

# Evolution of the pN/pS in tumors: scenarios with different amount of loci under positive and negative selection under the McFarland growth model

Andrés Pérez-Figueroa

25/12/2017

This report documents the results of the run for different scenarios in the simulations using *OncosimulR* (Díaz-Uriarte 2017). Here we used the McFarland et al. (2013) growth model, where the effects of drivers (non-synonymous mutations with positive effect) contribute to the numerator of the birth rate, and those of the deleterious (NS mutations with negative effect) passengers to the denominator as:

$$b = \frac{(1 + s^+)^{n_{drivers}}}{(1 - s^-)^{n_{deleterious}}}$$

For death rate, *oncoSimulR* uses the expression that McFarland et al. (2013) for large cancers (grown to  $10^6$  cells)":

$$d = \log(1 + N/K)$$

where  $N$  here is the current tumor size and  $K$  is the initial equilibrium population size. As the authors explain, for large  $N/K$  the above expression recapitulates Gompertzian dynamics observed experimentally for large tumors. Under the current neutral scenario, this means that population size would be buffered by the changes in the death rate.

We arbitrarily created a genome of 10000 synonymous sites (S). The number of non-synonymous sites (N) used is 27600, to fit the estimation of  $N/S = 2.76$  obtained in human soma based on codon usage and mutational spectrum (Milholland et al. 2017). This yield a total number of 37600 sites. The mutation rate is set in  $2.66 \times 10^{-9}$  as representative of human somatic mutation (Milholland et al. 2017). A starting population of 100000 cells, all with the same genotype, is evolved for 1000 time units, taking samples every 5 time units (so, 201 samples taken). For each sample we obtain both all the variants present in the population or just those above a threshold frequency of 0.05. This simulation setting is replicated 100 times. Different scenarios (Table 1) have different distribution of the N sites regarding its selective effect (positive:  $s=0.1$ ; negative:  $s=-0.1$ ; neutral:  $s=0$ ). The parameters used were chosen by trial and error. We have varied the value for number of nonsynonymous loci with positive selection until finding values that allow us to run the simulations in a reasonable time, with a manageable amount of results and that we would ensure that tumors were obtained that would make the population grow by (at least) an order of magnitude.

Table 1: Parameters used in each scenario. The number of replicates is the obtained after some of them failed

scenario	N positive	N negative	N neutral	replicates
neutral	0	0	27600	100
symmetric	500	500	26600	87
10x negative	500	5000	22100	90
mostly negative	500	27000	100	87
all positive	500	0	27100	89
all negative	0	27000	100	100

