S1 Text
Mathematical modeling the order of driver gene mutations in colorectal cancer

by Lingling Li, Yulu Hu, Yunshan Xu, Sanyi Tang
S1 Supplementary Figures

For the sequence KRAS → APC → TP53 (A), the detailed gene alterations are depicted in Fig A. Because the net growth rate ($\lambda_j$) remains the same whether APC comes first or TP53 comes first, only five models are showed based on the clonal expansion of cells in Fig A. In fact, there are eight pathways for the order with KRAS → APC → TP53, and Figs A3-A5 represent two pathways, respectively. It is similar for other orders. The detailed gene alterations are depicted in Figs B–F for the sequences KRAS → TP53 → APC (B), APC → TP53 → KRAS (C), APC → KRAS → TP53 (D), TP53 → APC → KRAS (E) and TP53 → KRAS → APC (F), respectively.

Fig A. All pathways of gene mutations for the case (A) in colorectal cancer.
Fig B. All pathways of gene mutations for the case (B) in colorectal cancer.

Fig C. All pathways of gene mutations for the case (C) in colorectal cancer.
Fig D. All pathways of gene mutations for the case (D) in colorectal cancer.

Fig E. All pathways of gene mutations for the case (E) in colorectal cancer.

Fig F. All pathways of gene mutations for the case (F) in colorectal cancer.
S2 Mathematical derivation

The derivation of approximate solution

For the model with five stages in Fig 1 of the text, we define the following probability generating function

\[ \Psi(x_1, x_2, x_3, x_4, y; 0, t) = \sum_{i_1, i_2, i_3, i_4} \text{prob}(P_1(t) = i_1, P_2(t) = i_2, P_3(t) = i_3, P_4(t) = i_4) \times x_1^{i_1} x_2^{i_2} x_3^{i_3} x_4^{i_4}. \] (S1)

Let \( P_{i_1,i_2,i_3,i_4,j}(t) = \text{prob}(P_1(t) = i_1, P_2(t) = i_2, P_3(t) = i_3, P_4(t) = i_4, M(t) = j) P_1(0) = 0, P_2(0) = 0, P_3(0) = 0, P_4(0) = 0, M(0) = 0 \), which satisfies the following Kolmogorov forward differential equation [1]

\[ \frac{d(P_{i_1,i_2,i_3,i_4,j}(t))}{dt} = -(N \mu_N + i_1(\alpha_{p_1} + \beta_{p_1} + \mu_{p_1}) P_{i_1,i_2,i_3,i_4,j}(t) + ((i_1-1)\alpha_{p_1} + N \mu_N) P_{i_1,i_2,i_3,i_4,j}(t)) \]
\[ + i_1 \mu_{p_1} P_{i_1,i_2-1,i_3,i_4,j}(t) \]
\[ - i_2(\alpha_{p_2} + \beta_{p_2} + \mu_{p_2}) P_{i_1,i_2,i_3,i_4,j}(t) \]
\[ + (i_2 + 1) \beta_{p_2} P_{i_1,i_2+1,i_3,i_4,j}(t) \]
\[ + i_3 \mu_{p_3} P_{i_1,i_2,i_3-1,i_4,j}(t) \]
\[ - i_3 \alpha_{p_3} P_{i_1,i_2,i_3+1,i_4,j}(t) \]
\[ + (i_3 - 1) \beta_{p_3} P_{i_1,i_2,i_3+1,i_4,j}(t) \]
\[ + i_4 \mu_{p_4} P_{i_1,i_2,i_3,i_4-1,j}(t) \]
\[ - i_4 \alpha_{p_4} P_{i_1,i_2,i_3,i_4+1,j}(t) \]
\[ + (i_4 - 1) \beta_{p_4} P_{i_1,i_2,i_3,i_4+1,j}(t) \]
\[ + i_4 \mu_{p_4} P_{i_1,i_2,i_3,i_4,j-1}(t). \] (S2)

