Somatic evolution of cancer genes on the sex chromosomes

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

[1]School of Life Sciences, University of Essex, Wivenhoe Park, Colchester CO43SQ, United Kingdom.

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

* Corresponding author: amarco@essex.ac.uk (AM)

† These authors contributed equally to the work.

ABSTRACT

Cancer progression arises through the somatic evolution of oncogenes (OGs) and tumour-suppressor genes (TSGs). According to Knudson's two-hit hypothesis, both alleles of a TSG must typically be inactivated for tumorigenesis to proceed, whereas a single activating mutation can drive an OG. However, this diploid framework overlooks the distinct evolutionary dynamics of sex chromosomes. In males, the hemizygous (functionally haploid) state of the X chromosome implies that only a single mutational "hit" is needed to disable an X-linked TSG, potentially accelerating its contribution to cancer development. Here, we integrate somatic population genetics models with whole-genome mutation data to investigate how ploidy and dominance shape somatic evolution. We show that X-linked TSGs in male tumours evolve significantly faster than autosomal TSGs, particularly in highly proliferative tissues, whereas in low-proliferation tumours autosomal OGs accumulate mutations more rapidly than X-linked OGs. We further demonstrate that fixation times of advantageous mutations differ systematically between OGs and TSGs, converging only for male X-linked genes, as predicted from mathematical models. Our findings reveal that haploidy of sex chromosomes fundamentally alters the dynamics predicted by the two-hit hypothesis, providing the first systematic evidence that chromosomal context biases cancer gene evolution. Incorporating these principles into cancer biology offers a more accurate framework for understanding sex differences in tumorigenesis and highlights new avenues for therapeutic strategies targeting genomic regions with accelerated mutational dynamics.

INTRODUCTION

Cancer results from the abnormal growth of populations of cells that escape multiple cellular control mechanisms [1,2]. Cancer development is a multistage process and typically requires one or more genes to be mutated at the somatic level [3–5]. A somatic mutation in a 'cancer gene', which is detrimental for the host, is actually beneficial to the somatic tumour as it may help it spread in the organism. Although this picture is much more complex, traditionally, two major types of cancer genes have been identified, depending on the type of mutation required to benefit the progressing tumour: oncogenes and tumour-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, tumour-suppressor genes (TSGs) usually require both copies to be inactive (by e.g. mutation, deletion, or silencing) to have an effect on 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 [6]. Intuitively, TSGs in a male X chromosome (mostly in haploidy) will only need one mutation. In general, we hypothesize that the chromosomal context of cancer genes (sex or autosomal) and the type of activating mutation (gain or loss of function) determines how fast these events leading to cancer accumulate in the genome.

To understand the dynamics of genetic mutations in cancer development, multiple mathematical models have been developed during the last decades. Cancer somatic evolutionary models are often based on the Luria-Delbrück model or its derivations (reviewed in [7]). These models often implement realistic parameters to account for cancer

complexity, such as explaining the different rates of cancer spread under different conditions. For instance, recent dynamic models implement multistage processes [8,9], compartmentalization [10], the joint impact of pro- and anti-cancer mutations [11], or even epistatic interactions [12]; see [13] for a historical overview of the development of cancer evolution models. These models of cancer evolution often assume full diploidy of the cells (or just ignore ploidy), which may have an impact when studying sex chromosomes. Instead, here we propose the application of traditional population genetics models that account for chromosomal ploidy and dominance of mutations in somatic evolution.

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 [14,15]. This is because the 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 [16]. However, some contradictory data related to the faster-X effect exist [17], as well as a number of confounding factors [18] that should be taken into account. In any case, these models assume diploidy and sexual reproduction so, in the context of somatic evolution in cancer, asexual

models for both haploids and diploids should be considered in order to understand a potential impact of the chromosomal context.

Some early models comparing haploid and diploid systems were developed by Crow and Kimura [19] based on earlier work in theoretical population genetics [20,21]. These models compared the evolution of genes in haploid and diploid (sexual) systems, concluding that sexual reproduction results in faster gene evolution. They also proposed that diploidy may confer some protection against damaging mutations in somatic tissues. This theory was further progressed by Orr and Otto [22] who, building upon the Crow-Kimura-65 model, specifically compared the rate of emergence of adaptive mutation between haploids and diploids in non-recombining systems. They conclude that, in somatic (asexual) evolution, the rate of adaptation depends, not only on the rate of emergence of new favorable mutations (ploidy) but also on the rate at which these mutations are fixed (dominance and population size).

We aim to investigate whether there is a difference in the rate of adaptive mutations in either TSGs and/or OGs between sex chromosomes and autosomes and the relative role of selective pressure and somatic mutation rates in their evolutionary rates during cancer progression. To do so, we build upon Crow-Kimura-65 and Orr-Otto-94 models, considering ploidy as a proxy for chromosome location, and evaluating our predictions by analyzing whole genome mutation screens from multiple cancer samples.

1

RESULTS

The rate of adaptation of cancer genes in a chromosomal context

To evaluate the rate of adaptation we consider the rate of emergence of a favorable mutation escaping stochastic loss per individual and per generation, *U*. Crow and Kimura [19] showed that, in asexual organisms, the average time to fixation of a favorable mutation is *1/U* and given by:

$$\frac{1}{U} = \int_0^g Np_i \, dt \qquad (1)$$

where p_i is the frequency of the advantageous mutation at time *t*, *g* is the average number of generations that it takes to reach p_t, and *N* is the effective population size. By assuming that the frequency of the advantageous mutation follows a logistic growth [23], they found that:

$$g = \frac{1}{s} ln\left[N\left(e^{s/UN} - 1\right) + 1\right] \qquad (2)$$

where *s* is the selective advantage of the mutant for asexual haploid populations, yet this term is *hs* for diploid asexual populations [22] where *h* is the dominance coefficient of the mutant with respect to the wild-type allele.