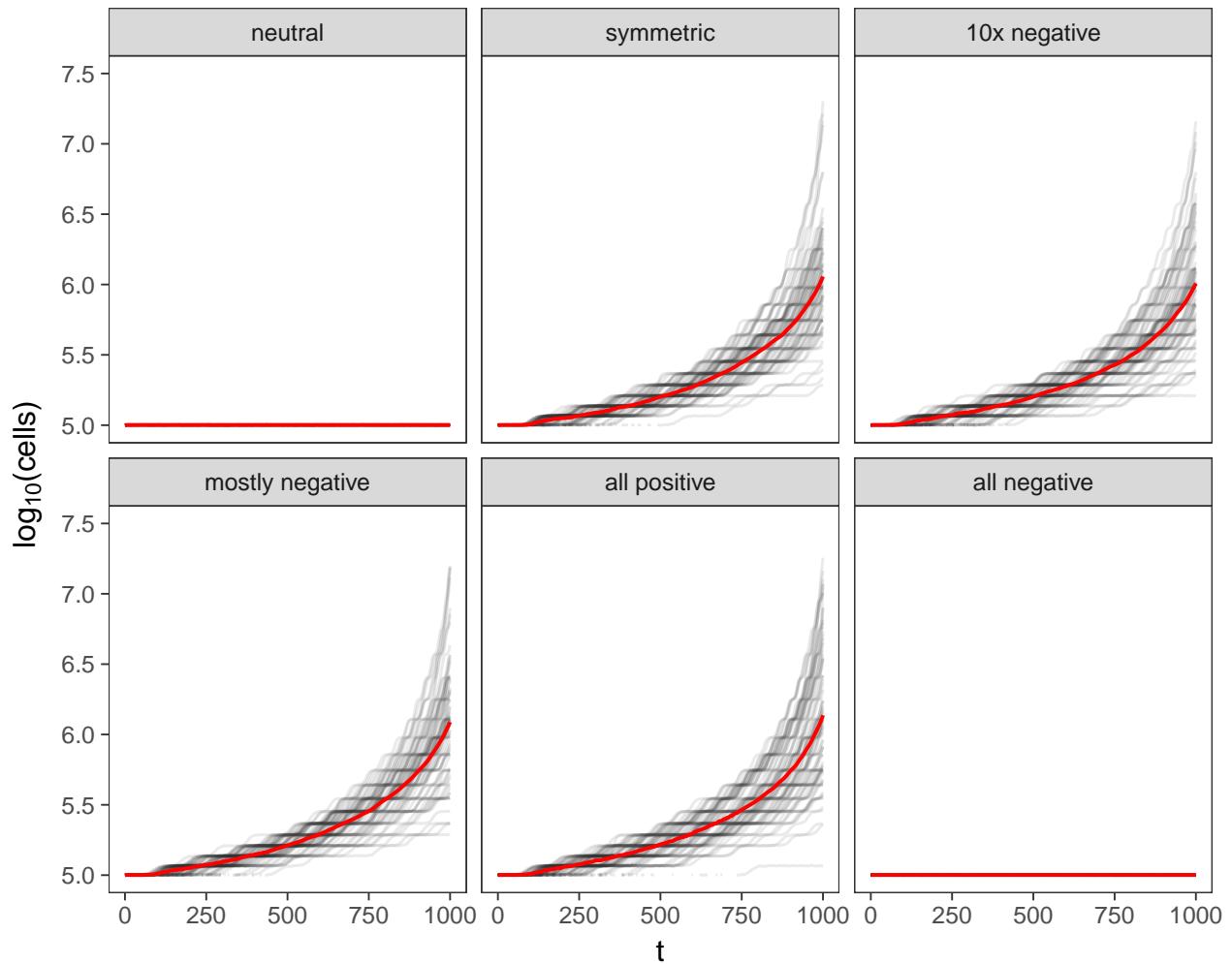


Figure 1: Population growth under the symmetric scenario measured as the  $\log_{10}$ -scaled number of cells by time unit. Grey lines show the individual replicates. Red line represents the mean across replicates

