Waiting times in a branching process model of colorectal cancer initiation

Ruibo Zhang[a], Obinna A. Ukogu[a], Ivana Bozic[a,*]

[a]*Department of Applied Mathematics, University of Washington, United States of America*

Abstract

We consider a multi-type branching process model for the initiation of colorectal cancer, which progresses through a complex sequence of driver mutations. The model incorporates the relative fitness advantages conferred by driver mutations, which leads us to consider non-dividing sub-populations, and the effects of neutral and advantageous mutations. We analyze large-time limits to approximate the population sizes of sub-clones founded by neutral or advantageous mutations, and use the results to compute the waiting time distributions for the incidence of novel mutations in colorectal cancer initiation. Analytic waiting time distributions, including the distribution of waiting time to malignancy, are in good agreement with results from Monte Carlo simulations of the model. The approach presented here can be extended to other multi-type branching process models where population growth rates of consecutive types are non-decreasing.

Keywords: Branching process, Waiting times, Cancer

1. Introduction

Following the seminal work by Armitage and Doll in the 1950s [1, 2], and Knudson [3] in the 1970s, cancer initiation has been identified as a multi-step process, whereby cells sequentially accumulate driver mutations required for malignant transformation. These works investigated age-dependent incidence rates for human cancers and argued that a multi-step model best explains the data across different cancers. Since the advent of genome sequencing, the molecular mechanisms of many cancer driver mutations have been described [4–8].

Multi-type branching processes, with types corresponding to genotypes or cell states [9, 10], have emerged as a viable model for studying cancer evolution in a population. Numerous processes have been modeled, including initiation [11], metastasis [12–14], and drug resistance to therapy [15, 16]. Evolutionary dynamics are closely related to the relative fitness advantages conferred by mutations. With strictly increasing clonal growth rates after mutation, Durrett and Moseley [10] modeled a pathway from healthy cells to cancerous cells, and computed distributions for clonal sizes and waiting times. Nicholson and Antal [16] studied a general framework wherein wild-type individuals have the largest fitness (growth

*Corresponding Author
Email address: ibozic@uw.edu (Ivana Bozic)

rate), which could be applied to cases involving drug resistance. Random fitness advantages have also been investigated [17].

Colorectal cancer (CRC) is one of the most common cancers in the United States [18]. It has been shown that mutations in just three driver genes, usually inactivation of two tumor suppressor genes and activation of an oncogene, are sufficient to trigger the initiation of CRC [19–21]. Typically, this involves a total of five genetic alterations, as inactivation of a tumor suppressor gene (TSG) requires inactivation of both alleles, and activation of an oncogene only requires a mutation in one allele of the gene [22]. Recent work [21] studied a five-step branching process model for the initiation of colorectal cancer that involves the three most commonly mutated driver genes in colorectal cancer: tumor suppressors *APC* and *TP53* and oncogene *KRAS*. The study found that, in the majority of cases, the driver mutations accrue in a specific order, with inactivation of *APC* followed by activation of *KRAS* and inactivation of the *TP53* gene.

Following [21], we study the single most likely mutational pathway to colorectal cancer. We model the dynamics using a multi-type branching process that starts from N wild-type crypts, small tubular assemblies of cells that line the intestinal epithelium [23, 24]. As the process evolves, individual crypts stochastically obtain driver mutations, with mutation rates determined by the genotype of the crypt and the driver gene in question. In contrast to previous studies, we allow growth rates of consecutive sub-populations to be equal. Furthermore, instead of founding a multi-type branching process with a single crypt, we consider the case where the process is initiated by a shrinking population of wild-type crypts. Using results from the theory of martingales, we compute the waiting time distributions associated with each step of the process. Finally, we compare analytic results for the waiting time distributions to those from Gillespie simulations.

The approach presented here can be extended to other multi-type branching process models in which the population growth rates of subsequent types are non-decreasing. For CRC, one can compute the waiting time distributions for other mutational pathways. The quantitative estimates provided here may also help better understand the lifetime risk of CRC.

2. Model

Let $N_i(t)$ be the stochastic process that counts the population of type-i crypts at time t. In a single step, a type-i crypt transforms into a type-$(i+1)$ crypt through either mutation or loss of heterozygosity (LOH) at rate u_i. Here we consider the most common mutational pathway observed (see Table 1) over the course of an 80-year human lifespan, i.e. $0 \leq t \leq 80$. We assume that independently from mutations, type-i crypts follow a pure birth process with rate λ_i. The growth rate of a crypt is determined by its genotype. For wild-type crypts, divisions are very rare [25] so we set their growth rate to zero. It has been shown that inactivation of *APC* and/or activation of *KRAS* provides a fitness advantage to mutated crypts, leading to clonal expansion [26, 27]. In contrast, under normal conditions *TP53* inactivation alone does not provide a fitness advantage [23]. These findings are reflected in the choice of growth parameters in our model (see Table 2).

Step	Rate	Biological process
$N_0 \to N_1$	u_1	*APC* LOH
$N_1 \to N_2$	u_2	*APC* mutation
$N_2 \to N_3$	u_3	*KRAS* mutation
$N_3 \to N_4$	u_4	*TP53* LOH
$N_4 \to N_5$	u_5	*TP53* mutation

Table 1: Most common pathway to CRC initiation.

(a) Non-decreasing crypt growth rates

crypts	N_0	N_1	N_2	N_3	N_4
birth rate	$\lambda_0 = 0$	$\lambda_1 = 0$	$\lambda_2 > 0$	$\lambda_3 > \lambda_2$	$\lambda_4 = \lambda_3$

(b) Biologically reasonable range of parameter values

number of wild-type crypts	N	$10^7 - 10^8$
birth rates	λ_i	$\lambda_2 = 0.2/y, \lambda_3 = 0.27/y$
transition rates	u_i	$10^{-7} - 10^{-4}/y$

Table 2: Parameter values used in simulations.

This branching process model can be summarized as

$$N_0(t) \xrightarrow{u_1} N_1(t) \xrightarrow{u_2} \begin{matrix} \circlearrowleft\text{divide at rate } \lambda_2 \\ N_2(t) \end{matrix} \xrightarrow{u_3} \begin{matrix} \circlearrowleft\text{divide at rate } \lambda_3 \\ N_3(t) \end{matrix} \xrightarrow{u_4} \begin{matrix} \circlearrowleft\text{divide at rate } \lambda_4 = \lambda_3 \\ N_4(t) \end{matrix} \xrightarrow{u_5} N_5(t)$$

For $N_i(t)$, the first generation of type-i crypts (all the type-i crypts produced directly from type-$(i-1)$ crypts) follows an inhomogeneous Poisson process, denoted by $K_i(t)$, with instantaneous intensity $u_i N_{i-1}(t)$. The transition times are the founding times of new type-i subclones. They are defined as

$$T_{i,j} = \inf\{t \geq 0 | K_i(t) = j\}.$$

At the j-th transition time, the newly formed type-i crypt founds a linear birth-death process with birth rate λ_i and death rate u_{i+1} (transition rate to the next type). Let $Z_{i,j}(t)$ denote this branching process where t is measured from the founding time. We also call $Z_{i,j}(t)$ the j-th lineage or j-th subclone of type-i individuals. Let $Z^{(i)}$ represent a single type-i individual. The transition scheme is

$$Z^{(i)} \to \begin{cases} Z^{(i)}\, Z^{(i)}, & \text{growth rate } \lambda_i \\ Z^{(i+1)}, & \text{transition rate } u_{i+1}. \end{cases}$$

Suppose $Z_i(t)$ is a linear birth-death process intiated by a single cell and $Z_i(t) = 0$ for $t < 0$, then measured from the starting time of CRC, $Z_{i,j}(t)$ is merely $Z_i(t)$ shifted by $T_{i,j}$. Thus

$$N_i(t) = \sum_{j=1}^{K_i(t)} Z_i\left(t - T_{i,j}\right),$$

$T_{i,j} \leq t$, for $1 \leq j \leq K_i(t)$. Here we note that since u_{i+1} is small, we treat all lineages of type-i individuals as pure birth processes when type-i has a positive birth rate. The system initially consists N wild-type (type-0) crypts, and we seek to estimate the waiting times for the first type-i crypt which is defined by

$$\tau_i = \inf\{t \geq 0 | N_i(t) > 0\}.$$

To verify our analytic results, we developed Monte Carlo simulations of a multi-type branching process model based on the Gillespie algorithm [28]. Parameter values for our model come from Paterson et al. [21]. Estimates of the values are listed in Table 2. To be consistent with their study, we initialize our simulations with 10^8 wild-type crypts.

3. General results for a multi-type branching process

In this section, we present and extend a common approach for computing the waiting time distribution of the first type-i individual in a multi-type branching process in which type-i individuals are produced from type-$(i-1)$ individuals at rate u_i, and each newly formed type-i individual starts a new pure birth process with growth rate λ_i. Even though multi-type models may contain many types, mutation from one type to the next can be treated as a two-type branching process. From this perspective, the process containing types $1, ..., n$ can be treated as a sequence of $n-1$ two-type branching processes, with the terminal state from one two-type branching process serving as the initial state for the next. We will show that when large time limits exist for type-i populations, it suffices to analyze two-type (initial and terminal type) branching processes $(N_i(t), N_{i+1}(t))$, leading to a unified approach. When a large time limit does not exist, we introduce an different approach (see Section 3.4).

In previous studies, the asymptotic population dynamics of two-type branching processes have been analyzed by decomposing the population of the initial type into the product of a deterministic, time-dependent exponential function and a time-invariant random variable [10, 14, 29]. This approximation greatly simplifies the computation of the waiting time distribution by making the stochastic component of the dynamics time-invariant. However, these studies typically consider the case of a wild type with a net growth rate λ_i strictly greater than zero. In contrast, the wild-type population in our model is not supercritical and in fact (slowly) declines. This, along with allowing $\lambda_i = \lambda_{i+1}$, causes non-exponential growth of mutant populations. In this section, we extend previous approaches [10, 14, 29] so that the time-deterministic component can belong to a class of non-exponential functions, which enables us to deal with the non-exponentially growing populations (type-0 and type-1). Furthermore, we review canonical expected value formulas that will be applied in later sections. Similarly to previous approaches [10, 14] we will assume that the mutation rates from one type to the next are small.

3.1. Expected value and Laplace transform

Throughout Section 3, we will be concerned with a multi-type branching process $(N_0(t), N_1(t), ...)_{t \geq 0}$, where $N_i(t)$ denotes the number of type-i individuals at time t. The process is started at time zero, with some positive number of type-0 individuals. Type-i individuals can divide with rate $\lambda_i \geq 0$ and mutate to type-$(i+1)$ with rate $u_i > 0$.

We start with a lemma that provides the Laplace transform of $N_i(t)$ conditional on the population of its precursor, $N_{i-1}(t)$.

Lemma 3.1. *Let $Z_i(t)$ be the number of type-i individuals in a pure-birth process that starts with $Z_i(0) = 1$ individuals at time $t = 0$. Then*

$$\mathbb{E}\left[e^{-\theta N_i(t)}\middle| N_{i-1}(s), s \leq t\right] = \exp\left(-u_i \int_0^t N_{i-1}(s)(1 - \phi_i(\theta, t-s))\right) \tag{1}$$

where $\phi_i(\theta, t) := \mathbb{E}\left[e^{-\theta Z_i(t)}\right]$.

This is a well-known result for multi-type branching processes. One can prove it by following the procedure of Lemma 2 in [10], replacing the start time with $s = 0$.

The expected value formula for the number of type-i individuals can be derived from Lemma 3.1.

