Why does the order matter for mutations in myeloproliferative neoplasm

Yue Wang[1]

[1]Department of Computational Medicine, University of California, Los Angeles, California, United States of America. yuew@g.ucla.edu

Abstract

We can find JAK2 V617F and TET2 mutations in some patients of myeloproliferative neoplasm. JAK2 V617F mutation has different effects on regulating gene expressions, depending on whether TET2 mutation is present or not. In addition, when JAK2 V617F and TET2 mutations are both present, which mutation appears first also affects certain properties of patients. We use nonlinear ordinary differential equation models and Markov chain models to explain such phenomena.

KEY WORDS: cancer, mutation, order, gene expression

1 Introduction

For a bone marrow cancer, myeloproliferative neoplasm, we can find different genetic mutations in different patients, and these mutations have different effects on cell behavior [24, 20, 27]. In this paper, we study the effects of two mutations in myeloproliferative neoplasm, JAK2 V617F (abbreviated as JAK2 in the following) and TET2.

JAK2 mutation and TET2 mutation introduce a growth advantage [3, 8], meaning that once JAK2 or TET2 mutation appears in some cells, such cells will proliferate and become common. However, the advantage might take years to become significant, and it is common to find cells with different numbers of mutations (0, 1, or 2). This type of gene-related growth patterns might lead to complicated phenomena [10, 15, 61]. For patients having cells with both JAK2 and TET2 mutations, an important question is to determine which mutation appears first. If we can find cells with just JAK2 mutation and cells with both JAK2 and TET2 mutations, but not cells with just TET2 mutation, we can determine that JAK2 mutation appears before TET2 mutation, and such patients are called JAK2-first. If we can find TET2-only cells but not JAK2-only cells, then such patients are TET2-first. If we can find JAK2-only cells, TET2-only cells and JAK2-TET2 cells, then at least two non-mutant cells undergo mutant events independently, and it is difficult to determine the order

of mutations. Such patients are not considered. For more mutations, there are different algorithms to determine possible orders of mutations from sequencing data [33, 36, 19, 11].

It has been noticed that the order of appearance for JAK2 and TET2 mutations can make a difference [31]. For JAK2-first patients and TET2-first patients, they have different clinical features, responses to therapy, biology of stem and progenitor cells, and even clonal evolution [23].

Ortmann et al. [31] compare the expression levels of various genes before and after the appearance of JAK2 and TET2 mutations. Consider the expression level x^* of a certain gene X with or without these two mutations: x^*_{O} means no JAK2 or TET2 mutation; x^*_{J} means with JAK2 but not TET2 mutation; x^*_{T} means with TET2 but not JAK2 mutation; x^*_{JT} means with JAK2 and TET2 mutations, and JAK2 appears first; x^*_{TJ} means with JAK2 and TET2 mutations, and TET2 appears first. Here are some important phenomena:
(1) Some genes are up-regulated (or down-regulated) by JAK2 if TET2 is not present, but not affected if TET2 is present, thus $x^*_{\mathrm{T}} = x^*_{\mathrm{TJ}}$, $x^*_{\mathrm{O}} > x^*_{\mathrm{J}}$ or $x^*_{\mathrm{O}} < x^*_{\mathrm{J}}$.
(2) Some genes are up-regulated (or down-regulated) by JAK2 if TET2 is present, but not affected if TET2 is not present, thus $x^*_{\mathrm{O}} = x^*_{\mathrm{J}}$, $x^*_{\mathrm{J}} > x^*_{\mathrm{TJ}}$ or $x^*_{\mathrm{J}} < x^*_{\mathrm{TJ}}$.
(3) Ten genes (AURKB, FHOD1, HTRA2, IDH2, MCM2, MCM4, MCM5, TK1, UQCRC1, WDR34) are up-regulated by JAK2 if TET2 is not present, but down-regulated by JAK2 if TET2 is present, thus $x^*_{\mathrm{O}} < x^*_{\mathrm{J}}$, $x^*_{\mathrm{T}} > x^*_{\mathrm{TJ}}$.
(4) Different appearance orders of JAK2 and TET2 have different effects on some genes, thus $x^*_{\mathrm{JT}} \neq x^*_{\mathrm{TJ}}$. This is phenomenon is inferred from other evidence (e.g., JAK2-first cells are more sensitive to ruxolitinib than TET2-first cells), not from direct measurements.

Phenomena (1-3) can be regarded as **non-additivity**, since the effect of JAK2 mutation differs with or without TET2 mutation, or in other words, $x^*_{\mathrm{J}} - x^*_{\mathrm{O}} \neq x^*_{\mathrm{TJ}} - x^*_{\mathrm{T}}$. Phenomenon (4) can be regarded as **non-commutativity** [23], since exchanging the order of mutations can lead to differences, or in other words,

$$x^*_{\mathrm{O}} + (x^*_{\mathrm{J}} - x^*_{\mathrm{O}}) + (x^*_{\mathrm{JT}} - x^*_{\mathrm{J}}) = x^*_{\mathrm{JT}} \neq x^*_{\mathrm{TJ}} = x^*_{\mathrm{O}} + (x^*_{\mathrm{T}} - x^*_{\mathrm{O}}) + (x^*_{\mathrm{TJ}} - x^*_{\mathrm{T}}).$$

In fact, if the gene expression is additive, namely $x^*_{\mathrm{J}} - x^*_{\mathrm{O}} = x^*_{\mathrm{TJ}} - x^*_{\mathrm{T}}$ and $x^*_{\mathrm{T}} - x^*_{\mathrm{O}} = x^*_{\mathrm{JT}} - x^*_{\mathrm{J}}$, then the gene expression is also commutative:

$$x^*_{\mathrm{JT}} = x^*_{\mathrm{J}} + x^*_{\mathrm{T}} - x^*_{\mathrm{O}} = x^*_{\mathrm{TJ}}.$$

Therefore, non-commutativity is a special case of non-additivity. JAK2 first or TET2 first cells are genetically identical, but have very different behaviors [51, 52, 47, 34].

On cell level or even macroscopic level, Ortmann et al. [31] also report two phenomena related to non-commutativity:
(5) For TET2-first patients, the percentage of cells with just one mutation (TET2) is significantly higher than the percentage of JAK2-only cells for JAK2-first patients.
(6) JAK2-first patients are significantly younger than TET2-first patients at diagnosis.

For myeloproliferative neoplasm, the order of JAK2 V617F and DNMT3A mutations can also affect cell behavior [28]. Such order of mutation phenomena also appear in other types of cancer. For instance, in adrenocortical carcinomas, if Ras mutation appears before p53 mutation, the tumor will be malignant and metastatic; if p53 mutation appears before Ras mutation, the tumor will be benign [12]. Similar phenomena can be found in other references [23, 44, 6].

There have been some explanations for the difference from order of mutation [31, 40, 38, 18, 9, 42, 41, 25, 43], and we summarize these works in Section 2. Then we build nonlinear ordinary differential equation (ODE) models to explain phenomena (1-3) and list experimental evidence that supports our model in Section 3. In Section 4, we build nonlinear ODE models and Markov chain models to explain phenomenon (4). In Section 5, we present a generalized Moran process model and three different mechanisms to explain phenomena (5, 6). We finish with some discussion in Section 6.

2 Previous works related to order of mutation

In this section, we summarize previous works that can be used to explain phenomena (1-6) and compare them with our results. Due to the complexity of gene expression [17, 48], how mutations affect gene expression is not quite clear. Thus such explanations are relatively theoretical.

Ortmann et al. [31] assume that TET2 mutation can significantly increase the proliferation rate of cancer stem cells, while JAK2 mutation only has a weak growth advantage. Therefore, for TET2-first patients, TET2-only cells first spread, and TET2-JAK2 cells (which do not have a significant growth advantage over TET2-only cells) do not dominate. For JAK2-first patients, JAK2-only cells do not spread that much, while JAK2-TET2 cells (once appear) can dominate. Therefore, TET2-first patients have a much higher percentage of cells with only one mutation. This explains phenomenon (5). See also the interpretation by Swanton [40]. We find that this mechanism can also explain phenomenon (6).

Kent and Green [18] propose two explanations for phenomenon (4): (A) Both JAK2 and TET2 mutations can participate in epigenetic regulation [13, 39], but the regulation mechanism might be incompatible. The first mutation might lead to the occupation of certain genomic regions, so that the second mutation cannot regulate. This explanation implies that $x^*_{\mathrm{JT}} = x^*_{\mathrm{J}} \neq x^*_{\mathrm{T}} = x^*_{\mathrm{TJ}}$, not applicable to explaining other phenomena. (B) For JAK2-first patients and TET2-first patients, before the appearance of the second mutation, different first mutations might produce different cell types and microenvironments. Therefore, the cell where the second mutation appears might be of different types and in different microenvironments.

Roquet et al. [38] consider the effect of recombinases (i.e., genetic recombination enzymes) on gene sequences. When applying different recombinases to gene sequences, different orders of recombinases can lead to different results. For example, consider a gene

sequence 12312 and two recombinases A, B. The effect of A is to delete genes between two "1", and the effect of B is to invert genes between two "2" (if there are not two "2", B does not work). If we add A before B, the gene sequence becomes

$$12312 \xrightarrow{\text{A}} 112 \xrightarrow{\text{B}} 112.$$

If we add A after B, the gene sequence becomes

$$12312 \xrightarrow{\text{B}} 12132 \xrightarrow{\text{A}} 1132.$$

Although this work is not about mutations, it provides a model for phenomenon (4).

Clarke et al. [9] model the gene regulatory network as a generalized boolean network. This network evolves under a given rule, and there are some fixed points and/or limit cycles. Each mutation fully activates or inhibits one gene (the expression level is fixed, like the do-operator in causal inference [4]). Certain orders of mutations lead to higher proliferation rates or apoptosis rates, making such orders more likely or less likely. There might be multiple fixed points and/or limit cycles, and different orders of mutations might lead to different final states. This model can be used to explain phenomenon (4).

Talarmain et al. [42, 41] apply the model in Clarke et al.'s paper to the order of JAK2 and TET2 problem and explain phenomenon (4). They find a concrete generalized boolean network. When there is no mutation or just one mutation, the system has one stable fixed point. When both JAK2 and TET2 mutations are present, the system has two stable fixed points, and different orders of mutations lead to different fixed points. Besides, they find that HOXA9 gene should be the source of phenomenon (4): HOXA9 is directly affected by different orders of JAK2 and TET2, and many genes are then affected by different levels of HOXA9.

Mazaya et al. [25] use a boolean network to model the effect of mutations. The model dynamics and the explanation of phenomenon (4) are similar to Clarke et al.'s model. Mazaya et al. further analyze this model to study when the network is more sensitive to different orders of mutations.

