Most cancers carry a substantial deleterious load due to Hill-Robertson interference

Susanne Tilk, Christina Curtis, Dmitri A Petrov, Christopher D McFarland

Declan Bennett

National University of Ireland, galway

d.bennett1@nuigalway.ie

October 24, 2019

IRISH RESEARCH COUNCIL
An Chomhairle um Thaighde in Éirinn

Somatic mutant clones colonize the human esophagus with age

Martincorena I.[1] Fowler JC.[1] *et al.* [1 2 3 4 5]

[1]Wellcome Sanger Institute, Hinxton, Cambridge, UK [2]MRC Unit Cancer unit, Hutchinson-MRC Research centre, University of Cambridge, UK [3]Dept Surgery and Cambridge NIHR Biomedical Research Centre, Biomedical Campus, University of Cambridge, Cambridge, UK [4]Department of Biochemistry, University of Oxford, South Parks Road, Oxford, UK [5]Department of Haematology, University of Cambridge, Cambridge, UK

Declan Bennett

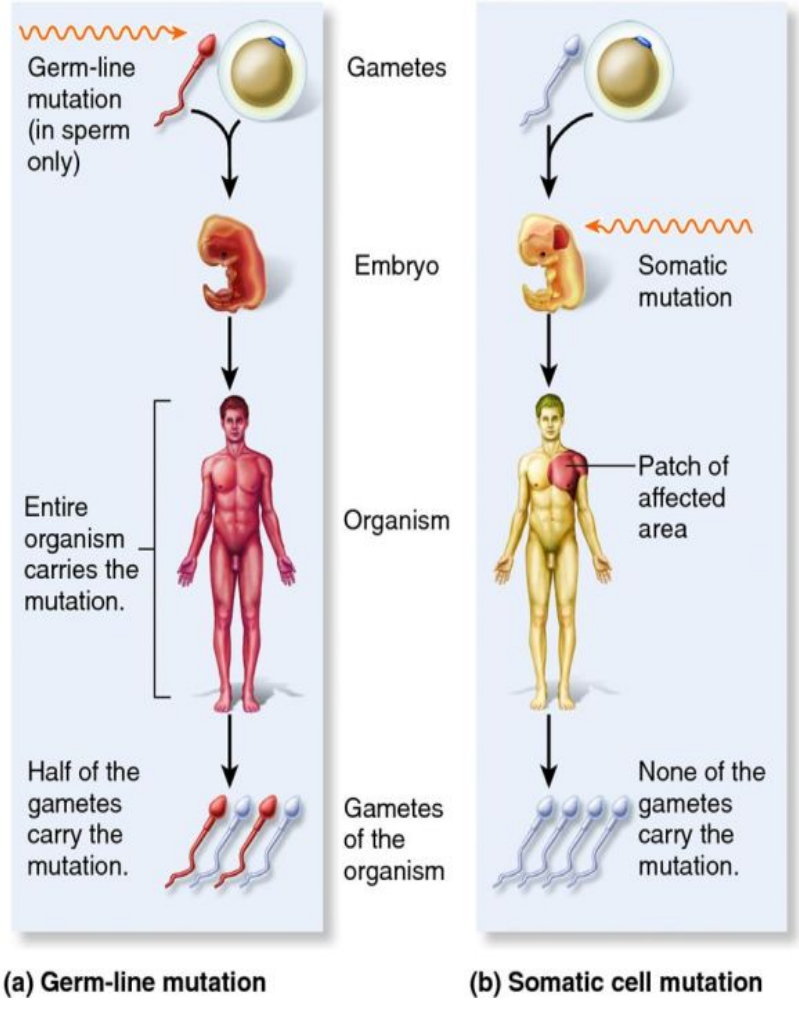
National University of Ireland, Galway, Ireland

November 8, 2018

IRISH RESEARCH COUNCIL
An Chomhairle um Thaighde in Éirinn

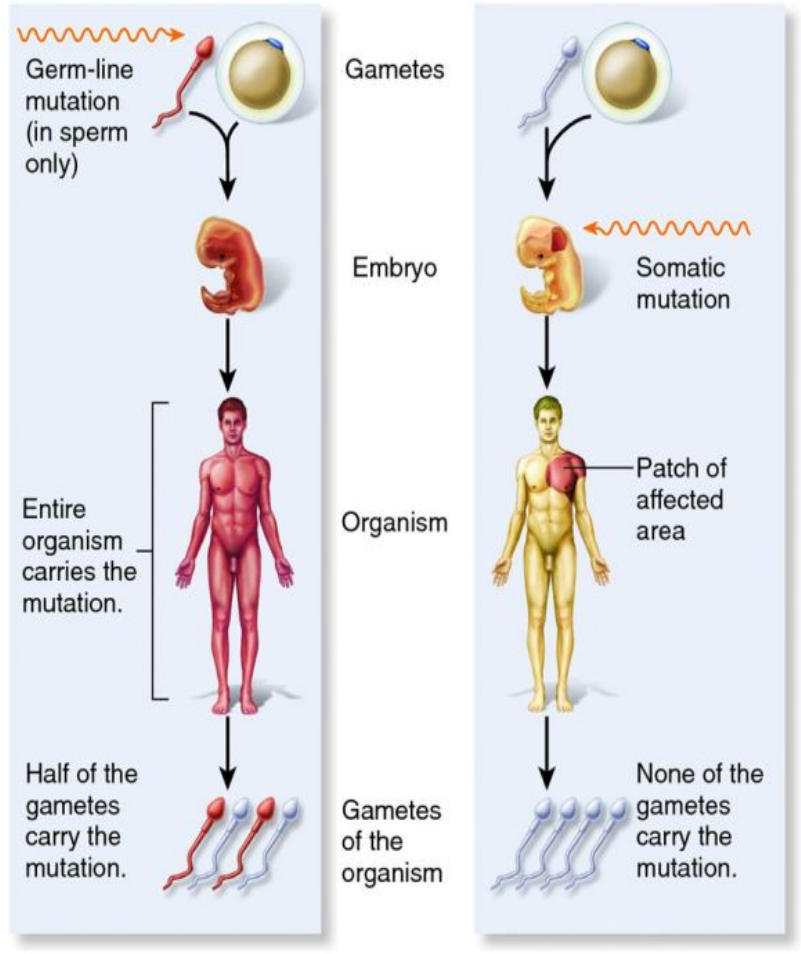
Overview

Somatic Mutation



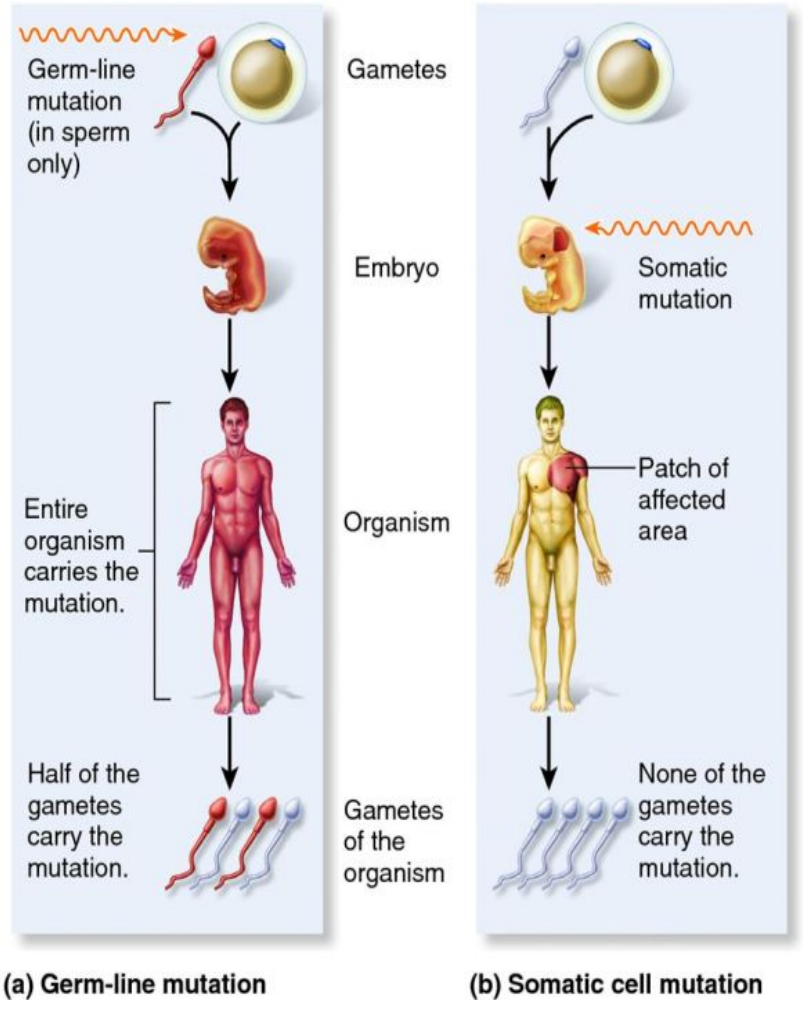
Somatic Mutation

- A somatic mutation is any mutation that occurs in non-germline cells and has no recombination