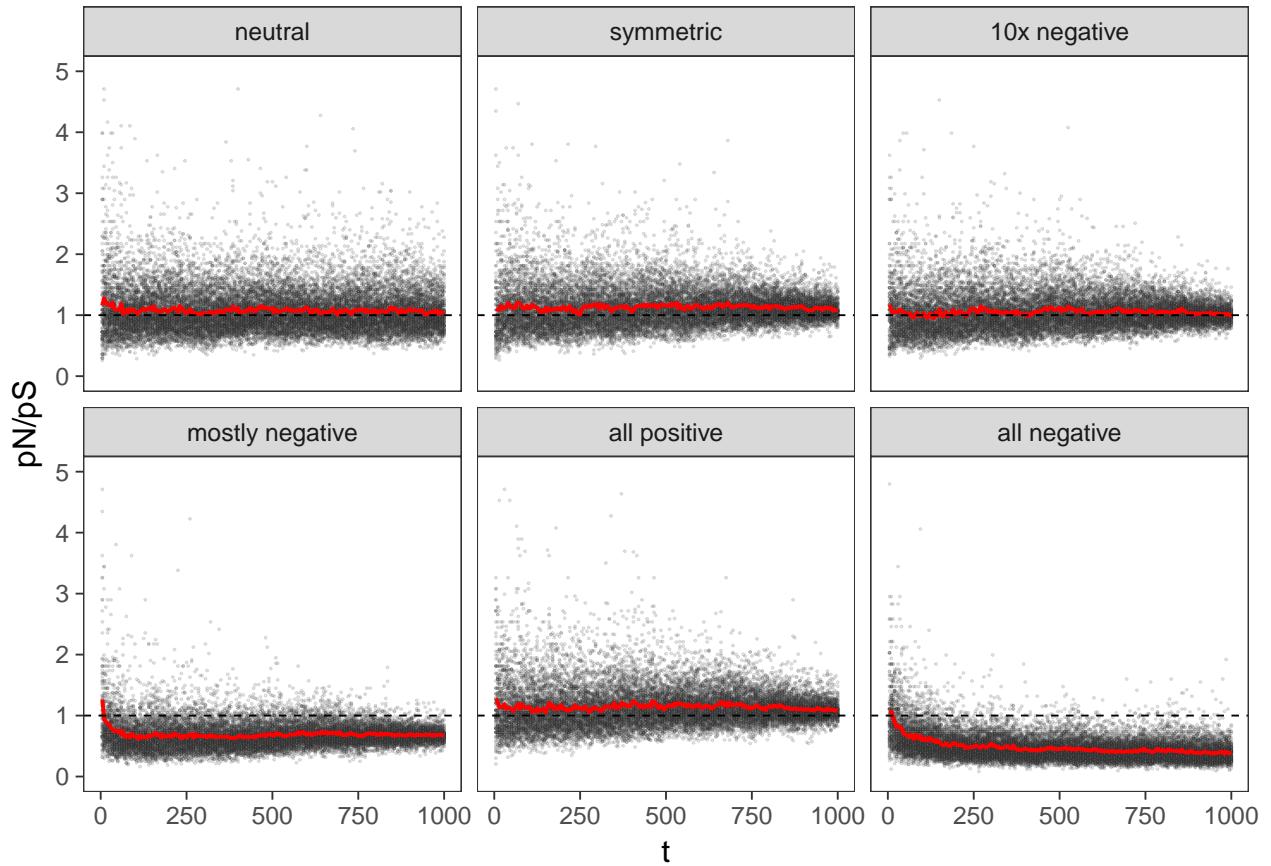


Figure 2: Evolution of the  $pN/pS$  under the symmetric scenario. Grey points show the values for the individual replicates in each sampled time unit. Red line represents the smoothed average across replicates. Dashed black line show the reference value  $pN/pS=1$ .