Theorem 3.2.

$$\mathbb{E}\left[N_i(t)\right] = u_i \int_0^t \mathbb{E}[N_{i-1}(s)] e^{\lambda_i(t-s)}\, \mathrm{d}s$$

Taking the derivative of equation (1) and evaluating at $\theta = 0$ gives the relationship between expected values.

3.2. Large time limits

Lemma 3.3. *Consider $(N_{i-1}(t), N_i(t))_{t\geq 0}$. Suppose $N_i(t)$ is the population of type-i crypts in the multi-type branching system where the population of type-$(i-1)$ crypts is given by $N_{i-1}(t) \geq 0$, then*

$$M(t) = e^{-\lambda_i t} N_i(t) - \int_0^t u_i N_{i-1}(s) e^{-\lambda_i s} \mathrm{d}s$$

is a martingale. If

$$I_i = \int_0^\infty u_i N_{i-1}(s) e^{-\lambda_i s} ds$$

has a finite expectation, then

$$e^{-\lambda_i t} N_i(t) \overset{a.s.}{\to} W_i,\ \mathbb{E}|W_i| < \infty,$$

as $t \to \infty$. Additionally, if $e^{-\lambda_i t} N_i(t)$ is uniform integrable, then

$$e^{-\lambda_i t} N_i(t) \overset{L^1}{\to} W_i.$$

This implies

$$\mathbb{E}[e^{-\lambda_i t} N_i(t)] \to \mathbb{E}[W_i] = \mathbb{E}[I_i].$$

If the first condition from the statement of the Lemma holds (i.e. if I_i has finite expectation) Lemma 3.3 provides a method of obtaining the long-term behavior of N_i using the limiting random variable W_i. In that case, we have $e^{-\lambda_i t} N_i(t) \to W_i$, and for large time t, $e^{\lambda_i t} W_i$ should be a good approximation of the stochastic process $N_i(t)$. The importance of

$N_i(t) \approx e^{\lambda_i t} W_i$ is that it separates a stochastic process into a deterministic function $e^{\lambda_i t}$ and a time-independent random variable W_i.

If, in addition, the second condition ($e^{-\lambda_i t} N_i(t)$ is uniform integrable) holds, then the expected value of the limiting random variable W_i is obtainable. In that case, we have $E[e^{-\lambda_i t} N_i(t)] \to E[W_i]$, which makes the large time approximation $N_i(t) \approx e^{\lambda_i t} W_i$ reasonable in terms of the first moment.

3.3. Estimating waiting times using large time limits

Let $\tau_i, (1 \leq i \leq n)$ be the waiting time of the first type-i individual in a multi-type branching process. At time $s \geq 0$, the arrival rate of type-i individuals is $u_i N_{i-1}(s)$. Conditional on the trajectory of $N_{i-1}(s)$ for $0 \leq s \leq t$, the probability of τ_i can be written as

$$P(\tau_i > t \mid N_{i-1}(s), 0 \leq s \leq t) = \exp\left(-u_i \int_0^t N_{i-1}(s) ds\right).$$

The functional form of $N_{i-1}(s)$ is generally unknown and potentially complicated. One way of evaluating this integral is to approximate $N_{i-1}(s)$ by the product of a deterministic time-dependent growth and a time independent random variable. For example, let $N_0(t) = Z_0(t)$, a pure birth two-type branching process that starts with a single individual. It is well-known that $e^{-\lambda_0 t} Z_0(t) \to W_0 \sim \text{Exponential}(1)$ [10]. A classical approximation is $N_0(s) \approx e^{\lambda_0 t} W_0$ where the deterministic time-dependent growth is characterized by $e^{\lambda_0 t}$ and time independent random variable is W_0. Applying this approximation yields

$$P(\tau_1 > t) \approx \mathbb{E}\left[\exp\left(-u_i \int_0^t e^{\lambda_0 s} W_0 ds\right)\right] \tag{2}$$

$$= \mathcal{L}_{W_0}\left(u_1 \frac{e^{\lambda_0 t} - 1}{\lambda_0}\right), \tag{3}$$

where $\mathcal{L}_{W_0}(\theta) = \frac{1}{1+\theta}$ is the Laplace transform of W_0. Let $f_0(t) = e^{\lambda_0 t}$ be the time deterministic function. Then the above approximation also holds if a sub-exponential term is added to f_0. In other words, we have many reasonable options for f_0. In later sections, we consider two specific deterministic functions,

$$f_{i0}(t) = e^{\lambda_i t}, \text{ and}$$
$$f_{i1}(t) = \frac{\mathbb{E}[N_i(t)]}{\mathbb{E}[W_i]}.$$

In the example mentioned above, we have

$$f_{01}(t) = \frac{\mathbb{E}[Z_0(t)]}{\mathbb{E}[W_0]} = e^{\lambda_0 t}$$
$$= f_{00}(t).$$

However, in our model when $i > 1$, $f_{i0} \neq f_{i1}$. Later we will see an approximation with f_{i1} is more precise than that with f_{i0}.

The second observation from this example is that the approximation (2) relies on the Laplace transform of W_0. However, one cannot obtain a closed form for the Laplace transform of $W_i, i \geq 2$. In this case, an alternative approach is to approximate W_i by another random variable V_i that has a a closed form Laplace transform. Such V_i can be found by approximating N_{i-1}, as described in Section 4.

Proposition 3.4. *Let $(N_{i-1}(t), N_i(t))$ be a two-type process such that $N_i(0) = 0$ and type-i individuals are produced by type-$(i-1)$ individuals with rate $u_i N_{i-1}(t), u_i > 0$. Suppose that $N_{i-1}(t) \approx N^*_{i-1}(t)$ such that $(f_{i-1}(t))^{-1} N^*_{i-1}(t) \to V_{i-1}$ as $t \to \infty$. Then the waiting time distribution of the first type-i individual can be approximated by*

$$P(\tau_i > t) \approx \mathcal{L}_{V_{i-1}} \left(u_i \int_0^t f_{i-1}(s) ds \right)$$

where $\mathcal{L}_{V_{i-1}}(\theta)$ is the Laplace transform of random variable V_{i-1}.

Applying the approximations gives us

$$\begin{aligned} P(\tau_i > t) &= \mathbb{E}\left[\exp\left(-\int_0^t u_i N_{i-1}(s) ds \right) \right] \\ &\approx \mathbb{E}\left[\exp\left(-V_{i-1} u_i \int_0^t f_{i-1}(s) ds \right) \right] \\ &= \mathcal{L}_{V_{i-1}} \left(u_i \int_0^t f_{i-1}(s) ds \right) . \end{aligned}$$

From Proposition 3.4, we see that the waiting time (τ_i) distribution depends on the Laplace transform of V_{i-1}. Here we present an iterative method for computing the Laplace transforms of V_i.

Lemma 3.5. *Let $\tilde{N}_i(t)$ be type-i individuals in the system with $N^*_{i-1}(t) = f_{i-1}(t) V_{i-1}$. Suppose $e^{-\lambda_i t} \tilde{N}_i(t) \overset{a.s.}{\to} V_i$, then*

$$\mathcal{L}_{V_i}(\theta) := \mathbb{E}\left[e^{-\theta V_i} \right] = \mathcal{L}_{V_{i-1}} \left(u_i \int_0^\infty \frac{\theta f_{i-1}(s)}{\theta + e^{\lambda_i s}} \, \mathrm{d}s \right).$$

3.4. An inhomogeneous Poisson process approximation

Proposition 3.4 estimates the arrival time of type-i individuals by the large time limit of the previous type (V_{i-1}). However, there are situations when the random variable V_{i-1} does not exist or when $f_{i-1} V_{i-1}$ fails to provide a precise first arrival time of N_i. More generally, we may only have a good large time limit V_j for some type-j where $j < i$, so that we do not have reliable limits for type-$(j+1)$ through type-$(i-1)$. To deal with this situation, we introduce a method that uses the large time limit of each independent lineage to 'skip' V_{j+1} through V_{i-1}. First, note that given V_j, we can approximate the birth rate of type-$(j+1)$ lineages, using the well-known fact that for each type-$(j+1)$ lineage,

$$e^{-\lambda_{j+1} t} Z_{j+1}(t) \overset{a.s.}{\to} V \sim \text{Exponential}(1). \tag{4}$$

Next, we want to find the likelihood of each type-$(j+1)$ lineage producing at least a single type-i individual. Recall that a type-$(j+1)$ lineage is a simple birth process initiated by a single type-$(j+1)$ individual that grows at rate λ_{j+1}. Suppose we have a type-$(j+1)$ lineage that was started by a single type-$(j+1)$ individual at time s. We define $p_0^{(j+1)\to i}(s,t)$ to be the probability that no type-i individual is produced by time t by this type-$(j+1)$ lineage.

At time s, type-$(j+1)$ crypts are producing at rate $u_{j+1}N_j(s) \approx u_{j+1}f_j(s)V_j$. Each type-$(j+1)$ lineage (present at time s) has a probability $1-p_0^{(j+1)\to i}(s,t)$ to produce at least a single type-i crypt (at time t). Finally, for fixed t, we approximate the process of producing type-i individuals as an inhomogeneous Possion process with rate $u_{j+1}N_j(s)(1-p_0^{(j+1)\to i}(s,t))$. This implies

$$\begin{aligned} P(\tau_i > t) &\approx \mathbb{E}\left[\exp\left(-V_j u_{j+1}\int_0^t f_j(s)\left(1-p_0^{(j+1)\to i}(s,t)\right)ds\right)\right] \\ &= \mathcal{L}_{V_j}\left(u_{j+1}\int_0^t f_j(s)\left(1-p_0^{(j+1)\to i}(s,t)\right)ds\right). \end{aligned}$$

This approach is summarized in the following proposition.

Proposition 3.6. *Let $(N_j(t), N_{j+1}(t), \cdots, N_i(t))$ be a $(i-j+1)$-type process such that $N_k(0) = 0, j < k \le i$ and type-k individuals are produced by type-$(k-1)$ individuals with rate $u_k N_{k-1}(t), u_k > 0$. Suppose that $N_j(t) \approx N_j^*(t)$ such that $(f_j(t))^{-1}N_j^*(t) \to V_j$. Then the waiting time distribution of type-i crypts can be approximated by*

$$P(\tau_i > t) \approx \mathcal{L}_{V_j}\left(u_{j+1}\int_0^t f_j(s)\left(1-p_0^{(j+1)\to i}(s,t)\right)ds\right),$$

where $\mathcal{L}_{V_j}(\theta)$ is the Laplace transform of V_j.

Proposition 3.4 is consistent with Proposition 3.6, and one can treat Proposition 3.4 as a single-step version of Proposition 3.6.

To compute $p_0^{j\to i}(s,t), j < i$, we use the iterative relationship between $p_0^{j\to i}(s,t)$ and $p_0^{(j+1)\to i}(s,t)$. This is provided by the following proposition.

Lemma 3.7. *For $i < j$,*

$$p_0^{j\to i}(s,t) \approx \int_0^t e^{-v}\exp\left(-v\int_s^t u_{j+1}e^{\lambda_j(r-s)}\left(1-p_0^{(j+1)\to i}(r,t)\right)dr\right)dv.$$

with $p_0^{i\to i}(s,t) = 0$.

4. Population dynamics of type-i crypts

In this section we derive the asymptotic growth dynamics of the clonal sub-populations (or types) along the path to CRC initiation. For each type i, the population size $N_i(t)$ is a counting process indexed by $t > 0$. The first two types, $N_0(t)$ and $N_1(t)$, are well approximated by time-deterministic functions. Starting with these, we sequentially derive the

large-time population dynamics for the later types, leveraging results for two-type branching processes, $(N_i, N_{i+1}), 0 \leq i \leq 4$. The success of this approach relies on our ability to approximate the large-time growth dynamics by the product of a time-deterministic function and a time-invariant random variable. The existence of such a decomposition depends on the martingale properties of the counting processes $N_i(t)$.