Teimouri and Kolomeisky [43] use a random walk model to study the acquisition of two mutations. The first stage is a random walk on $0, 1, \ldots, n$, representing the number of cells with the first mutation. The first stage starts at 0, and finishes when reaching n, meaning that all cells have the first mutation. The process terminates if reaching 0 again before reaching n. The second stage is a random walk on $n, n+1, \ldots, 2n$, representing the number of cells with the second mutation plus n. The second stage starts at n, and finishes when reaching $2n$, meaning that all cells have both mutations. If the process reaches $2n$, we count the total time and take the expectation. They prove that if the first mutation has a higher fitness than the second mutation, then the tumor formation probability is higher, but the time for tumor formation is longer. This explains phenomenon (6).

Phenomena (1, 2) are a classical scheme in biology: some quantity can be changed if and only if both conditions are met. This scheme is very common, and we do not list

related references. We build a simple nonlinear ODE model to explain phenomena (1, 2). The essential idea is a threshold which can be passed if and only if both conditions are met. This model is a building block of explaining phenomenon (3).

For phenomenon (3), we do not find any explanations. We explain this with a nonlinear ODE model, and we find two candidates for the hidden factor in this model. Some regulatory relations in our model have been verified experimentally, and we propose experiments to examine other regulatory relations.

Kent and Green [18] explain phenomenon (4) without mathematical details. Roquet et al. [38] just propose a mathematical space where operators are not commutative, and it is difficult to link this to describe genetic mutations. Clarke et al. [9], Talarmain et al. [42, 41], and Mazaya et al. [25] all use (generalized) boolean netowrks to explain phenomenon (4). The state space is discrete, and the dynamics is deterministic. Besides, the dynamics is artificially chosen without further justifications. On single-cell level, gene expression level is discrete but stochastic. On bulk level, gene expression level is deterministic but continuous. Therefore, we propose a nonlinear ODE model (deterministic, continuous-state) and a Markov chain model (stochastic, discrete-state) to explain phenomenon (4).

For phenomenon (5), Ortmann et al. [31] assume that different mutations provide different proliferation advantages. Ortmann et al. only describe the mechanism without studying it mathematically.

For phenomenon (6), the only explanation we can find is by Teimouri and Kolomeisky [43]. In their model, the second mutation can appear if and only if the first mutation has dominated the whole cell population. Also, the explanation of phenomenon (6) depends on a setting: if in a simulation, the process reaches 0 again, this simulation is abandoned. If we continue this simulation, it will reach $2n$ with a longer time. Therefore, this model is not quite realistic.

To explain phenomena (5, 6), we construct a generalized Moran process model, which is relatively more realistic for describing the population dynamics of hematopoietic stem cells. We consider three different possible mechanisms. (I) The same as Ortmann et al., assume cells with different mutations have different growth rates. (II) Similar to Teimouri and Kolomeisky, assume JAK2 and TET2 have different mutation rates. (III) Assume that JAK2 mutation can induce TET2 mutation. In simulations, we verify that each of these three mechanisms can produce phenomena (5, 6). Since it is natural to assume that different mutations have different mutation rates, mechanism (II) should be more persuasive.

3 Models for non-additivity in gene expression

In this section, we build models to explain phenomena (1, 2, 3) that the effect of one mutation is not additive for gene expression, based on experimental evidence.

3.1 Mathematical preparations

We consider ordinary differential equation (ODE) models for gene expression and regulation.

For a gene X with expression level x, the simplest model only considers synthesis and degradation with constant rates b, d:

$$\mathrm{d}x/\mathrm{d}t = b - dx.$$

Therefore, the stationary state for x is $x^* = b/d$.

If other genes regulate the expression of X, we can add boolean terms (for the presence of certain mutations) or linear terms (for the expression levels of certain genes) in this ODE:

$$\mathrm{d}x/\mathrm{d}t = b + b_{\mathrm{Y}}\mathbb{1}_{\mathrm{Y}} + b_{\mathrm{Z}}z - dx.$$

Here $\mathbb{1}_{\mathrm{Y}} = 1$ if mutation Y is present, and $\mathbb{1}_{\mathrm{Y}} = 0$ otherwise; z is the expression level of gene Z; $b_{\mathrm{Y}}, b_{\mathrm{Z}}$ are constant regulatory rates for Y, Z.

This linear ODE cannot explain the above phenomena (1-3), since the regulation effect of each gene (mutation) is additive: regardless of the status of other genes and mutations, the presence of one mutation always has the same effect. Therefore, we need to add nonlinearity into the ODE.

One choice is to consider a "thin" sigmoid function $F(b_0)$, defined as

$$F(b_0) = \frac{1}{1 + e^{-50b_0}}. \tag{1}$$

See Fig. 1 for the plot of $F(b_0)$. This function has a threshold at 0: if b_0 is smaller than 0, $F(b_0) \approx 1$; if b_0 is larger than 0, $F(b_0) \approx 2$.

Consider an ODE

$$\mathrm{d}x/\mathrm{d}t = F(b_0) - x, \tag{2}$$

where $b_0 = b + b_{\mathrm{Y}}\mathbb{1}_{\mathrm{Y}} + b_{\mathrm{Z}}\mathbb{1}_{\mathrm{Z}}$, and $F(b_0)$ is defined in Eq. 1. If $b_0 > 0$, the stationary state is $x^* \approx 2$; if $b_0 < 0$, the stationary state is $x^* \approx 1$. Therefore, Y and Z affect the stationary state only if their change make b_0 cross the threshold at 0. Such dependence means that the model is non-additive: the effect of one mutation cannot be directly added in, since it depends on other factors.

3.2 Models for phenomena (1, 2)

We build models to explain why some genes have $x^*_{\mathrm{O}} \neq x^*_{\mathrm{J}}$, but $x^*_{\mathrm{T}} = x^*_{\mathrm{TJ}}$ (and vice versa). We consider different variants of Eq. 2. In the following representations, J means the JAK2 mutation, and T means the TET2 mutation. X and Y are the target genes that are regulated by J and T.

Consider Eq. 2, $\mathrm{d}x/\mathrm{d}t = F(b_0) - x$ with $b_0 = b + b_{\mathrm{J}}\mathbb{1}_{\mathrm{J}} + b_{\mathrm{T}}\mathbb{1}_{\mathrm{T}}$.

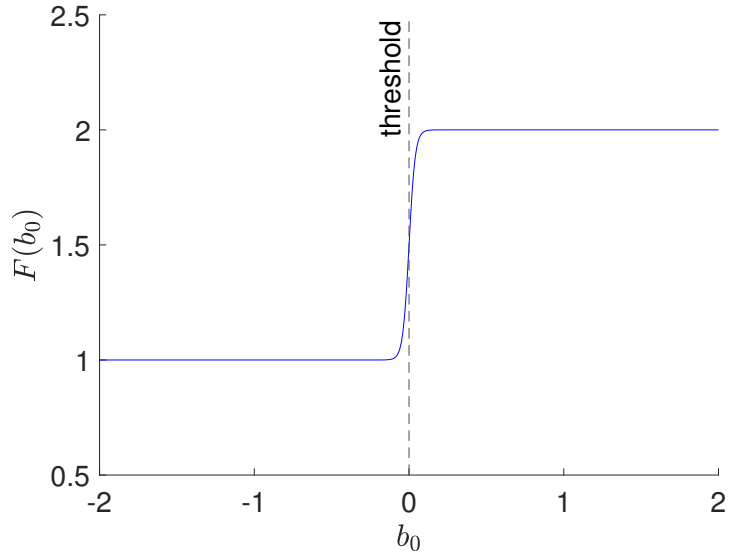

Figure 1: The plot of $F(b_0)$ as a function of b_0. There is a threshold at $b_0 = 0$. Below this threshold, $F(b_0) \approx 1$; above this threshold, $F(b_0) \approx 2$.

Assume $b = -1.5$, and $b_{\mathrm{J}} = b_{\mathrm{T}} = 1$. With no mutation, $b_0 < 0$, and the system is at the low state $x^*_{\mathrm{O}} = 1$. Consider the case that $\mathbb{1}_{\mathrm{T}} = 0$, $\mathbb{1}_{\mathrm{J}} = 1$, meaning that JAK2 mutation is present, but not TET2 mutation (or vice versa). Then $b_0 < 0$, and the system is at the low state $x^*_{\mathrm{J}} = 1$ (also $x^*_{\mathrm{T}} = 0$). If both JAK2 and TET2 mutations are present, then $\mathbb{1}_{\mathrm{T}} = \mathbb{1}_{\mathrm{J}} = 1$, $b_0 > 0$, and the system is at the high state $x^*_{\mathrm{TJ}} = 2$. We have $x^*_{\mathrm{O}} = x^*_{\mathrm{J}}$ but $x^*_{\mathrm{T}} < x^*_{\mathrm{TJ}}$. Therefore, in this case, J up-regulates X only if T is present. See Fig. 2 for an illustration of this situation.

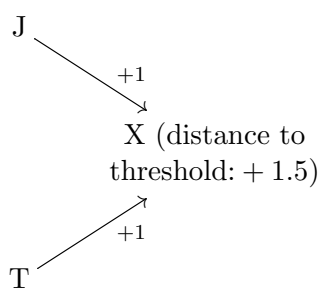

Figure 2: The model diagram that explains $x^*_{\mathrm{O}} = x^*_{\mathrm{J}}$ but $x^*_{\mathrm{T}} < x^*_{\mathrm{TJ}}$ for the level of X. "$\rightarrow$" means up-regulation. We have $x^*_{\mathrm{O}} = 1$, $x^*_{\mathrm{J}} = 1$, $x^*_{\mathrm{T}} = 1$, $x^*_{\mathrm{TJ}} = 2$.

Assume $b = 1.5$, $b_{\mathrm{J}} = b_{\mathrm{T}} = -1$. If $\mathbb{1}_{\mathrm{T}} = 0$ (no T), then $b_0 > 0$, and J itself does not

affect X ($x^*_{\mathrm{J}} = x^*_{\mathrm{O}} = 2$). If $\mathbb{1}_{\mathrm{T}} = 1$, then changing $\mathbb{1}_{\mathrm{J}}$ from 0 to 1 will make $b_0 < 0$, and the stationary state moves down from $x^*_{\mathrm{T}} = 2$ to $x^*_{\mathrm{TJ}} = 1$. We have $^*_{\mathrm{O}} =^*_{\mathrm{J}}$ but $^*_{\mathrm{T}} >^*_{\mathrm{TJ}}$. Thus J down-regulates X only if T is present. See Fig. 3 for an illustration of this situation.