Let’s multiply both sides of above equation by \( x_1^{i_1} x_2^{i_2} x_3^{i_3} x_4^{i_4} \), we can obtain,

\[ \frac{d\Psi(x_1, x_2, x_3, x_4, y; 0, t)}{dt} = (x_1 - 1)\mu_N N \Psi(x_1, x_2, x_3, x_4, y; 0, t) + [\mu_{p_1} x_1 + \alpha_{p_1} x_1^2 + \beta_{p_1} - (\alpha_{p_1} + \beta_{p_1} + \mu_{p_1}) x_1] \frac{d\Psi(x_1, x_2, x_3, x_4, y; 0, t)}{dx_1} \]
\[ + [\mu_{p_2} x_2 + \alpha_{p_2} x_2^2 + \beta_{p_2} - (\alpha_{p_2} + \beta_{p_2} + \mu_{p_2}) x_2] \frac{d\Psi(x_1, x_2, x_3, x_4, y; 0, t)}{dx_2} \]
\[ + [\mu_{p_3} x_3 + \alpha_{p_3} x_3^2 + \beta_{p_3} - (\alpha_{p_3} + \beta_{p_3} + \mu_{p_3}) x_3] \frac{d\Psi(x_1, x_2, x_3, x_4, y; 0, t)}{dx_3} \]
\[ + [\mu_{p_4} x_4 + \alpha_{p_4} x_4^2 + \beta_{p_4} - (\alpha_{p_4} + \beta_{p_4} + \mu_{p_4}) x_4] \frac{d\Psi(x_1, x_2, x_3, x_4, y; 0, t)}{dx_4}. \] (S3)

By the definition of the probability generating function (S1), we found that \( \frac{d\Psi(x_1, x_2, x_3, x_4, y; 0, t)}{dx_i} \) is the expectation at time \( t \) for the number of the premalignant cells, \( P_i \). From the equation (S3), we get

\[
\begin{align*}
\frac{dE[P_1(t)]}{dt} &= \lambda_{p_1} E[P_1(t)] + \mu_N N, \\
\frac{dE[P_2(t)]}{dt} &= \lambda_{p_2} E[P_2(t)] + \mu_{p_1} E[P_1(t)], \\
\frac{dE[P_3(t)]}{dt} &= \lambda_{p_3} E[P_3(t)] + \mu_{p_2} E[P_2(t)], \\
\frac{dE[P_4(t)]}{dt} &= \lambda_{p_4} E[P_4(t)] + \mu_{p_3} E[P_3(t)].
\end{align*}
\] (S4)
where \( \lambda_{pi} = \alpha_{pi} - \beta_{pi} \).

The approximate solution of hazard function can be given by

\[
 h(t) \approx \frac{N_{pi}}{\lambda_{pi} \lambda_{p4} (\lambda_{p3} - \lambda_{p2})} \left[ \lambda_{pi}^3 \exp(\lambda_{p4} (t - T_{lag})) - \frac{\lambda_{p2}^2 (\lambda_{p4} - \lambda_{p2})^2}{2} (t - T_{lag})^2 \right. \\
\left. + \lambda_{p1} \lambda_{p2} (\lambda_{p4} - \lambda_{p2}) (2 \lambda_{p2} - \lambda_{p4}) (t - T_{lag}) + 3 \lambda_{p2}^2 \lambda_{p4} - 3 \lambda_{p4} \lambda_{p2}^2 + \lambda_{p4}^3 \right].
\]

For TSG, only both alleles are inactivated, TSG is inactivated [2]. Hence, the cells do not grow abnormally when only one one allele is inactivated. As a consequence, we need to consider the alternations in TSG. Inactivation of each allel of TSG has two ways, one way is mutation, another is a loss of heterozygosity (LOH). Studies showed that crypts carrying one of APC and KRAS alterations (that is the inactivation of APC and activation of KRAS) confers a selective growth advantage to the cell [3, 4]. Therefore, all possible orders for the alterations of three genes in Figure A-F can be summarized as the five-stage models with different clonal expansion of premalignant cells by the clonal expansion of premalignant cells, which is displayed as follows