4.1. Type-0

Wild-type human crypts (i.e. type-0 crypts) rarely divide [25], hence we set the growth rate of the wild-type crypts to zero ($\lambda_0 = 0$). These wild-type crypts transition into type-1 crypts at rate $u_1 N_0(t)$ after losing one copy of the *APC* gene.

Proposition 4.1. *$N_0(t)$ is a pure death process with death rate u_1 and initial condition $N_0(0) = N$. Its generating function at time t is*

$$G_0(s,t) = \mathbb{E}[s^{N_0(t)}] = \left(se^{-tu_1} + 1 - e^{-tu_1}\right)^N,$$

and the expectation and variance are

$$\mathbb{E}[N_0(t)] = Ne^{-u_1 t},$$
$$\mathrm{Var}[N_0(t)] = e^{-2u_1 t} N(e^{u_1 t} - 1).$$

For a detailed description of a single-type birth-death process, one can refer to chapter 3, section 5 in [30]. Typical biologically realistic parameter values are $u_1 t \sim 10^{-2} \ll 1$ and $N \sim 10^8$, which leads to

$$\frac{\sqrt{\mathrm{Var}[N_0(t)]}}{\mathbb{E}[N_0(t)]} = \sqrt{\frac{e^{tu_1} - 1}{N}} \sim 10^{-5}, \quad \mathbb{E}[N_0(t)] = Ne^{-u_1 t} \approx N.$$

Motivated by this result, we approximate the population of the wild-type crypts by its expectation

$$N_0(t) \approx Ne^{-u_1 t} \approx N,$$

in which the second approximation is justified because $u_1 t \ll 1$. Therefore, the waiting time of $N_1(t)$, $\tau_1 \sim \mathrm{Exponential}(u_1 N)$.

4.2. Type-1

A type-1 crypt is produced when a healthy crypt (type-0) loses one copy of the *APC* gene. This mutation does not lead to clonal expansion [26], so the growth rate of type-1 crypts, $\lambda_1 = 0$. While being produced by wild-type crypts, type-1 crypts mutate into type-2 at rate $u_2 N_1(t)$, losing both copies of the *APC* gene. We assume that the population outflows are negligible. Initially, there are no type-1 crypts, i.e., $N_1(0) = 0$.

Proposition 4.2. *$N_1(t) = N - N_0(t)$ has the following expectation and variance:*

$$\mathbb{E}[N_1(t)] = N(1 - e^{-u_1 t}),$$
$$\mathrm{Var}[N_1(t)] = e^{-2u_1 t} N(e^{u_1 t} - 1).$$

This can be derived by observing $N_0(t) + N_1(t) = N$, since the outflows from type-1 to type-2 are negligible. Therefore $\mathbb{E}[N_1(t)] = \mathbb{E}[N - N_0(t)] = N - \mathbb{E}[N_0(t)], \mathrm{Var}[N_1(t)] = \mathrm{Var}[N_0(t)]$.

To approximate $N_1(t)$, we consider the case when the wild-type population is effectively constant. In this case, $N_1(t)$ can be treated as a Poisson process with intensity $u_1 N$. According to the properties of a Poisson process, the approximated expectation and variance are

$$\begin{aligned}\mathbb{E}[N_1(t)] &\approx u_1 N t,\\ \mathrm{Var}[N_1(t)] &\approx u_1 N t.\end{aligned}$$

Notice that the approximated expected value and variance are consistent with the true expected value and variance. Since $u_1 t \sim 10^{-2} \ll 1$ is small, we have

$$\begin{aligned}\mathbb{E}[N_1(t)] &= N(1 - e^{-u_1 t}) \approx u_1 N t,\\ \mathrm{Var}[N_1(t)] &= e^{-2u_1 t} N(e^{u_1 t} - 1) = N(e^{-u_1 t} - e^{-2u_1 t}) \approx u_1 N t.\end{aligned}$$

Additionally, observe that

$$\frac{\sqrt{\mathrm{Var}[N_1(t)]}}{\mathbb{E}[N_1(t)]} \approx \frac{1}{\sqrt{u_1 N t}} \sim 10^{-2},$$

and so we approximate $N_1(t)$ by a deterministic function

$$N_1(t) \approx u_1 N t.$$

Thus the waiting time distribution of type-2 crypts is simply estimated as

$$P(\tau_2 \le t) = 1 - \exp\left(-u_2 \int_0^t N_1(s) ds\right) \approx 1 - \exp\left(-\frac{1}{2} u_1 u_2 N t^2\right).$$

4.3. Type-2

After losing both copies of the *APC* genes, type-1 crypts mutate into type-2 crypts. The *APC* inactivation provides a fitness advantage to type-2 crypts [26]. Thus for type-2, we have a positive growth rate $\lambda_2 > 0$. Meanwhile, the *KRAS* oncogenes are activated at rate $u_3 N_2(t)$, mutating type-2 crypts into type-3 crypts. These outflows are ignored when considering the population of $N_2(t)$. At time $t = 0$, there are no type-2 crypts, i.e $N_2(0) = 0$. We begin with the expectation of $N_2(t)$. Using Theorem 3.2, we compute

$$\mathbb{E}[N_2(t)] \approx \frac{N u_1 u_2 (e^{\lambda_2 t} - \lambda_2 t - 1)}{\lambda_2^2}.$$

The approximation is obtained by noting that $u_1 \sim 10^{-4} \ll \lambda_2 \sim 10^{-1}$ and $u_1 t \sim 10^{-2} \ll 1$. The following large-time asymptotic limit exists for $N_2(t)$.

Theorem 4.3. *$e^{-\lambda_2 t} N_2(t) \to W_2$ a.s. and in L^1 with*

$$\mathbb{E}[W_2] = \frac{N u_1 u_2}{\lambda_2(\lambda_2 + u_1)} \approx \frac{N u_1 u_2}{\lambda_2^2}.$$

To characterize the distribution of the limiting random variable, we consider $\tilde{N}_2(t)$, the population of type-2 crypts produced by $N_1^*(t) = u_1 N t$.

Theorem 4.4. *$e^{-\lambda_2 t}\tilde{N}_2(t) \to V_2$ a.s. and in L^1 with*

$$\mathbb{E}[V_2] = \frac{N u_1 u_2}{\lambda_2^2}$$

and

$$\mathcal{L}_{V_2}(\theta) = \mathbb{E}\left[e^{-\theta V_2}\right] = \exp\left(\frac{N u_1 u_2 \mathrm{PolyLog}(2, -\theta)}{\lambda_2^2}\right).$$

Corollary 4.5. *$\frac{\tilde{N}_2(t)}{f_2(t)} \to V_2$ a.s. and in L^1 for all*

$$f_2(t) \in \{f \in C(\mathbb{R}) | \lim_{t\to\infty} e^{-\lambda_2 t} f_2(t) = 1\}.$$

4.4. Type-3

Type-3 crypts are produced by type-2 crypts through activation of the *KRAS* oncogene, which increases the fission rate of mutated cells [27]. Thus the birth rate has a positive increment, i.e. $\lambda_3 > \lambda_2$. At a small rate $u_4 N_3(t)$, type-3 crypts lose one copy of the TSG *TP53* and mutate into type-4 crypts. The initial population is $N_3(0) = 0$. Using Theorem 3.2, the expected value of $N_3(t)$ is

$$\mathbb{E}[N_3(t)] \approx \frac{N u_1 u_2 u_3 \left(\lambda_2^2\left(e^{\lambda_3 t} - \lambda_3 t - 1\right) - \lambda_3^2\left(e^{\lambda_2 t} - \lambda_2 t - 1\right)\right)}{\lambda_2^2\left(\lambda_3 - \lambda_2\right)\lambda_3^2},$$

where the approximation is made by observing $\lambda_i + u_1 \approx \lambda_i$ and $u_1 t \sim 10^{-2}$.

Theorem 4.6. *$e^{-\lambda_3 t} N_3(t) \to W_3$ a.s. and in L^1 with*

$$\mathbb{E}[W_3] = \frac{N u_1 u_2 u_3}{(\lambda_3 - \lambda_2)\lambda_3(\lambda_3 + u_1)} \approx \frac{N u_1 u_2 u_3}{(\lambda_3 - \lambda_2)\lambda_3^2}.$$

Consider a system with $N_2^*(t) := f_{20}(t) V_2 = e^{\lambda_2 t} V_2$ and let $\tilde{N}_3(t)$ denote the number of type-3 crypts in this system. The following large-time limit holds.

Theorem 4.7. *$e^{-\lambda_3 t}\tilde{N}_3(t) \to V_3$ a.s. and in L^1 with*

$$\mathbb{E}[V_3] = \frac{N u_1 u_2 u_3}{\lambda_2^2(\lambda_3 - \lambda_2)}.$$

For other potential $f_{2,k}$'s, the Laplace transform changes accordingly.

Corollary 4.8. *Let $N_2^*(t) := f_2(t) V_2$ where*

$$f_2(t) \in \{f \in C(\mathbb{R}) | \lim_{t\to\infty} e^{-\lambda_2 t} f_2(t) = 1\},$$

and let $\tilde{N}_3(t)$ denote type-3 crypts in this system. Then $\frac{\tilde{N}_3(t)}{f_3(t)} \to V_3$ a.s. and in L^1 with

$$\mathbb{E}[V_3] = \frac{N u_1 u_2 u_3}{\lambda_2^2} \int_0^\infty f_2(t) e^{-\lambda_3 s} \mathrm{d}s,$$

and

$$\mathcal{L}_{V_3}(\theta) = \mathbb{E}[e^{-\theta V_3}] = \mathcal{L}_{V_2}\left(u_i \int_0^\infty \frac{\theta f_2(s)}{\theta + e^{\lambda_i s}}\,\mathrm{d}s\right),$$

for all

$$f_3(t) \in \{f \in C(\mathbb{R}) | \lim_{t\to\infty} e^{-\lambda_3 t} f_3(t) = 1\}.$$

Note that $f_{i1}(t) = \mathbb{E}[N_i]/\mathbb{E}[W_i]$ is more precise than $f_{i0}(t) = e^{\lambda_i t}$. By Corollary 4.8, $\mathbb{E}[V_3] = \mathbb{E}[I_3^*] = \int_0^\infty u_3 f_2(t)\mathbb{E}[V_2]e^{-\lambda_3 s}\mathrm{d}s$, depends on $f_2(t)$ and $\mathbb{E}[V_2]$. Since $\mathbb{E}[W_3] = \int_0^\infty u_3\mathbb{E}[N_2(s)]e^{-\lambda_3 s}\mathrm{d}s$ and $\mathbb{E}[V_2] = \mathbb{E}[W_2]$, setting $f_2(t) = \mathbb{E}[N_2(t)]/\mathbb{E}[W_2]$ equates $\mathbb{E}[V_3]$ and $\mathbb{E}[W_3]$. Indeed, when using $f_2(t) = f_{20}(t) = e^{\lambda_2 t}$, the resulting random variable V_3 has an expectation different from W_3.