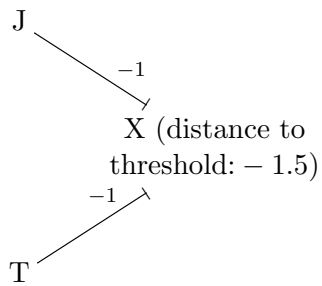

Figure 3: The model diagram that explains $x^*_{\mathrm{O}} = x^*_{\mathrm{J}}$ but $x^*_{\mathrm{T}} > x^*_{\mathrm{TJ}}$ for the level of X. "$\dashv$" means down-regulation. We have $x^*_{\mathrm{O}} = 2$, $x^*_{\mathrm{J}} = 2$, $x^*_{\mathrm{T}} = 2$, $x^*_{\mathrm{TJ}} = 1$.

Assume $b^* - b = 0.5$, $b_{\mathrm{J}} = 1$, $b_{\mathrm{T}} = -1$. If T is not present, then J up-regulates X, since $x^*_{\mathrm{O}} = 1$, $x^*_{\mathrm{J}} = 2$. With the presence of T, J does not affect X, since $x^*_{\mathrm{T}} = 1$, $x^*_{\mathrm{TJ}} = 1$. See Fig. 4 for an illustration of this situation.

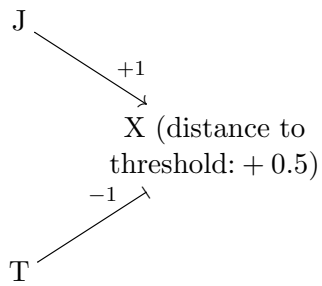

Figure 4: The model diagram that explains $x^*_{\mathrm{O}} < x^*_{\mathrm{J}}$ but $x^*_{\mathrm{T}} = x^*_{\mathrm{TJ}}$ for the level of X. "$\rightarrow$" means up-regulation; "$\dashv$" means down-regulation. We have $x^*_{\mathrm{O}} = 1$, $x^*_{\mathrm{J}} = 2$, $x^*_{\mathrm{T}} = 1$, $x^*_{\mathrm{TJ}} = 1$.

Assume $b^* - b = -0.5$, $b_{\mathrm{J}} = -1$, $b_{\mathrm{T}} = 1$. If T is not present, then J down-regulates X, since $x^*_{\mathrm{O}} = 2$, $x^*_{\mathrm{J}} = 1$. With the presence of T, J does not affect X, since $x^*_{\mathrm{T}} = 2$, $x^*_{\mathrm{TJ}} = 2$. See Fig. 5 for an illustration of this situation.

3.3 Model for phenomenon (3)

To explain phenomenon (3) that O $\rightarrow$ J and T $\rightarrow$ TJ have opposite effects, we need a more complicated variant of Eq. 2.

Consider a gene Y with $\mathrm{d}y/\mathrm{d}t = F(b_0) - y$, where $b_0 = -0.5 + \mathbb{1}_{\mathrm{J}} - \mathbb{1}_{\mathrm{T}}$. We have $y^*_{\mathrm{O}} = y^*_{\mathrm{T}} = y^*_{\mathrm{TJ}} = 1$, and $y^*_{\mathrm{J}} = 2$.

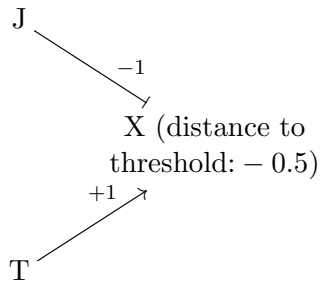

Figure 5: The model diagram that explains $x^*_{\mathrm{O}} > x^*_{\mathrm{J}}$ but $x^*_{\mathrm{T}} = x^*_{\mathrm{TJ}}$ for the level of X. "$\rightarrow$" means up-regulation; "$\dashv$" means down-regulation. We have $x^*_{\mathrm{O}} = 2$, $x^*_{\mathrm{J}} = 1$, $x^*_{\mathrm{T}} = 2$, $x^*_{\mathrm{TJ}} = 2$.

Consider a gene X with linear dynamics:

$$\mathrm{d}x/\mathrm{d}t = 2 - \mathbb{1}_{\mathrm{J}} + 2(y-1) - x.$$

X has a base synthesis rate 2, death rate 1. J can down-regulate X with strength 1. If Y is in its high state $y^* = 2$, Y has a positive effect to X with strength 2; otherwise, when $y^* = 1$, Y has no effect on X.

See Fig. 6 for an illustration of this model. Without J and T, $b_0 = -0.5$, under the threshold 0. In this case, y^*_{O} is in its low state 1, so that X is not affected by J or Y, and the stationary state of X is $x^*_{\mathrm{O}} = 2$. With J but not T, $b_0 = 0.5$, above the threshold 0. In this case, Y is in its high state $y^*_{\mathrm{J}} = 2$. X is affected by both J and Y, and $x^*_{\mathrm{J}} = 3$. With T but not J, $b_0 = -1.5$, below the threshold 0. In this case, Y is in its low state $y^*_{\mathrm{T}} = 1$. X is not affected by J or Y, and $x^* = 2$. With J and T, $b_0 = -0.5$, under the threshold 0. In this case, Y is in its low state $y^*_{\mathrm{TJ}} = 1$. X is only affected by J, and $x^*_{\mathrm{TJ}} = 1$. Therefore, without TET2 mutation, JAK2 mutation up-regulates X (from $x^*_{\mathrm{O}} = 2$ to $x^*_{\mathrm{J}} = 3$); with TET2 mutation, JAK2 mutation down-regulates X (from $x^*_{\mathrm{T}} = 2$ to $x^*_{\mathrm{TJ}} = 1$).

With this simple model, we can explain some genes with $x^*_{\mathrm{O}} < x^*_{\mathrm{J}}$ but $x^*_{\mathrm{T}} > x^*_{\mathrm{TJ}}$, i.e., non-additivity. Here X can be AURKB, MCM2, MCM4, MCM5, or TK1. The role of Y can be played by E2F1 and/or PRMT5. See Fig. 7 for the gene regulatory network that explains this mechanism.

(a): For myeloproliferative neoplasm cells with the JAK2 V617F mutation, the expression of PRMT5 is increased [32].

(b): PRMT5 inhibition reduced the expression of E2F1, and thus PRMT5 up-regulates E2F1 [32].

(c): The expression of E2F1 induces all genes of the endogenous MCM family [30]; E2F1 is a transcriptional activator of AURKB [66]; E2F1 can up-regulate AURKB and MCM5 [37]; overexpressing E2F1 alone results in the up-regulation of MCM5 and TK1 [22]. In sum, E2F1 up-regulates AURKB, MCM2, MCM4, MCM5, and TK1.

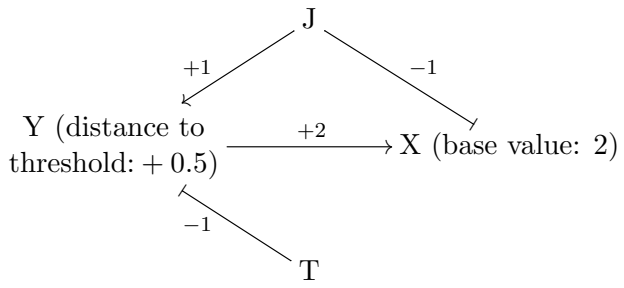

Figure 6: The model diagram that explains $x^*_{\mathrm{O}} < x^*_{\mathrm{J}}$ but $x^*_{\mathrm{T}} > x^*_{\mathrm{TJ}}$. "→" means up-regulation; "⊣" means down-regulation. For the stationary level of Y, we have $y^*_{\mathrm{O}} = 1$, $y^*_{\mathrm{J}} = 2$, $y^*_{\mathrm{T}} = 1$, $y^*_{\mathrm{TJ}} = 1$. For the stationary level of X, we have $x^*_{\mathrm{O}} = 2$, $x^*_{\mathrm{J}} = 3$, $x^*_{\mathrm{T}} = 2$, $x^*_{\mathrm{TJ}} = 1$.

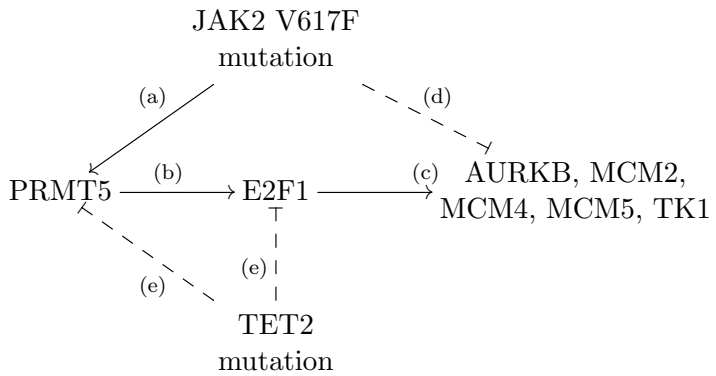

Figure 7: The gene regulatory network that explains $x^*_{\mathrm{O}} < x^*_{\mathrm{J}}$ but $x^*_{\mathrm{T}} > x^*_{\mathrm{TJ}}$ for certain genes. "→" means up-regulation; "⊣" means down-regulation. Solid line means verified regulation; dashed line means hypothesized regulation.

(d): From (a)-(c), JAK2 mutation indirectly up-regulates AURKB, MCM2, MCM4, MCM5, and TK1 through PRMT5 and E2F1. We guess that JAK2 mutation can weakly down-regulate these genes directly. This can be verified by an experiment: after the knock-down or knockout of PRMT5 or E2F1, introducing JAK2 mutation can decrease the expression of AURKB, MCM2, MCM4, MCM5, and TK1.

(e): We guess that TET2 mutation can down-regulate E2F1 directly or through PRMT5. This down-regulation is effective only if JAK2 mutation is present. This means E2F1 (and possibly PRMT5) has $y^*_{\mathrm{J}} > y^*_{\mathrm{JT}}$.

In sum, without TET2 mutation, JAK2 mutation can up-regulate PRMT5 and E2F1, which up-regulate AURKB, MCM2, MCM4, MCM5, and TK1; this covers the weak down-regulation of JAK2 mutation on these genes, meaning that (a-c)>(d). Thus we have

$x^*_{\mathrm{O}} < x^*_{\mathrm{J}}$. With TET2 mutation, the up-regulation of JAK2 mutation on PRMT5 and E2F1 is covered by the down-regulation of TET2 mutation. In this case, PRMT5 and E2F1 are locked in low levels, so that the only effective regulation is (d). Therefore, we have $x^*_{\mathrm{T}} > x^*_{\mathrm{TJ}}$.