Orr and Otto [22] build upon this model and approximate forms of *g* for haploid and diploids separately and, since the rate of adaptive evolution (*K*) is inversely proportional to the average time to fixation (*g*) they derived that:

$$\frac{K_{diploid}}{K_{haploid}} \approx h \frac{ln\left[N\left(e^{1/(2Nv)} - 1\right)\right]}{ln\left[N\left(e^{1/(4Nv)} - 1\right)\right]} \qquad (3)$$

where *v* is the "rate at which favorable alleles appear by mutation per haploid genome" [22]. From here we will refer to the ratio K_{dip}/K_{hap} as *R*, the relative rate of adaptive evolution of

diploids with respect to haploids. Hence, a value of R greater than 1 indicates a faster evolution in a diploid context, and a value less than 1 indicates a faster evolution in a haploid context.

In the context of cancer somatic evolution, cancer genes on the X chromosome in male tissues are modeled as haploids, whilst autosomal genes (and X genes in female tissues, which some caveats that we will explore later) are modeled as diploids. Further, oncogenes (OGs) are characterized for becoming activated by gain-of-function dominant mutations (we explore this below), and hence we can assume partial or full dominance: $h > 0.5$. On the other hand, and following Hudson's two-hit model, Tumour Suppressor Genes (TSGs) are characterized for loss-of-function recessive mutations, so we assume either full or partial recessiveness: $h < 0.5$. Orr and Otto gave approximate solutions for the case of rare favorable mutations ($R \approx 2h$ for $Nv \ll 1$) and for more frequent favorable mutations in the population ($R \approx h$ for $Nv \geq 1$).

Estimation of the parameter *Nv* in human tissues

It is challenging to estimate Nv in human somatic tissues and tumours, yet we can attempt a rough approximation. The estimation of somatic mutation rates varies form study to study, yet most agree that it is in the range of 10^{-8} to 10^{-9} mutations per nucleotide per cell division [24–26]. Assuming that ~1.5% of the nucleotides of the human genome are part of codons, we can set up an upper limit (as not all mutations in codons are non-synonymous and beneficial for cancer progression) at 0.015 as the proportion of mutations that are beneficial to tumour

growth. Hence, v may be in the range of 1.5×10^{-10} to 1.5×10^{-11} beneficial substitutions per nucleotide per cell division.

The estimation of N is also problematic as different tissues have different sizes, and also only a fraction of a tissue is usually proliferative. We can start by considering skin, the largest organ that has a high proliferative rate. It is estimated that there are approximately 2.7 million proliferative cells per cm^2 of skin under normal conditions [27]. Given that the average size of the skin surface in an adult is around 1.6 m^2 [28], that gives us a total of 43.2 billion of proliferative cells. Given these estimates, the Nv in skin should be between 0.65 and 6.5. We can consider now a smaller organ with a lower proliferative rate, like in the case of the pancreas. Estimates from healthy individuals report an average number of acinar cells of 1.22×10^{11} cells per adult pancreas [29]; acinar cells are largely responsive for the development of carcinomas [30]. Although less than 2% of these cells are proliferative at the same time in an adult pancreas [31,32], this still gives an estimate of Nv between 3.66 and 36.60.

However, these rough estimates are more indicative of the rate of emergence of pioneer/driver mutations in healthy tissues. Once a tumour clone is established, the actual value of Nv during cancer progression will depend on the number of cells of the specific tumour. In this context, low proliferative tumours will have lower Nv values. For instance, in Clear Cell Renal Cell Carcinoma (ccRCC) is has been estimated that there are about 1.13×10^5 cells per mm^3 of tumour tissue [33]. For 1 cm^3 of a developing tumoral clonal population, the estimate of of Nv will range between 0.0017 and 0.0170, assuming that the tumour is not undergoing an accelerated mutation rate. This indicates that genes in low proliferation tumours, according to the model described above, may have different evolutionary dynamics

with respect to sex chromosomes compared to high proliferation tumours, as we will also explore in this work.

If the model discussed holds for somatic cancer evolution, we can make the following precise predictions:

(1) In male mid- and high- proliferative tumours ($Nu \geq 1$), advantageous mutations accumulate (collectively) faster in the X chromosome than in autosomes, with this effect being stronger in TSGs ($h<1/2$).

(2) In male low proliferative tumours ($Nu << 1$), advantageous mutations in TSGs accumulate faster in the X chromosome than in the autosomes ($h<1/2$) and vice versa in OGs ($h>1/2$).

Time to fixation

In the above model, the Poisson process with rate U, the time to advantageous mutation occurrence is exponentially distributed [34] and, therefore, the expected (average) arrival time is $1/U$. This is precisely where the term $1/U$ in equation (1) comes from [19]. That means that the expected time for an advantageous mutation to occur in haploids is:

$$\frac{1}{2vs}$$

and for diploids it is:

$$\frac{1}{4vhs}$$

That means that the time to advantageous mutation arrival is dependent on the coefficient of dominance in diploids, where for dominant mutations (h>0.5) the time to arrival is smaller than for recessive mutations (h<0.5), as it can be derived from the equation above. This allows us to make a third prediction from the model, in addition to the two made in the previous subsection:

3) Advantageous mutations arrise, on average, earlier in OGs than in TSGs, except for X chromosomes in males, where both arrise, on average, at the same time.