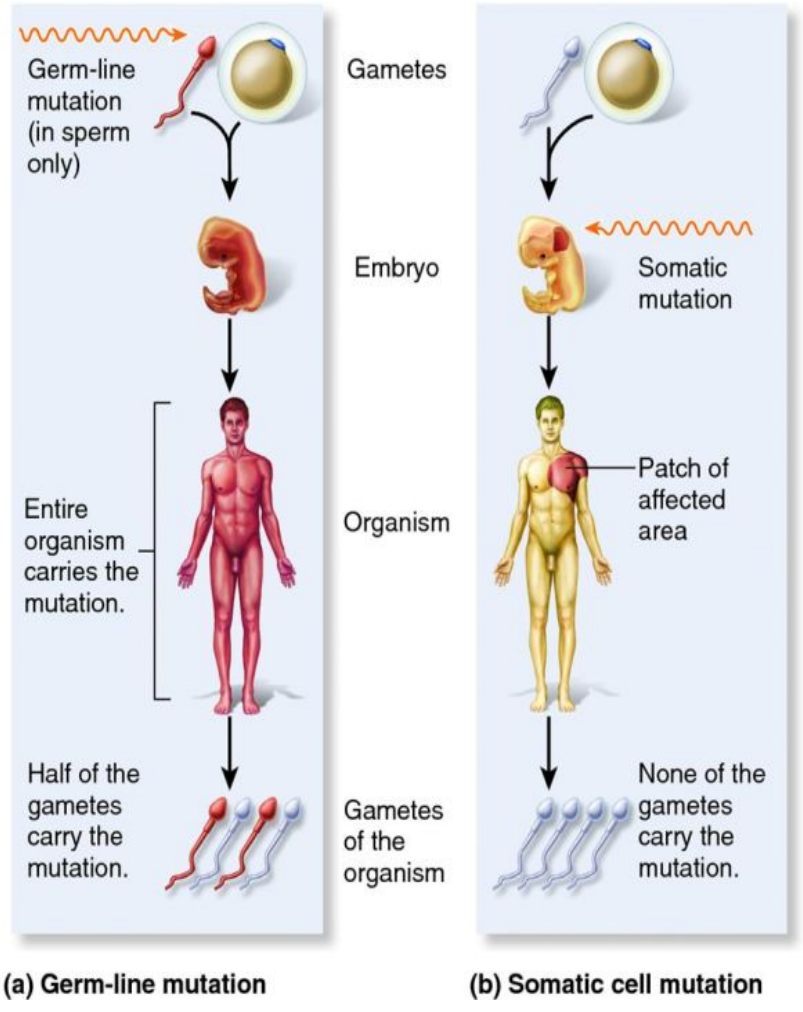
Somatic Mutation

- A somatic mutation is any mutation that occurs in non-germline cells and has no recombination
- Typically occur at a much higher rate than germline mutations



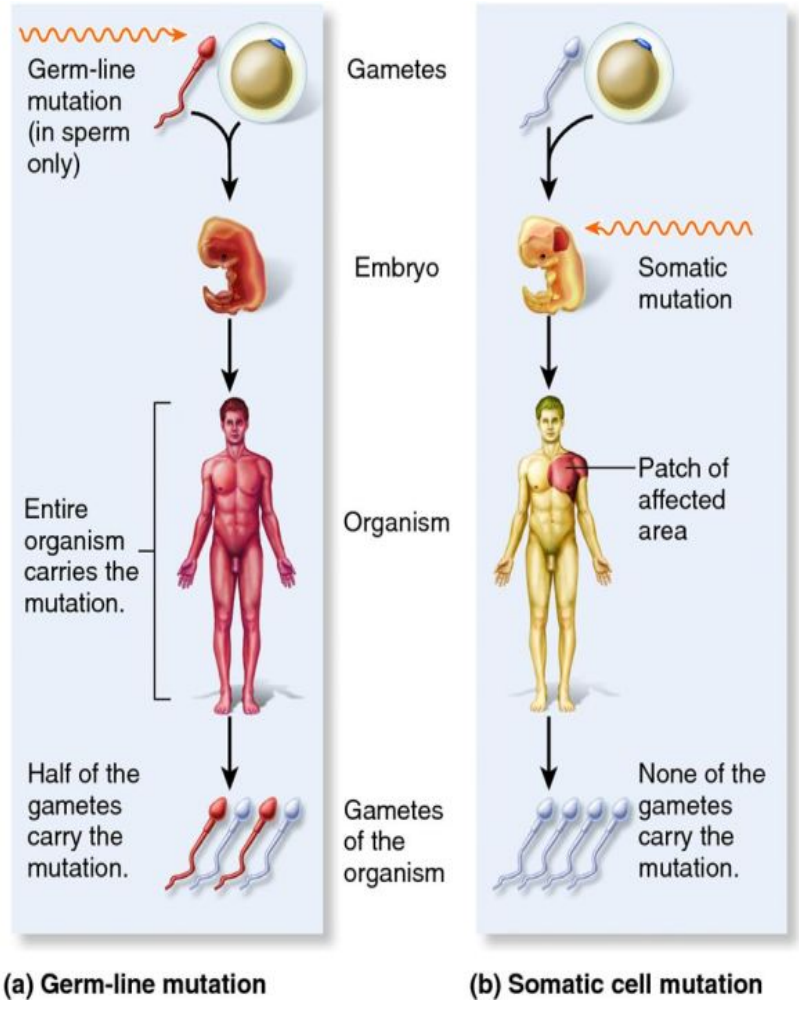
Somatic Mutation

- A somatic mutation is any mutation that occurs in non-germline cells and has no recombination
- Typically occur at a much higher rate than germline mutations
- Germline mutation rate of $\sim 3.3 \times 10^{-11}$ per bp per division