Ortmann et al. [31] reported ten genes that $x^*_{\mathrm{O}} < x^*_{\mathrm{J}}$ but $x^*_{\mathrm{T}} > x^*_{\mathrm{TJ}}$: AURKB, FHOD1, HTRA2, IDH2, MCM2, MCM4, MCM5, TK1, UQCRC1, WDR34. This model can explain five of them (AURKB, MCM2, MCM4, MCM5, TK1) with the same pathway JAK2→PRMT5→E2F1→AURKB/MCM2/MCM4/MCM5/TK1.

This explanation also has some predictions: JAK2 mutation has weak direct down-regulation on these genes with $x^*_{\mathrm{O}} < x^*_{\mathrm{J}}$ but $x^*_{\mathrm{T}} > x^*_{\mathrm{TJ}}$; E2F1 and possibly PRMT5 have $y^*_{\mathrm{J}} > y^*_{\mathrm{JT}}$.

Certainly, the pathway JAK2→PRMT5→E2F1→ $\cdots$ is just one possibility. Some evidence indicates the role of p53 in phenomenon (3): JAK2 V617F negatively regulates p53 stabilization [26]; p53 can regulate AURKB and MCM5 [37]; AURKB protein can phosphorylate p53 and alter its transcriptional activity [21]. The complete gene regulatory network should be determined by certain inference methods based on gene expression data [56, 5, 50].

Genes in phenomenon (3) have other correlations, which might imply other related pathways. After the knockdown of transcription factor ZNF143, AURKB, MCM2, MCM4, MCM5 are down-regulated [14]. In hepatocyte-like cells and primary human hepatocytes, AURKB and MCM4 are both up-regulated [64]. AURKB and MCM2 are both positively correlated with TK1 [63]. MCM2 and FHOD1 are both hypermethylated in neuroblastoma tumors [2].

4 Models for non-commutativity in gene expression

In this section, we build models to explain phenomenon (4) that the effect of two mutations are not commutative for gene expression.

4.1 Non-linear ODE model

The model in Eq. 2 can explain different effects of one mutation when another mutation is present or not (non-additivity). However, the effect of different orders of mutations (non-commutativity) cannot be explained, since in that model, we only consider whether one mutation is present or not, while order is not counted. We need another model. This model should have bistability.

Consider an ODE

$$\mathrm{d}x/\mathrm{d}t = b_0 - (x-5)^3 + 3(x-5), \tag{3}$$

where $b_0 = b + b_{\mathrm{Y}}\mathbb{1}_{\mathrm{Y}} + b_{\mathrm{Z}}\mathbb{1}_{\mathrm{Z}}$. See Fig. 8 for the values of fixed points x^* under different b_0. When $-2 < b_0 < 2$, there are two stable branches of fixed points (high and low)

and an unstable middle branch of fixed points. When $b_0 < -2$, there is one stable fixed point, continued from the stable low branch. When $b_0 > 2$, there is one stable fixed point, continued from the stable high branch.

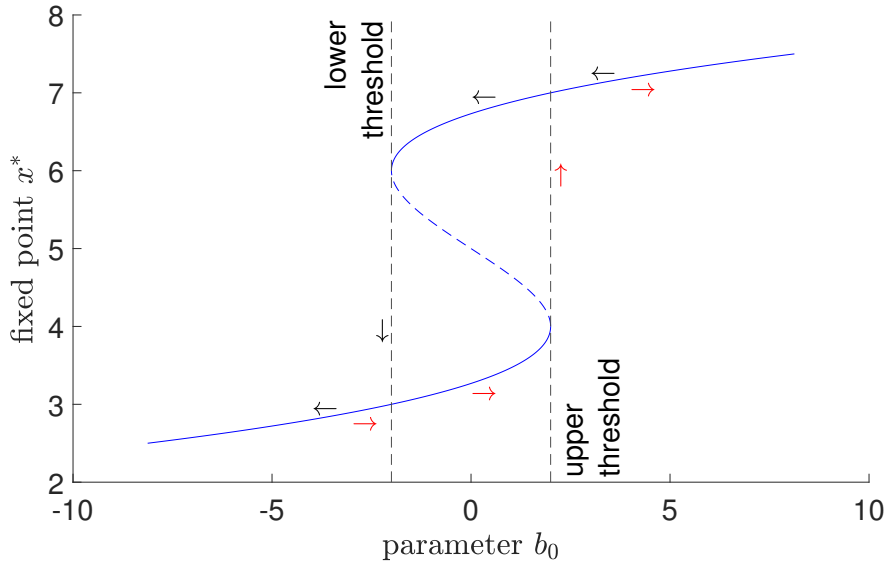

Figure 8: The fixed points of Eq. 3 for different values of b_0. The solid blue line is the stable fixed point, and the dashed blue line is the unstable fixed point. If the system starts at $b_0 < -2$, as the increase of b_0, the system moves along the red arrow in the low branch, until reaching the upper threshold at $b_0 = 2$, at which the system jumps to the high branch. If the system starts at $b_0 > 2$, as the decrease of b_0, the system moves along the black arrow in the high branch, until reaching the lower threshold at $b_0 = -2$, at which the system jumps to the low branch.

When we start from $b_0 < -2$, the system is on the only low branch. When we increase b_0 to $-2 < b_0 < 2$, although there are two stable branches, the system stays at the low branch. When we further increase b_0 until $b_0 > 2$, the stable low branch and the unstable middle branch collide and disappear (saddle-node bifurcation), and the system jumps to the stable high branch. If we start with $b_0 > 2$, the system is at the high branch. When we decrease b_0 to $-2 < b_0 < 2$, the system stays at this stable high branch, until $b_0 < -2$, when the stable high branch and the unstable middle branch collide and disappear, and the system jumps to the low branch. In this model, when we adjust the parameter b_0 in different orders, the stationary state might differ. For example, if the value of b_0 is changed as $0 \to -3 \to 0$, the final state is $x^* \approx 3.268$; if the value of b_0 is changed as $0 \to 3 \to 0$, the final state is $x^* \approx 6.732$.

To explain why TJ and JT have different effects, namely $x^*_{\mathrm{TJ}} \neq x^*_{\mathrm{JT}}$, consider an ODE

as Eq. 3: $\mathrm{d}x/\mathrm{d}t = b_0 - (x-5)^3 + 3(x-5)$, where $b_0 = 3 \times \mathbb{1}_{\mathrm{J}} - 3 \times \mathbb{1}_{\mathrm{T}}$.

With J but not T, $b_0 = 3$, and X is in its only high stationary state $x^*_{\mathrm{J}} \approx 7.104$; if T appears after J, then $b_0 = 0$, and X is in its high branch with stationary state $x^*_{\mathrm{JT}} = 6$. With T but not J, $b_0 = -3$, and X is in its only low stationary state $x^*_{\mathrm{T}} \approx 2.896$; if J appears after T, then $b_0 = 0$, and X is in its low branch with stationary state $x^*_{\mathrm{TJ}} = 4$. See Fig. 8 for detailed explanations. For patients, if the order is JT, final X is high ($x^*_{\mathrm{JT}} = 6$); if the order is TJ, final X is low ($x^*_{\mathrm{TJ}} = 4$). See Fig. 9 for an illustration of this model.

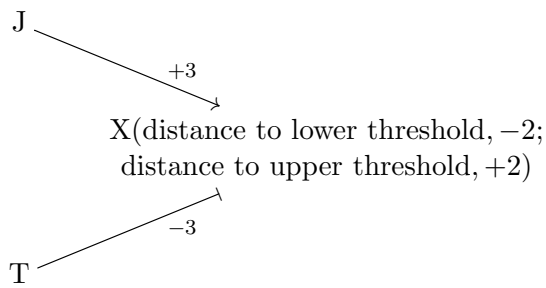

Figure 9: The model diagram that explains $x^*_{\mathrm{TJ}} < x^*_{\mathrm{JT}}$. "$\rightarrow$" means up-regulation; "$\dashv$" means down-regulation. If the input is larger than 2, the system is forced to be on the high branch; if the input is smaller than 2, the system is forced to be on the low branch.

Consider $\mathrm{d}x/\mathrm{d}t = b_0 - (x-5)^3 + 3(x-5)$, where $b_0 = -3 \times \mathbb{1}_{\mathrm{J}} + 3 \times \mathbb{1}_{\mathrm{T}}$. If the order is JT, final X is low ($x^*_{\mathrm{JT}} = 4$); if the order is TJ, final X is high ($x^*_{\mathrm{TJ}} = 6$). See Fig. 10 for an illustration of this model.

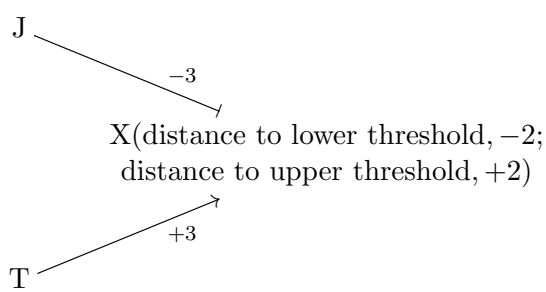

Figure 10: The model diagram that explains $x^*_{\mathrm{JT}} < x^*_{\mathrm{TJ}}$. "$\rightarrow$" means up-regulation; "$\dashv$" means down-regulation. If the input is larger than 2, the system is forced to be on the high branch; if the input is smaller than -2, the system is forced to be on the low branch.

4.2 Markov chain model

In the above ODE model, for a cell with no mutation and a cell with both mutations, the landscapes of gene expression are the same. Since the gene expression on the single-cell

level is essentially stochastic, we can build a Markov chain model [53], and the landscape of gene expression changes with the appearance of each mutation.