The nature of mutations in cancer genes

The classification of cancer genes into OG or TSG is a simplification. However, it often captures the dominance of the most likely driver mutations. From the COSMIC dataset here used we identified the number of OG and TSGs that are classified as having recessive or dominant mutations in cancer screening (Table 1). We also include genes classified as both TSG and OG and also those reported to be both dominant and recessive. From the data available it is evident that oncogenes are mostly characterized by hosting dominant (probably gain-of-function) somatic mutations, while TSGs are more likely to host recessive mutations ($p < 0.001$; Fisher's Exact test).

Table 1. Reported somatic mutations in COSMIC for census cancer genes

	Dominance		
	Dominant	**Recessive**	**Both**

OG	205	2	0
TSG	72	116	2
POG and TSG	40	17	5

If we assume that dominant mutations are mostly gain-of-function and recessive are loss-of-function, the nature of the mutations themselves must be different. Looking at coding genes, gain-of-function mutations like those expected in OGs should mostly involve a change in an amino acid (although this is not always the case, see Discussion). On the other hand, although a change in amino acid may imply a loss of function, we will expect an enrichment of nonsense mutations (gain of a STOP codon) as loss-of-function somatic mutations and, therefore, in TSGs. In Fig 1 we compare the number of nonsense and missense (nonsynonymus) mutations identified in genomic screens (see Methods). As expected, the number of nonsynonymous mutations is higher that the number of nonsense mutations in OGs (slope = 0.73), and the opposite is true for TSGs (slope = 1.17). Both regression lines are statistically independent ($p < 0.001$; regression interaction). For the study of TSGs, despite losing some statistical power, it is convenient to consider only nonsense mutations to avoid including gain-of-function nonsynonymous mutations that may bias the outcome. We conclude then that, as a good approximation, TSGs are characterised by loss-of-function recessive nonsense mutations whilst OGs are characterised by gain-of-function dominant nonsynonymous mutations.

Fig 1. Comparison between nonsynonymous and nonsense somatic mutations in cancer genes. Scatter-plot of the normalised number of non-synonymous somatic mutations versus the number of nonsense (STOP codon gain) somatic mutations mutation in the Genome

screens in the COSMIC database. Data shown for proto-oncogenes (POG) in red and for tumour-suppressor genes (TSG) in red. Lines of the same colour are the linear model fit, including the error for each line.

Evolutionary rates at the sex chromosomes

As a proxy for the somatic evolutionary rate of cancer genes we compute the log-odds ratio (LOR) as the logarithm (natural) of the ratio of the fraction of function changing mutations over potential function changing sites with respect to the fraction of synonymous mutations over synonymous sites (see Methods for details). For TSGs we considered STOP codon gain mutations are change of function, while for OG we consider nonsynonymous mutations (see rationale in the previous section). The genomic screen studies were classified according to whether the cancer samples were retrieved from females or male samples, and the analysis were performed separately for both sexes.

In TSGs, the LOR (approx. evolutionary rate) in female cancer samples was comparable between autosomes and the X chromosome (Fig 2A, $p = 0.387$; Mann-Whitney test). However, in female samples, and in concordance with the theoretical model described above, the evolutionary rate was higher in genes in the X chromosome compared to genes in autosomes (Fig 2B, $p = 0.01663$). The effect size of this difference is notable (difference between the medians = 0.9985). In OGs, like with TSGs, the evolutionary rate of genes in autosomes is comparable to that of genes in the X chromosome in female cancer samples (Fig 3A, $p = 0.8502$). In male samples, and again in agreement with the theoretical model, genes in the X chromosome evolve faster compared to genes in the autosomes (Fig 3B, $p = 0.0363$).

A specific prediction of the model is that the relative accelerated evolution of genes in the X chromosome with respect to autosomes should be stronger in TSG compared to OG. Indeed, the effect size of the differences in evolutionary rate in OGs is smaller, as predicted (difference of the medians = 0.3997) than in the case of TSGs. We conclude that the empirical data available from the study of cancer samples agrees with the first prediction we made from the theoretical model.

Fig 2. Nonsense mutations in TSGs for males and females in autosomes and the X chromosome. Boxplot of the scaled mutation rate (log-odds ratio of the number of nonsense mutations, see Methods) of TSGs located in any autosome (A) or in the X chromosome (X). Data is shown separately for female (A) and for male (B) cancer samples.

Fig 3. Non-synonymous mutations in OGs for males and females in autosomes and the X chromosome. Boxplot of the scaled mutation rate (log-odds ratio of the number of non-synonymous mutations, see Methods) of OGs located in any autosome (A) or in the X chromosome (X). Data is shown separately for female (A) and for male (B) cancer samples.

A second prediction of the model is that, for low proliferation tumours, the pattern for OGs will be reverse, and genes in the autosomes will evolve faster than genes in the X chromosome. To test that, we restricted the analysis of OGs to samples from low proliferative tumours (see Methods for details). Like in the previous analysis, there is no noticeable difference in the evolutionary rates between chromosomes in female samples (Fig 4A, p = 0.1348). On the other hand, and in agreement with the theoretical model, OGs in autosomes

evolve faster than genes in the X although in male samples, given the limited sample size, the difference did not quite meet the significance threshold (Fig 4B, p = 0.0895).

Fig 4. Non-synonymous mutations in low-proliferative cancer samples in OGs located in autosomes and the X chromosome. Boxplox of the scaled mutation rate (log-odds ratio of the number of non-synonymous mutations, see Methods) of OGs located in any autosome (A) or in the X chromosome (X) of low-proliferative cancer samples. Data is shown separately for female (A) and for male (B) cancer samples.