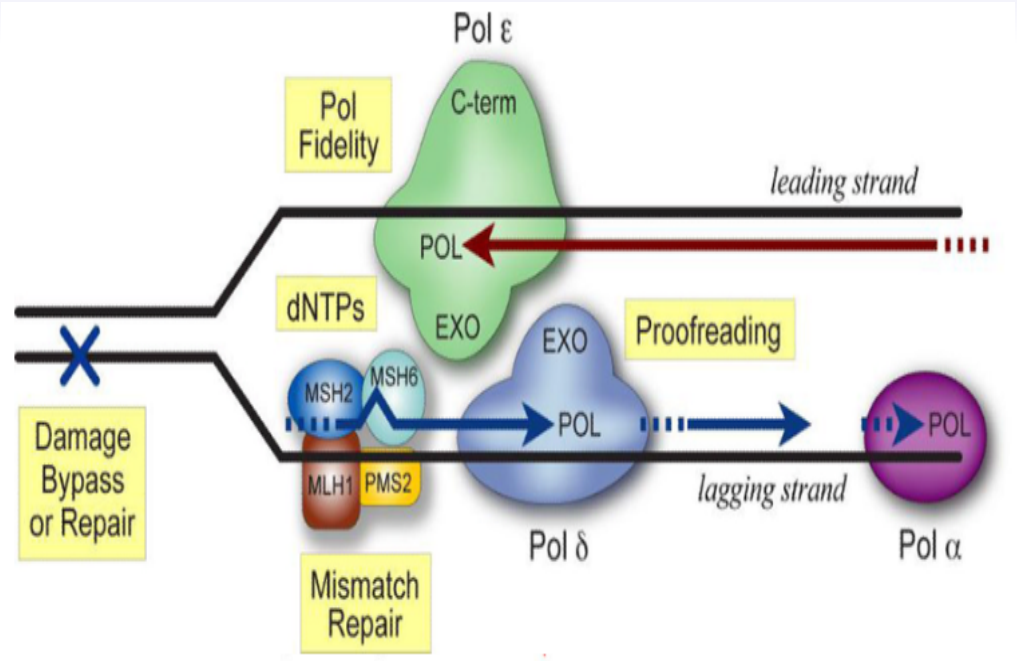


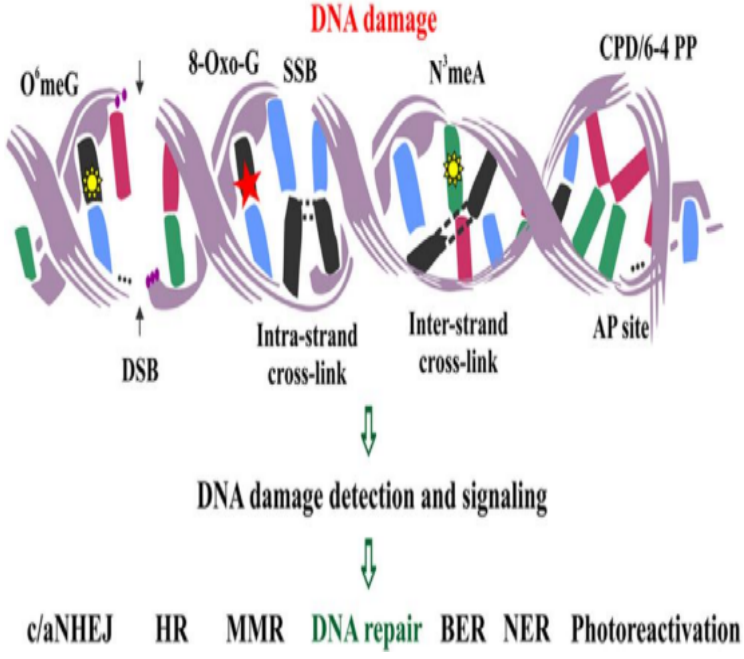
Somatic Mutation

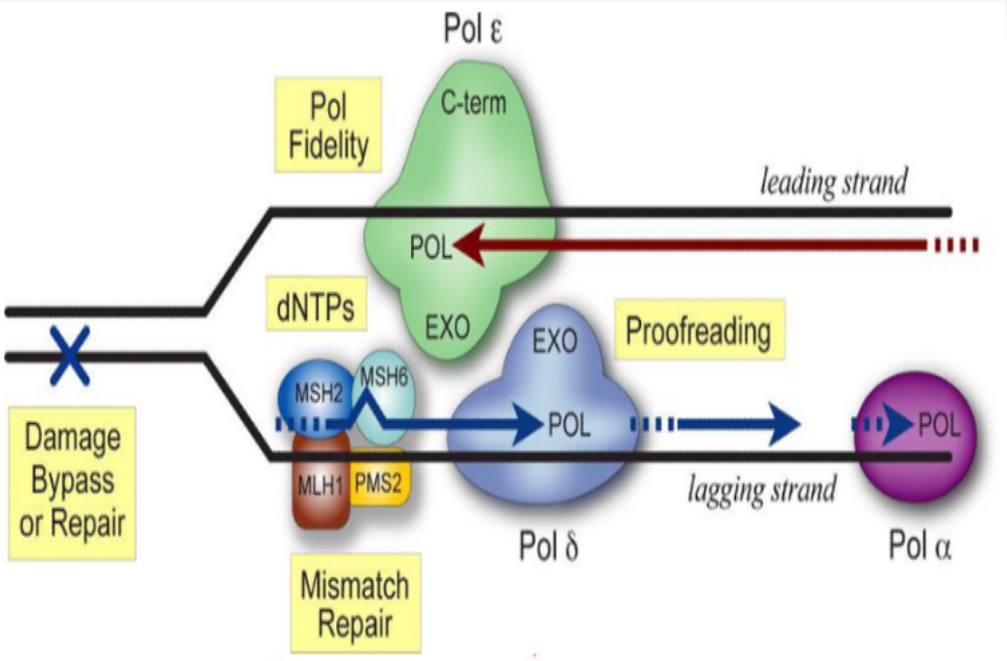
- A somatic mutation is any mutation that occurs in non-germline cells and has no recombination
- Typically occur at a much higher rate than germline mutations
- Germline mutation rate of $\sim 3.3 \times 10^{-11}$ per bp per division
- Somatic mutation rate of $\sim 2.8 \times 10^{-7}$ per bp per division

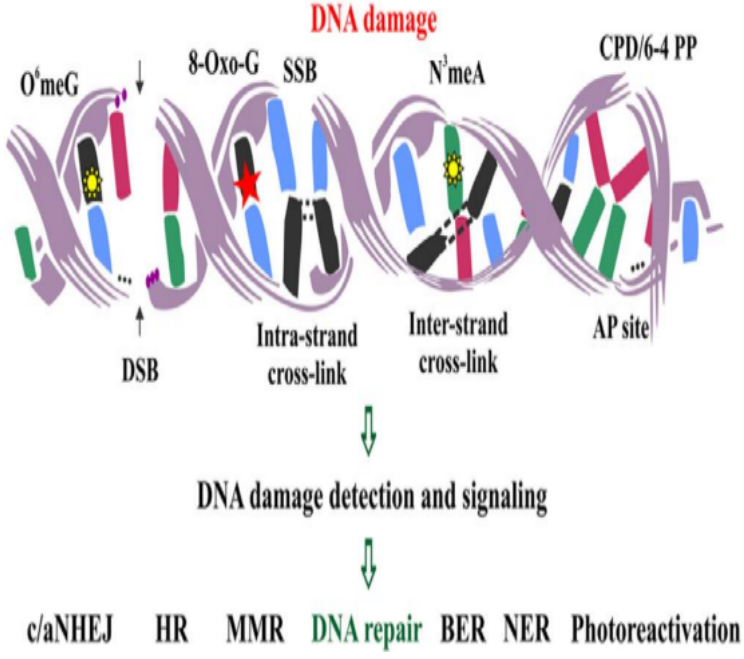


Milholland et al. 2017

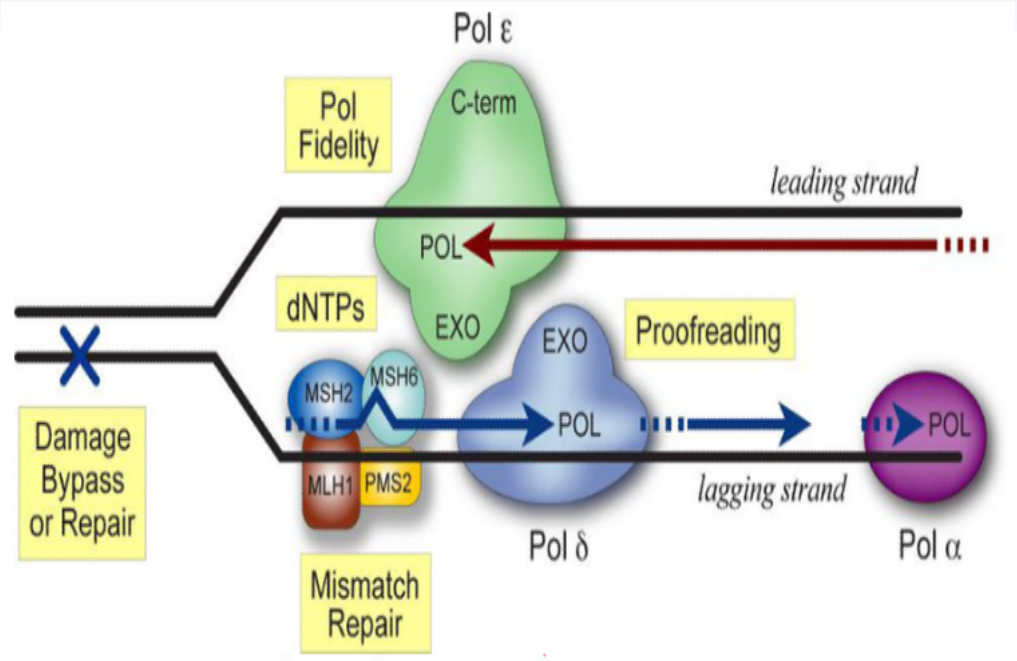

Pol ε
Pol Fidelity
C-term
leading strand
POL
EXO
dNTPs
EXO
Proofreading
MSH2
MSH6
POL
POL
Damage Bypass or Repair
MLH1
PMS2
lagging strand
Pol δ
Pol α
Mismatch Repair