4.5. Type-4

Since only one copy of the *TP53* gene has been lost, type-4 does not confer a growth advantage [23]. Thus the division rate of type-4 crypts $\lambda_4 = \lambda_3$. The dynamics of type-4 crypts also include negligible outflows at rate $u_5 N_4(t)$. And the starting condition is $N_4(0) = 0$. In this case, we note that using $e^{-\lambda_3 t}$ to scale the growth process fails to provide a large-time limiting random variable. Even though

$$M_t = e^{-\lambda_3 t}N_4(t) - \int_0^t u_4 N_3(s)e^{-\lambda_3 s}\mathrm{d}s,$$

is still a martingale, Lemma 3.3 does not apply. We have seen that $\mathbb{E}[N_3(t)] = O(e^{\lambda_3 t})$ as $t \to \infty$, and so

$$u_4\mathbb{E}[N_3(s)]e^{-\lambda_3 s} = O(e^{\lambda_3 s})e^{-\lambda_3 s} = O(1).$$

This implies the first condition in Lemma 3.3 is invalid, that is as $t \to \infty$

$$I_4(t) = \int_0^t u_4 N_3(s)e^{-\lambda_3 s}\mathrm{d}s = O(t) \to \infty.$$

Therefore we cannot derive a large time limit of type-4 individuals. Instead, we will estimate the waiting time of type-5 crypts by the inhomogeneous Poisson approximation discussed in section 3.4. To this end, we compute $p_0^{4\to5}(s,t)$ and $p_0^{3\to5}(s,t)$, corresponding to estimating τ_5 by V_3 and V_2 respectively,

$$p_0^{4\to5}(s,t) = \frac{1}{1 + u_5\frac{\exp(\lambda_3(t-s))-1}{\lambda_3}},$$

$$p_0^{3\to5}(s,t) = \left(1 + \frac{u_4 u_5(\lambda_3 - u_5 + e^{\lambda_3(t-s)}(u_5 - \lambda_3 + \lambda_3 \log\left(\frac{\lambda_3 e^{\lambda_3(t-s)}}{\lambda_3 + u_5(e^{\lambda_3(t-s)}-1)}\right)}{(\lambda_3 - u_5)^2\lambda_3}\right)^{-1}.$$

5. Estimating waiting times for colorectal cancer initiation

In this section, we compare results obtained from exact computer simulations of the multi-type process with our analytic results. We define $p_i(t)$ as our estimate of the cumulative distribution function of τ_i, the waiting time to the first type-i crypt. For all figures in this section, parameter values are given in Table 3, and follow estimates from Paterson et al [21].

Crypt N_i	N_0	N_1	N_2	N_3	N_4
Initial population (crypts)	$N = 1 \times 10^8$	0	0	0	0
Birth rate λ_i (crypts/year)	0	0	0.2	0.27	0.27
Transition rate u_{i+1} (crypts/year)	2.86×10^{-4}	1.06×10^{-5}	9.00×10^{-7}	1.36×10^{-4}	4.56×10^{-7}

Table 3: Estimates of parameter values for colorectal cancer initiation from Paterson et al. [21].

5.1. Waiting times to type-1 and type-2

For τ_1 and τ_2, the analytic results are listed in the previous section. Recall that the population of wild-type (type-0) crypts are estimated to have a fixed population N. Thus,

$$P(\tau_1 \leq t) \approx p_1(t) = 1 - \exp\left(-u_1 N t\right).$$

For τ_2, recall that we approximate the population of type-1 crypts by a time deterministic function $u_1 N t$.

$$P(\tau_2 \leq t) \approx p_2(t) = 1 - \exp\left(-\frac{1}{2} u_1 u_2 N t^2\right). \tag{5}$$

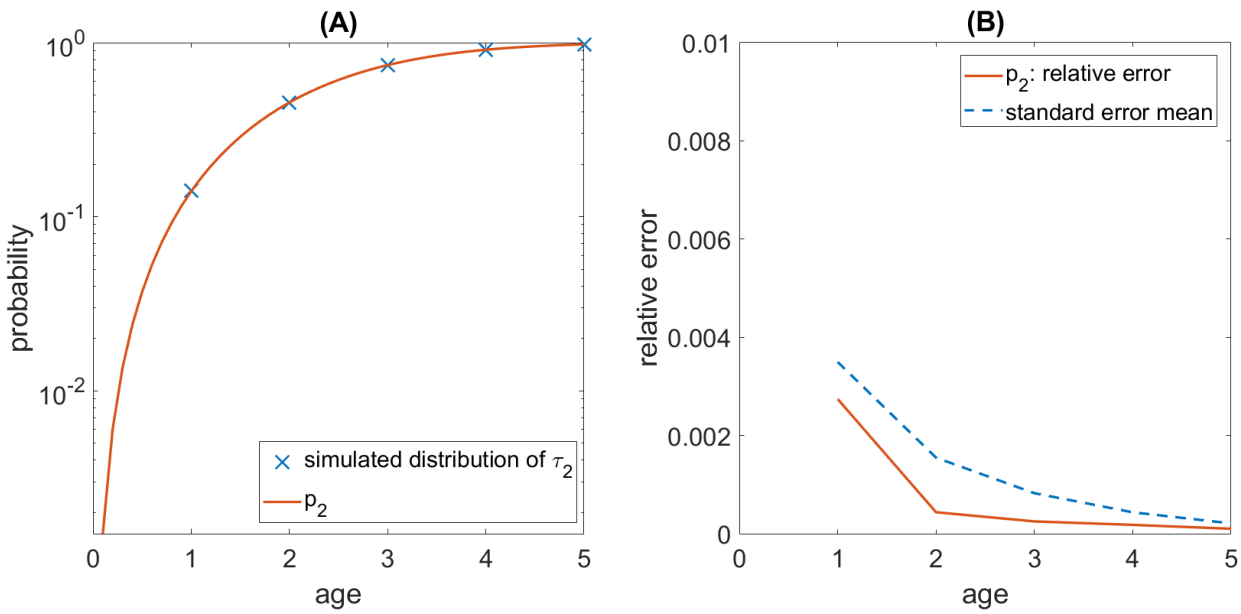

Figure 1: (A) Comparison of the analytic cumulative distribution function of τ_2, the waiting time to the first type-2 crypt (p_2, equation (5)), and the simulated distribution of τ_2 across 5×10^5 runs. (B) Dashed line shows the standard error of the mean obtained from simulations. Solid line is the relative error of the analytic result.

5.2. Waiting time to type-3

To compute the distribution functions of τ_3, τ_4 and τ_5, we need to apply the approximation in Proposition 3.4. Hence, we want to find f_2, f_3. Recall that we consider two functions for each sub-population:

$$f_{i0}(t) := e^{\lambda_i t}, \;\; f_{i1}(t) := \frac{\mathbb{E}[N_i(t)]}{\mathbb{E}[V_i]}.$$

The approximation for the distribution τ_3 is

$$P(\tau_3 \leq t) \approx 1 - \mathcal{L}_{V_2}\left(u_3 \int_0^t f_2(s)ds\right),$$

where $f_2(t) \in \{f \in C(\mathbb{R}) : f^{-1}(t)e^{\lambda_2 t} \to 1\}$. First, consider $f_{20}(t) = e^{\lambda_2 t}$. The corresponding distribution function is

$$p_{30}(t) = 1 - \exp\left(\frac{Nu_1u_2}{\lambda_2^2}\text{PolyLog}\left(2, -u_3\frac{e^{\lambda_2 t}-1}{\lambda_2}\right)\right). \tag{6}$$

In the second case,

$$f_{21}(t) = \frac{\mathbb{E}[N_2(t)]}{\mathbb{E}[V_2]} = e^{\lambda_2 t} - \lambda_2 t - 1,$$

and the corresponding distribution is

$$p_{31}(t) = 1 - \exp\left(\frac{Nu_1u_2}{\lambda_2^2}\text{PolyLog}\left(2, -u_3\left(\frac{e^{\lambda_2 t}-1}{\lambda_2} - \frac{1}{2}t^2 - t\right)\right)\right). \tag{7}$$

By design, the first moment of $N_{21}^*(t)$ is identical to that of $N_2(t)$,

$$\mathbb{E}[N_{21}^*(t)] = \mathbb{E}[f_{21}(t)V_i] = \mathbb{E}[N_2(t)].$$

Both (6) and (7) agree with the simulated distributions for $t > 40$ (See Fig 2). However, in the intermediate regime where t is small, one can observe that p_{31} is more accurate than p_{30}.

5.3. Waiting time to type-4

Considering Proposition 3.6, we have several options to compute τ_4, the waiting time to the first type-4 crypt. Specifically, we may compute the distribution of τ_4 in two ways: (i) compute the distribution directly using $\mathcal{L}_{V_3}(\theta)$; and (ii) skip V_3 and compute the distribution using $\mathcal{L}_{V_2}(\theta)$ and $p_0^{3\to 4}(s,t)$. In these approaches, by Corollary 4.8, the time deterministic growth functions f_2, f_3 can be any functions that have large time limits identical to the corresponding exponential functions, $e^{\lambda_2 t}$ and $e^{\lambda_3 t}$. Among different choices of the growth functions, we typically consider two specific cases, $f_i = f_{i0} = e^{\lambda_i t}$ and $f_i = f_{i1} = \mathbb{E}[N_i]/\mathbb{E}[V_i]$.

If V_3 is not skipped, $\mathcal{L}_{V_3}$ depends on the choice of f_2. Since f_{21} leads to more accurate computations of τ_2, we use f_{21} to compute $\mathcal{L}_{V_3}(\theta)$, that is

$$\mathcal{L}_{V_3}(\theta) = \mathcal{L}_{V_2}\left(u_3 \int_0^\infty \frac{f_{21}(s)\theta}{\theta + e^{\lambda_3 s}}\mathrm{d}s\right).$$

For f_3, we consider two candidate functions,

$$f_{30}(t) = e^{\lambda_3 t},$$
$$f_{31}(t) = \frac{\mathbb{E}[N_3(t)]}{\mathbb{E}[V_3]} = e^{\lambda_3 t} - \lambda_3 t - 1 - \frac{\lambda_3^2}{\lambda_2^2}(e^{\lambda_2 t} - \lambda_2 t - 1).$$

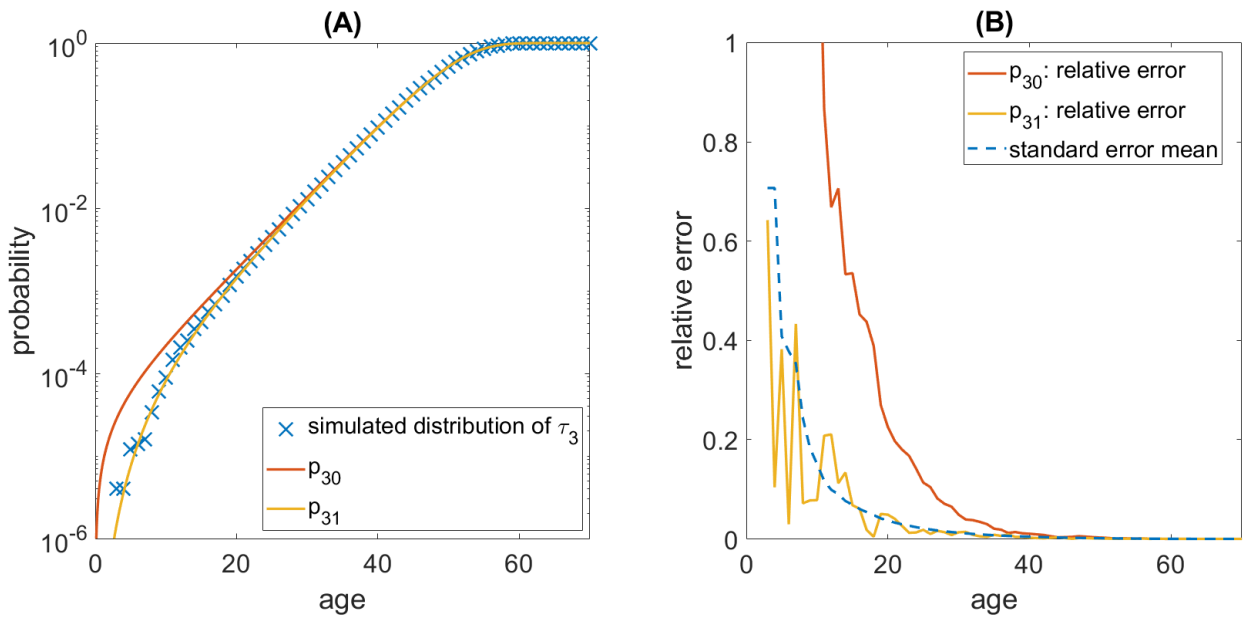

Figure 2: (A) Comparison of the analytic cumulative distribution functions of τ_3, the waiting time to the first type-3 crypt, and the distribution of τ_3 across 5×10^5 simulation runs. p_{30} and p_{31} are distribution functions obtained from equations (6) and (7) respectively. (B) Dashed line shows the standard error of the mean of the simulation. Solid lines are the relative errors of the analytic results.