It would be of interest to evaluate whether genes silenced by X chromosome inactivation behave as haploids in the model. However, there are two important issues. First, if we select genes that are described as X silenced or not silenced, among our TSG we only identify 4 and 5 respectively, limiting any statistical analysis. Second, gene specific X chromosome inactivation may be very different in cancer cells compared to healthy somatic tissues, and any outcome should be taken with precaution. In any case, the median evolutionary rate (LOR) of documented X inactivated TSGs is -1.22, higher than that evolutionary rate of non-inactivated X genes and autosomes, being -1.56 and -1.49 respectively.

Arrival times

A last prediction we made from the model is that the time of fixation of advantageous mutations is smaller in OG than in TSG except in male X chromosomes (haploids). To evaluate this we considered samples for which only mutations at either TSG or OG have been reported, and counted how many of these samples are per chromosome, sex and either TSG

or OG. A shorter fixation time should be, therefore, reflected in a higher number of mutations in a set of genes compared to a gene for which a longer fixation time is expected. As expected from the model predictions, we observed a significantly higher number of mutations (shorter fixation time) in OG with respect to those in TSG for all samples except for male X chromosome genes (Table 2). That is, arrival times for (cancer) beneficial mutations in TSG and OG are comparable for X-linked genes, as predicted by the model.

Table 2. Expected and observed mutations found in OG and TSG

Chromosome	Sex	O(TSG)	E(TSG)	O(OG)	E(OG)	P[1] (Bonferroni)
A	**female**	300	804.7	2206	1701.3	<0.001
X	**female**	21	32.1	29	17.9	0.006
A	**male**	350	893.9	2434	1890.1	<0.001
X	**male**	31	37.8	28	21.2	0.275

[1] G-test (Likelihood Ratio test) associated p-value corrected for multiple testing with Bonferroni correction.

In summary, we showed that mathematical models of somatic evolution predict that X-linked inactivation of TSGs will evolve faster than those in the autosomes, and the available empirical data strongly support this prediction.

DISCUSSION

Our results demonstrate that the chromosomal context and dominance relationships of mutations profoundly shape the somatic evolution of cancer genes. By combining population genetics theory with large-scale cancer genomic data, we show that the hemizygous state of the male X chromosome accelerates the accumulation of mutations in tumour-suppressor genes (TSGs), consistant with predictions derived from the Knudson's two-hit hypothesis. In this framework, TSGs typically require inactivation of both alleles to drive tumorigenesis [6]. Our results provide strong empirical support for this prediction and extend the scope of the two-hit hypothesis by incorporating chromosomal context and ploidy into models of somatic cancer evolution.

Although the model used treat cancer cells as haploid or diploid depending on chromosomal location, real tumours are often characterized by pervasive genomic instability, including whole-chromosome gains, losses, and local copy-number alterations [4]. These alterations may shift the effective ploidy of cancer genes, modulating their evolutionary dynamics. For example, loss of heterozygosity (LOH) is a well-established mechanism by which diploid TSGs effectively become hemizygous, facilitating their inactivation with a single additional mutation [6,35]. Conversely, chromosomal gains can buffer the effect of recessive mutations, delaying the fixation of loss-of-function alleles. Notably, recurrent aneuploidies are observed across multiple cancer types, and these may accelerate or decelerate the dynamics predicted by our models depending on whether they involve dosage-sensitive oncogenes (OGs) or TSGs. Incorporating variable ploidy and chromosomal instability into mathematical models therefore represents an important extension of the framework presented here.

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 [36]. 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 [37]. This type of selection-driven movement of genes across chromosomes has been extensively studied, such as in the demasculinization of the X chromosome [38]. 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.

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 [39]. Our analysis also suggest that the evolutionary rate of inactivated TSG in females is higher than that of non-inactivated (escapees) TSG in X and those in autosomes. However, the data is limited and we could not perform adequate statistical evaluations.

A further limitation of our work is considering dominance solely in terms of recessive versus dominant mutations. In reality, there are cases where heterozygosity itself provides a selective advantage, a phenomenon known as overdominance or heterozygote advantage. For example, partial loss of function in DNA repair genes may increase mutation rates, fostering

genetic diversity and clonal expansion, while complete loss can be deleterious to tumour cell viability [5,36]. Similarly, certain p53 variants can act in a dominant-negative fashion, where heterozygous states are selectively favored over either homozygous wild-type or complete loss [40]. Mathematically, overdominance is modelled by allowing values of h greater than 1. Thus, according to the model, overdominance will result in a systematic higher evolutionary rate of both OG and TSG in diploidy (autosomes or X in females). The available data on overdominance mutation is limited, so we could not evaluate this prediction, but we can anticipate that, if the prediction of the models are correct, future results may confirm that over dominant mutations accumulate more slowly in male X chromosomes compared to chromosomes in somatic diploidy.

Taken together, our findings highlight the importance of considering chromosomal context in cancer evolution. The faster evolution of X-linked TSGs in males provides a mechanistic explanation for observed sex differences in cancer incidence and progression [39]. Moreover, our study demonstrates that population genetics models originally developed for organismal evolution can be adapted to somatic evolution, offering an alternative framework for understanding mutation dynamics in cancer. By extending these models to account for variable ploidy and chromosomal instability we may gain deeper insights into the evolutionary logic of tumour development.