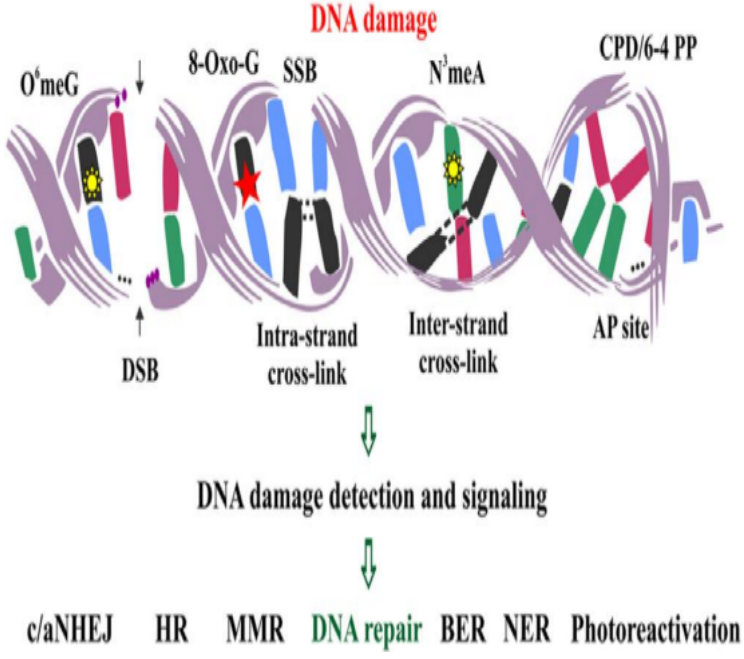

DNA damage
O⁶meG
8-Oxo-G
SSB
N³meA
CPD/6-4 PP
DSB
Intra-strand cross-link
Inter-strand cross-link
AP site
DNA damage detection and signaling
c/aNHEJ
HR
MMR
DNA repair
BER
NER
Photoreactivation



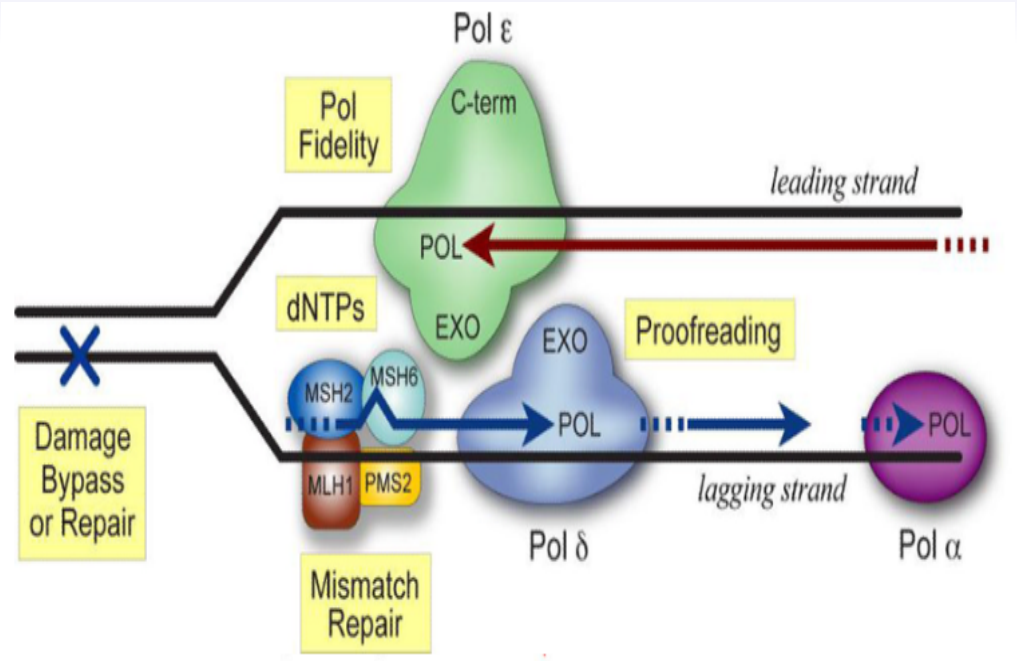


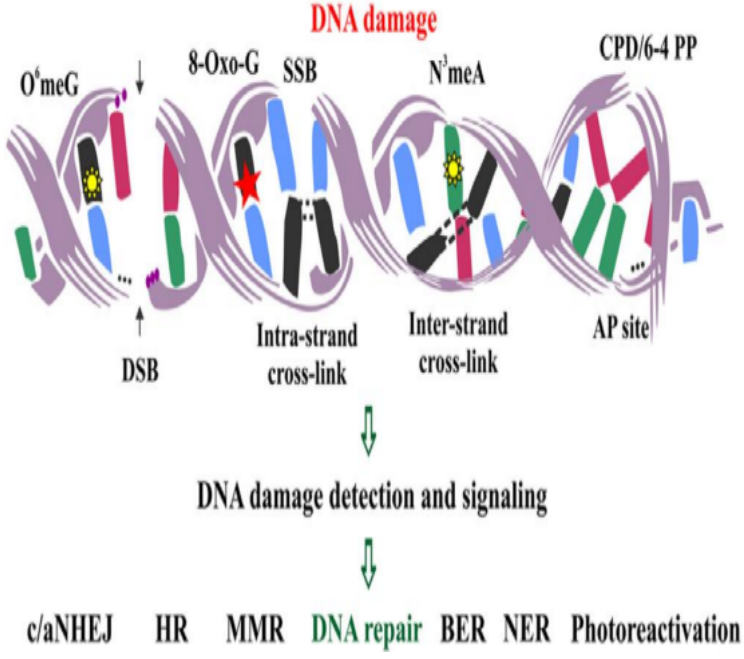
- Polymerase error is the main contributor





- Polymerase error is the main contributor
- Environmental factors also contribute to DNA damage



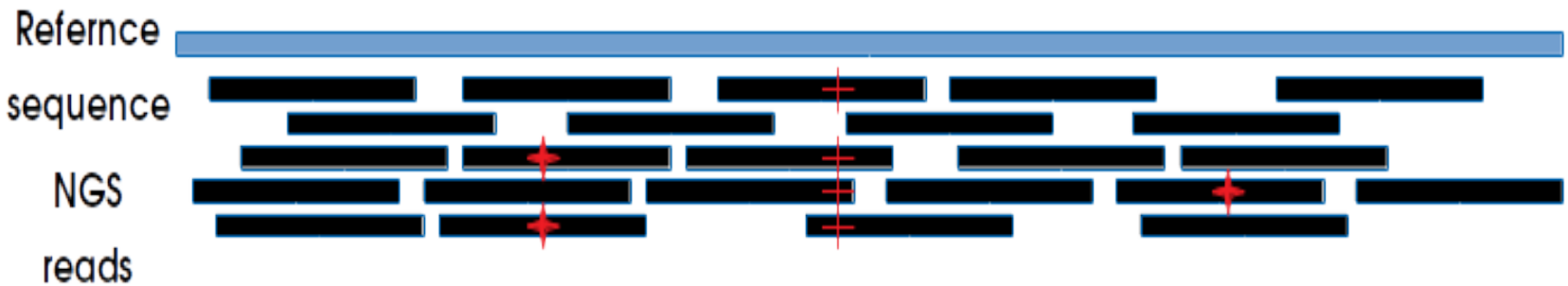


- Polymerase error is the main contributor
- Environmental factors also contribute to DNA damage
- Other intrinsic processes such as Apobec deamination, also drive mutatagenesis