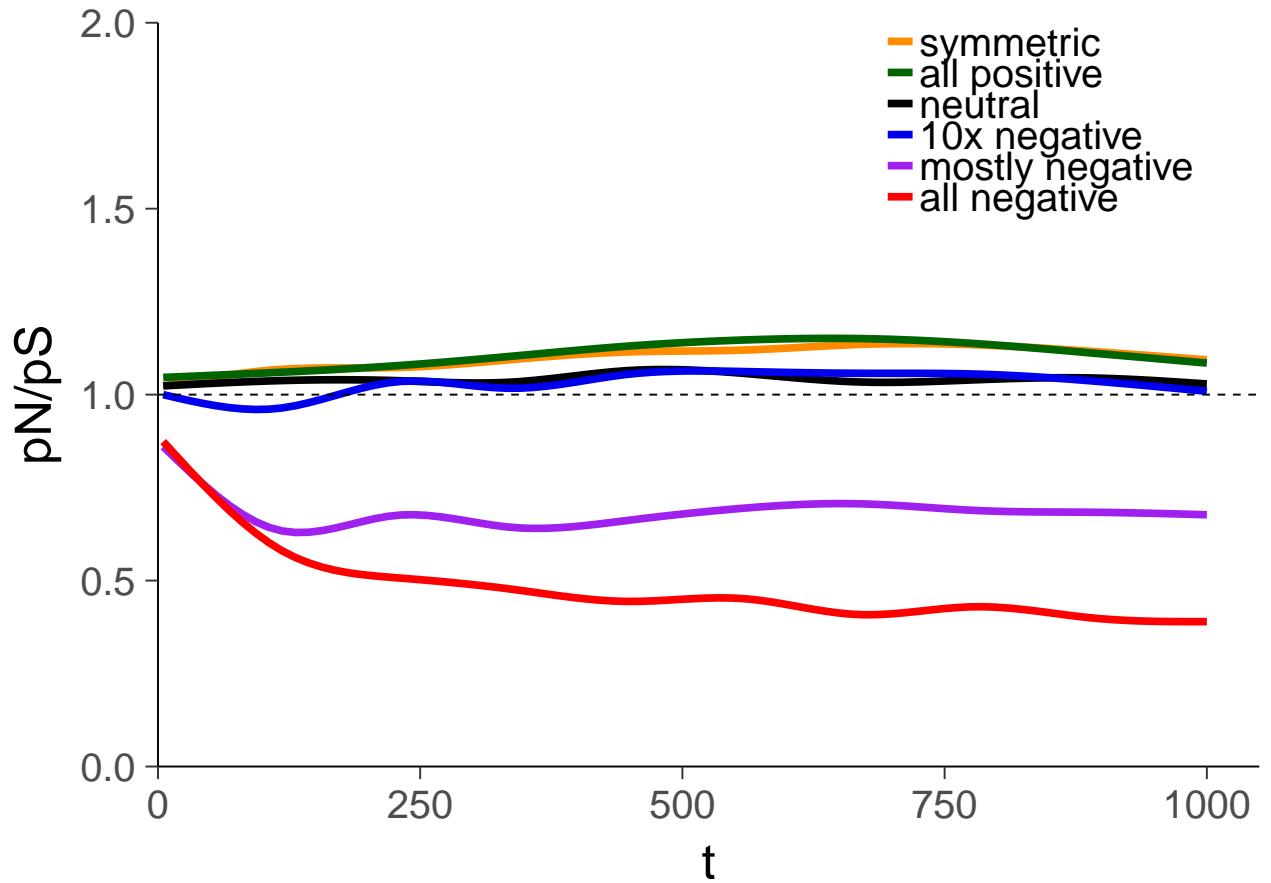


Figure 3: Evolution of the  $pN/pS$  parameter under the different scenarios simulated. Solid lines represent the smoothed average across replicates for each scenario. Dashed black line show the reference value  $pN/pS=1$ .