Consider a gene X. For a single cell, the expression level (protein count) of X is a random variable X defined on $0, 1, 2, \ldots$. Define

$$A_1(x, c_1, \mu_1, \sigma_1) = \frac{c_1}{\sqrt{2\pi\sigma_1^2}} e^{\frac{(x-\mu_1+1)^2}{2\sigma_1^2}},$$

$$A_2(x, c_1, \mu_1, \sigma_1) = \frac{c_1}{\sqrt{2\pi\sigma_1^2}} e^{\frac{(x-\mu_1)^2}{2\sigma_1^2}},$$

$$B_1(x, c_2, \mu_2, \sigma_2) = \frac{c_2}{\sqrt{2\pi\sigma_2^2}} e^{\frac{(x-\mu_2+1)^2}{2\sigma_2^2}},$$

$$B_2(x, c_2, \mu_2, \sigma_2) = \frac{c_2}{\sqrt{2\pi\sigma_2^2}} e^{\frac{(x-\mu_2)^2}{2\sigma_2^2}}.$$

X follows a continuous-time Markov chain on $0, 1, 2, \ldots$, where the transition rate from x to $x+1$ is

$$r_{x \to x+1} = A_1(x) + B_1(x),$$

and the transition rate from $x+1$ to x is

$$r_{x+1 \to x} = A_2(x) + B_2(x).$$

In all cases, we set $\mu_1 = 1000$, $\mu_2 = 2000$. If JAK2 mutation is not present, we set $c_1 = 1$ and $\sigma_1 = 400$ in $A_1(x, c_1, \mu_1, \sigma_1)$ and $A_2(x, c_1, \mu_1, \sigma_1)$; otherwise, when JAK2 mutation is present, we set $c_1 = 5$ and $\sigma_1 = 80$. Similarly, if TET2 mutation is not present, we set $c_2 = 1$ and $\sigma_2 = 400$ in $B_1(x, c_2, \mu_2, \sigma_2)$ and $B_2(x, c_2, \mu_2, \sigma_2)$; otherwise, when TET2 mutation is present, we set $c_2 = 5$ and $\sigma_2 = 80$.

Since this Markov chain has no cycle, the detailed balance condition is satisfied [54], and we can directly calculate the stationary probability distribution $\mathbb{P}(X = x)$ through

$$\mathbb{P}(X = x) r_{x \to x+1} = \mathbb{P}(X = x+1) r_{x+1 \to x}.$$

(1) If no mutation is present, the stationary distribution is rather flat, with two low peaks near $X = 1000$ and $X = 2000$. (2) If only JAK2 mutation is present, the stationary distribution is mostly concentrated in a sharp peak near $X = 1000$, with a small flat probability near $X = 2000$. (3) If TET2 mutation appears after JAK2 mutation, the small flat probability near $X = 2000$ first shrinks to a sharp peak near $X = 2000$; after a sufficiently long time (e.g., several thousands of years), the heights of two peaks near $X = 1000$ and $X = 2000$ become equal. (4) If only TET2 mutation is present, the stationary distribution is mostly concentrated in a sharp peak near $X = 2000$, with a small

flat probability near $X = 1000$. (5) If JAK2 mutation appears after TET2 mutation, the small flat probability near $X = 1000$ first shrinks to a sharp peak near $X = 1000$; after a sufficiently long time (e.g., several thousands of years), the heights of two peaks near $X = 1000$ and $X = 2000$ become equal.

We can define the potential at $X = x$ as the negative logarithm of the stationary distribution:

$$P(x) = -\log \mathbb{P}(X = x).$$

(1) If no mutation is present, the potential has two shallow wells near $X = 1000$ and $X = 2000$. The system can jump freely between these two wells. (2) If only JAK2 mutation is present, there is a deep well near $X = 1000$ and a shallow well near $X = 2000$, so that the system is very likely to fall into the deep well near $X = 1000$. Here it is easy to jump from the shallow well into the deep well, but not inversely. (3) If TET2 mutation appears after JAK2 mutation, the potential has two deep wells near $X = 1000$ and $X = 2000$. Since the probability is concentrated in the well near $X = 1000$ before the appearance of TET2 mutation, initially, there is no significant change. The only difference is that the flat small probability near $X = 2000$ is reshaped due to the change of the potential function. The reason is that both wells are deep, so that in a relatively short time scale, it is extremely difficult to jump from one well into another. (4) After sufficiently long time (possibly longer than the life span of human being), two wells become equally likely again. (5) If only TET2 mutation is present, the system is likely to stay in the deep well near $X = 2000$. (6) If JAK2 mutation appears after TET2 mutation, the system will firstly stay in the well near $X = 2000$, but finally get equally distributed.

In this model, different orders of mutations (JT and TJ) finally lead to the same stationary distribution. However, different history leads to different wells in a relative short time scale. If this mesoscopic time scale is comparable to the life span of human being, then the final stationary distribution is de facto inaccessible, since it might take thousands of years. This explains the difference between patients with different orders of mutations.

5 Models for non-commutativity in cell population and age

In this section, we build models to explain phenomena (5, 6) that the effect of two mutations are not commutative for cell population and the age at diagnosis.

5.1 Different mechanisms for explaining phenomena (5, 6)

Since myeloproliferative neoplasm is the result of certain mutations, to simplify the discussion, we assume that the age of diagnosis is highly correlated to the age when the first cell with both JAK2 and TET2 mutations appears. Also, we assume that the percentage of cells with only one mutation is similar for the time when the first cell with both JAK2

and TET2 mutations appears and the time when this percentage is measured. Therefore, we actually discuss two other phenomena:
(5') For TET2-first patients, at the time when the first TET2-JAK2 cell appears, the percentage of TET2-only cells is significantly higher than the percentage of JAK2-only cells at the time when the first JAK2-TET2 cell appears for JAK2-first patients.
(6') For JAK-first patients, the time when the first JAK2-TET2 cell appears is significantly earlier than the time when the first TET2-JAK2 cell appears for TET2-first patients.

There are three mechanisms that can be used to explain phenomena (5', 6').
(I) Ortmann et al. [31] propose that cells with JAK2 mutation have a mild proliferation advantage, and cells with TET2 mutation (whether JAK2 is present or not) have a significant proliferation advantage. This is used to explain phenomenon (5'). If JAK2 mutation first appears, such JAK2-only cells grow slower. Thus when TET2 mutation appears in a JAK2-only cell, the number of JAK2-only cells is not large. (Notice that if TET2 mutation appears in a non-mutant cell, we cannot determine the order of mutation, and this patient is not considered.) If TET2 mutation first appears, such TET2-only cells grow faster. Thus when JAK2 mutation appears in a TET2-only cell, the number of JAK2-only cells is large. In the following simulations, we find that this mechanism can also be used to explain phenomenon (6').
(II) Teimouri and Kolomeisky [43] assume that two mutations have different mutation rates, and prove phenomenon (6') in a specific model. We find that in a more general model, this assumption can still be used to explain phenomena (5', 6'). If the mutation rate of JAK2 is lower than that of TET2, then when JAK2 appears first, JAK2-only cells have a shorter time to proliferate. This explains phenomenon (5'). The explanation for phenomenon (6') is in Subsection 5.4
(III) We propose another mechanism that for cells with JAK2 mutation, the mutation rate for TET2 is higher. In other words, JAK2 mutation can induce TET2 mutation. Therefore, if JAK2 mutation appears first, TET2 mutation can appear in a shorter time, which explains phenomenon (6'). Besides, this means JAK2-only cells do not have much time to proliferate before the appearance of TET2 mutation, which explains phenomenon (5').

5.2 Generalized Moran process model

Cell population dynamics with state transition has been widely studied [67, 29, 7, 1]. To mathematically study phenomena (5', 6'), we consider a generalized Moran process model [35] for cell population dynamics with mutations. Unlike branching processes [16], in this process, the total number of cells is fixed, which makes sense since the number of hematopoietic stem cells is relatively stable. Cells can be in five states: non-mutant, JAK2-only, TET2-only, JAK2-TET2, TET2-JAK2. Each type of cell has its own birth coefficient and death coefficient. The process is in discrete time. At each time step, one cell is randomly picked out (weighted by its death coefficient) to die. Then one cell is

randomly picked out to divide (weighted by its birth coefficient). After the division, one child cell will keep the original state, and the other child cell has a chance to acquire a new mutation (depending on the mutation probability).

In the following, the suffix O means non-mutant cells; the suffix J means JAK2-only cells; the suffix T means TET2-only cells; the suffix JT means JAK2-TET2 cells; the suffix TJ means TET2-JAK2 cells.

The number of non-mutant cells is n_{O}. The birth coefficient of non-mutant cells is b_{O}. The death coefficient of non-mutant cells is d_{O}. These three quantities are also defined for suffixes J, T, JT, TJ.

We have four mutation probabilities $m_{\mathrm{O}\to\mathrm{J}}$, $m_{\mathrm{O}\to\mathrm{T}}$, $m_{\mathrm{J}\to\mathrm{JT}}$, $m_{\mathrm{T}\to\mathrm{TJ}}$. For example, $m_{\mathrm{O}\to\mathrm{J}}$ means the probability of having JAK2 mutation for a non-mutant cell if it is chosen to divide.

The state space of this process is $(n_{\mathrm{O}}, n_{\mathrm{J}}, n_{\mathrm{T}}, n_{\mathrm{JT}}, n_{\mathrm{TJ}})$. For each time point, during the death phase, the transition probability for the death of a non-mutant cell is

$$\begin{aligned}&\mathbb{P}[(n_{\mathrm{O}}, n_{\mathrm{J}}, n_{\mathrm{T}}, n_{\mathrm{JT}}, n_{\mathrm{TJ}}) \to (n_{\mathrm{O}}-1, n_{\mathrm{J}}, n_{\mathrm{T}}, n_{\mathrm{JT}}, n_{\mathrm{TJ}})]\\ =&\frac{n_{\mathrm{O}}d_{\mathrm{O}}}{n_{\mathrm{O}}d_{\mathrm{O}}+n_{\mathrm{J}}d_{\mathrm{J}}+n_{\mathrm{T}}d_{\mathrm{T}}+n_{\mathrm{JT}}d_{\mathrm{JT}}+n_{\mathrm{TJ}}d_{\mathrm{TJ}}}.\end{aligned}$$

The transition probability of other death events can be defined similarly. During the birth phase, the transition probability for the division of a non-mutant cell without introducing new mutation is

$$\begin{aligned}&\mathbb{P}[(n_{\mathrm{O}}, n_{\mathrm{J}}, n_{\mathrm{T}}, n_{\mathrm{JT}}, n_{\mathrm{TJ}}) \to (n_{\mathrm{O}}+1, n_{\mathrm{J}}, n_{\mathrm{T}}, n_{\mathrm{JT}}, n_{\mathrm{TJ}})]\\ =&\frac{n_{\mathrm{O}}b_{\mathrm{O}}}{n_{\mathrm{O}}b_{\mathrm{O}}+n_{\mathrm{J}}b_{\mathrm{J}}+n_{\mathrm{T}}b_{\mathrm{T}}+n_{\mathrm{JT}}b_{\mathrm{JT}}+n_{\mathrm{TJ}}b_{\mathrm{TJ}}}(1-m_{\mathrm{O}\to\mathrm{J}}-m_{\mathrm{O}\to\mathrm{T}}).\end{aligned}$$