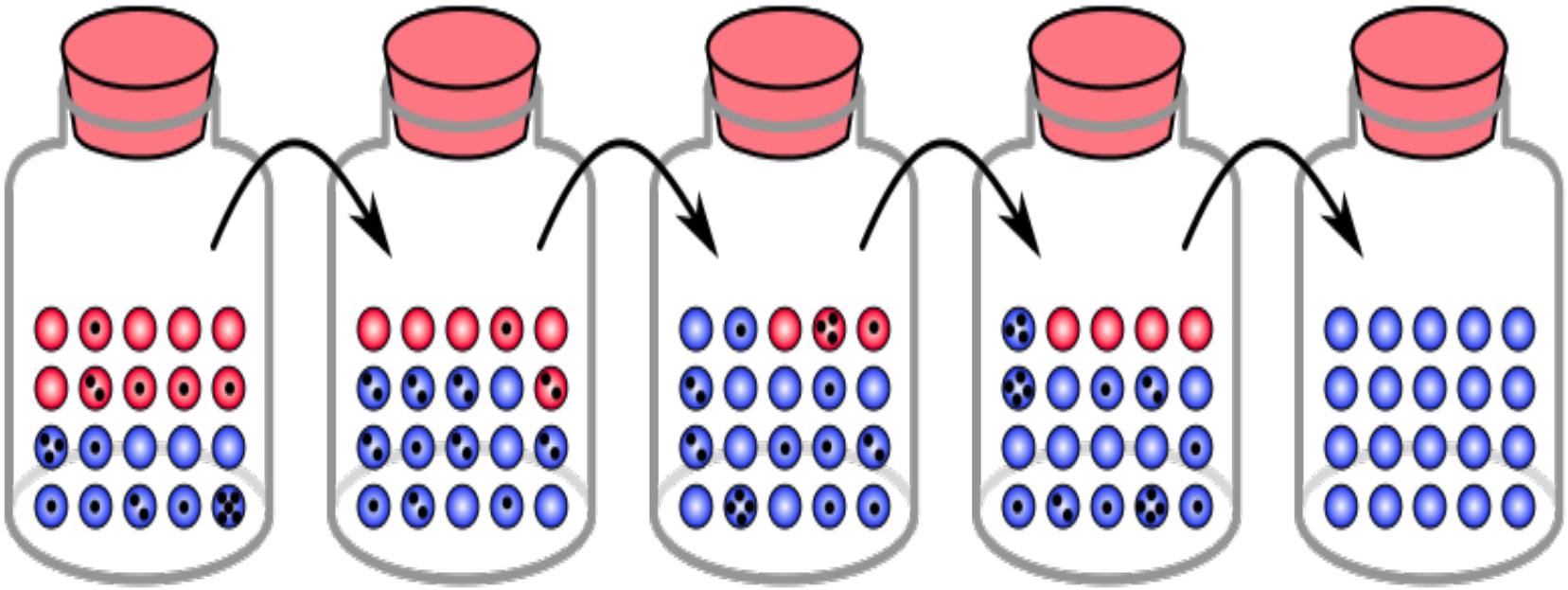
Variant allelic fraction

Given sufficient read depth clonality can be estimated from

$$VAF = \frac{\text{Number of reads with mutated loci}}{\text{Total number of reads covering the mutated loci}}$$



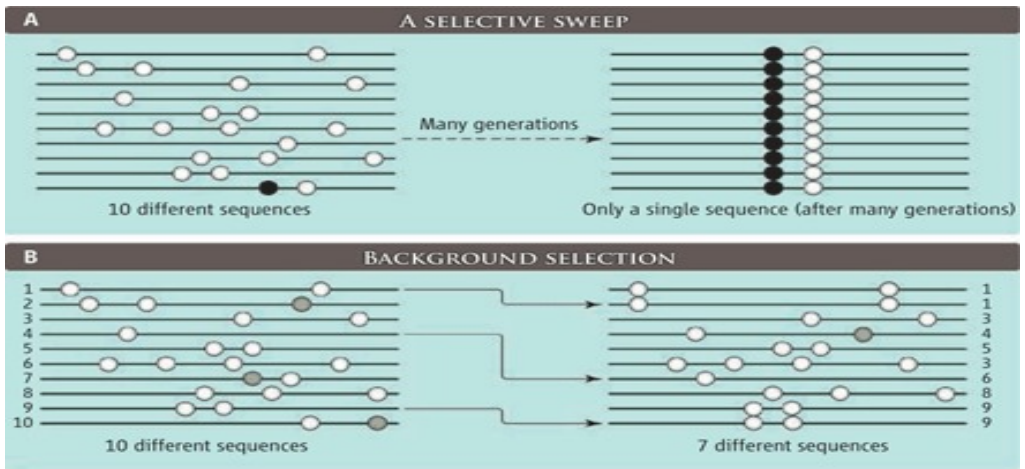
Genetic Drift

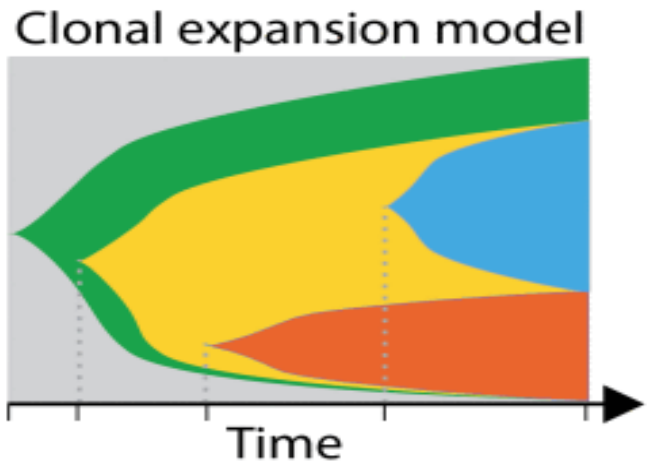


$$\binom{N}{k} \cdot p^k q^{N-k}$$

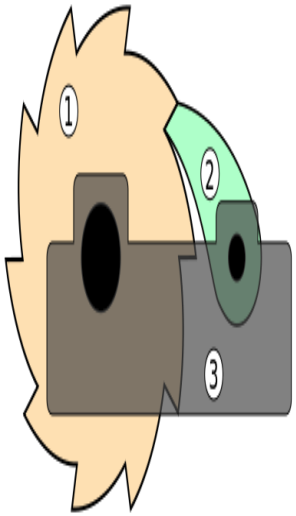
Genetic Draft

- AKA genetic hitchiking or selective sweep





Hill-Robertson Interference



Selection

Assumption

In coding regions, somatic mutations affect the three nucleotide codon resulting in either, no change to the amino acid, encoding a new a new amino acid or signalling for premature termination of translation. These can be grouped into 2 categories; synonymous, d_s and non-synonymous, d_n

We assume that there are no selective pressures acting on synonymous mutations. We can then infer the presence of positive,negative or neutral selection using the ratio of non-synonymous to synonymous mutations

$$\omega = \frac{d_n}{d_s}$$

$$\omega \begin{cases} > 1 & \text{loci is under positive selection} \\ = 1 & \text{loci is evolving neutrally} \\ < 1 & \text{loci is under negative selection} \end{cases} \quad (1)$$

The expected number of nonsynonymous mutations, $\mathrm{E}[d_N]$, is given by:

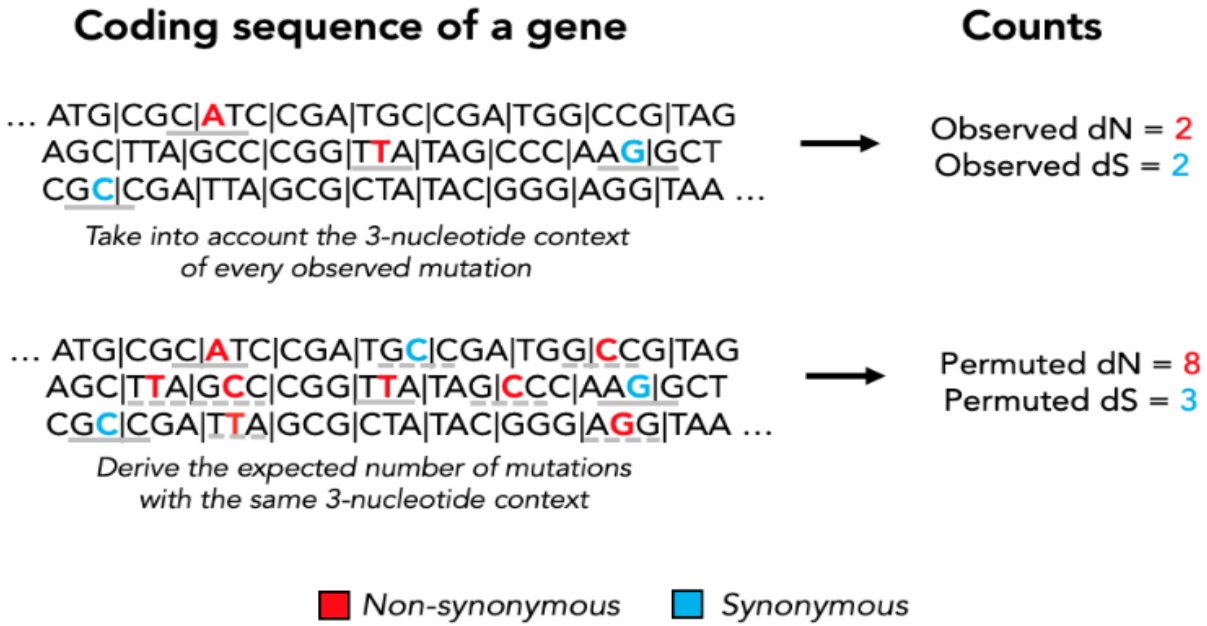
$$\mathrm{E}[d_N] = \omega_{gt} \sum_i M_{igt} N_{ig}$$

The expected number of synonymous mutations, $\mathrm{E}[d_S]$, is given by:

$$\mathrm{E}[d_S] = \sum_i M_{igt} S_{ig}$$

Permutation

We need to normalise the observed number of mutations



Nonparametric null model of selection

Given:

$$\frac{E\left[d_N\right]}{E\left[d_S\right]} = \frac{\omega_{gt}\sum_i M_{igt}N_{ig}}{\sum_i M_{igt}S_{ig}} = \omega_{gt}\frac{< M_{igt}, N_{igt} >}{< M_{igt}, S_{igt} >} = \omega_{gt}\frac{\rho_{MN}\left\|M_{gt}\right\|\left\|N_{gt}\right\|}{\rho_{MS}\left\|M_{gt}\right\|\left\|S_{gt}\right\|} = \omega_{gt}\frac{\rho_{MN}\left\|N_{gt}\right\|}{\rho_{MS}\left\|S_{gt}\right\|}$$

And..