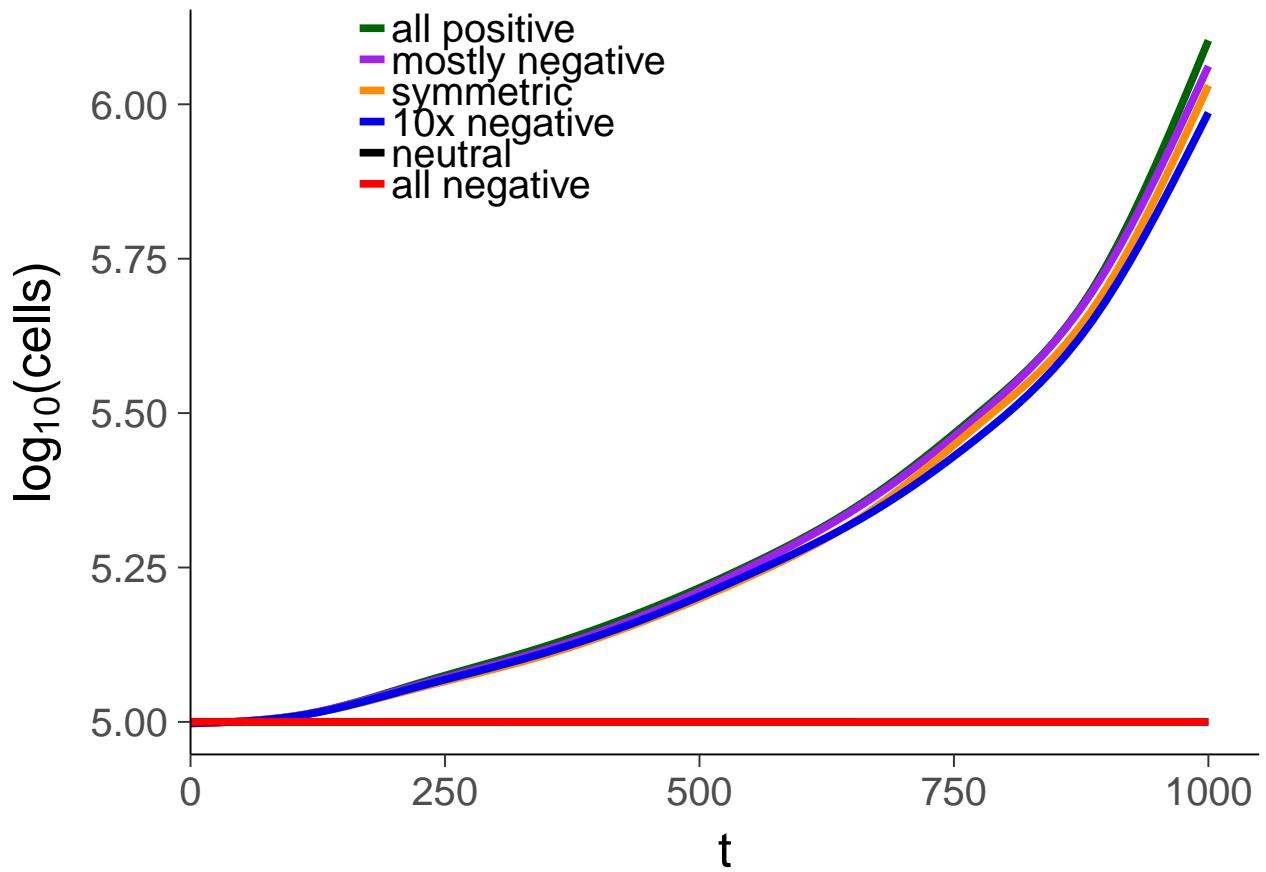


Figure 4: Population growth under the different scenarios measured as the number of cells by time unit. Solid lines represent the smoothed average across replicates for each scenario.