1. the five-stage model with \( \lambda_{p2} = \lambda_{p1} \) and \( \lambda_{p4} = \lambda_{p3} \) for the models (A1) and (B1);
2. the five-stage model with \( \lambda_{p3} = \lambda_{p2} = \lambda_{p1} \) for the models (A3) and (B3);
3. the five-stage model with \( \lambda_{p1} = 0 \) and \( \lambda_{p3} = \lambda_{p2} \) for the models (A4), (B4), (C1), (D2), (E1) and (F2);
4. the five-stage model with \( \lambda_{p1} = 0 \) and \( \lambda_{p4} = \lambda_{p3} \) for the models (A2), (B2), (D1) and (F1);
5. the five-stage model with \( \lambda_{p1} = 0 \) and \( \lambda_{p2} = 0 \) for the models (A5), (B5), (C2), (D3), (E2) and (F3).

Then, the hazard function for the case with \( \lambda_{p3} = \lambda_{p2} = \lambda_{p1} \) can be written as

\[
 h(t) \approx \frac{N_{pi}}{\lambda_{p2} \lambda_{p4} (\lambda_{p4} - \lambda_{p2})} \left[ \lambda_{p2}^3 \exp(\lambda_{p4} (t - T_{lag})) - \frac{\lambda_{p2}^2 (\lambda_{p4} - \lambda_{p2})^2}{2} (t - T_{lag})^2 \right. \\
\left. + \lambda_{p1} \lambda_{p2} (\lambda_{p4} - \lambda_{p2}) (2 \lambda_{p2} - \lambda_{p4}) (t - T_{lag}) + 3 \lambda_{p2}^2 \lambda_{p4} - 3 \lambda_{p4} \lambda_{p2}^2 + \lambda_{p4}^3 \right].
\]

The hazard function for the case with \( \lambda_{p1} = 0 \) and \( \lambda_{p4} = \lambda_{p2} \) can be written as

\[
 h(t) \approx \frac{N_{pi}}{\lambda_{p2} \lambda_{p4} (\lambda_{p4} - \lambda_{p2})} \left[ \lambda_{p2} \lambda_{p4}^2 (\lambda_{p4} - \lambda_{p2}) (2 \lambda_{p4} - \lambda_{p2}) (t - T_{lag}) + \lambda_{p4}^3 \right. \\
\left. \exp(\lambda_{p4} (t - T_{lag})) - (\lambda_{p2} \lambda_{p4}^2 (\lambda_{p4} - \lambda_{p2}) (t - T_{lag}) + 3 \lambda_{p4}^2 \lambda_{p2} - 2 \lambda_{p4}^3) \right. \\
\left. \exp(\lambda_{p2} (t - T_{lag})) - (\lambda_{p2} \lambda_{p4} (\lambda_{p4} - \lambda_{p2}) (t - T_{lag}) + \lambda_{p4}^2 (3 \lambda_{p2} - 2 \lambda_{p4})) \right. \\
\left. + (\lambda_{p2} - \lambda_{p4})^2 (2 \lambda_{p4} + \lambda_{p2}) \right].
\]

For the case with \( \lambda_{p1} = 0 \) and \( \lambda_{p3} = \lambda_{p2} \), the hazard function yields

\[
 h(t) \approx \frac{N_{pi}}{\lambda_{p2} \lambda_{p4} (\lambda_{p4} - \lambda_{p2})} \left[ \lambda_{p2} \lambda_{p4}^2 (\lambda_{p4} - \lambda_{p2})^2 (t - T_{lag}) + \lambda_{p4} \lambda_{p2} (2 \lambda_{p4}^2 - 3 \lambda_{p2}^2 \lambda_{p4}) \right. \\
\left. \exp(\lambda_{p4} (t - T_{lag})) + \lambda_{p4}^2 \exp(\lambda_{p2} (t - T_{lag})) - (\lambda_{p2} \lambda_{p4}^2 (\lambda_{p4} - \lambda_{p2})^2 (t - T_{lag}) \right. \\
\left. + \lambda_{p4}^2 + 2 \lambda_{p2} \lambda_{p4} - 3 \lambda_{p4}^2 \lambda_{p2}^2) \right].
\]

