NESCC: A Novel Approach to Incorporate Cancer "Virality" in Cancer Phylogenetics

1. Background

Current cancer phylogenetic techniques employ birth-death models to infer speciation and extinction of lineages as well as traits overlaid on a phylogenetic tree. Such models need to simplify the complex biology of the microenvironment out of necessity, however, emerging modeling techniques give us the opportunity to incorporate some of the complexities of these systems into evolutionary modeling. One notable nuance of this system are cancer exosomes, vesicles composed of miRNA and proteins, which can change the phenotypic presentation of neighboring cells.

Through this mechanism, cancerous cells can alter the transcriptome of the noncancerous cells surrounding the cancerous cell therefore changing the noncancerous cell's phenotypic presentation, resulting in cancerous behavior [1]. This process is facilitated by Dicer among other proteins needed for miRNA in cancer exosomes to silence genes within normal neighboring cells [2]. When these cells were injected into mice they resulted in tumors, suggesting the normal cells had become cancerous through exposure to cancer cell exosomes. The typical process of cancer cell development starts with a normal cell with a mutation that results in rapid proliferation which shortens telomeres, a sequence of buffering DNA used to prevent coding DNA loss during DNA replication, to a critical state [3]. This critical state results in senescence, a stage in which the cell does not undergo birth or death processes [4]. In some instances, these cells can escape into crisis, a cell state characterized by rapid proliferation and rapid cell death that entails chromosome end fusion which results in chromosome instability [5]. Throughout this process, cells can undergo mutations that can result in a cancerous cell type [3].

Using conventional cancer phylogenetic modeling does not account for this "virality" of cancer and therefore can underestimate the exponential occurrence of cancerous phenotypes within a microenvironment over time. Such modeling can also underestimate the birth rates within cancerous cells in their microenvironment.

2. Aims

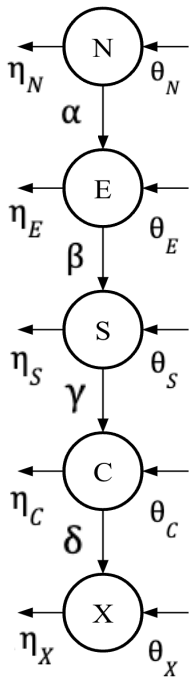
In this study, we aim to incorporate "infection" of normal cells and the trajectory they undergo to more accurately represent the characteristics of cancer evolution within a microenvironment as well as document the ways in which this novel modeling technique impacts patterns we see in cancer phylogenies relative to conventional birth-death models.

3. Model

We are proposing a modified SIR model coupled with a birth-death model to account for the duality of cancer evolution within a singular site. Instead of having susceptible, infected, and recovered stages we will instead have normal, exposed, senescent, crisis, and cancerous or NESCC. It should be noted that we are making the assumption that the "infected" normal cell undergoes a similar trajectory to a normal cell becoming a cancerous cell. In a conventional SIR model, it is assumed the population is stagnant which does not reflect the cancer microenvironment and as such are adding birth and death events within the

differential equations for each population and including the BDS model within our likelihood function [6]. The transition between the exposed to senescent phenotypes is directly dependent on the birth rate as per the biological mechanism behind the transition and as such was incorporated in the transition expression between these two cell states.

The markov chain is as follows:



$$\frac{dN(t)}{dt} = \frac{-\alpha N(t)X(t)}{n} + N(t) * (\theta_N - \eta_N) \quad \text{(Equation 1)}$$

$$\frac{dE(t)}{dt} = \frac{\alpha N(t)X(t)}{n} - \beta^{\theta_E} E(t) + E(t) * (\theta_E - \eta_E) \quad \text{(Equation 2)}$$

$$\frac{dS(t)}{dt} = \beta^{\theta_E} E(t) - \gamma S(t) + S(t) * (\theta_S - \eta_S) \quad \text{(Equation 3)}$$

$$\frac{dC(t)}{dt} = \gamma S(t) - \delta C(t) + C(t) * (\theta_C - \eta_C) \quad \text{(Equation 4)}$$

$$\frac{dX(t)}{dt} = \delta C(t) + X(t) * (\theta_X - \eta_X) \quad \text{(Equation 5)}$$

Figure 1. The following figure depicts the markov chain in the model with N for noncancerous cells, E for exposed cells, S for senescent cells, C for crisis cells, and X for cancerous cells. θ and η represent the birth and death rates respectively and are used in the differential equations to derive the true birth rate which is used to determine the change in each cell population. In Equations 2 and 3 the transition rate β is taken to the power of θ_E as the transition from exposed to senescent cells is dependent on the number of birth events undergone by the cell. In Equations 1 and 2 the expression $\frac{X(t)}{n}$, with n referring to the number of total cells in the system, can be assumed to be 1 as we make the assumption that a significant number of the cells in the microenvironment are cancer cells.

3.A. Hidden States

Most single-cell data is genome data which presents an issue as some of the phenotypes within our markov chain can only be distinguished using transcriptomic data [2][7][8].

To mitigate this issue, we propose two discrete observed traits, normal and cancer, with hidden states with normal phenotype having the hidden states normal and exposed and the cancer phenotype having the hidden states senescent, crisis, and cancerous [9]. Each of the hidden states will have their own respective birth and death rates which we will derive from the phylogenetic analysis using strong priors based on the known characteristics of the birth and death rates for each of these phenotypes.

4. Comparing to Conventional Birth-Death Approach

We hope to compare the results of our model to birth-death models on simulated data as well as single-cell sequence data composed of 44 cancer cells and 11 normal cells from adjacent tissue [10]. The results we aim to obtain from this comparison include differences in the lengths of cancerous lineages (which will include senescent and crisis cells as current birth-death models do not account for the heterogeneity within the genotypically presenting cancer cells), birth and death rates of different cell types within the NESCC model versus the birth-death model, etc.