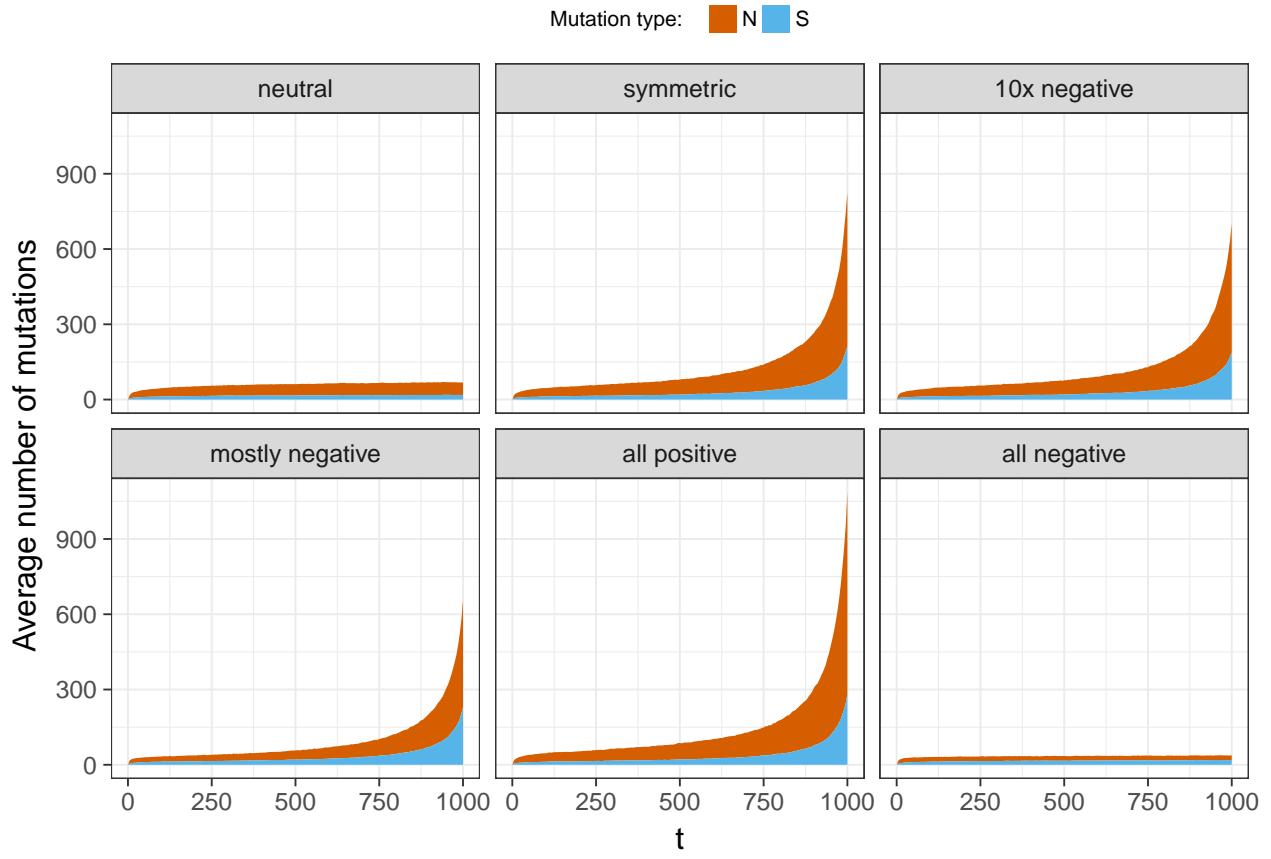


Figure 5: Accumulation of synonymous and nonsynonymous mutations across the time for the different scenarios simulated averaged by replicates. Blue areas represent the number of synonymous mutations, and red areas represent the number of nonsynonymous mutations stacked to the the number of synonymous. The total colored area represent then the total number of mutations