The hazard function for the case with \( \lambda_{p1} = 0 \) and \( \lambda_{p2} = 0 \) can given by

\[
 h(t) \approx \frac{N_{pi}}{\lambda_{p2} \lambda_{p4} (\lambda_{p4} - \lambda_{p2})} \left[ \lambda_{p2}^3 \exp(\lambda_{p4} (t - T_{lag})) - \lambda_{p2}^3 \exp(\lambda_{p2} (t - T_{lag})) \right. \\
\left. + \lambda_{p2} \lambda_{p4} (\lambda_{p4} - \lambda_{p2}) (t - T_{lag}) + \lambda_{p2} \lambda_{p4} (\lambda_{p4}^2 - \lambda_{p2}) (t - T_{lag}) \right. \\
\left. + (\lambda_{p4}^3 - \lambda_{p2}^3) \right].
\]

\[ (S5) \]

\[ (S6) \]

\[ (S7) \]

\[ (S8) \]

\[ (S9) \]
The derivation of Kolmogorov backward equations

Here, we mainly give the differential equation for Φ₁, the derivations of Kolmogorov backward equation for other probability generating functions can be obtained similarly. We consider the Φ₁(x₁, x₂, x₃, x₄, y; 0, t + Δ), where Δ is very small increment in time. By the Chapman-Kolmogorov equations, the Φ₁(x₁, x₂, x₃, x₄, y; 0, t + Δ) can be written as

\[
\Phi₁(x₁, x₂, x₃, x₄, y; 0, t + Δ) = \sum_{i₁, i₂, i₃, i₄} \sum_{n₁, n₂, n₃, n₄, m} \text{prob}\{P₁(t + Δ) = i₁, P₂(t + Δ) = i₂, P₃(t + Δ) = i₃, P₄(t + Δ) = i₄, M(t + Δ) = j|P₁(Δ) = n₁, P₂(Δ) = n₂, P₃(Δ) = n₃, P₄(Δ) = n₄, M(Δ) = m}\}
\]

\[
P₁(Δ) = n₁, P₂(Δ) = n₂, P₃(Δ) = n₃, P₄(Δ) = n₄, M(Δ) = m|P₁(0) = 1,
\]

\[
P₂(0) = 0, P₃(0) = 0, P₄(0) = 0, M(0) = 0 \}
\]

\[
x₁^{i₁}x₂^{i₂}x₃^{i₃}x₄^{i₄}y^{j}. \quad (S10)
\]

It is well known that at most one event occurs in a sufficiently small amount of time. Thus the probabilities of the events that can occur in [0, Δ] are approximately \(α_PΔ\) (i = 1, 2, 3, 4) for the birth of a state \(P_i\) cell, \(β_PΔ\) for the death of a state \(P_i\) cell, \(µ_PΔ\) for the transformation from state \(P_i\) to \(P_{i+1}\) (from state \(P_i\) to \(M\) if \(i = 4\)) and \(1 - α_PΔ - β_PΔ - µ_PΔ\) for no change in the state \(P_i\), respectively. Then, we get

\[
\Phi₁(x₁, x₂, x₃, x₄, y; 0, t + Δ) = \sum_{i₁, i₂, i₃, i₄} \left\{ \alpha_{p₃}\Delta \text{prob}\{P₁(t + Δ) = i₁, P₂(t + Δ) = i₂, P₃(t + Δ) = i₃, P₄(t + Δ) = i₄, M(t + Δ) = j|P₁(Δ) = 2, P₂(Δ) = 0, P₃(Δ) = 0, P₄(Δ) = 0, M(Δ) = 0\} \right\},
\]

