# Evolution of the pN/pS in tumors: neutral scenario

*Andres Perez-Figueroa*

*20/12/2017*

This report documents the results of the run for the **neutral case** in the simulations using *OncoSimulR* (Díaz-Uriarte 2017). The objective of this scenario is to present a null case and test the simulator. Here, all mutations have not effect on the fitness (birth rate) of the carrying cells. Under this scenario we expect that net population growth would be 0 and that pN/pS would be around 1, at least when the system reaches some equilibrium.

## Description of the model

*OncoSimulR* can use an use exponential growth model or a model with carrying capacity that follows McFarland et al. (2013). For the exponential growth model, the death rate is fixed at one whereas in the model with carrying capacity death rate increases with population size.

### Exponential model

The default model in *OncoSimulR*. The death rate is fixed to $d = 1$ and the birth rate is

$$b = \prod (1 + s_i)$$

being $s_i$ the selective coefficient for each mutation. Given that in this scenario $s_i = 0$ for all the loci then $b = 1$. However, this does not mean that the population size remains constant over time. There is stochasticity in the birth-death process and when the initial population size is not very large and we start from the wild-type, it is not uncommon for simulations to become extinct (or growing too far).

### McFarland (Logistic) model

Following the model of McFarland et al. (2013), 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:

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

However, **in this scenario all the N (non-synonymous) mutations have no effect**, all the $s$ are 0, and **then birth rate is fixed to 1**.

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.

### Simulation settings

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 (Milhollland et al. 2017). This yield a total number of 37600 sites. In this scenario all the N sites are neutral. 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 10000 cells, all with the same genotype, is evolved for 50000 time units, taking samples every 500 time units. This simulation setting is replicated 1000 times.

Table 1: Parameter values used in this simulation

| Parameter | Value |
|---|---|
| Total loci (sites) | 37600 |
| .N loci | 27600 |
| ..N loci (drivers) | 0 |
| ..N loci (deleterious) | 0 |
| ..N loci (neutral) | 27600 |
| .S loci | 10000 |
| mutation rate | 2.66e-09 |
| sampling interval | 500 t.u. |
| Final time | 50000 t.u. |
| Initial size | 10000 cells |
| replicates | 1000 |

## Results

### Evolution of the population size

As expected, under the McFarland model the population size (number of cells) remains quite stable around the initial size of 10000 cells (Figure 1). This size is the equilibrium value and if the population size increases then the death rate also increases, and viceversa. This helped to keep the population size very stable within a very limited range of ~ 9500-10500 cells.

With **the exponential model**, however, the stochasticity was very high and several replicates went extinct and others grew up to one magnitude order (Figure 2). With this model, a neutral population could behave as "tumor" by growing unlimited.

### Evolution of the pN/pS

When considering all the variants already present in the population, the pN/pS quickly tends to 1 (Figure 3). When we only measure those variants above a (detection) threshold, the pN/pS is usually below 1 during long time. This is particularly true for a threshold of 5%, the usual frequency for a variant be detected by NGS.

Figure 4 show the distribution of the pN/pS under the different sampling regimes in three time points. The biggest differences are very early in the evolution of the population but the distributions are getting closer after ¿long? time.

Figure 5 shows the absolute number of mutations detected under the different sampling regimes and their proportion into N and S. Here is the explanation for the differences in the distributions of pN/pS. The overall number of mutations detected is very low at early times, and for different sampling regimes are obviously lower. This produces some noise in the pN/pS as in early times there are replicates without N (pN/pS=0) or without S (pN/pS=NA). In this context, where there are on average ~12 mutations segregating in the population, the calculation of pN/pS is very noisy.