For the right hand side, the first term is the probability that a non-mutant cell is chosen to divide, and the second term is the probability that no mutation is introduced. Similarly,

$$\begin{aligned}&\mathbb{P}[(n_{\mathrm{O}}, n_{\mathrm{J}}, n_{\mathrm{T}}, n_{\mathrm{JT}}, n_{\mathrm{TJ}}) \to (n_{\mathrm{O}}, n_{\mathrm{J}}+1, n_{\mathrm{T}}, n_{\mathrm{JT}}, n_{\mathrm{TJ}})]\\ =&\frac{n_{\mathrm{O}}b_{\mathrm{O}}}{n_{\mathrm{O}}b_{\mathrm{O}}+n_{\mathrm{J}}b_{\mathrm{J}}+n_{\mathrm{T}}b_{\mathrm{T}}+n_{\mathrm{JT}}b_{\mathrm{JT}}+n_{\mathrm{TJ}}b_{\mathrm{TJ}}}m_{\mathrm{O}\to\mathrm{J}}\\ &+\frac{n_{\mathrm{J}}b_{\mathrm{J}}}{n_{\mathrm{O}}b_{\mathrm{O}}+n_{\mathrm{J}}b_{\mathrm{J}}+n_{\mathrm{T}}b_{\mathrm{T}}+n_{\mathrm{JT}}b_{\mathrm{JT}}+n_{\mathrm{TJ}}b_{\mathrm{TJ}}}(1-m_{\mathrm{J}\to\mathrm{JT}}).\end{aligned}$$

$$\begin{aligned}&\mathbb{P}[(n_{\mathrm{O}}, n_{\mathrm{J}}, n_{\mathrm{T}}, n_{\mathrm{JT}}, n_{\mathrm{TJ}}) \to (n_{\mathrm{O}}, n_{\mathrm{J}}, n_{\mathrm{T}}+1, n_{\mathrm{JT}}, n_{\mathrm{TJ}})]\\ =&\frac{n_{\mathrm{O}}b_{\mathrm{O}}}{n_{\mathrm{O}}b_{\mathrm{O}}+n_{\mathrm{J}}b_{\mathrm{J}}+n_{\mathrm{T}}b_{\mathrm{T}}+n_{\mathrm{JT}}b_{\mathrm{JT}}+n_{\mathrm{TJ}}b_{\mathrm{TJ}}}m_{\mathrm{O}\to\mathrm{T}}\\ &+\frac{n_{\mathrm{T}}b_{\mathrm{T}}}{n_{\mathrm{O}}b_{\mathrm{O}}+n_{\mathrm{J}}b_{\mathrm{J}}+n_{\mathrm{T}}b_{\mathrm{T}}+n_{\mathrm{JT}}b_{\mathrm{JT}}+n_{\mathrm{TJ}}b_{\mathrm{TJ}}}(1-m_{\mathrm{T}\to\mathrm{TJ}}).\end{aligned}$$

$$
\begin{aligned}
&\mathbb{P}[(n_{\mathrm{O}}, n_{\mathrm{J}}, n_{\mathrm{T}}, n_{\mathrm{JT}}, n_{\mathrm{TJ}}) \to (n_{\mathrm{O}}, n_{\mathrm{J}}, n_{\mathrm{T}}, n_{\mathrm{JT}}+1, n_{\mathrm{TJ}})] \\
=&\frac{n_{\mathrm{J}} b_{\mathrm{J}}}{n_{\mathrm{O}} b_{\mathrm{O}} + n_{\mathrm{J}} b_{\mathrm{J}} + n_{\mathrm{T}} b_{\mathrm{T}} + n_{\mathrm{JT}} b_{\mathrm{JT}} + n_{\mathrm{TJ}} b_{\mathrm{TJ}}} m_{\mathrm{J}\to\mathrm{JT}} \\
&+ \frac{n_{\mathrm{JT}} b_{\mathrm{JT}}}{n_{\mathrm{O}} b_{\mathrm{O}} + n_{\mathrm{J}} b_{\mathrm{J}} + n_{\mathrm{T}} b_{\mathrm{T}} + n_{\mathrm{JT}} b_{\mathrm{JT}} + n_{\mathrm{TJ}} b_{\mathrm{TJ}}}.
\end{aligned}
$$

$$
\begin{aligned}
&\mathbb{P}[(n_{\mathrm{O}}, n_{\mathrm{J}}, n_{\mathrm{T}}, n_{\mathrm{JT}}, n_{\mathrm{TJ}}) \to (n_{\mathrm{O}}, n_{\mathrm{J}}, n_{\mathrm{T}}, n_{\mathrm{JT}}, n_{\mathrm{TJ}}+1)] \\
=&\frac{n_{\mathrm{T}} b_{\mathrm{T}}}{n_{\mathrm{O}} b_{\mathrm{O}} + n_{\mathrm{J}} b_{\mathrm{J}} + n_{\mathrm{T}} b_{\mathrm{T}} + n_{\mathrm{JT}} b_{\mathrm{JT}} + n_{\mathrm{TJ}} b_{\mathrm{TJ}}} m_{\mathrm{T}\to\mathrm{TJ}} \\
&+ \frac{n_{\mathrm{TJ}} b_{\mathrm{TJ}}}{n_{\mathrm{O}} b_{\mathrm{O}} + n_{\mathrm{J}} b_{\mathrm{J}} + n_{\mathrm{T}} b_{\mathrm{T}} + n_{\mathrm{JT}} b_{\mathrm{JT}} + n_{\mathrm{TJ}} b_{\mathrm{TJ}}}.
\end{aligned}
$$

After each time point, the total population $n_{\mathrm{O}} + n_{\mathrm{J}} + n_{\mathrm{T}} + n_{\mathrm{JT}} + n_{\mathrm{TJ}}$ is fixed, which is denoted as n.

5.3 Simulation results

In the simulations, for all three mechanisms, we set $n = 100$, $b_{\mathrm{O}} = 1$, $d_{\mathrm{O}} = 1$, $d_{\mathrm{J}} = 1$, $d_{\mathrm{T}} = 1$, $d_{\mathrm{JT}} = 1$, $d_{\mathrm{TJ}} = 1$.

In mechanism (I), all mutations have the same probability, and cells with TET2 mutation grows faster than JAK2-only cells. We set $b_{\mathrm{J}} = 2$, $b_{\mathrm{T}} = 4$, $b_{\mathrm{JT}} = 4$, $b_{\mathrm{TJ}} = 4$, $m_{\mathrm{O}\to\mathrm{J}} = 0.1$, $m_{\mathrm{O}\to\mathrm{T}} = 0.1$, $m_{\mathrm{J}\to\mathrm{JT}} = 0.1$, $m_{\mathrm{T}\to\mathrm{TJ}} = 0.1$.

In mechanism (II), all cells with at least one mutation have the same birth coefficient, and the mutation probability of TET2 is higher than that of JAK2 mutation. We set $b_{\mathrm{J}} = 2$, $b_{\mathrm{T}} = 2$, $b_{\mathrm{JT}} = 2$, $b_{\mathrm{TJ}} = 2$, $m_{\mathrm{O}\to\mathrm{J}} = 0.1$, $m_{\mathrm{O}\to\mathrm{T}} = 0.2$, $m_{\mathrm{J}\to\mathrm{JT}} = 0.2$, $m_{\mathrm{T}\to\mathrm{TJ}} = 0.1$.

In mechanism (III), all cells with at least one mutation have the same birth coefficient, and JAK2 mutation can induce TET2 mutation, correspondint to a larger $m_{\mathrm{J}\to\mathrm{JT}}$. We set $b_{\mathrm{J}} = 2$, $b_{\mathrm{T}} = 2$, $b_{\mathrm{JT}} = 2$, $b_{\mathrm{TJ}} = 2$, $m_{\mathrm{O}\to\mathrm{J}} = 0.1$, $m_{\mathrm{O}\to\mathrm{T}} = 0.1$, $m_{\mathrm{J}\to\mathrm{JT}} = 0.2$, $m_{\mathrm{T}\to\mathrm{TJ}} = 0.1$.

Since we use this model to study phenomena (5', 6'), the process is stopped once $n_{\mathrm{JT}} = 1$ or $n_{\mathrm{TJ}} = 1$. If $n_{\mathrm{J}} > 0$ and $n_{\mathrm{T}} > 0$, then we cannot determine the order of mutations in practice, and this simulation result is abandoned. If $n_{\mathrm{JT}} = 1$ and $n_{\mathrm{T}} = 0$, we record the corresponding n_{J} and the current time point T. This correspond to a JAK2-first patient. If $n_{\mathrm{TJ}} = 1$ and $n_{\mathrm{J}} = 0$, we record the corresponding n_{T} and the current time point T. This correspond to a TET2-first patient. To verify phenomenon (5'), we compare $\mathbb{E}(n_{\mathrm{J}} \mid n_{\mathrm{JT}} = 1, n_{\mathrm{T}} = 0)$ and $\mathbb{E}(n_{\mathrm{T}} \mid n_{\mathrm{TJ}} = 1, n_{\mathrm{J}} = 0)$. To verify phenomenon (6'), we compare $\mathbb{E}(T \mid n_{\mathrm{JT}} = 1, n_{\mathrm{T}} = 0)$ and $\mathbb{E}(T \mid n_{\mathrm{TJ}} = 1, n_{\mathrm{J}} = 0)$. We also use t-test to examine whether the difference in mean is significant. For each mechanism, we run the simulation 10^6 times.

For mechanism (I),

$$\mathbb{E}(n_{\mathrm{J}} \mid n_{\mathrm{JT}} = 1, n_{\mathrm{T}} = 0) = 4.02 < 5.41 = \mathbb{E}(n_{\mathrm{T}} \mid n_{\mathrm{TJ}} = 1, n_{\mathrm{J}} = 0),$$

and the p-value from t-test is smaller than 10^{-200}.

$$\mathbb{E}(T \mid n_{\mathrm{JT}} = 1, n_{\mathrm{T}} = 0) = 24.15 < 25.93 = \mathbb{E}(T \mid n_{\mathrm{TJ}} = 1, n_{\mathrm{J}} = 0),$$

and the p-value from t-test is $3.55 * 10^{-40}$. We have a guess for why this mechanism produces phenomenon (6'): Since $b_{\mathrm{T}} > b_{\mathrm{J}}$, the probability that $n_{\mathrm{JT}} = 1, n_{\mathrm{T}} = 0$ is smaller than $n_{\mathrm{TJ}} = 1, n_{\mathrm{J}} = 0$. Those fewer simulations with $n_{\mathrm{JT}} = 1, n_{\mathrm{T}} = 0$ generally mean that JAK2 mutation appears fast, and JAK2-only cells happen to divide more often.