MATERIALS AND METHODS

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 (except for the analysis in Table 1). We identified mutations described in multiple genome screens cataloged in COSMIC, tracking the individual studies and the sex of the donor. The log-odds ratio (LOR) was computed as:

$$LOR = log \frac{\frac{N_{changing}}{N_{possible}}}{\frac{N_{syn}}{N_{synSites}}}$$

where $N_{changing}$ refers to either STOP gain codon (nonsense) mutations or nonsynonymous (missense) mutations, depending on the specific analysis (see main text), and in accordance $N_{possible}$ refers to the number of possible nonsense of missense mutations. N_{syn} and $N_{synSites}$ refer to the number of synonymous mutations and the number of synonymous sites respectively. The number of synonymous and nonsynonymous sites was computed using the Nei-Gojobori (NG86) approach [41].

We classified studies from TCGA into high and low proliferative tumour samples. To do so, we consider high proliferative tumours those with a median proliferation score of over 0 according to Thorsson et al. [42]. X-chromosome inactivated genes and escapees were retrieved from the compilation by Slavney et al. [43]

All statistical analysis and graph plotting was done with R 4.5.0 [44] and tidyverse 2.0.0 [45]. All the code, with annotations, is available on GitHub at https://github.com/antoniomarco/CancerSomaticEvolutionX

1

2

REFERENCES

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

2. Ruddon RW. Cancer Biology. Oxford University Press, USA; 2007.

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

4. Greaves M, Maley CC. Clonal evolution in cancer. Nature. 2012;481: 306–313. doi:10.1038/nature10762

5. Martincorena I, Campbell PJ. Somatic mutation in cancer and normal cells. Science. 2015;349: 1483–1489. doi:10.1126/science.aab4082

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

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

8. Frank SA. Somatic evolutionary genomics: Mutations during development cause highly variable genetic mosaicism with risk of cancer and neurodegeneration. Proceedings of the National Academy of Sciences. 2010;107: 1725–1730. doi:10.1073/pnas.0909343106

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

10. Frank SA, Nowak MA. Problems of somatic mutation and cancer. Bioessays. 2004;26: 291–299. doi:10.1002/bies.20000

11. Michor F, Frank SA, May RM, Iwasa Y, Nowak MA. Somatic selection for and against cancer. J Theor Biol. 2003;225: 377–382. doi:10.1016/s0022-5193(03)00267-4

12. Alfaro-Murillo JA, Townsend JP. Pairwise and higher-order epistatic effects among somatic cancer mutations across oncogenesis. Mathematical Biosciences. 2023;366: 109091. doi:10.1016/j.mbs.2023.109091

13. Cairns J. Mutation and cancer: the antecedents to our studies of adaptive mutation. Genetics. 1998;148: 1433–1440. doi:10.1093/genetics/148.4.1433

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

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

16. Charlesworth B, Coyne JA, Barton NH. The Relative Rates of Evolution of Sex Chromosomes and Autosomes. The American Naturalist. 1987;130: 113–146. doi:10.2307/2461884

17. Avila V, de Procé SM, Campos JL, Borthwick H, Charlesworth B, Betancourt AJ. Faster-X effects in two Drosophila lineages. Genome Biol Evol. 2014. doi:10.1093/gbe/evu229

18. Fraïsse C, Puixeu Sala G, Vicoso B. Pleiotropy Modulates the Efficacy of Selection in Drosophila melanogaster. Mol Biol Evol. 2019;36: 500–515. doi:10.1093/molbev/msy246

19. Crow JF, Kimura M. Evolution in Sexual and Asexual Populations. The American Naturalist. 1965;99: 439–450. doi:10.1086/282389

20. Fisher RA. The genetical theory of natural selection. Oxford Clarendon Press; 1930. Available: http://archive.org/details/geneticaltheoryo00fishuoft

21. Muller HJ. Some Genetic Aspects of Sex. The American Naturalist. 1932;66: 118–138. doi:10.1086/280418

22. Orr HA, Otto SP. Does diploidy increase the rate of adaptation? Genetics. 1994;136: 1475–1480. doi:10.1093/genetics/136.4.1475

23. Michor F, Iwasa Y, Nowak MA. Dynamics of cancer progression. Nat Rev Cancer. 2004;4: 197–205. doi:10.1038/nrc1295

24. Iwasa Y, Michor F, Komarova NL, Nowak MA. Population genetics of tumor suppressor genes. Journal of Theoretical Biology. 2005;233: 15–23. doi:10.1016/j.jtbi.2004.09.001

25. Maynard Smith J. Evolution in Sexual and Asexual Populations. The American Naturalist. 1968;102: 469–473. doi:10.1086/282559

26. Crow JF, Kimura M. Evolution in Sexual and Asexual Populations: A Reply. The American Naturalist. 1969;103: 89–91. doi:10.1086/282585

27. Orr HA. Somatic mutation favors the evolution of diploidy. Genetics. 1995;139: 1441–1447. doi:10.1093/genetics/139.3.1441

28. Chandrashekar P, Ahmadinejad N, Wang J, Sekulic A, Egan JB, Asmann YW, et al. Somatic selection distinguishes oncogenes and tumor suppressor genes. Bioinformatics. 2020;36: 1712–1717. doi:10.1093/bioinformatics/btz851

29. Wang X, Hu W, Li X, Huang D, Li Q, Chan H, et al. Single-Hit Inactivation Drove Tumor Suppressor Genes Out of the X Chromosome during Evolution. Cancer Research. 2022;82: 1482–1491. doi:10.1158/0008-5472.CAN-21-3458