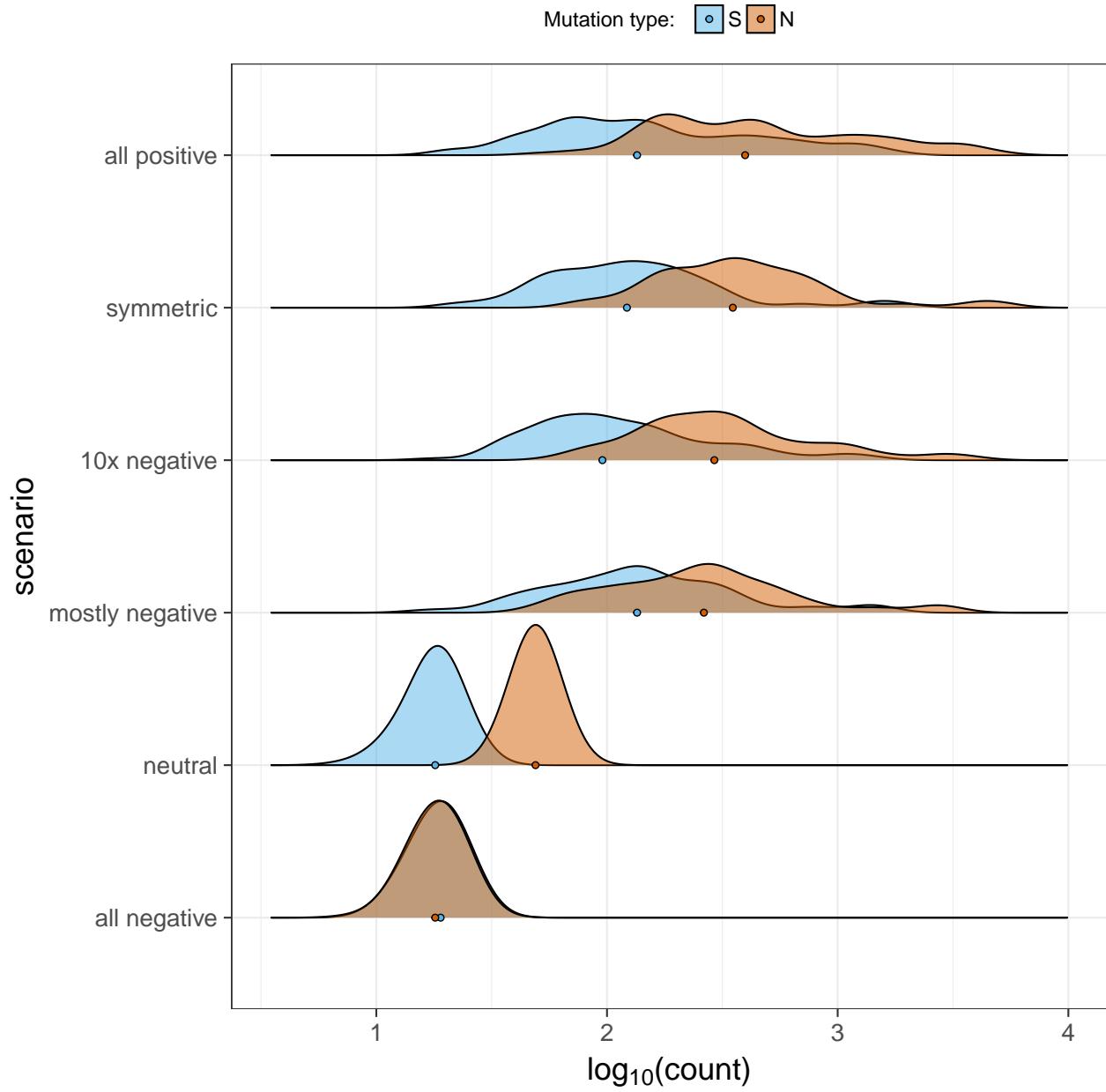


Figure 6: Distribution of the number of synonymous (S) and nonsynonymous (N) mutations across replicates at the end of the simulations under the different scenarios. Colored areas represent the kernel density estimation, with bandwidth = 0.1, for the number of mutations of a specific type. Points represent the mean of the distribution. Red colors stand for nonsynonymous mutations while blue stand for synonymous. Scenarios in the y-axis are ordered by increasing average of nonsynonymous mutations.