For mechanism (II),

$$\mathbb{E}(n_{\mathrm{J}} \mid n_{\mathrm{JT}} = 1, n_{\mathrm{T}} = 0) = 2.37 < 6.51 = \mathbb{E}(n_{\mathrm{T}} \mid n_{\mathrm{TJ}} = 1, n_{\mathrm{J}} = 0),$$

and the p-value from t-test is smaller than 10^{-200}.

$$\mathbb{E}(T \mid n_{\mathrm{JT}} = 1, n_{\mathrm{T}} = 0) = 11.52 < 24.75 = \mathbb{E}(T \mid n_{\mathrm{TJ}} = 1, n_{\mathrm{J}} = 0),$$

and the p-value from t-test is smaller than 10^{-200}.

For mechanism (III),

$$\mathbb{E}(n_{\mathrm{J}} \mid n_{\mathrm{JT}} = 1, n_{\mathrm{T}} = 0) = 3.59 < 4.27 = \mathbb{E}(n_{\mathrm{T}} \mid n_{\mathrm{TJ}} = 1, n_{\mathrm{J}} = 0),$$

and the p-value from t-test is $1.50 * 10^{-192}$.

$$\mathbb{E}(T \mid n_{\mathrm{JT}} = 1, n_{\mathrm{T}} = 0) = 22.79 < 25.97 = \mathbb{E}(T \mid n_{\mathrm{TJ}} = 1, n_{\mathrm{J}} = 0),$$

and the p-value from t-test is $4.35 * 10^{-133}$.

We can see that all three mechanisms can produce phenomena (5', 6') in this model, and the differences are all statistically significant.

5.4 Theoretical analysis of mechanism (II) for phenomenon (6')

Our generalized Moran process model is difficult to study analytically. To explain why mechanism (II) produces phenomenon (6'), we consider a simplified model, which is a limiting case of the generalized Moran process model when $b_{\mathrm{J}} = b_{\mathrm{T}} = b_{\mathrm{JT}} = b_{\mathrm{TJ}} \gg b_{\mathrm{O}}$.

Assume the appearance of JAK2 mutation needs a waiting time T_1 that satisfies an exponential distribution with parameter λ_1, meaning that $\mathbb{E}(T_1) = 1/\lambda_1$. TET2 mutation needs an exponential waiting time T_2 with parameter λ_2. Here T_1 and T_2 are independent, and $\lambda_1 < \lambda_2$, meaning that the mutation rate of TET2 is higher. Define $T = \max\{T_1, T_2\}$ to be the time that both mutations appear. Since the growth of cells with at least one

mutation is much faster than non-mutant cells, this means that we can find a cell with both mutations. We can calculate that

$$\mathbb{E}(T \mid T_1 > T_2) = \frac{\int_0^{+\infty}\int_0^{t_1} t_1\lambda_1 e^{-\lambda_1 t_1}\lambda_2 e^{-\lambda_2 t_2}\mathrm{d}t_2\mathrm{d}t_1}{\int_0^{+\infty}\int_0^{t_1} \lambda_1 e^{-\lambda_1 t_1}\lambda_2 e^{-\lambda_2 t_2}\mathrm{d}t_2\mathrm{d}t_1} = \frac{2\lambda_1\lambda_2 + \lambda_2^2}{\lambda_1\lambda_2(\lambda_1+\lambda_2)},$$

$$\mathbb{E}(T \mid T_1 \leq T_2) = \frac{\int_0^{+\infty}\int_0^{t_2} t_2\lambda_1 e^{-\lambda_1 t_1}\lambda_2 e^{-\lambda_2 t_2}\mathrm{d}t_1\mathrm{d}t_2}{\int_0^{+\infty}\int_0^{t_2} \lambda_1 e^{-\lambda_1 t_1}\lambda_2 e^{-\lambda_2 t_2}\mathrm{d}t_1\mathrm{d}t_2} = \frac{\lambda_1^2 + 2\lambda_1\lambda_2}{\lambda_1\lambda_2(\lambda_1+\lambda_2)}.$$

Since $\lambda_1 < \lambda_2$, we have $\mathbb{E}(T \mid T_1 > T_2) > \mathbb{E}(T \mid T_1 \leq T_2)$, meaning that for JAK2-first patients ($T_1 \leq T_2$), the waiting time is shorter than TET2-first patients ($T_1 > T_2$). One illustrative explanation is that when $T_1 < T_2$, it is more likely that T_1 is exceptionally shorter, not that T_2 is exceptionally longer.

6 Discussion

In this paper, we consider two genetic mutations in myeloproliferative neoplasm: JAK2 and TET2. The effect of one mutation depends on whether the other mutation is present. Besides, when both mutations are present, the order of appearance also affects gene expression. We build several models to explain such phenomena. Specifically, in Subsection 3.3, we find experimental evidence that partially verifies our model. There are also some conjectures that can be examined by experiments, and we can use our results to guide experiment design [57]. This paper could motivate more related experiments, and more experimental data could lead to more theoretical works inspired by models in this paper. In this paper, we only use ordinary differential equations models and Markov chain models. Other mathematical tools, such as dynamical system [65, 62] and partial differential equations [49], can also be applied. Models in this paper can also be used to explain some phenomena in other cancer types, not necessarily for myeloproliferative neoplasm [60]. Besides, our results might provide insights for cancer treatment [45], which might needs complicated policy optimization [58, 59]. Besides building models to illustrate what mechanisms are possible, we can also think reversibly to find what mechanisms are impossible [55, 46].

References

[1] Angelini, E., Wang, Y., Zhou, J. X., Qian, H., and Huang, S. A model for the intrinsic limit of cancer therapy: Duality of treatment-induced cell death and treatment-induced stemness. *PLOS Computational Biology 18*, 7 (2022), e1010319.

[2] Aygun, N. Biological and genetic features of neuroblastoma and their clinical importance. *Current Pediatric Reviews 14*, 2 (2018), 73–90.

[3] Baik, R., Wyman, S. K., Kabir, S., and Corn, J. E. Genome editing to model and reverse a prevalent mutation associated with myeloproliferative neoplasms. *PLOS ONE 16*, 3 (2021), e0247858.

[4] Benferhat, S., and Smaoui, S. Possibilistic causal networks for handling interventions: A new propagation algorithm. In *Proceedings of the National Conference on Artificial Intelligence* (2007), vol. 22, Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999, p. 373.

[5] Bocci, F., Zhou, P., and Nie, Q. spliceJAC: transition genes and state-specific gene regulation from single-cell transcriptome data. *Molecular Systems Biology 18*, 11 (2022), e11176.

[6] Caravagna, G., Giarratano, Y., Ramazzotti, D., Tomlinson, I., Graham, T. A., Sanguinetti, G., and Sottoriva, A. Detecting repeated cancer evolution from multi-region tumor sequencing data. *Nature Methods 15*, 9 (2018), 707–714.

[7] Chen, X., Wang, Y., Feng, T., Yi, M., Zhang, X., and Zhou, D. The overshoot and phenotypic equilibrium in characterizing cancer dynamics of reversible phenotypic plasticity. *Journal of Theoretical Biology 390* (2016), 40–49.

[8] Chiba, S. Dysregulation of tet2 in hematologic malignancies. *International Journal of Hematology 105* (2017), 17–22.

[9] Clarke, M. A., Woodhouse, S., Piterman, N., Hall, B. A., and Fisher, J. Using state space exploration to determine how gene regulatory networks constrain mutation order in cancer evolution. *Automated Reasoning for Systems Biology and Medicine* (2019), 133–153.

[10] Dessalles, R., Pan, Y., Xia, M., Maestrini, D., D'Orsogna, M. R., and Chou, T. How naive T-cell clone counts are shaped by heterogeneous thymic output and homeostatic proliferation. *Frontiers in Immunology 12* (2021).

[11] Gao, Y., Gaither, J., Chifman, J., and Kubatko, L. A phylogenetic approach to inferring the order in which mutations arise during cancer progression. *PLOS Computational Biology 18*, 12 (2022), e1010560.

[12] Herbet, M., Salomon, A., Feige, J.-J., and Thomas, M. Acquisition order of Ras and p53 gene alterations defines distinct adrenocortical tumor phenotypes. *PLOS Genetics 8*, 5 (2012), e1002700.

[13] Ito, S., D'Alessio, A. C., Taranova, O. V., Hong, K., Sowers, L. C., and Zhang, Y. Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature 466*, 7310 (2010), 1129–1133.

[14] Izumi, H., Wakasugi, T., Shimajiri, S., Tanimoto, A., Sasaguri, Y., Kashiwagi, E., Yasuniwa, Y., Akiyama, M., Han, B., Wu, Y., et al. Role of ZNF143 in tumor growth through transcriptional regulation of DNA replication and cell-cycle-associated genes. *Cancer Science 101*, 12 (2010), 2538–2545.

[15] Jacobs, M., Kim, I., and Tong, J. Tumor growth with nutrients: Regularity and stability. *arXiv preprint arXiv:2204.07572* (2022).

[16] Jiang, D.-Q., Wang, Y., and Zhou, D. Phenotypic equilibrium as probabilistic convergence in multi-phenotype cell population dynamics. *PLOS ONE 12*, 2 (2017), e0170916.

[17] Kang, Y., Gu, C., Yuan, L., Wang, Y., Zhu, Y., Li, X., Luo, Q., Xiao, J., Jiang, D., Qian, M., et al. Flexibility and symmetry of prokaryotic genome rearrangement reveal lineage-associated core-gene-defined genome organizational frameworks (vol 5, e01867, 2014). *MBIO 6*, 1 (2015).

[18] Kent, D. G., and Green, A. R. Order matters: the order of somatic mutations influences cancer evolution. *Cold Spring Harbor Perspectives in Medicine 7*, 4 (2017), a027060.

[19] Khakabimamaghani, S., Ding, D., Snow, O., and Ester, M. Uncovering the subtype-specific temporal order of cancer pathway dysregulation. *PLOS Computational Biology 15*, 11 (2019), e1007451.

[20] Klampfl, T., Gisslinger, H., Harutyunyan, A. S., Nivarthi, H., Rumi, E., Milosevic, J. D., Them, N. C., Berg, T., Gisslinger, B., Pietra, D., et al. Somatic mutations of calreticulin in myeloproliferative neoplasms. *New England Journal of Medicine 369*, 25 (2013), 2379–2390.

[21] Koczor, C. A., Lee, E. K., Torres, R. A., Boyd, A., Vega, J. D., Uppal, K., Yuan, F., Fields, E. J., Samarel, A. M., and Lewis, W. Detection of differentially methylated gene promoters in failing and nonfailing human left ventricle myocardium using computation analysis. *Physiological Genomics 45*, 14 (2013), 597–605.