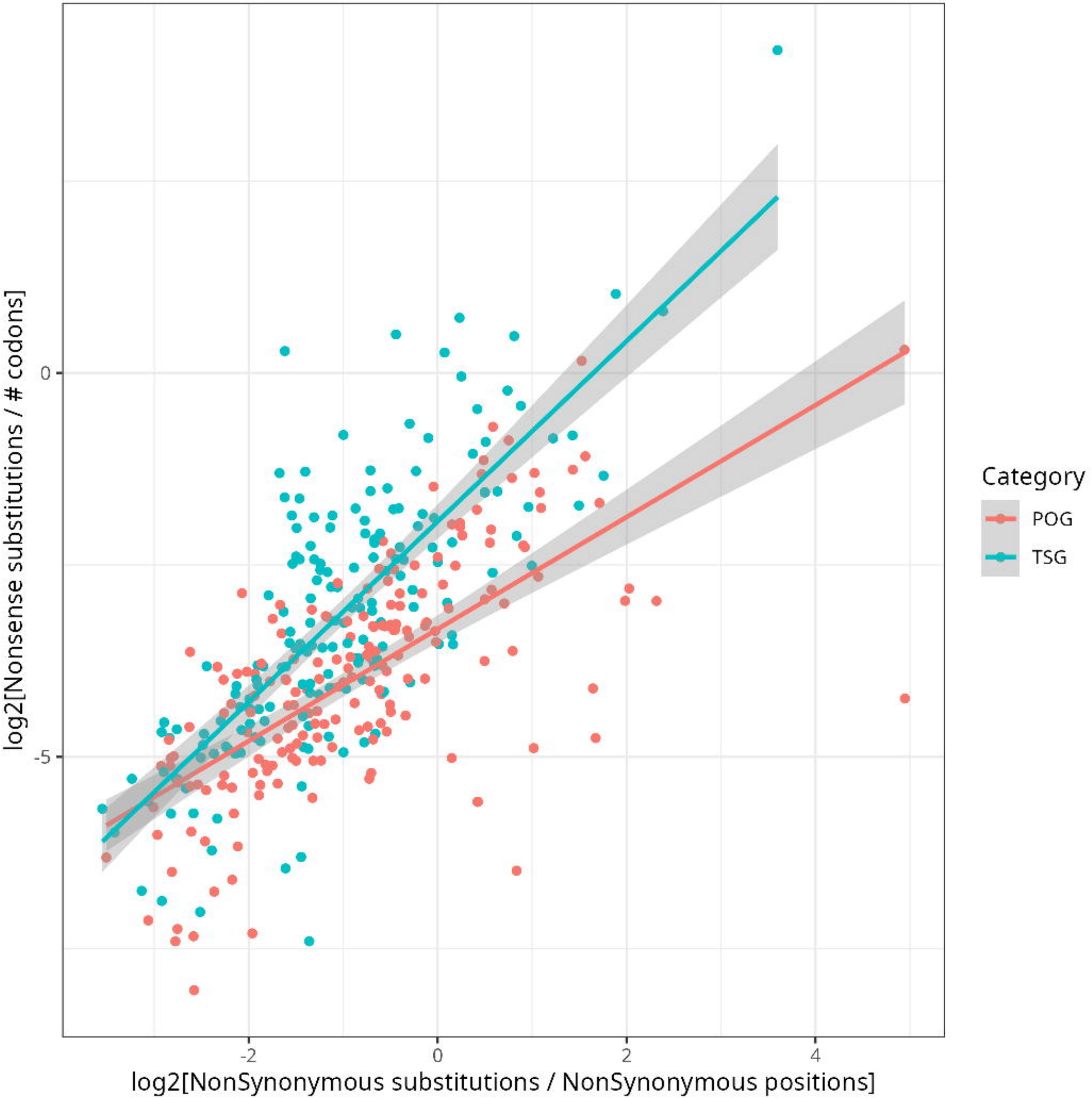
30. Sturgill D, Zhang Y, Parisi M, Oliver B. Demasculinization of X chromosomes in the Drosophila genus. Nature. 2007;450: 238–241. doi:10.1038/nature06330