\[
+ β_{p₂}\Delta \text{prob}\{P₁(t + Δ) = i₁, P₂(t + Δ) = i₂, P₃(t + Δ) = i₃, P₄(t + Δ) = i₄, M(t + Δ) = j|P₁(Δ) = 2, P₂(Δ) = 0, P₃(Δ) = 0, P₄(Δ) = 0, M(Δ) = 0\} \right\},
\]

\[
+ µ_{p₁}\Delta \text{prob}\{P₁(t + Δ) = i₁, P₂(t + Δ) = i₂, P₃(t + Δ) = i₃, P₄(t + Δ) = i₄, M(t + Δ) = j|P₁(Δ) = 2, P₂(Δ) = 0, P₃(Δ) = 0, P₄(Δ) = 0, M(Δ) = 0\} \right\},
\]

\[
+(1 - α_{p₃}Δ - β_{p₂}Δ - µ_{p₁}Δ) \text{prob}\{P₁(t + Δ) = i₁, P₂(t + Δ) = i₂, P₃(t + Δ) = i₃, P₄(t + Δ) = i₄, M(t + Δ) = j|P₁(Δ) = 2, P₂(Δ) = 0, P₃(Δ) = 0, P₄(Δ) = 0, M(Δ) = 0\} \right\},
\]

\[
\}
\]

\[
x₁^{i₁}x₂^{i₂}x₃^{i₃}x₄^{i₄}y^{j}. \quad (S11)
\]

We let \(Φ₁(t + Δ)\) to denote the \(Φ₁(x₁, x₂, x₃, x₄, y; 0, t + Δ)\). By the definition for \(Φ₁(t)\), Markov property and relationship formula (3) in reference [5], the formula (S11) can be written as

\[
Φ₁(t + Δ) = α_{p₃}\Delta Φ₁(t) + β_{p₂}\Delta + µ_{p₁}ΔΦ₁(t)Φ₂(t)
\]

\[
+(1 - α_{p₃}Δ - β_{p₂}Δ - µ_{p₁}Δ)Φ₁(t).
\]

\[\text{(S12)}\]
Subtracting $\Phi_1(t)$ from both sides of above formula and divide by $\Delta$, then we obtain

$$\frac{d\Phi_1}{dt}(t) = -(\alpha_{p_1} + \beta_{p_1} + \mu_{p_1})\Phi_1(t) + \alpha_{p_1}\Phi_1^2(t) + \mu_{p_1}\Phi_1(t)\Phi_2(t) + \beta_{p_1}. \quad (S13)$$

$\frac{d\Phi_2}{dt}(t)$ and $\frac{d\Phi_4}{dt}(t)$ are obtained similarly, where $k = 2, 3, 4$.

**The expected number of mutant cell**

For the orders of genetic mutations with $KRAS - APC - TP53$, $APC - TP53 - KRAS$ and $APC - KRAS - TP53$ that can explain the incidence rate of colorectal cancer, the number of premalignant cells with only $KRAS^+$, $APC^{-/-}$, $KRAS^+APC^{-/-}$, $APC^{-/-}KRAS^+$ and $APC^{-/-}TP53^{-/-}$ can be calculated by the equations (S4). The results are as follows:

$$E[n_{KRAS^+}(t)] = \frac{N\mu_{p_2}(e^{\lambda_{p_1}t} - 1)}{\lambda_{p_2} \lambda_{p_1} (e^{\lambda_{p_2}t} - \lambda_{p_2}t - 1)},$$

$$E[n_{APC^{-/-}}(t)] = \frac{N\mu_{p_2} \mu_{p_3} \lambda_{p_2} \lambda_{p_3}}{\lambda_{p_2} \lambda_{p_1} (e^{\lambda_{p_2}t} - \lambda_{p_2}t - 1)} \frac{1}{\lambda_{p_1} (\lambda_{p_3} - \lambda_{p_1})} (t - \frac{1}{\lambda_{p_1}}),$$