[22] Koushyar, S., Economides, G., Zaat, S., Jiang, W., Bevan, C. L., and Dart, D. The prohibitin-repressive interaction with E2F1 is rapidly inhibited by androgen signalling in prostate cancer cells. *Oncogenesis 6*, 5 (2017), e333–e333.

[23] Levine, A. J., Jenkins, N. A., and Copeland, N. G. The roles of initiating truncal mutations in human cancers: the order of mutations and tumor cell type matters. *Cancer Cell 35*, 1 (2019), 10–15.

[24] Levine, R. L., and Gilliland, D. G. JAK-2 mutations and their relevance to myeloproliferative disease. *Current Opinion in Hematology 14*, 1 (2007), 43–47.

[25] Mazaya, M., Trinh, H.-C., and Kwon, Y.-K. Effects of ordered mutations on dynamics in signaling networks. *BMC Medical Genomics 13* (2020), 1–12.

[26] Nakatake, M., Monte-Mor, B., Debili, N., Casadevall, N., Ribrag, V., Solary, E., Vainchenker, W., and Plo, I. JAK2V617F negatively regulates p53 stabilization by enhancing MDM2 via La expression in myeloproliferative neoplasms. *Oncogene 31*, 10 (2012), 1323–1333.

[27] Nangalia, J., Massie, C. E., Baxter, E. J., Nice, F. L., Gundem, G., Wedge, D. C., Avezov, E., Li, J., Kollmann, K., Kent, D. G., et al. Somatic CALR mutations in myeloproliferative neoplasms with nonmutated JAK2. *New England Journal of Medicine 369*, 25 (2013), 2391–2405.

[28] Nangalia, J., Nice, F. L., Wedge, D. C., Godfrey, A. L., Grinfeld, J., Thakker, C., Massie, C. E., Baxter, J., Sewell, D., Silber, Y., et al. Dnmt3a mutations occur early or late in patients with myeloproliferative neoplasms and mutation order influences phenotype. *Haematologica 100*, 11 (2015), e438.

[29] Niu, Y., Wang, Y., and Zhou, D. The phenotypic equilibrium of cancer cells: From average-level stability to path-wise convergence. *Journal of Theoretical Biology 386* (2015), 7–17.

[30] Ohtani, K., Iwanaga, R., Nakamura, M., Ikeda, M.-a., Yabuta, N., Tsuruga, H., and Nojima, H. Cell growth-regulated expression of mammalian MCM5 and MCM6 genes mediated by the transcription factor E2F. *Oncogene 18*, 14 (1999), 2299–2309.

[31] Ortmann, C. A., Kent, D. G., Nangalia, J., Silber, Y., Wedge, D. C., Grinfeld, J., Baxter, E. J., Massie, C. E., Papaemmanuil, E., Menon, S., et al. Effect of mutation order on myeloproliferative neoplasms. *New England Journal of Medicine 372*, 7 (2015), 601–612.

[32] Pastore, F., Bhagwat, N., Pastore, A., Radzisheuskaya, A., Karzai, A., Krishnan, A., Li, B., Bowman, R. L., Xiao, W., Viny, A. D., et al. PRMT5 Inhibition Modulates E2F1 Methylation and Gene-Regulatory Networks Leading to Therapeutic Efficacy in JAK2V617F-Mutant MPNPRMT5 Inhibition in MPN. *Cancer Discovery 10*, 11 (2020), 1742–1757.

[33] Pellegrina, L., and Vandin, F. Discovering significant evolutionary trajectories in cancer phylogenies. *Bioinformatics 38*, Supplement_2 (2022), ii49–ii55.

[34] Qian, H., and Cheng, Y.-C. Counting single cells and computing their heterogeneity: from phenotypic frequencies to mean value of a quantitative biomarker. *Quantitative Biology 8*, 2 (2020), 172–176.

[35] Quan, J., and Wang, X.-J. Evolutionary games in a generalized moran process with arbitrary selection strength and mutation. *Chinese Physics B 20*, 3 (2011), 030203.

[36] Ramazzotti, D., Graudenzi, A., De Sano, L., Antoniotti, M., and Caravagna, G. Learning mutational graphs of individual tumour evolution from single-cell and multi-region sequencing data. *BMC Bioinformatics 20*, 1 (2019), 1–13.

[37] Reyes, M. E., Ma, J., Grove, M. L., Ater, J. L., Morrison, A. C., and Hildebrandt, M. A. RNA sequence analysis of inducible pluripotent stem cell-derived cardiomyocytes reveals altered expression of DNA damage and cell cycle genes in response to doxorubicin. *Toxicology and Applied Pharmacology 356* (2018), 44–53.

[38] Roquet, N., Soleimany, A. P., Ferris, A. C., Aaronson, S., and Lu, T. K. Synthetic recombinase-based state machines in living cells. *Science 353*, 6297 (2016), aad8559.

[39] Shih, A. H., Abdel-Wahab, O., Patel, J. P., and Levine, R. L. The role of mutations in epigenetic regulators in myeloid malignancies. *Nature Reviews Cancer 12*, 9 (2012), 599–612.

[40] Swanton, C. Cancer evolution constrained by mutation order. *New England Journal of Medicine 372*, 7 (2015), 661–663.

[41] Talarmain, L. *Modelling timing in blood cancers*. Ph.D. thesis, University of Cambridge, 2021.

[42] Talarmain, L., Clarke, M. A., Shorthouse, D., Cabrera-Cosme, L., Kent, D. G., Fisher, J., and Hall, B. A. HOXA9 has the hallmarks of a biological switch with implications in blood cancers. *Nature Communications 13*, 1 (2022), 5829.

[43] Teimouri, H., and Kolomeisky, A. B. Temporal order of mutations influences cancer initiation dynamics. *Physical Biology 18*, 5 (2021), 056002.

[44] Turajlic, S., Xu, H., Litchfield, K., Rowan, A., Horswell, S., Chambers, T., O'Brien, T., Lopez, J. I., Watkins, T. B., Nicol, D., et al. Deterministic evolutionary trajectories influence primary tumor growth: TRACERx renal. *Cell 173*, 3 (2018), 595–610.

[45] Wang, Y. *Some Problems in Stochastic Dynamics and Statistical Analysis of Single-Cell Biology of Cancer*. Ph.D. thesis, University of Washington, 2018.

[46] Wang, Y. Impossibility results about inheritance and order of death. *PLOS ONE 17*, 11 (2022), e0277430.

[47] Wang, Y. Two metrics on rooted unordered trees with labels. *Algorithms for Molecular Biology 17*, 1 (2022), 1–17.

[48] Wang, Y. Longest common subsequence algorithms and applications in determining transposable genes. *arXiv preprint arXiv:2301.03827* (2023).

[49] Wang, Y., Dessalles, R., and Chou, T. Modelling the impact of birth control policies on China's population and age: effects of delayed births and minimum birth age constraints. *Royal Society Open Science 9*, 6 (2022), 211619.

[50] Wang, Y., and He, S. Inference on autoregulation in gene expression. *arXiv preprint arXiv:2201.03164* (2022).

[51] Wang, Y., Kropp, J., and Morozova, N. Biological notion of positional information/value in morphogenesis theory. *International Journal of Developmental Biology 64*, 10-11-12 (2020), 453–463.

[52] Wang, Y., Minarsky, A., Penner, R., Soulé, C., and Morozova, N. Model of morphogenesis. *Journal of Computational Biology 27*, 9 (2020), 1373–1383.

[53] Wang, Y., Mistry, B. A., and Chou, T. Discrete stochastic models of selex: Aptamer capture probabilities and protocol optimization. *The Journal of Chemical Physics 156*, 24 (2022), 244103.

[54] Wang, Y., and Qian, H. Mathematical representation of Clausius' and Kelvin's statements of the second law and irreversibility. *Journal of Statistical Physics 179*, 3 (2020), 808–837.

[55] Wang, Y., and Wang, L. Causal inference in degenerate systems: An impossibility result. In *International Conference on Artificial Intelligence and Statistics* (2020), PMLR, pp. 3383–3392.

[56] Wang, Y., and Wang, Z. Inference on the structure of gene regulatory networks. *Journal of Theoretical Biology 539* (2022), 111055.

[57] Wang, Y., Zhang, B., Kropp, J., and Morozova, N. Inference on tissue transplantation experiments. *Journal of Theoretical Biology 520* (2021), 110645.

[58] Wang, Y., and Zheng, Z. Measuring policy performance in online pricing with offline data. *Available at SSRN 3729003* (2021).

[59] Wang, Y., Zheng, Z., and Shen, Z.-J. M. Online pricing with polluted offline data. *Available at SSRN 4320324* (2023).

[60] Wang, Y., Zhou, J. X., Pedrini, E., Rubin, I., Khalil, M., Qian, H., and Huang, S. Cell population growth as manifestation of stochastic heterogeneity of cell phenotype. *bioRxiv* (2023), 2023–02.

[61] Wang, Z., Lincoln, S., Nguyen, A. D., Li, W., and Young, M. W. Chronic sleep loss disrupts rhythmic gene expression in drosophila. *Frontiers in Physiology 13* (2022), 2396.

[62] Xia, M., Li, X., and Chou, T. An age-structured Lotka-Volterra model and the emergence of overcompensation. *Bulletin of the American Physical Society* (2023).

[63] Xie, H., Guo, L., Wang, Z., Peng, S., Ma, Q., Yang, Z., Shang, Z., and Niu, Y. Assessing the potential prognostic and immunological role of TK1 in prostate cancer. *Frontiers in Genetics 13* (2022).

[64] Xu, Z., He, X., Shi, X., Xia, Y., Liu, X., Wu, H., Li, P., Zhang, H., Yin, W., Du, X., et al. Analysis of differentially expressed genes among human hair follicle–derived iPSCs, induced hepatocyte-like cells, and primary hepatocytes. *Stem Cell Research & Therapy 9*, 1 (2018), 1–15.

[65] Ye, F. X.-F., Wang, Y., and Qian, H. Stochastic dynamics: Markov chains and random transformations. *Discrete & Continuous Dynamical Systems-B 21*, 7 (2016), 2337.

[66] Yu, Z., Sun, Y., She, X., Wang, Z., Chen, S., Deng, Z., Zhang, Y., Liu, Q., Liu, Q., Zhao, C., et al. SIX3, a tumor suppressor, inhibits astrocytoma tumorigenesis by transcriptional repression of AURKA/B. *Journal of Hematology & Oncology 10* (2017), 1–16.

[67] Zhou, D., Wang, Y., and Wu, B. A multi-phenotypic cancer model with cell plasticity. *Journal of Theoretical Biology 357* (2014), 35–45.