Competitive Selection: to do

“Clonal interference” appears to be the term for competing clones limiting each other’s fitness.

Most of the existing theory seems to be on equilibria though (i.e. fixation, etc.), whereas in our system we are clearly not in equilibrium.

# Data

* Mitchell
  + 10 patients aged 0-81
  + ~350 sequenced colonies per patient
  + whole genome per colony
  + single timepoint per patient
* Fabre
  + 385 patients, 55-93
  + Targeted sequencing of 56 genes (17 selected genes) from peripheral blood
  + Up to 5 timepoints per patient

## Simulations

* Implement “non-competing” system, where stem cell pool can grow in size.

## Quantities to investigate

* Clonal hematopoiesis prevalence (def?) with age
* Average clone size (in clones >1-2%?) with age
* Average number of clones >1% with age
* Fraction of simulations where a fixation occurs within 1 lifetime (i.e. can be considered cancer)
* Proportion of compartment covered by expanded (>1-2%) clones
* Slope of fitness decrease
* Distribution of estimated fitness values?

# Theoretical

* Maximum number of clones (>1%) present at age in variable parameter space
* Shannon diversity index
* Does variance of (inferred) clone fitness decrease with age? (possibly needs to be checked with multi-fitness simulations
* Are near-neutral fitness clones small in size? (I.e. more likely to simply be neutral)
* Check whether small-size-low-growth-rate clones could have reasonably expanded through neutral drift.
* Create plot of all fitness values over time, across multiple simulations (parallel data plot)

# Random thoughts, observations, ideas, etc.

* Using exponential distribution, even with low fitness, clones fit enough to take over the entire compartment within a single lifetime seem to occur too often.
  + Incorporating double hits could potentially circumvent this problem: complete fitness distribution becomes much steeper
  + Could potentially use distribution with double hits directly? Probably less realistic, but easier to calculate with. Though if running simulations, doesn’t really matter too much I guess…

# Figures

1. Conceptual figure?
2. Data + simulations?
   1. Data: logistic fitness +