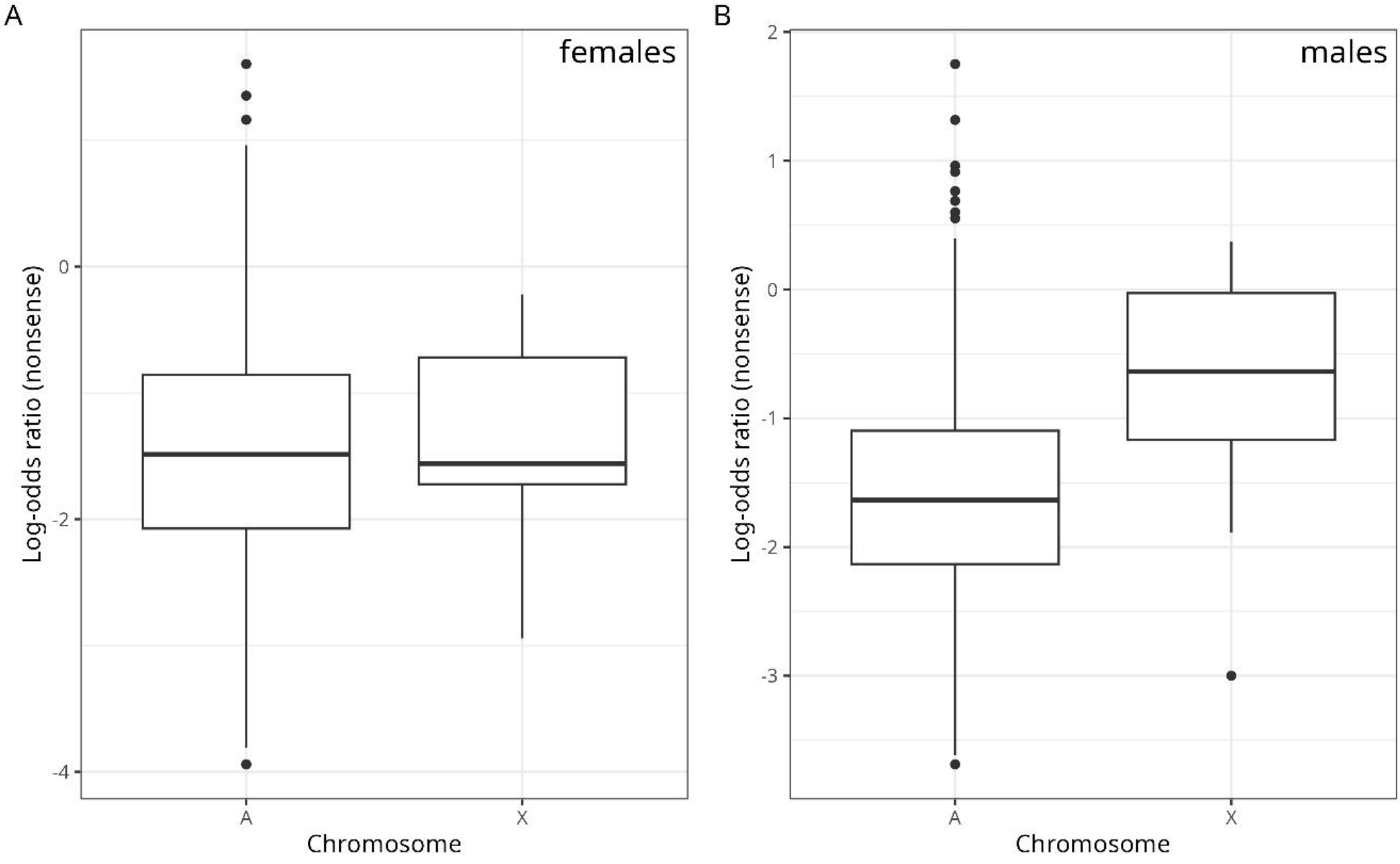
31. Tomlinson I, Bodmer W. Selection, the mutation rate and cancer: ensuring that the tail does not wag the dog. Nat Med. 1999;5: 11–12. doi:10.1038/4687

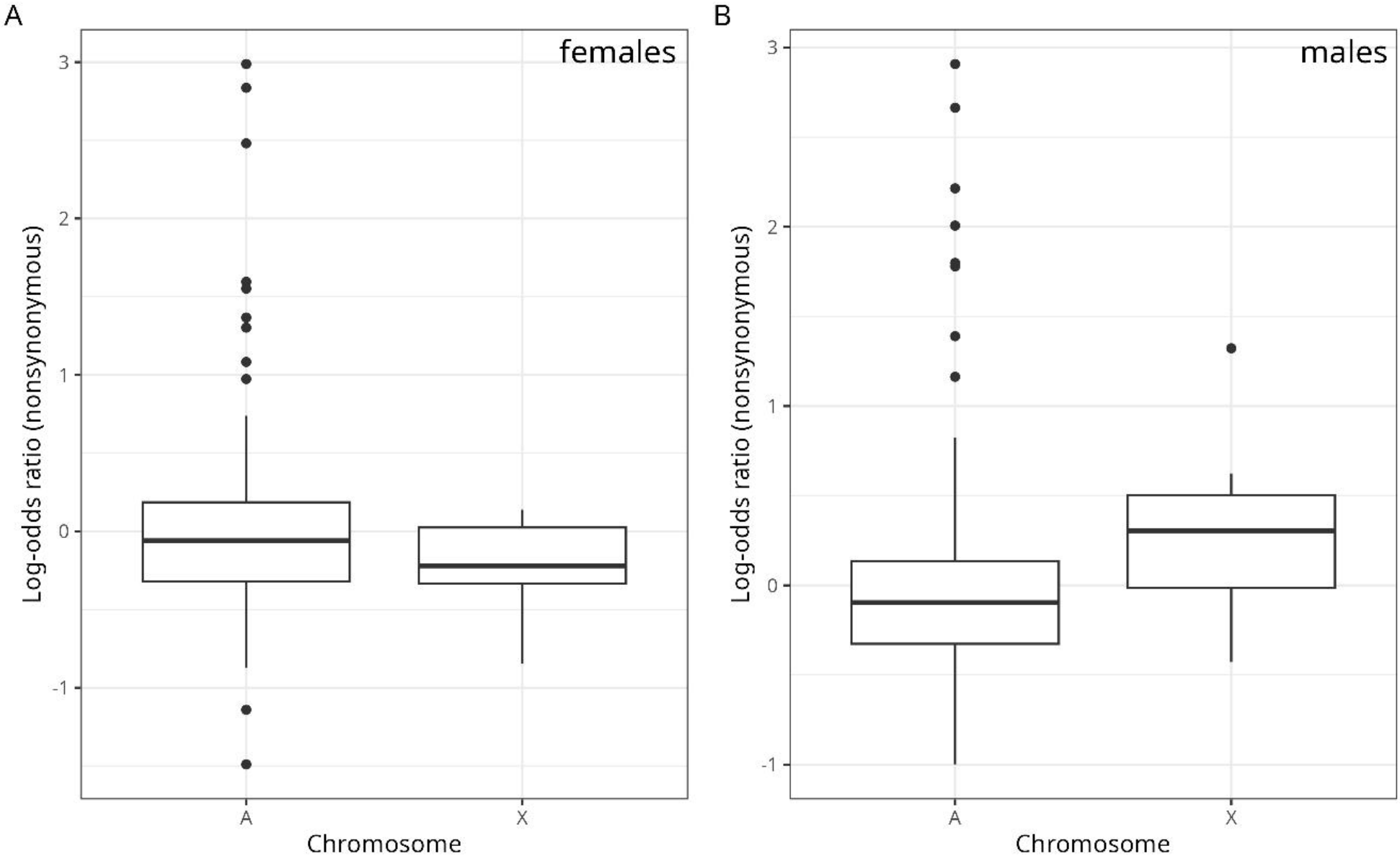
32. Lynch M, Bürger R, Butcher D, Gabriel W. The mutational meltdown in asexual populations. J Hered. 1993;84: 339–344. doi:10.1093/oxfordjournals.jhered.a111354