The first, $f_{30}(t)$, has a simpler form corresponding to the leading order of $\mathbb{E}[N_3(t)]$, with

$$\mathbb{E}[N_{30}^*(t)] = \mathbb{E}[f_{30}(t)V_3] \neq \mathbb{E}[N_3(t)],$$

and the second, $f_{31}(t)$, matches the expected value of $\mathbb{E}[N_3(t)]$

$$\mathbb{E}[N_{31}^*(t)] = \mathbb{E}[f_{31}(t)V_3] = \mathbb{E}[N_3(t)].$$

Next, we define

$$p_{40}(t) = 1 - \mathcal{L}_{V_3}\left(u_4 \int_0^t f_{30}(s)ds\right), \tag{8}$$

$$p_{41}(t) = 1 - \mathcal{L}_{V_3}\left(u_4 \int_0^t f_{31}(s)ds\right). \tag{9}$$

On the other hand, if we skip V_3 and use the inhomogeneous approximation in section 3.4, we can consider two types of f_2's:

$$p_{40}^{s3}(t) = 1 - \mathcal{L}_{V_2}\left(u_4 \int_0^t f_{20}(s)(1 - p_0^{3\to 4}(s,t))ds\right), \tag{10}$$

$$p_{41}^{s3}(t) = 1 - \mathcal{L}_{V_2}\left(u_4 \int_0^t f_{21}(s)(1 - p_0^{3\to 4}(s,t))ds\right). \tag{11}$$

We note that p_{40}^{s3} and p_{41}^{s3} have closed form expressions, which are omitted here and listed in the Appendix. The results are shown in Fig 3. We observe that skipping V_3 improves the accuracy of the results for large time t, while including more terms in f_i generally improves the accuracy of the approximation.

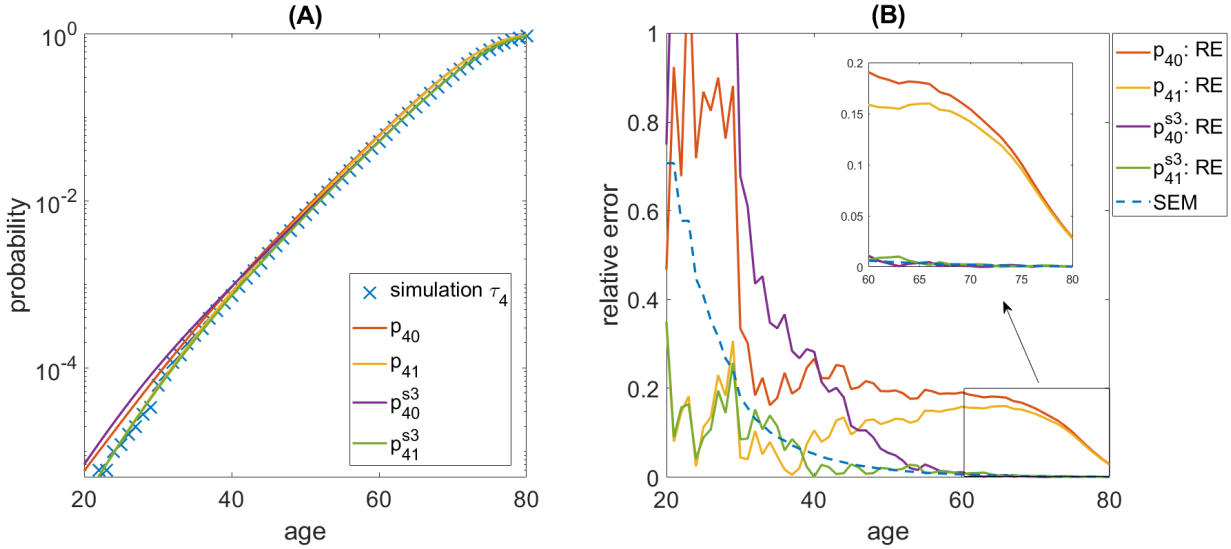

Figure 3: (A) Comparison of analytic cumulative distribution functions of τ_4, the waiting time to the first type-4 crypt, and the distribution of τ_4 across 5×10^5 simulation runs. p_{40} (equation (8)) and p_{41} (equation (9)) are analytic distribution functions derived from the Laplace transform V_3. p_{40}^{s3} (equation (10)) and p_{41}^{s3} (equation (11)) are analytic distribution functions derived by skipping V_3 and using the Laplace transform of V_2. (B) Dashed line shows the standard error of the mean of the simulation. Solid lines are the relative errors of the analytic results.

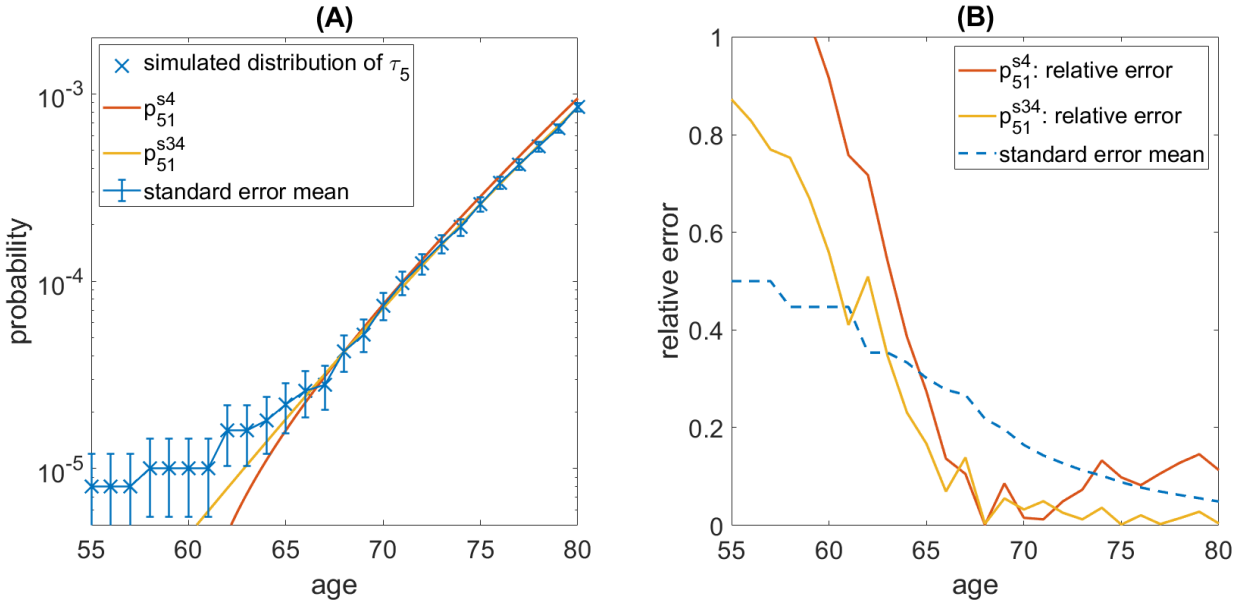

Figure 4: (A) Comparison of the analytic cumulative distribution functions of τ_5, the waiting time to the first type-5 crypt, and the distribution of τ_5 across 5×10^5 simulation runs. p_{51}^{s3} (equation (12)) is derived by skipping type-4 and using the Laplace transform of V_3. p_{51}^{s34} (equation (13)) is derived by using the Laplace transform of V_2 and skipping the Laplace transforms of type-4 and type-3. The error bars represent the standard error of the mean of the simulation. (B) Dashed line shows the standard error of the mean of the simulation. Solid lines are the relative errors of the analytic results.

5.4. Waiting time to type-5

For τ_5, the waiting time to the first type-5 crypt, we consider two approaches: (i) compute the distribution using the large-time limit of type-3 crypts, $\mathcal{L}_{V_3}(\theta)$, and $p_0^{4\to5}(s,t)$ (effectively skipping the large-time limit of type-4 crypts); and (ii) skip large-time limits of both type-4 and type-3 crypts, and compute the distribution using $\mathcal{L}_{V_2}(\theta)$ and $p_0^{3\to5}(s,t)$. Note that the time-deterministic growth functions f_2 and f_3 which appear in the long-time limits of type-2 and type-3 crypts can vary across these approaches (see Corollary 4.8). However, the most accurate results are obtained when $f_i = f_{i1} = \mathbb{E}[N_i]/\mathbb{E}[V_i]$.

The corresponding approximations of the distribution of τ_5 resulting from approaches (i) and (ii) are

$$p_{51}^{s4}(t) = 1 - \mathcal{L}_{V_3}\left(u_4 \int_0^t f_{31}(s)(1 - p_0^{4\to5}(s,t))ds\right), \tag{12}$$

$$p_{51}^{s34}(t) = 1 - \mathcal{L}_{V_2}\left(u_3 \int_0^t f_{21}(s)(1 - p_0^{3\to5}(s,t))ds\right). \tag{13}$$

The first result, $p_{51}^{s4}(t)$, has an explicit form given by equation (B.1), while the second result, p_{51}^{s34} (given by equation (B.3)), to the best of our knowledge, is not an elementary function or a standard special function. Comparison of these two formulas with the waiting time for the first type-5 crypt obtained from exact computer simulations is shown in Fig 4. We observe that p_{51}^{s34} achieves higher accuracy compared to p_{51}^{s3}, especially at later times (above age 70). In other words, compared with the result incorporating V_3, skipping this stage gives more accurate results. The intuition behind this is that approximating each subclone by its large time limit is more accurate than approximating the total population by its overall large time limit.

Appendix

A. Proofs

Proof of Lemma 3.3.

Proof. The proof follows that of Theorem 2 in [10]. The only difference is that we have derived a new bound on the martingale. By Lemma 1 in [10],

$$M(t) = e^{-\lambda_i t} N_i(t) - \int_0^t u_i N_{i-1}(s) e^{-\lambda_i s} \mathrm{d}s$$

is a martingale. If I_i has a finite expectation, then by the martingale convergence theorem (Theorem 4.2.11 in [31]), the submartingale $X(t) = -M(t)$ converges a.s. to some integrable limit X as $t \to \infty$. Since

$$I_i(t) = \int_0^t u_i N_{i-1}(s) e^{-\lambda_i s} \mathrm{d}s \overset{a.s.}{\to} I_i,$$

we also have

$$e^{-\lambda_i t} N_i(t) \overset{a.s.}{\to} W_i.$$

The martingale starts at zero (i.e. $M(0) = 0$), which implies

$$\mathbb{E}[e^{-\lambda_i t} N_i(t)] = \mathbb{E}[I_i(t)].$$

Suppose $e^{-\lambda_i t} N_i(t)$ is uniform integrable, we have (Theorem 4.6.3 in [31])

$$e^{-\lambda_i t} N_i(t) \xrightarrow{L^1} W_i,$$

which guarantees

$$\mathbb{E}[I_i(t)] = \mathbb{E}[e^{-\lambda_i t} N_i(t)] \to \mathbb{E}[W_i].$$

Thus, we have

$$\mathbb{E}[I_i(t)] \to \mathbb{E}[I_i],$$

and

$$\mathbb{E}[W_i] = \mathbb{E}[I_i].$$

□

Proof of Lemma 3.5.

