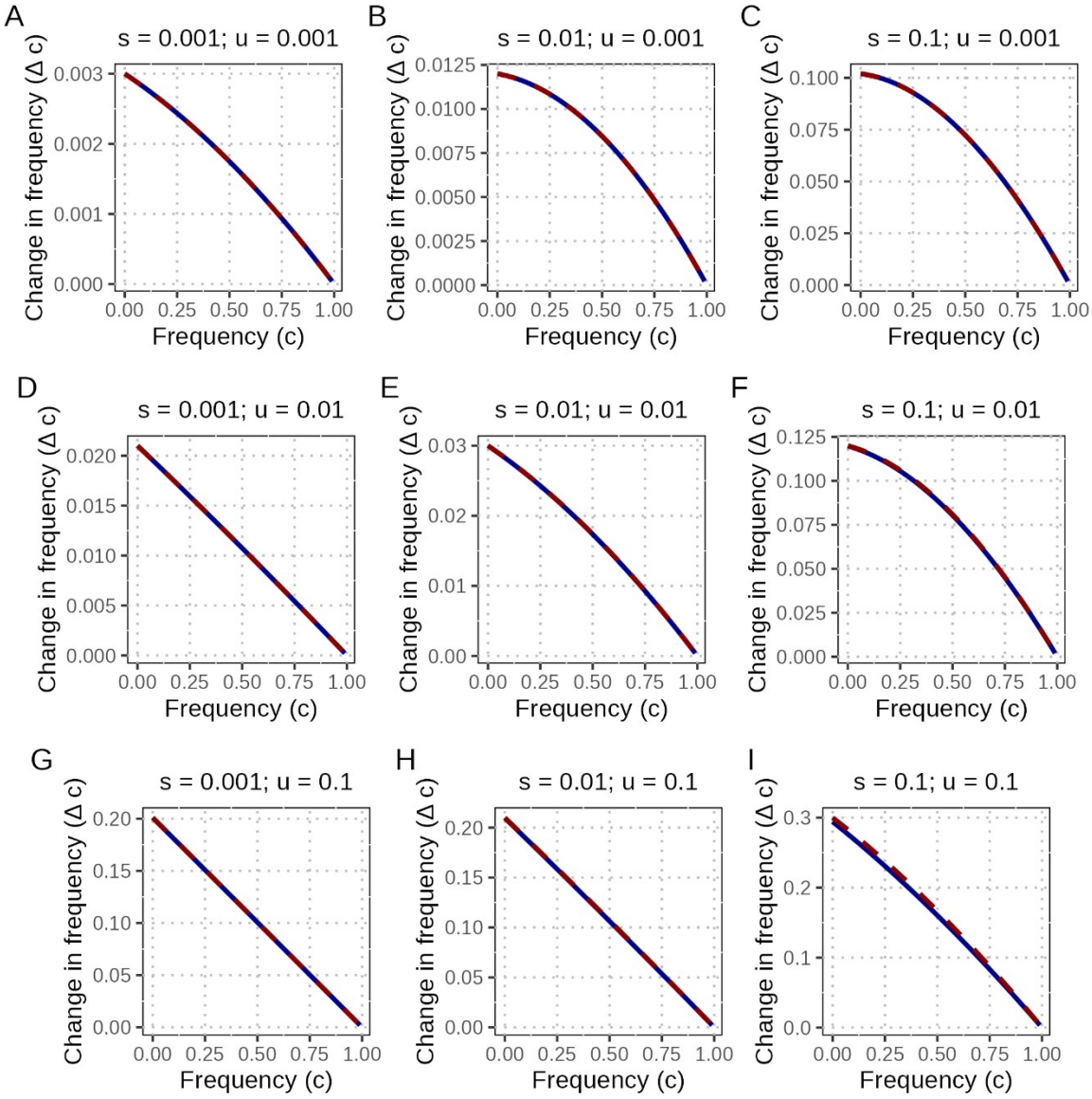
1 **Figure 7**


A
OG-BC
X
A
10^{-3}
$10^{-2.5}$
10^{-2}
$10^{-1.5}$
10^{-1}
Mutations per nucleotide
B
TSG-BC
DDX3
X
A
10^{-3}
$10^{-2.5}$
10^{-2}
$10^{-1.5}$
10^{-1}
Mutations per nucleotide
C
OG-PCa
X
A
10^{-3}
$10^{-2.5}$
10^{-2}
$10^{-1.5}$
10^{-1}
$10^{-0.5}$
Mutations per nucleotide
D
TSG-PCa
X
A
$10^{-3.5}$
10^{-3}
$10^{-2.5}$
10^{-2}
$10^{-1.5}$
Mutations per nucleotide