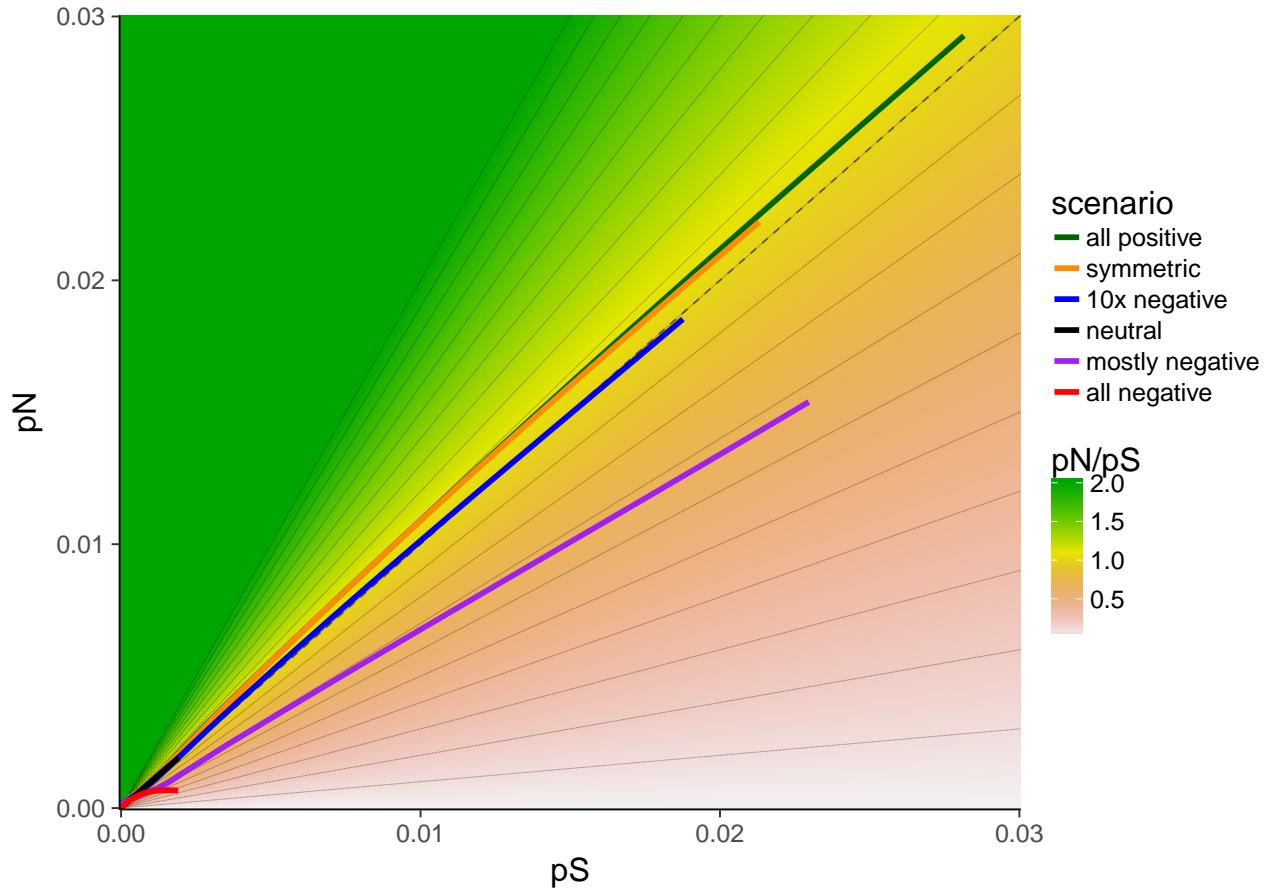


Figure 7: Relationship between the proportion of nonsynonymous mutated sites ( $pN$ ) and the proportion of synonymous mutated sites ( $pS$ ). Background gradient represent the  $pN/pS$  ratio space. Solid lines represent the smoothed regression for  $pN$  over  $pS$  under the different scenarios. Dashed diagonal is the reference for  $pN/pS = 1$ , while the dotted lines represent the reference for  $pN/pS$  values between 0.1 to 2 with 0.1 increase.

## Results

### Effect of the detection threshold and clonal mutations

In general, almost all mutations were found at very low frequency, except for some drivers. Some of these (after visual inspection in some replicas) become clonal (fixed) but generally no more than 1 or 2. Therefore, when sampling those mutations with a frequency higher than 0.05 we find almost no mutations, especially there were no synonymous mutations above the threshold (Table 2). This may be due to the fact that our tumor, given the simulated parameters, is “not very aggressive” and most drivers appear independently, competing with each other.

Table 2: Average number of nonsynonymous and synonymous mutations with frequency above 0.05 on those replicates with at least one synonymous mutation over that threshold

scenario	N	S	replicates
symmetric	12.8	1.0	6
10x negative	15.0	1.0	5
mostly negative	17.8	1.0	4
all positive	13.8	1.2	6

## References

- Diaz-Uriarte, R. (2017). OncoSimulR: Genetic simulation with arbitrary epistasis and mutator genes in asexual populations. *Bioinformatics*, 33(12), 1898–1899. <http://doi.org/10.1093/bioinformatics/btx077>
- McFarland, C. D., Korolev, K. S., Kryukov, G. V., Sunyaev, S. R., & Mirny, L. A. (2013). Impact of deleterious passenger mutations on cancer progression. *Proceedings of the National Academy of Sciences*, 110(8), 2910–2915. <http://doi.org/10.1073/pnas.1213968110>
- Milholland, B., Dong, X., Zhang, L., Hao, X., Suh, Y., & Vijg, J. (2017). Differences between germline and somatic mutation rates in humans and mice. *Nature Communications*, 8(May), 1–8. <http://doi.org/10.1038/ncomms15183>