$$E[n_{KRAS^+APC^{-/-}}(t)] = N\mu_{p_2} \mu_{p_3} \lambda_{p_2} \lambda_{p_3} \lambda_{p_1} \left[ \lambda_{p_2} e^{\lambda_{p_2}t} - \lambda_{p_2} e^{\lambda_{p_2}t} + \lambda_{p_2} \lambda_{p_3} \left( \lambda_{p_3} - \frac{1}{\lambda_{p_1}} \right) \right],$$

$$E[n_{APC^{-/-}KRAS^+}(t)] = \frac{N\mu_{p_2} \mu_{p_3} \lambda_{p_2} \lambda_{p_3}}{\lambda_{p_2} \lambda_{p_1} (e^{\lambda_{p_2}t} - \lambda_{p_2}t - 1)} \left[ \lambda_{p_2} \lambda_{p_4} (\lambda_{p_4} - \lambda_{p_2}) (2\lambda_{p_4} - \lambda_{p_2})^t \right. + \lambda_{p_4}^2 e^{\lambda_{p_4}t} - \left. \lambda_{p_2} \lambda_{p_4} (\lambda_{p_4} - \lambda_{p_2})^t + \lambda_{p_4}^2 (3\lambda_{p_4} - 2\lambda_{p_2}) \lambda_{p_4} (\lambda_{p_4} - \lambda_{p_2})^t \right],$$

$$E[n_{APC^{-/-}TP53^{-/-}}(t)] = \frac{N\mu_{p_2} \mu_{p_3} \lambda_{p_2} \lambda_{p_3}}{\lambda_{p_2} \lambda_{p_1} (e^{\lambda_{p_2}t} - \lambda_{p_2}t - 1)} \left[ \lambda_{p_2} \lambda_{p_4} (\lambda_{p_4} - \lambda_{p_2})^t \right. + \lambda_{p_4}^2 e^{\lambda_{p_4}t} - \lambda_{p_2} \lambda_{p_4} (\lambda_{p_4} - \lambda_{p_2})^t \right. \left. + \lambda_{p_4}^2 (3\lambda_{p_4} - 2\lambda_{p_2}) \lambda_{p_4} (\lambda_{p_4} - \lambda_{p_2})^t \right],$$

$$E[n_{APC^{-/-}TP53^{-/-}}(t)] = \frac{N\mu_{p_2} \mu_{p_3} \lambda_{p_2} \lambda_{p_3}}{\lambda_{p_2} \lambda_{p_1} (e^{\lambda_{p_2}t} - \lambda_{p_2}t - 1)} \left[ \lambda_{p_2} \lambda_{p_4} (\lambda_{p_4} - \lambda_{p_2})^t \right. + \lambda_{p_4}^2 e^{\lambda_{p_4}t} - \lambda_{p_2} \lambda_{p_4} (\lambda_{p_4} - \lambda_{p_2})^t \right. \left. + \lambda_{p_4}^2 (3\lambda_{p_4} - 2\lambda_{p_2}) \lambda_{p_4} (\lambda_{p_4} - \lambda_{p_2})^t \right],$$

$$E[n_{APC^{-/-}TP53^{-/-}}(t)] = \frac{N\mu_{p_2} \mu_{p_3} \lambda_{p_2} \lambda_{p_3}}{\lambda_{p_2} \lambda_{p_1} (e^{\lambda_{p_2}t} - \lambda_{p_2}t - 1)} \left[ \lambda_{p_2} \lambda_{p_4} (\lambda_{p_4} - \lambda_{p_2})^t \right. + \lambda_{p_4}^2 e^{\lambda_{p_4}t} - \lambda_{p_2} \lambda_{p_4} (\lambda_{p_4} - \lambda_{p_2})^t \right. \left. + \lambda_{p_4}^2 (3\lambda_{p_4} - 2\lambda_{p_2}) \lambda_{p_4} (\lambda_{p_4} - \lambda_{p_2})^t \right],$$
References