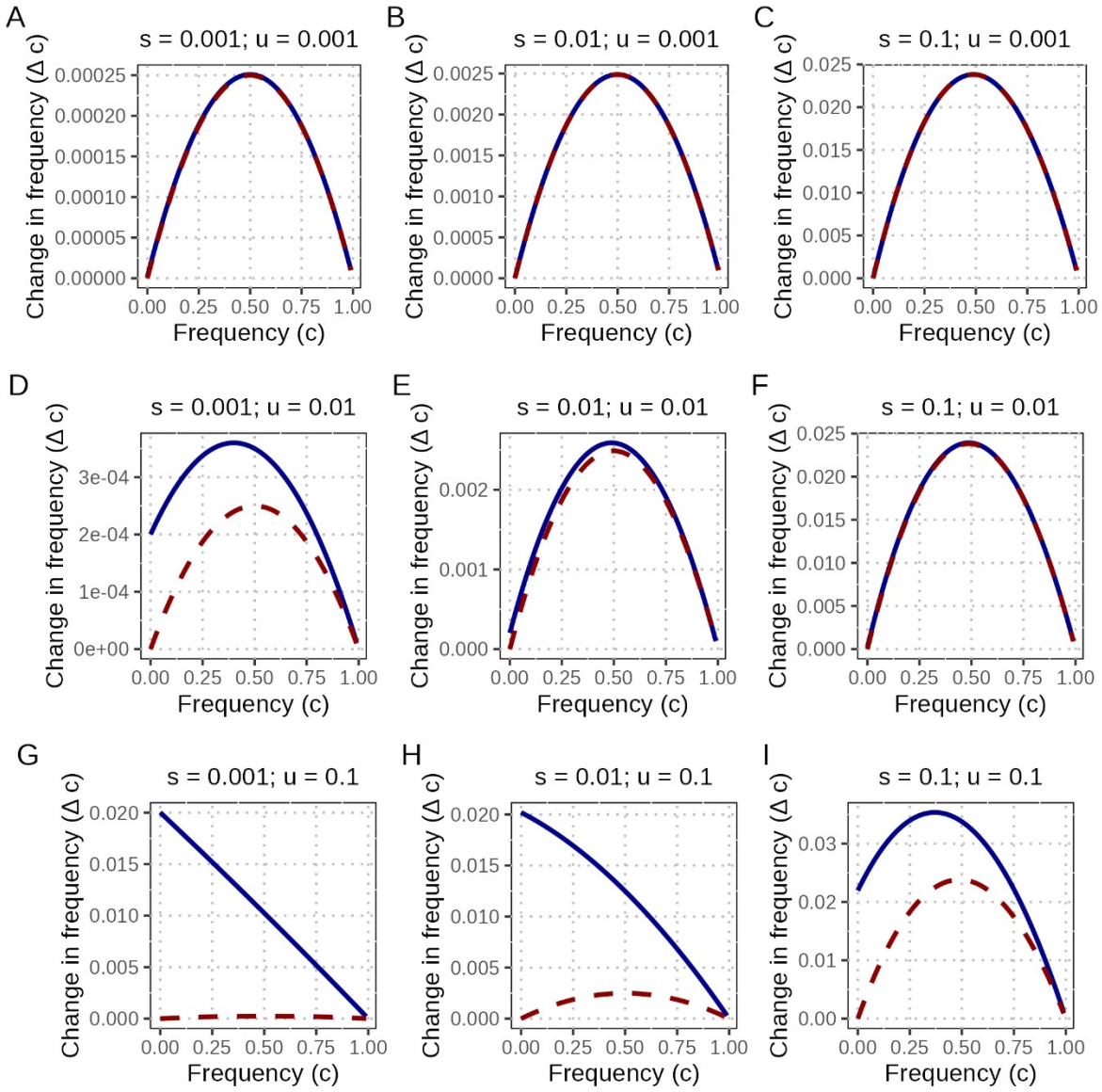
2



Supplementary Figure 1. Evaluation of the approximation for the change in mutant cells frequency in haploids. Each panel shows the result for combination of mutation and selective pressures, ranging from 0.001 to 0.1 each. The solid line represents the outcome from the full equation of change in frequency (3) while the dashed line is the outcome of the approximated equation (4), both as a function of the frequency.



Supplementary Figure 2. Evaluation of the approximation for the change in mutant cells frequency in diploid for dominant mutations. Each panel shows the result for combination of mutation and selective pressures, ranging from 0.001 to 0.1 each. The solid line represents the outcome from the full equation of change in frequency (5) while the dashed line is the outcome of the approximated equation (6), both as a function of the frequency.



Supplementary Figure 3. Evaluation of the approximation for the change in mutant cells frequency in diploid for recessive mutations. Each panel shows the result for combination of mutation and selective pressures, ranging from 0.001 to 0.1 each. The solid line represents the outcome from the full equation of change in frequency (7) while the dashed line is the outcome of the approximated equation (8), both as a function of the frequency.