Proof. To get the Laplace transform of V_i, we start from the star process approximation

$$N^*_{i-1}(t) = f_{i-1}(t) V_{i-1}.$$

Applying Lemma 3.1 to the 2-type process $\left(N^*_{i-1}(t), \tilde{N}_i(t)\right)$ yields

$$\mathbb{E}[e^{-\theta \tilde{N}_i(t)} | V_{i-1}] = \exp\left(-u_i \int_0^t f_{i-1}(s) V_{i-1} (1 - \phi_i(\theta, t-s)) \mathrm{d}s\right).$$

Replacing θ with $\theta e^{-\lambda_i t}$,

$$\mathbb{E}[e^{-\theta e^{-\lambda_i t} \tilde{N}_i(t)} | V_{i-1}] = \exp\left(-u_i \int_0^t f_{i-1}(s) V_{i-1} (1 - \phi_i(e^{-\lambda_i t}\theta, t-s)) \mathrm{d}s\right), \tag{A.1}$$

then for each subclone of N_i, by equation (4), we have

$$Z_i(t-s) e^{-\lambda_i (t-s)} \to \text{Exponential}(1) \ a.s.$$

Thus, it follows

$$Z_i(t-s) e^{-\lambda_i t} \to \text{Exponential}(e^{\lambda_i s}).$$

Now consider the following limit involving terms on the right hand side of (A.1),

$$\begin{aligned}
\lim_{t\to\infty} 1 - \phi_i(\theta e^{-\lambda_i t}, t-s) &= 1 - \int_0^\infty e^{-\theta x} e^{\lambda_i s} \exp(-x e^{\lambda_i s}) \mathrm{d}x \\
&= \int_0^\infty (1 - e^{-\theta x}) e^{\lambda_i s} \exp(-x e^{\lambda_i s}) \mathrm{d}x \\
&= \frac{\theta}{\theta + e^{\lambda_i s}}.
\end{aligned}$$

And note that as $t \to \infty$, the left hand side of (A.1) yields

$$\begin{aligned}\lim_{t\to\infty} \mathbb{E}[e^{-\theta e^{-\lambda_i t}\tilde{N}_i(t)}|V_{i-1}] &= \mathbb{E}[\lim_{t\to\infty} e^{-\theta e^{-\lambda_i t}\tilde{N}_i(t)}|V_{i-1}] \\ &= \mathbb{E}[e^{-\theta V_i}|V_{i-1}],\end{aligned}$$

in which switching the limit and the integration is allowed by the dominated convergence theorem. Thus, we can write

$$\mathbb{E}[e^{-\theta V_i}|V_{i-1}] = \exp\left(-u_i V_{i-1} \int_0^\infty \frac{\theta f_{i-1}(s)}{\theta + e^{\lambda_i s}} \mathrm{d}s\right).$$

Taking expectation on both sides gives us

$$\mathbb{E}[e^{-\theta V_i}] = \mathcal{L}_{V_{i-1}}\left(u_i \int_0^\infty \frac{\theta f_{i-1}(s)}{\theta + e^{\lambda_i s}}\, \mathrm{d}s\right).$$

□

Proof of Lemma 3.7

Proof. For a type-j lineage that appears at time s, we approximate it by its large time limit. At time $r > s$, the population of this lineage would be $Z_j(r) \approx e^{\lambda_j(r-s)}V$, where $V \sim$ Exponential(1). Thus, type-$(j+1)$ individuals that mutated from this lineage are produced at rate

$$u_{j+1} Z_j(r) \approx u_{j+1} e^{\lambda_j(r-s)} V.$$

The probability for a type-$(j+1)$ individual to produce at least a type-i individual is $1 - p_0^{(j+1)\to i}(r,t)$. Thus, conditional on V, the expected number of type-i individuals that were produced by a type-j lineage that appeared at time s is

$$\begin{aligned}\Lambda^{j\to i}(s,t,V) &= \int_s^t u_{j+1} Z_j(r) \left(1 - p_0^{(j+1)\to i}(r,t)\right) \mathrm{d}r \\ &\approx \int_s^t u_{j+1} e^{\lambda_j(r-s)}\, V \left(1 - p_0^{(j+1)\to i}(r,t)\right) \mathrm{d}r.\end{aligned}$$

Let $X^{j\to i}(s,t)$ be the number of type-i individuals that are produced by a type-j subclone appeared at s. In the time period $[s,t]$, $X^{j\to i}(\cdot,t)$ follows an inhomogeneous Possion process with mean $\Lambda(\cdot,t,V)$. Thus the probability that no type-i crypt is produced from this particular type-$(j+1)$ crypt by time t is

$$\begin{aligned}P(X^{j\to i}(s,t) = 0|V) &= \exp(-\Lambda^{j\to i}(s,t,V)) \\ &= \exp\left(-V \int_s^t u_{j+1} e^{\lambda_j(r-s)} \left(1 - p_0^{(j+1)\to i}(r,t)\right) \mathrm{d}r\right).\end{aligned}$$

This implies

$$\begin{aligned}p_0^{j\to i}(r,t) &:= P(X^{j\to i}(s,t) = 0) \\ &= \mathbb{E}\left[P(X^{j\to i}(s,t) = 0|V)\right] \\ &= \int_0^t e^{-v} \exp\left(-v \int_s^t u_{j+1} e^{\lambda_j(r-s)} \left(1 - p_0^{(j+1)\to i}(r,t)\right) dr\right) dv.\end{aligned}$$

Finally, for $i = j$, notice that the founding individual of a type-j lineage is of type-j. This guarantees that at any time t greater than the founding time s, the probability of having at least one type-j individual is 1. Thus $p_0^{j\to j}(s,t) = 0$. □

Proof of Theorem 4.3.

Proof. By Lemma 3.3, consider

$$I_2 = \int_0^\infty u_2 N_1(s) e^{-\lambda_2 s} \mathrm{d}s.$$

Since

$$\begin{aligned}
\mathbb{E}[I_2] =& \mathbb{E}\left[\int_0^\infty u_2 e^{-\lambda_2 s} N_1(s) \mathrm{d}s\right] \\
\text{(By Tonelli's theorem)} =& \int_0^\infty u_2 \mathbb{E}[N_1(s)] e^{-\lambda_2 s} \mathrm{d}s \\
\leq& \int_0^\infty u_2 u_1 N s e^{-\lambda_2 s} \mathrm{d}s \\
=& N u_1 u_2 \int_0^\infty s e^{-\lambda_2 s} \mathrm{d}s \\
=& \frac{N u_1 u_2}{\lambda_2^2} < \infty,
\end{aligned}$$

there exists W_2 s.t.

$$e^{-\lambda_2 t} N_2(t) \to W_2 \ a.s. \ \text{ as } t \to \infty.$$

Next, we show uniform integrability so that the $\mathbb{E}[W_2]$ is well-defined. We prove this for N_1, N_2 and N_3 in Lemma A.1. Since L^1 convergence is guaranteed, $\mathbb{E}[W_2] = \mathbb{E}[I_2]$, so that

$$\begin{aligned}
\mathbb{E}[W_2] &= \int_0^\infty u_2 \mathbb{E}[N_1(s)] e^{-\lambda_2 s} \mathrm{d}s \\
&= \int_0^\infty u_2 N (1 - e^{-u_1 s}) e^{-\lambda_2 s} \mathrm{d}s \\
&= \frac{N u_1 u_2}{\lambda_2(\lambda_2 + u_1)}.
\end{aligned}$$

□

Proof of Theorom 4.4.

Proof. Note that I_2^* is deterministic and has a finite expected value

$$\mathbb{E}[I_2^*] = I_2^* = \int_0^\infty u_2 u_1 N s e^{-\lambda_2 s} \mathrm{d}s = \frac{N u_1 u_2}{\lambda_2^2} < \infty.$$

By Lemma 3.3, we must have $e^{-\lambda_2 t} \tilde{N}_2(t) \overset{a.s.}{\to} V_2$. Next, we show uniform integrability which guarantees L^1 convergence. It is shown in Lemma A.2 that all tilde processes in this paper are uniform integrable. This implies $e^{-\lambda_2 t} \tilde{N}_2(t) \overset{L^1}{\to} V_2$ and $\mathbb{E}[V_2] = \mathbb{E}[I_2^*] = \frac{N u_1 u_2}{\lambda_2^2}$. Finally, to compute the Laplace transform of V_2, we plug $f_2(t) = e^{\lambda_2 t}$ into the formula in 3.5. □

Proof of Corollary 4.5.

Proof. The convergence directly follows Theorem 4.4 by the fact that

$$\forall\ \omega \in \{\omega : \lim_{t\to\infty} e^{-\lambda_2 t}\tilde{N}_2(\omega,t) = V_2(\omega)\},$$

$$\lim_{t\to\infty} \frac{\tilde{N}_2(\omega,t)}{f_2(t)} = \lim_{t\to\infty} e^{-\lambda_2 t} f_2(t) \lim_{t\to\infty} \frac{\tilde{N}_2(\omega,t)}{f_2(t)} = V_2(\omega).$$

Observe that for $t > 0$ sufficiently large, we have $f_2(t) > 0$ and $e^{\lambda_2 t}/f_2(t) < M$. Note that

$$\sup_t \mathbb{E}\left|\frac{\tilde{N}_2(t)}{f_2(t)}\right| \le \sup_t \left(e^{\lambda_2 t}/f_2(t)\right) \sup_t E\left[\tilde{N}_2(t)e^{-\lambda_2 t}\right],$$

and by Lemma A.2 $\sup_t \mathbb{E}\left[\tilde{N}_2(t)e^{-\lambda_2 t}\right]$ is bounded. Hence $\frac{\tilde{N}_2(t)}{f_2(t)}$ is uniform integrable and the convergence is in L^1. □

Proof of Theorem 4.6.

Proof. By Lemma 3.3 and Lemma A.1, we need to verify that I_3 has finite expectation.

$$\begin{aligned}
\mathbb{E}[I_3] &= \int_0^\infty u_3 \mathbb{E}[N_2(s)] e^{-\lambda_3 s} \mathrm{d}s \\
&\le \int_0^\infty u_3 \frac{u_1 u_2 N}{\lambda_2^2}(e^{\lambda_2 s} - \lambda_2 s - 1)e^{-\lambda_3 s}\mathrm{d}s \\
&= \frac{N u_1 u_2 u_3}{\lambda_2^2}\int_0^\infty (e^{(\lambda_2-\lambda_3)s} - \lambda_2 s e^{-\lambda_3 s} - e^{-\lambda_3 s})\mathrm{d}s \\
&= \frac{N u_1 u_2 u_3}{(\lambda_3-\lambda_2)\lambda_3^2} < \infty.
\end{aligned}$$

And the expected value of W_3 is given by the expected value of I_3,

$$\begin{aligned}
\mathbb{E}[W_3] &= \mathbb{E}[I_3] \\
&= \int_0^\infty u_3 \mathbb{E}[N_2(s)] e^{-\lambda_3 s}\mathrm{d}s \\
&= \int_0^\infty u_3 N u_2 \frac{u_1(e^{\lambda_2 s} - 1) + \lambda_2(e^{-u_1 s} - 1)}{\lambda_2(\lambda_2 + u_1)} e^{-\lambda_3 s}\mathrm{d}s \\
&= \frac{N u_1 u_2 u_3}{(\lambda_3 - \lambda_2)\lambda_3(\lambda_3 + u_1)}.
\end{aligned}$$

□

Proof of Theorem 4.7.