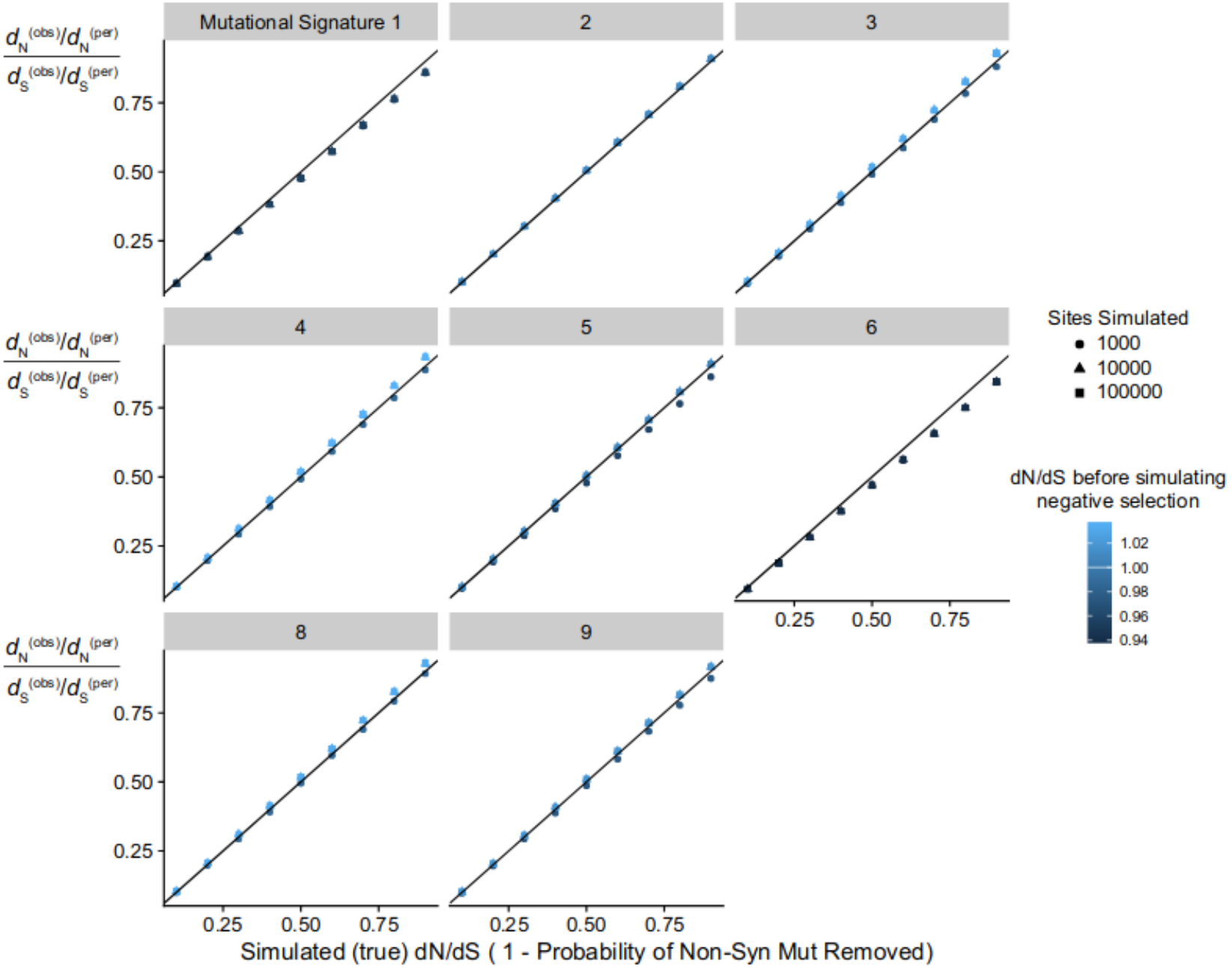
$$d_N^{permuted} = \sum_i \left(d_N^{\text{observed}}\ N_i + d_S^{\text{observed}}\ N_i\right) \tag{2}$$

then..

$$\frac{E\left[d_N^{\text{permuted}}\right]}{E\left[d_S^{\text{permulted}}\right]} = \frac{\sum_i\left(\omega_{gt}M_{igt}N_{igt}^2 + M_{igt}N_{igt}S_{igt}\right)}{\sum_i\left(\omega_{gt}M_{igt}N_{igt}S_{igt} + M_{igt}S_{igt}^2\right)} = \frac{\omega_{gt}\rho_{MN}\|M_{gt}\|\|N_{gt}\|^2 + \rho_{MN}\|M_{gt}\|\|N_{gt}\|\|S_{gt}\|}{\omega_{gt}\rho_{MS}\|M_{gt}\|\|S_{gt}\|\|N_{gt}\| + \rho_{MS}\|M_{gt}\|\|S_{gt}\|^2} = \frac{\rho_{MN}\|N_{gt}\|}{\rho_{MS}\|S_{gt}\|} \tag{3}$$

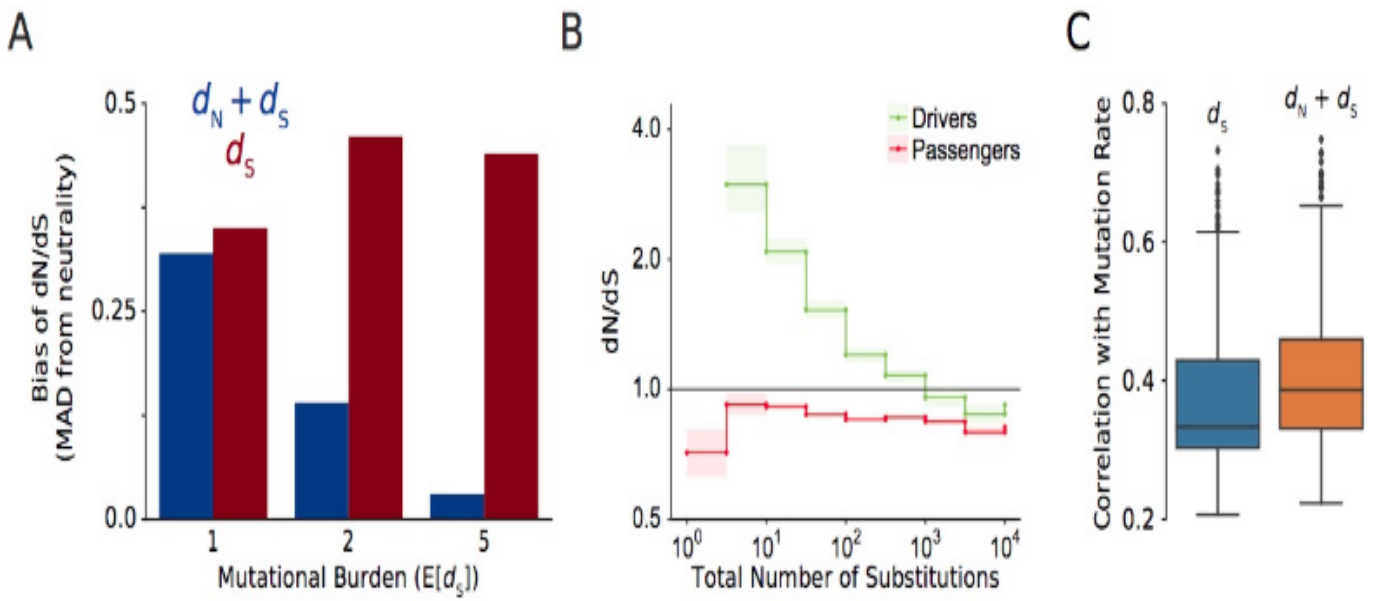
following on..

$$\frac{dN}{dS} = \frac{d_N^{(\text{observed})}/d_N^{(\text{permuted})}}{d_S^{(\text{observed})}/d_S^{(\text{permuted})}} \tag{4}$$


Mutational Signature 1
2
3
4
5
6
8
9
$\frac{d_N^{(obs)}/d_N^{(per)}}{d_S^{(obs)}/d_S^{(per)}}$
0.25
0.50
0.75
Simulated (true) dN/dS (1 - Probability of Non-Syn Mut Removed)
Sites Simulated
1000
10000
100000
dN/dS before simulating negative selection
1.02
1.00
0.98
0.96
0.94

Mutational Burden calculation

The total number of substitutions, $d_n + d_s$, is used to calculate mutational burden. Here however, $d_n + d_s$ is also used to calculate ω and may bias the relationship between selection and mutation rate



Estimating selection on CNAs

$$\frac{dE}{dI}_{i,M} = \frac{\sum_m^M \sum_g^G T_{i,g} C_{m,g}}{\sum_g^G T_{i,g}} \tag{5}$$

where dE is the fraction of CNA overlapping exonic regions, dI is the fraction of CNA overlapping intronic/intergenic regions, i is the gene set, T is genomic track, C is the length of the CNA and m is the mutational burden.