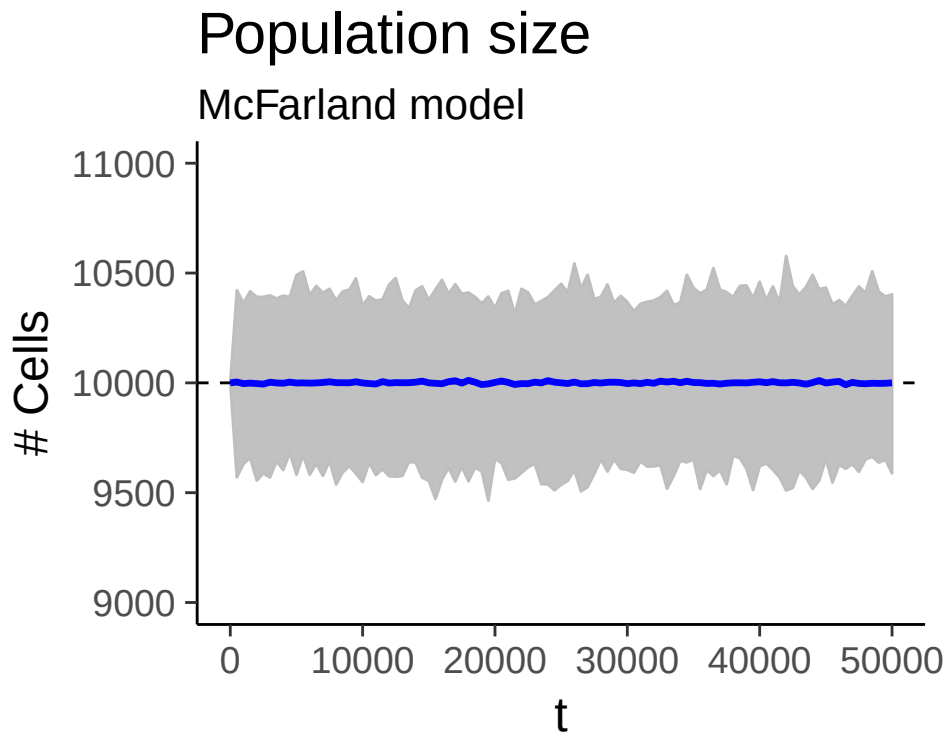


Figure 1: Population size across time under the neutral scenario with the MacFarland model. Gray ribbon represents the range individual values for each of the 1000 replicates. Blue line shows the mean number of cells. Dashed black line shows the equilibrium size.