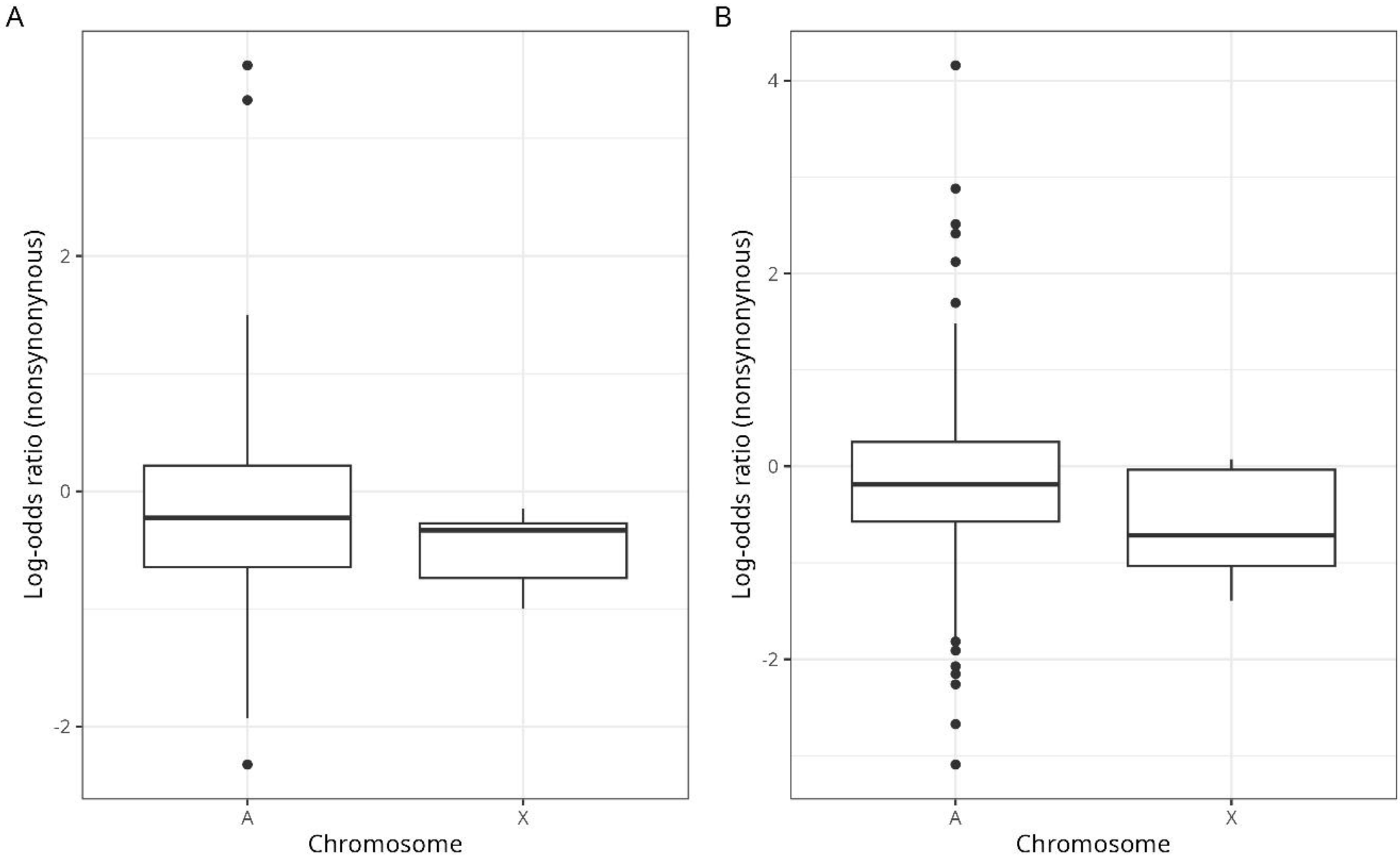
33. Dunford A, Weinstock DM, Savova V, Schumacher SE, Cleary JP, Yoda A, et al. Tumor-suppressor genes that escape from X-inactivation contribute to cancer sex bias. Nat Genet. 2017;49: 10–16. doi:10.1038/ng.3726

1


log2[Nonsense substitutions / # codons]
0
-5
-2
0
2
4
log2[NonSynonymous substitutions / NonSynonymous positions]
Category
POG
TSG


A
females
Log-odds ratio (nonsense)
0
-2
-4
A
X
Chromosome
B
males
Log-odds ratio (nonsense)
2
1
0
-1
-2
-3
A
X
Chromosome


A
females
Log-odds ratio (nonsynonymous)
3
2
1
0
-1
A
X
Chromosome
B
males
Log-odds ratio (nonsynonymous)
3
2
1
0
-1
A
X
Chromosome


A
Log-odds ratio (nonsynonynous)
2
0
-2
A
X
Chromosome
B
Log-odds ratio (nonsynonynous)
4
2
0
-2
A
X
Chromosome