Tilk et al. hypothesis

Tilk et al. hypothesis

- Negative selection is largely absent in cancer

Tilk et al. hypothesis

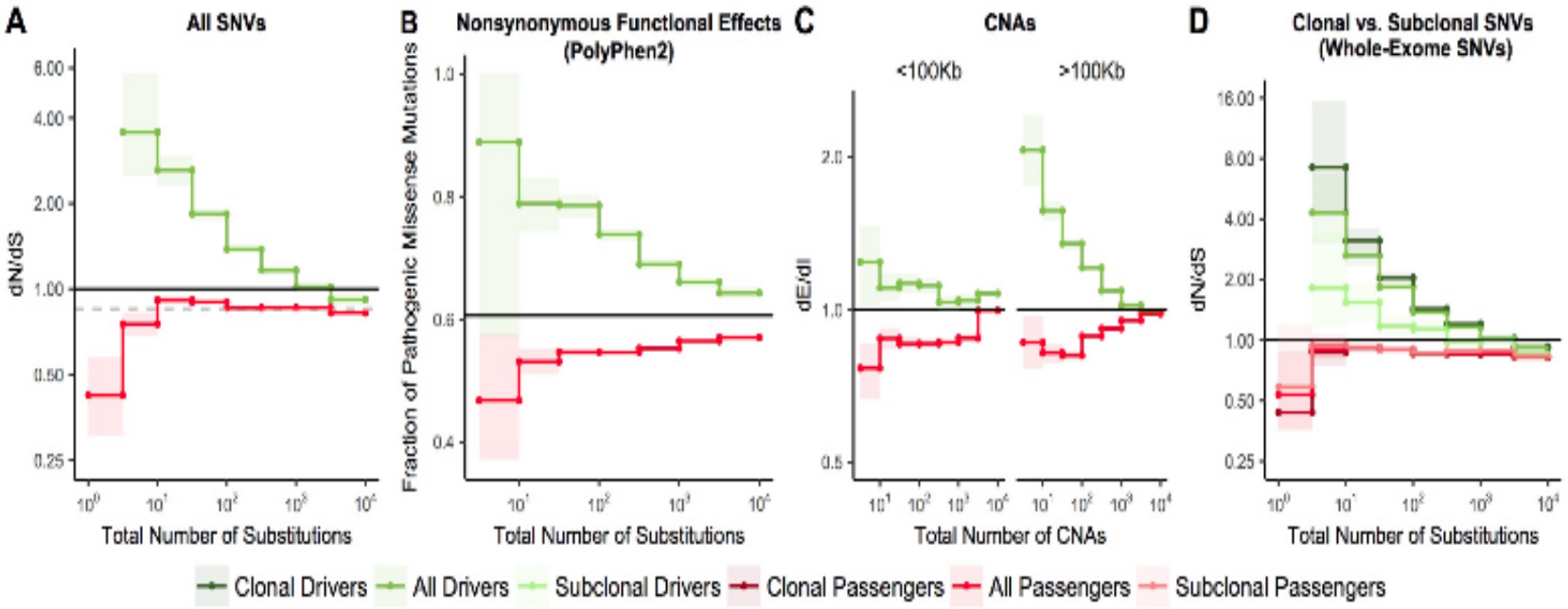
- Negative selection is largely absent in cancer
- As cancer evolves asexually under linkage constraints mutations are unable to be removed via selection

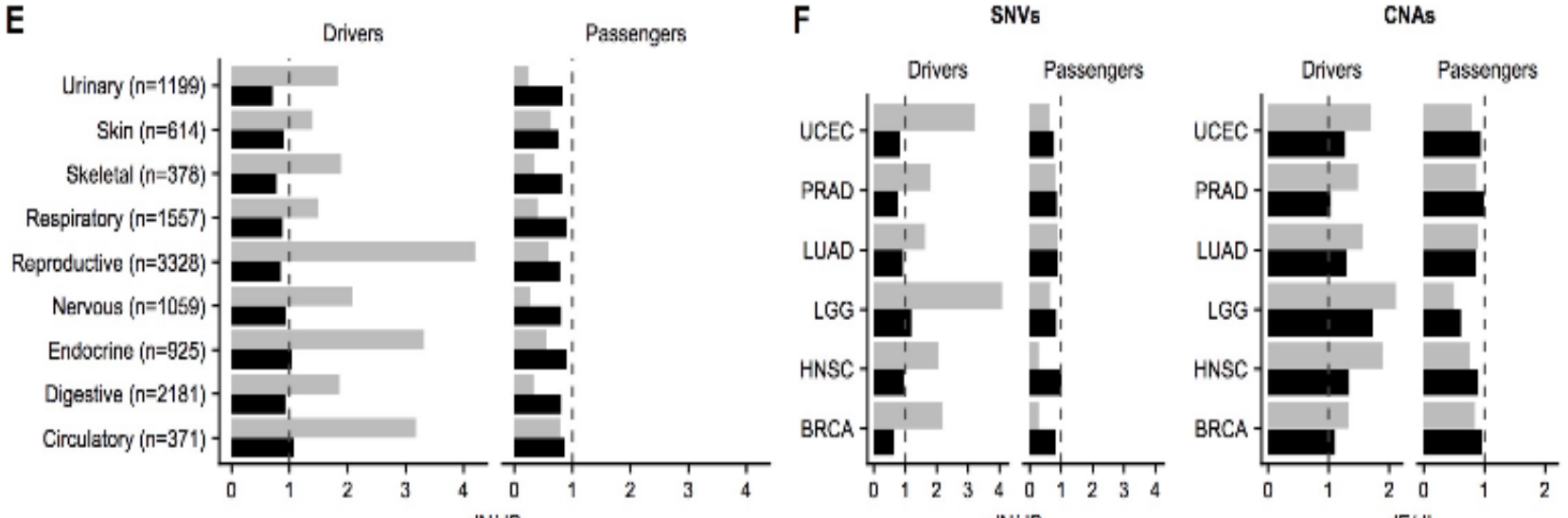
Tilk et al. hypothesis

- Negative selection is largely absent in cancer
- As cancer evolves asexually under linkage constraints mutations are unable to be removed via selection
- Given linkage-effects increase with mutation rate increase then tumours with high mutational burden should have decreased selection efficacy over low mutational burden tumours due to hitchhiking and Mullers ratchet

Tilk et al. hypothesis

- Negative selection is largely absent in cancer
- As cancer evolves asexually under linkage constraints mutations are unable to be removed via selection
- Given linkage-effects increase with mutation rate increase then tumours with high mutational burden should have decreased selection efficacy over low mutational burden tumours due to hitchhiking and Mullers ratchet
- Test $\frac{d_n}{d_s}$ in tumours stratified by mutational burden


A
All SNVs
dN/dS
Total Number of Substitutions
B
Nonsynonymous Functional Effects (PolyPhen2)
Fraction of Pathogenic Missense Mutations
Total Number of Substitutions
C
CNAs
<100Kb
>100Kb
dE/dI
Total Number of CNAs
D
Clonal vs. Subclonal SNVs (Whole-Exome SNVs)
dN/dS
Total Number of Substitutions
Clonal Drivers
All Drivers
Subclonal Drivers
Clonal Passengers
All Passengers
Subclonal Passengers