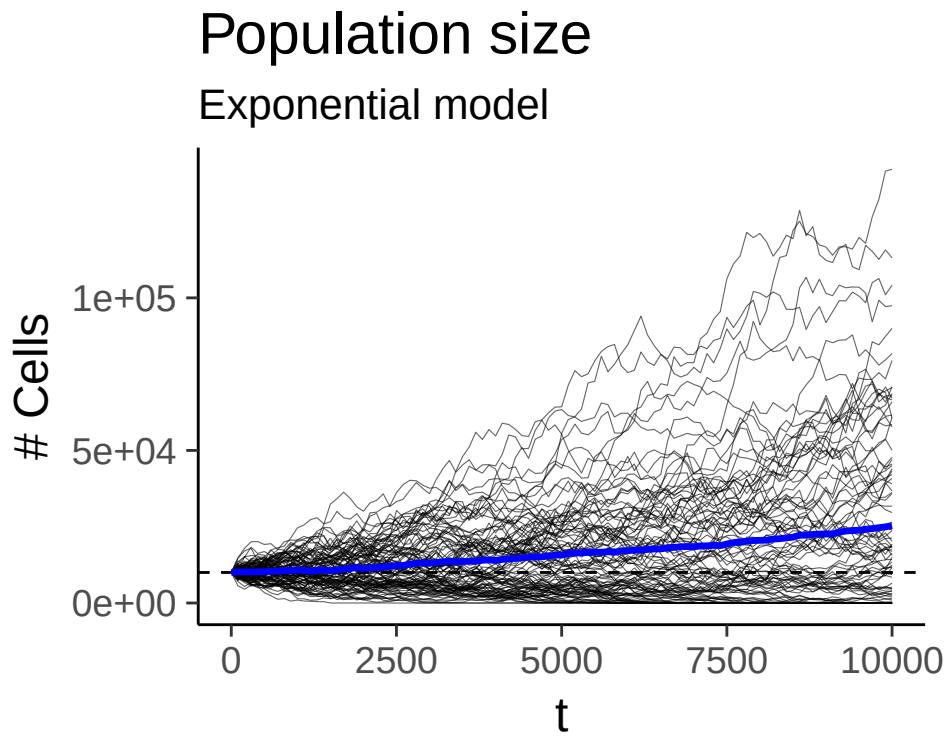


Figure 2: Population size across time under the neutral scenario with the Exponential model. Gray lines represent different replicates (only 100). Blue line shows the mean number of cells in all replicates. Dashed black line shows the initial size.

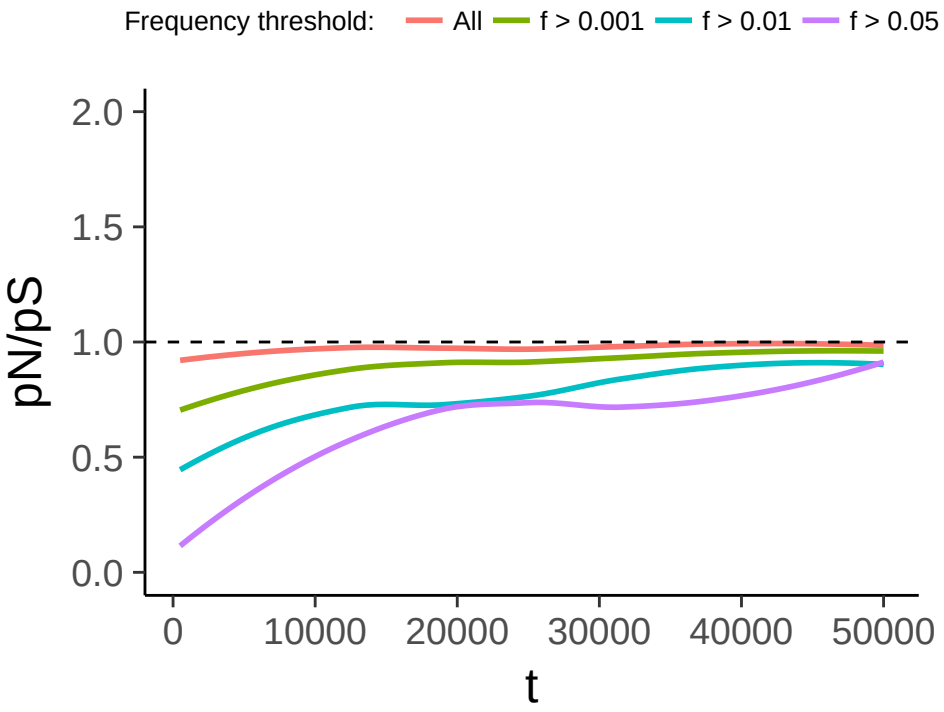


Figure 3: pN/pS replicates across time under the neutral case. Each line represents the loess regression of the median for 1000 replicates by sampling those variants with frequencies above the threshold.