Proof. By Lemma 3.3 and Lemma A.2, we compute

$$\begin{aligned}\mathbb{E}[I_3^*] &= \int_0^\infty u_3 E[N_2^*(s)]e^{-\lambda_3 s}\mathrm{d}s \\ &= \int_0^\infty u_3 \frac{u_1 u_2 N}{\lambda_2^2} e^{\lambda_2 s} e^{-\lambda_3 s}\mathrm{d}s \\ &= \frac{N u_1 u_2 u_3}{\lambda_2^2}\int_0^\infty e^{(\lambda_2-\lambda_3)s}\mathrm{d}s \\ &= \frac{N u_1 u_2 u_3}{\lambda_2^2(\lambda_3-\lambda_2)} < \infty.\end{aligned}$$

□

Proof of Corollary 4.8.

Proof. Since $f_2(t)e^{-\lambda_3 t}$ is integrable, the validity of the convergence and expectation follow directly from Theorem 4.7. Lemma 3.5 gives the Laplace transform of V_3. □

Lemma A.1. *For* $i \le 3, \sup_t \mathbb{E}\left[(e^{-\lambda_i t}N_i(t))^2\right] < \infty$.

Proof. By Lemma 5 in [10], we know that inductively, if $\sup_t \mathbb{E}[e^{-\lambda_i t}N_i(t)]^2 < \infty$ and $\lambda_i < \lambda_{i+1}$ holds, then $\sup_t (e^{-\lambda_{i+1}t}N_{i+1})^2 < \infty$. In this case, since $0 = \lambda_0 = \lambda_1 < \lambda_2 < \lambda_3$, we only need to show that

$$\sup_t \mathbb{E}(N_0(t))^2 < \infty \text{ and, } \sup_t \mathbb{E}(N_1(t))^2 < \infty.$$

Note that by our transition scheme, one can have

$$N_0(t) + N_1(t) \le N$$

. Therefore

$$\max\left(\sup_t N_0^2(t), \sup_t N_1^2(t)\right) < N^2 < \infty.$$

□

Lemma A.2. *For* $i \in \{2,3\}$, $\sup_t \mathbb{E}\left[(e^{-\lambda_i t}\tilde{N}_i(t))^2\right] < \infty$.

Proof. For $i = 2$, recall that $\tilde{N}_2(t)$ is the second type in a two-type branching process where the first type is $N_1^*(t) = u_1 N t$. And $\tilde{N}_2(t)$ is produced at rate $u_2 N_1^*(t)$. By manipulating the master equation of this two-type process, we obtain the following differential equation of $\mathbb{E}\left[\tilde{N}_2(t)^2\right]$,

$$\frac{\mathrm{d}\mathbb{E}\left[\tilde{N}_2(t)^2\right]}{dt} = 2\lambda_2 \mathbb{E}\left[\tilde{N}_2(t)^2\right] + (\lambda_2 + 2u_2u_1Nt)\mathbb{E}[\tilde{N}_2(t)] + u_2u_1Nt$$

subject to $\mathbb{E}\left[\tilde{N}_2(0)^2\right] = 0$. The solution is

$$\mathbb{E}\left[\tilde{N}_2(t)^2\right] = e^{2\lambda_2 t}\int_0^t e^{-2\lambda_2 s}\left((\lambda_2 + 2u_2u_1Ns)\mathbb{E}[\tilde{N}_2(s)] + u_2u_1Ns\right)\mathrm{d}s.$$

Note that

$$\mathbb{E}[\tilde{N}_2(t)] = \frac{Nu_1u_2(e^{\lambda_2 t} - \lambda_2 t - 1)}{\lambda_2^2} \leq \frac{u_2u_1N}{\lambda_2^2}e^{\lambda_2 t}.$$

Thus,

$$\mathbb{E}\left[\tilde{N}_2(t)^2\right] \leq e^{2\lambda_2 t}\int_0^t \left(\frac{2u_2^2u_1^2N^2}{\lambda_2^2}se^{-\lambda_2 s} + \frac{u_1u_2N}{\lambda_2}e^{-\lambda_2 s} + u_2u_1Nse^{-2\lambda_2 s}\right)\mathrm{d}s.$$

Using the fact that for $a > 0$, $\int_0^t se^{-as}\mathrm{d}s \leq 1/a^2$ and $\int_0^t e^{-as}\mathrm{d}s \leq 1/a$, we get $\sup_t \mathbb{E}\left[(e^{-\lambda_2 t}\tilde{N}_2(t))^2\right] < \infty$.

For $i = 3$, $\tilde{N}_3(t)$ is produced at rate $u_3N_2^*(t) = u_3e^{\lambda_2 t}V_2$. By the master equation we get the following differential equation:

$$\frac{\mathrm{d}\mathbb{E}\left[\tilde{N}_3(t)^2|V_2\right]}{dt} = 2\lambda_3\mathbb{E}\left[\tilde{N}_3(t)^2|V_2\right] + (\lambda_3 + 2u_3e^{\lambda_2 t}V_2)\mathbb{E}[\tilde{N}_3(t)|V_2] + u_3e^{\lambda_2 t}V_2$$

subject to $\mathbb{E}\left[\tilde{N}_3(0)^2|V_2\right] = 0$. The solution is

$$\mathbb{E}\left[\tilde{N}_3(t)^2|V_2\right] = e^{2\lambda_3 t}\int_0^t e^{-2\lambda_3 s}\left((\lambda_3 + 2u_3e^{\lambda_2 s}V_2)\mathbb{E}[\tilde{N}_3(s)|V_2] + u_3e^{\lambda_2 s}V_2\right)\mathrm{d}s.$$

Note that

$$\mathbb{E}[\tilde{N}_3(t)|V_2] = \frac{u_3V_2}{\lambda_3 - \lambda_2}(e^{\lambda_3 t} - e^{\lambda_2 t}) \leq \frac{u_3V_2}{\lambda_3 - \lambda_2}e^{\lambda_3 t}.$$

Thus we have

$$\begin{aligned} E\left[\tilde{N}_3(t)^2|V_2\right] &\leq e^{2\lambda_3 t}\int_0^t \left(\frac{2u_3^2V_2^2}{\lambda_3 - \lambda_2}e^{-(\lambda_3-\lambda_2)s} + \frac{\lambda_3u_3V_2}{\lambda_3 - \lambda_2}e^{-\lambda_3 s} + u_3e^{-(2\lambda_3-\lambda_2)s}V_2\right)\mathrm{d}s \\ &\leq e^{2\lambda_3 t}\left(\frac{2u_3^2V_2^2}{(\lambda_3 - \lambda_2)^2} + \frac{\lambda_3u_3V_2}{\lambda_3(\lambda_3 - \lambda_2)} + \frac{u_3V_2}{(2\lambda_3 - \lambda_2)}\right). \end{aligned}$$

By Lemma 3.5, one can compute the moments of V_2 by it's Laplace transform. Then we get

$$\mathbb{E}[V_2] = \frac{Nu_1u_2}{\lambda_2^2}, \ \ \mathbb{E}[V_2^2] = \frac{Nu_1u_2(2Nu_1u_2 + \lambda_2^2)}{\lambda_2^4}$$

Hence we conclude that $\sup_t \mathbb{E}\left[(e^{-\lambda_3 t}\tilde{N}_3(t))^2\right] < \infty$. □

B. Closed from formulas of analytic distributions

In the main text, we have omitted a few cumbersome formulas to increase readability. Here we present their closed form expressions. We begin with the analytic probability distribution functions of τ_4.

$$\begin{aligned} p_{40}(t) &= 1 - \mathcal{L}_{V_3}\left(u_4\int_0^t f_{30}(s)ds\right) = 1 - \mathcal{L}_{V_3}\left(\frac{u_4}{\lambda_3}(e^{\lambda_3 t} - 1)\right). \\ p_{41}(t) &= 1 - \mathcal{L}_{V_3}\left(u_4\int_0^t f_{31}(s)ds\right) \\ &= 1 - \mathcal{L}_{V_3}\left(u_4\left(\frac{1}{\lambda_3}(e^{\lambda_3 t} - 1) - \frac{\lambda_3^2}{2}t^2 - t - \frac{\lambda_3^2}{\lambda_2^2}(\frac{1}{\lambda_2}(e^{\lambda_2 t} - 1) - \frac{\lambda_2}{2}t^2 - t)\right)\right). \end{aligned}$$

where

$$\begin{aligned}\mathcal{L}_{V_3}(\theta) &= \mathcal{L}_{V_2}\left(u_3\int_0^\infty \frac{f_{21}(s)\theta}{\theta+e^{\lambda_3 s}}\mathrm{d}s\right)\\ &= \exp\left(\frac{Nu_1u_2}{\lambda_2^2}\mathrm{PolyLog}\left(2,-u_3\left(\frac{1}{\lambda_3-\lambda_2}\theta F(1-\frac{\lambda_2}{\lambda_3},1;2-\frac{\lambda_2}{\lambda_3};-\theta)\right.\right.\right.\\ &\quad\left.\left.\left.-\lambda_2\frac{\mathrm{PolyLog}(2,-\theta)}{\lambda_3^2}-\frac{1}{\lambda_3}\log(1+\theta)\right)\right)\right)\end{aligned}$$

In the equation above, $F(a,b;c;z)$ is the hypergeometric function.

The two results of skipping type-4 are

$$\begin{aligned}p_{40}^{s3}(t) &= 1-\mathcal{L}_{V_2}\left(u_4\int_0^t f_{20}(s)(1-p_0^{3\to4}(s,t))ds\right)\\ &= 1-\mathcal{L}_{V_2}\left(\frac{u_4}{\lambda_2(u_4-\lambda_3)}\left(-u_4+\lambda_3F(1,\frac{\lambda_2}{\lambda_3},\frac{\lambda_2}{\lambda_3}+1,\frac{u_4-\lambda_3}{u_4}e^{-\lambda_2T})\right.\right.\\ &\qquad\left.\left.+e^{\lambda_2T}(u_4-\lambda_3F(1,\frac{\lambda_2}{\lambda_3},\frac{\lambda_2}{\lambda_3}+1,1-\frac{\lambda_3}{u_4}))\right)\right),\\ p_{41}^{s3}(t) &= 1-\mathcal{L}_{V_2}\left(u_4\int_0^t f_{21}(s)(1-p_0^{3\to4}(s,t))ds\right)\\ &= 1-\mathcal{L}_{V_2}\left(u_4\left(L_{\lambda_2}^{(4)}(t)-\lambda_2L_l^{(4)}(t)-L_c^{(4)}(t)\right)\right)\end{aligned}$$

where

$$\begin{aligned}p_0^{3\to4}(s,t) &= \frac{1}{1+u_4\frac{\exp(\lambda_3(t-s))-1}{\lambda_3}}\\ \mathcal{L}_{V_2}(\theta) &= \exp\left(Nu_1u_2\frac{\mathrm{PolyLog}(2,-\theta)}{\lambda_2^2}\right)\\ L_{\lambda_2}^{(4)}(t) &:= \int_0^t e^{\lambda_2 s}(1-p_0^{3\to4}(s,t))ds = \frac{1}{\lambda_2(u_4-\lambda_3)}\left(-u_4+\lambda_3F(1,\frac{\lambda_2}{\lambda_3};\frac{\lambda_2}{\lambda_3}+1;\frac{u_4-\lambda_3}{u_4}e^{-\lambda_2t})\right.\\ &\qquad\left.+e^{\lambda_2t}(u_4-\lambda_3F(1,\frac{\lambda_2}{\lambda_3};\frac{\lambda_2}{\lambda_3}+1;1-\frac{\lambda_3}{u_4}))\right)\\ L_l^{(4)}(t) &:= \int_0^t s(1-p_0^{3\to4}(s,t))ds\\ &= \frac{t^2}{2}+\frac{-\lambda_3t\log(\frac{\lambda_3}{u_4})+\mathrm{PolyLog}(2,\frac{-(\lambda_3-u_4)e^{-\lambda_3t}}{u_4})-\mathrm{PolyLog}(2,1-\frac{\lambda_3}{u_4})}{\lambda_3(\lambda_3-u_4)}\\ L_c^{(4)}(t) &:= \int_0^t(1-p_0^{3\to4}(s,t))ds\\ &= \frac{\log(u_4(e^{\lambda_3t}-1)+\lambda_3)-\log(\lambda_3)-u_4t}{\lambda_3-u_4}.\end{aligned}$$