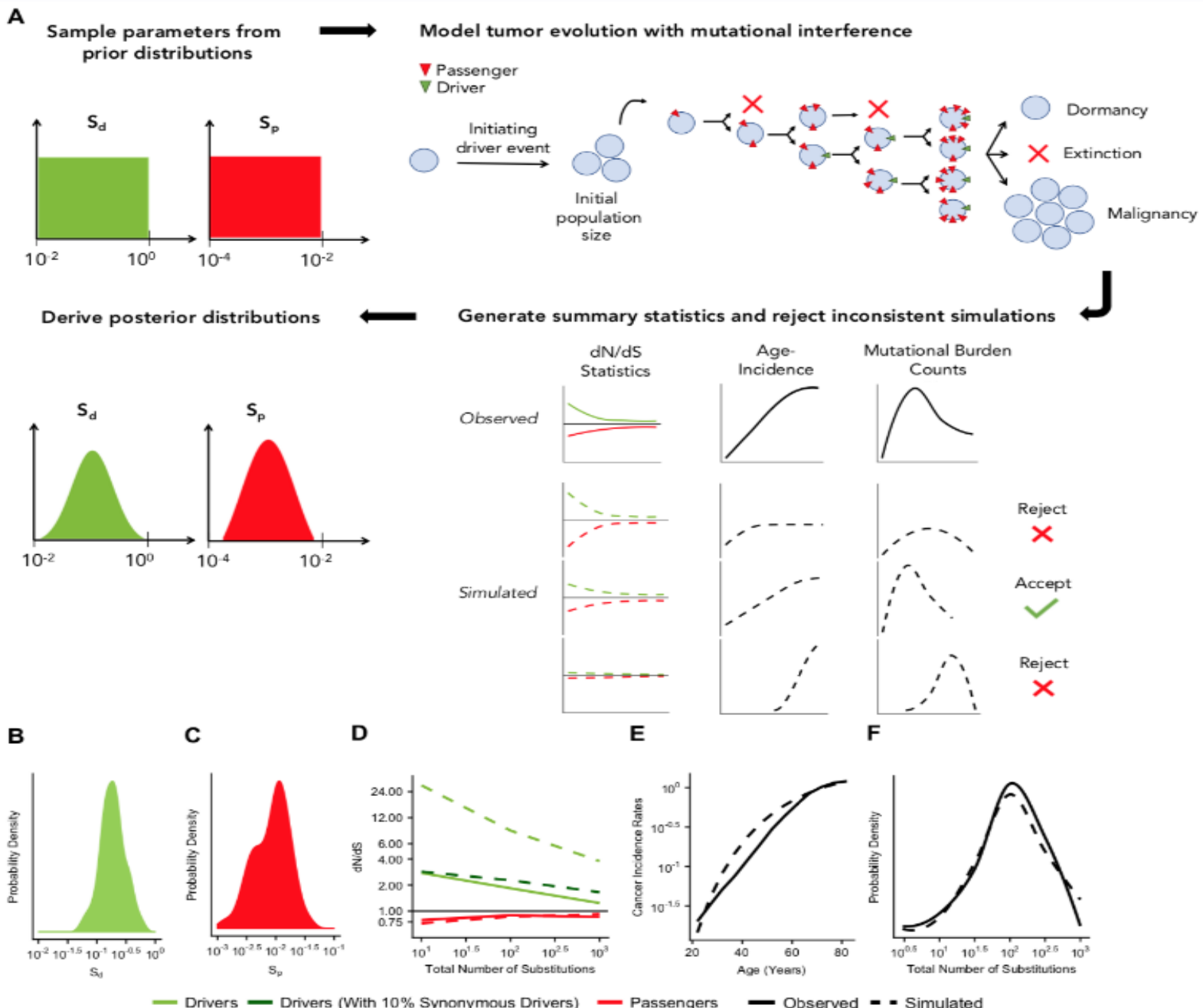

E
Drivers
Passengers
Urinary (n=1199)
Skin (n=614)
Skeletal (n=378)
Respiratory (n=1557)
Reproductive (n=3328)
Nervous (n=1059)
Endocrine (n=925)
Digestive (n=2181)
Circulatory (n=371)
F
SNVs
CNAs
Drivers
Passengers
UCEC
PRAD
LUAD
LGG
HNSC
BRCA

Evolutionary model incorporating Hill-Robertson interference

Based on a first order Gillespie Algorithm with 5 parameters $\mu T_{\mathrm{d}}, \mu T_{\mathrm{p}}, S_{\mathrm{d}}, S_{\mathrm{p}},$ and N^0

Individual cells can stochastically divide and die

Obtain MLE estimates of S_{d} and S_{p}


A
Sample parameters from prior distributions
Model tumor evolution with mutational interference
Passenger
Driver
Initiating driver event
Initial population size
Dormancy
Extinction
Malignancy
Derive posterior distributions
Generate summary statistics and reject inconsistent simulations
dN/dS Statistics
Age-Incidence
Mutational Burden Counts
Observed
Simulated
Reject
Accept
Reject
B
C
D
E
F
Probability Density
dN/dS
Cancer Incidence Rates
Age (Years)
Total Number of Substitutions
Drivers
Drivers (With 10% Synonymous Drivers)
Passengers
Observed
Simulated

Results

- Negative selection is absent in most tumours due to the inability of selection to remove mutations under linkage constraints

Results

- Negative selection is absent in most tumours due to the inability of selection to remove mutations under linkage constraints
- This result holds across SNVs, CNAs and across broad and tumour sub-types

Results

- Negative selection is absent in most tumours due to the inability of selection to remove mutations under linkage constraints
- This result holds across SNVs, CNAs and across broad and tumour sub-types
- Passenger mutations convey an individual selective cost of $\approx 1\%$ while drivers convey a selective advantage of $\approx 20\%$

Results

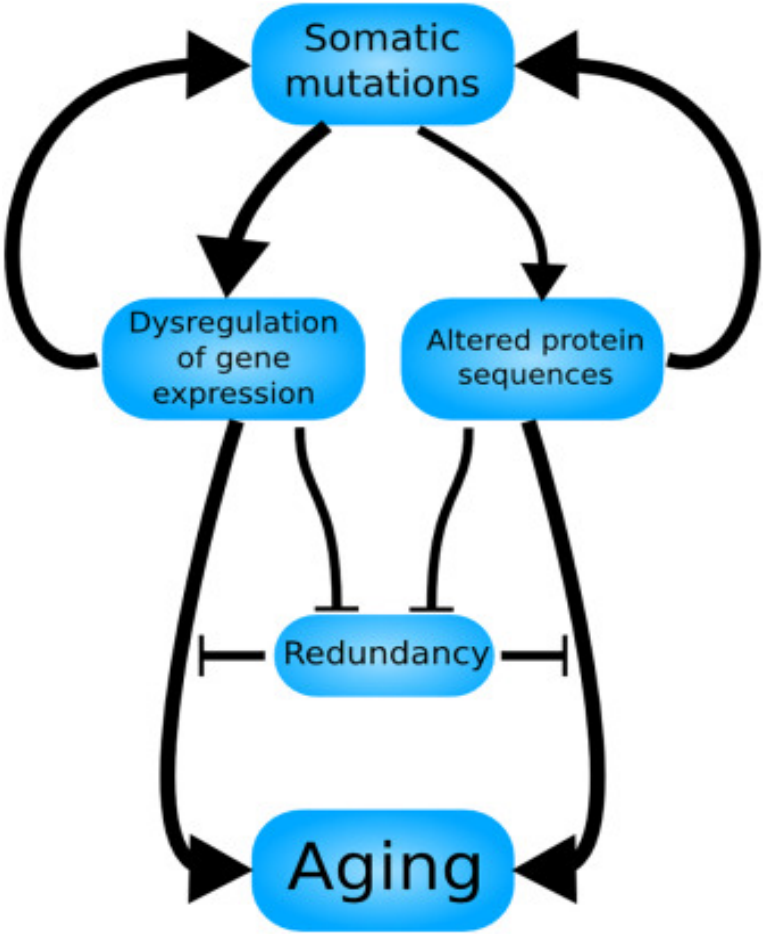
- Negative selection is absent in most tumours due to the inability of selection to remove mutations under linkage constraints
- This result holds across SNVs, CNAs and across broad and tumour sub-types
- Passenger mutations convey an individual selective cost of $\approx 1\%$ while drivers convey a selective advantage of $\approx 20\%$
- Most cancers harbour a large mutational load with median fitness cost of $\approx 40\%$ acquire ≈ 5 drivers with fitness benefit of $\approx 130\%$

Further notes...

- The buffering effect of having multiple copies of genes

Further notes...

- The buffering effect of having multiple copies of genes



Further notes...

- The buffering effect of having multiple copies of genes
- A large number of genes are indispensable in somatic lineages for indispensable genes d_N are neutral

Further notes...

- The buffering effect of having multiple copies of genes
- A large number of genes are indispensable in somatic lineages for indispensable genes d_N are neutral
- Differences in stem cell population size and structure can increase the stochasticity of genetic drift

Further notes...

- The buffering effect of having multiple copies of genes
- A large number of genes are indispensable in somatic lineages for indispensable genes d_N are neutral
- Differences in stem cell population size and structure can increase the stochasticity of genetic drift
- Very little negative selection is observed in healthy somatic tissues.

The End