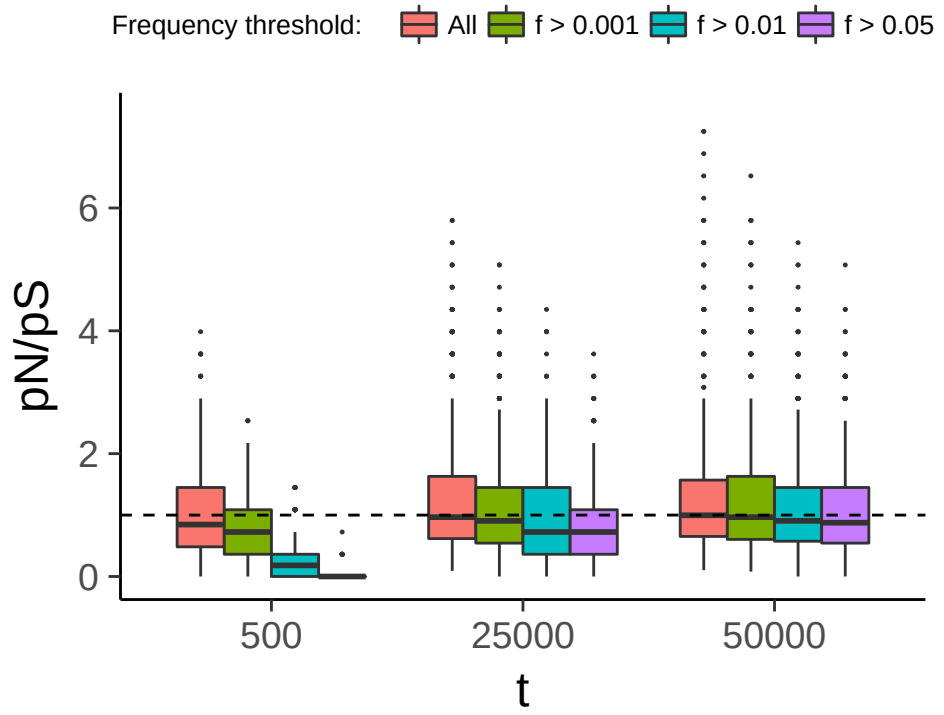


Figure 4: Distribution of pN/pS at different time points

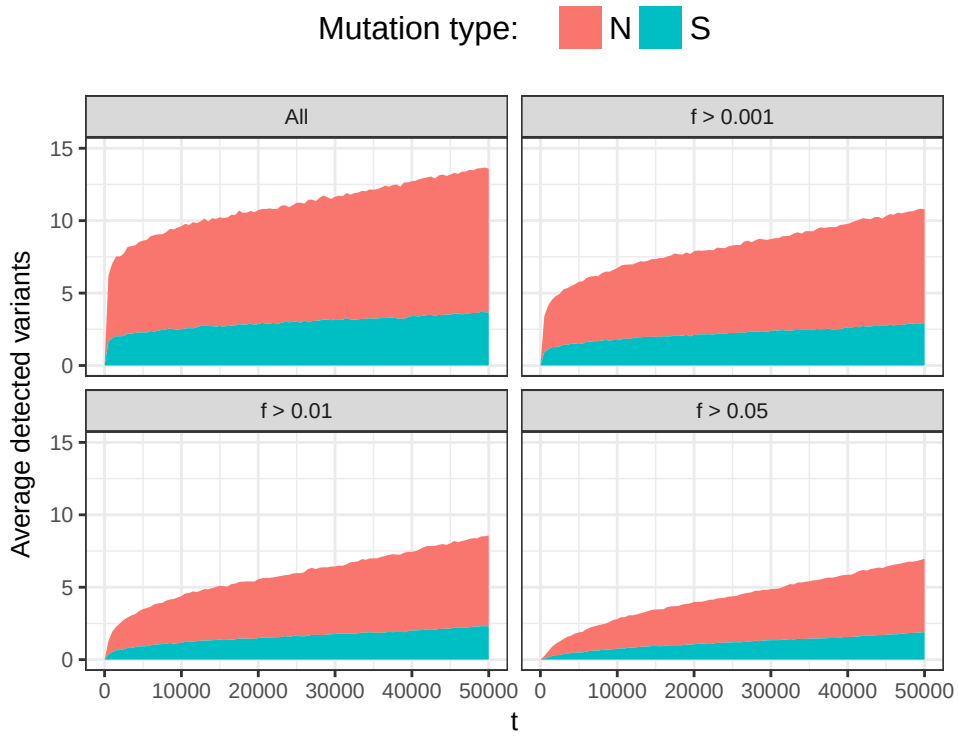


Figure 5: Average number of detected variants across the time. Colors represent the proportion of N and S mutations.

## 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