For waiting time distributions of the first type-5 crypt, our estimations are $p_{51}^{s4}(t)$ and

$p_{51}^{s34}(t)$. $p_{51}^{s4}(t)$ can be expressed explicitly.

$$p_{51}^{s4}(t) = 1 - \mathcal{L}_{V_3}\left(u_4 \int_0^t f_{31}(s)(1 - p_0^{4\to 5}(s,t))ds\right) \tag{B.1}$$

$$= \mathcal{L}_{V_3}\left(u_4\left(L_{\lambda_3}^{(5)}(t) - CL_{\lambda_2}^{(5)}(t) - (\lambda_3 - C\lambda_2)L_l^{(5)}(t) + (C-1)L_c^{(5)}(t)\right)\right) \tag{B.2}$$

where

$$\begin{aligned}
L_{\lambda_3}^{(5)}(t) &:= \int_0^t e^{\lambda_3 s}(1 - p_0^{4\to 5}(s,t))ds \\
&= \frac{u_5}{\lambda_3(\lambda_3 - u_5)^2}\left(\lambda_3 - u_5 + e^{\lambda_3 t}\left(-\lambda_3 + u_5 + \lambda_3 \log(\lambda_3 e^{\lambda_3 t}) - \lambda_3 \log(\lambda_3 + u_5(e^{\lambda_3 t} - 1))\right)\right) \\
L_{\lambda_2}^{(5)}(t) &:= \int_0^t e^{\lambda_2 s}(1 - p_0^{4\to 5}(s,t))ds = \frac{1}{\lambda_2(u_5 - \lambda_3)}\left(-u_5 + \lambda_3 F(1, \frac{\lambda_2}{\lambda_3}; \frac{\lambda_2}{\lambda_3} + 1; \frac{u_5 - \lambda_3}{u_5}e^{-\lambda_2 t})\right. \\
&\qquad \left. + e^{\lambda_2 t}(u_5 - \lambda_3 F(1, \frac{\lambda_2}{\lambda_3}; \frac{\lambda_2}{\lambda_3} + 1; 1 - \frac{\lambda_3}{u_5}))\right) \\
L_l^{(5)}(t) &:= \int_0^t s(1 - p_0^{4\to 5}(s,t))ds \\
&= \frac{t^2}{2} + \frac{-\lambda_3 t \log(\frac{\lambda_3}{u_4}) + \mathrm{PolyLog}(2, \frac{-(\lambda_3 - u_4)e^{-\lambda_3 t}}{u_4}) - \mathrm{PolyLog}(2, 1 - \frac{\lambda_3}{u_4})}{\lambda_3(\lambda_3 - u_4)} \\
L_c^{(5)}(t) &:= \int_0^t (1 - p_0^{4\to 5}(s,t))ds \\
&= \frac{\log(u_4(e^{\lambda_3 t} - 1) + \lambda_3) - \log(\lambda_3) - u_4 t}{\lambda_3 - u_4} \\
C &:= \frac{\lambda_3^2}{\lambda_2^2}.
\end{aligned}$$

Define $I(t) =$

$$\int_0^t (e^{\lambda_2 s} - \lambda_2 s - 1)\left(1 - \left(1 + \frac{u_4 u_5(\lambda_3 - u_5 + e^{\lambda_3(t-s)}(u_5 - \lambda_3 + \lambda_3 \log(\frac{\lambda_3 e^{\lambda_3(t-s)}}{\lambda_3 + u_5(e^{\lambda_3(t-s)} - 1)}))}{(\lambda_3 - u_5)^2 \lambda_3}\right)^{-1}\right) ds.$$

Then we can have

$$p_{51}^{s34}(t) = \mathcal{L}_{V_2}(u_3 I(t)) \tag{B.3}$$

Unfortunately, we didn't find a way to solve the integral $I(t)$ explicitly. Nevertheless, we can compute this value numerically.

References

[1] P. Armitage, R. Doll, The age distribution of cancer and a multi-stage theory of carcinogenesis, Br. J. Cancer 8 (1954) 1–12. doi:https://doi.org/10.1038/bjc.1954.1.

[2] P. Armitage, R. Doll, A two-stage theory of carcinogenesis in relation to the age distribution of human cancer, Br. J. Cancer 11 (1957) 161–169. doi:https://doi.org/10.1038/bjc.1957.22.

[3] A. G. Knudson, Mutation and cancer: statistical study of retinoblastoma, Proc. Natl. Acad. Sci. U.S.A. 68 (1971) 820–823. doi:https://doi.org/10.1073/pnas.68.4.820.

[4] L. D. Wood, D. W. Parsons, S. Jones, J. Lin, T. Sjoblom, et al., The genomic landscapes of human breast and colorectal cancers, Science 318 (2007) 1108–1113. doi:https://doi.org/10.1126/science.1145720.

[5] E. Hodis, I. Watson, G. Kryukov, S. Arold, M. Imielinski, et al., A landscape of driver mutations in melanoma, Cell 150 (2012) 251–263. doi:https://doi.org/10.1016/j.cell.2012.06.024.

[6] K. Wang, S. T. Yuen, J. Xu, et al., Whole-genome sequencing and comprehensive molecular profiling identify new driver mutations in gastric cancer, Nat. Genet. 46 (2014) 573–582. doi:https://doi.org/10.1038/ng.2983.

[7] C. W. Brennan, R. G. Verhaak, A. McKenna, B. Campos, H. Noushmehr, et al., The somatic genomic landscape of glioblastoma, Cell 155 (2013) 462–477. doi:https://doi.org/10.1016/j.cell.2013.09.034.

[8] S. Nik-Zainal, H. Davies, J. Staaf, et al., Landscape of somatic mutations in 560 breast cancer whole-genome sequences, Nature 534 (2016) 47–54. doi:https://doi.org/10.1038/nature17676.

[9] T. Antal, P. Krapivsky, Exact solution of a two-type branching process: models of tumor progression, J. Stat. Mech. Theory Exp. 2011 (2011) P08018. doi:https://doi.org/10.1088/1742-5468/2011/08/p08018.

[10] R. Durrett, S. Moseley, Evolution of resistance and progression to disease during clonal expansion of cancer, Theor. Popul. Biol. 77 (2010) 42–48. doi:https://doi.org/10.1016/j.tpb.2009.10.008.

[11] I. Bozic, T. Antal, et al., Accumulation of driver and passenger mutations during tumor progression, Proc. Natl. Acad. Sci. U.S.A. 107 (2010) 18545–18550. doi:https://doi.org/10.1073/pnas.1010978107.

[12] J. Foo, K. Leder, Dynamics of cancer recurrence, Ann. Appl. Probab. 23 (2013) 1437–1468. doi:https://doi.org/10.1214/12-AAP876.

[13] S. Avanzini, T. Antal, Cancer recurrence times from a branching process model, PLoS Comput. Biol. 15 (2019) e1007423. doi:https://doi.org/10.1371/journal.pcbi.1007423.

[14] K. Danesh, R. Durrett, L. J. Havrilesky, E. Myers, A branching process model of ovarian cancer, J. Theor. Biol. 314 (2012) 10–15. doi:https://doi.org/10.1016/j.jtbi.2012.08.025.

[15] I. Bozic, M. A. Nowak, Timing and heterogeneity of mutations associated with drug resistance in metastatic cancers, Proc. Natl. Acad. Sci. U.S.A. 111 (2014) 15964–15968. doi:https://doi.org/10.1073/pnas.1412075111.

[16] M. D. Nicholson, T. Antal, Competing evolutionary paths in growing populations with applications to multidrug resistance, PLoS Comput. Biol. 15 (2019) e1006866. doi:https://doi.org/10.1371/journal.pcbi.1006866.

[17] J. Foo, K. Leder, J. Zhu, Escape times for branching processes with random mutational fitness effects, Stoch. Process. Their Appl. 124 (2014) 3661–3697. doi:https://doi.org/10.1016/j.spa.2014.06.003.

[18] R. L. Siegel, K. D. Miller, H. E. Fuchs, A. Jemal, Cancer statistics, 2021, CA Cancer J. Clin. 71 (2021) 7–33. doi:https://doi.org/10.3322/caac.21654.

[19] C. Tomasetti, L. Marchionni, M. A. Nowak, G. Parmigiani, B. Vogelstein, Only three driver gene mutations are required for the development of lung and colorectal cancers, Proc. Natl. Acad. Sci. U.S.A. 112 (2015) 118–123. doi:https://doi.org/10.1073/pnas.1421839112.

[20] E. R. Fearon, Molecular genetics of colorectal cancer, Annu. Rev. Pathol. Mech. Dis. 6 (2011) 479–507. doi:https://doi.org/10.1146/annurev-pathol-011110-130235.

[21] C. Paterson, H. Clevers, I. Bozic, Mathematical model of colorectal cancer initiation, Proc. Natl. Acad. Sci. U.S.A. 117 (2020) 20681–20688. doi:https://doi.org/10.1073/pnas.2003771117.

[22] C. J. Sherr, Principles of tumor suppression, Cell 116 (2004) 235–246. doi:https://doi.org/10.1016/S0092-8674(03)01075-4.

[23] L. Vermeulen, E. Morrissey, M. van der Heijden, A. M. Nicholson, A. Sottoriva, et al., Defining stem cell dynamics in models of intestinal tumor initiation, Science 342 (2013) 995–998. doi:https://doi.org/10.1126/science.1243148.

[24] N. Barker, R. A. Ridgway, J. H. van Es, et al., Crypt stem cells as the cells-of-origin of intestinal cancer, Nature 457 (2009) 608–611. doi:https://doi.org/10.1038/nature07602.

[25] A. M. Nicholson, C. Olpe, A. Hoyle, A.-S. Thorsen, T. Rus, et al., Fixation and spread of somatic mutations in adult human colonic epithelium, Cell Stem Cell 22 (2018) 909–918.e8. doi:https://doi.org/10.1016/j.stem.2018.04.020.

[26] H. Lamlum, A. Papadopoulou, M. Ilyas, A. Rowan, C. Gillet, et al., APC mutations are sufficient for the growth of early colorectal adenomas, Proc. Natl. Acad. Sci. U.S.A. 97 (2000) 2225–2228. doi:https://doi.org/10.1073/pnas.040564697.

[27] H. J. Snippert, A. G. Schepers, J. H. van Es, B. D. Simons, H. Clevers, Biased competition between lgr5 intestinal stem cells driven by oncogenic mutation induces clonal expansion, EMBO Rep. 15 (2014) 62–69. doi:https://doi.org/10.1002/embr.201337799.

[28] D. T. Gillespie, Exact stochastic simulation of coupled chemical reactions, J. Phys. Chem. 81 (1977) 2340–2361. doi:https://doi.org/10.1021/j100540a008.

[29] D. Cheek, T. Antal, Mutation frequencies in a birth–death branching process, Ann. Appl. Prob. 28 (2018) 3922–3947. doi:https://doi.org/10.1214/18-AAP1413.

[30] K. B. Athreya, P. E. Ney, Branching processes, Dover Publications, New York, 2004.

[31] R. Durrett, Probability: Theory and Examples, fifth ed., Cambridge University Press, Cambridge; New York, 2019.