esiCancer: Evolutionary in Silico Cancer Simulator

Darlan Conterno Minussi¹,²; Bernardo Henz³; Mariana dos Santos Oliveira¹,²; Eduardo C. Filippi-Chiela³; Manuel M. Oliveira²; Guido Lenz¹,⁴
¹Departamento de Biofísica, ²Instituto de Informática, ³Departamento de Ciências Morfológicas, ⁴Centro de Biotecnologia, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, RS, Brazil
* Corresponding author: E-mail: lenz@ufrgs.br
¶ These authors contributed equally to this work

Abstract

The evolution of cancer is inferred mainly from samples taken at discrete points that represent glimpses of the complete process. In this study, we present esiCancer as a cancer-evolution simulator. It uses a branching process, randomly applying events to a diploid oncogenome, altering probabilities of proliferation and death of the affected cells. Multiple events that occur over hundreds of generations lead to a gradual change in cell fitness and the establishment of a fast-growing population. esiCancer provides a platform to study the impact of several factors on tumor evolution, including dominance, fitness, event rate, and interactions among genes as well as factors affecting the tumor microenvironment. The output of esiCancer can be used to reconstruct clonal composition and Kaplan-Meier-like survival curves of multiple evolutionary stories. esiCancer is an open-source, standalone software to model evolutionary aspects of cancer biology.

Introduction

Despite immense advances in the study of the molecular biology of cancer (1-3), it remains dependent on biopsies, restricted to specific time points and to a fraction of the whole tumor. Thus, they fail to capture a complete picture of cancer heterogeneity offering only a snapshot of tumor evolution. The evolutionary story of a normal cell to a heterogeneous population of billions of cells is complex and, therefore, requires new theoretical insights to better understand the process.

Several models have been developed to study cancer in silico (4), each focusing on a specific characteristic of cancer biology. These include the hallmarks of cancer (5), the rate of clonal expansion (6), stem-cell driven tumor initiation (7), the effect of cell migration on tumor growth (8), and the impact of the microenvironment on tumor evolution (9). With esiCancer, we provide a fully-customizable tool designed to help stitch together genetic events during cancer clonal evolution. esiCancer follows a stochastic branching model. Its simulations generate evolutionary paths with events that modify the fitness of cancer cells leading to the selection of the fittest cells.

Methodology

esiCancer simulates a population of esiCells, each containing a diploid representation of its genome as two independent lists, a probability of death, a probability of division, and a maximum number of divisions (Fig. 1a). This genome can be hit by genetic events, representing point mutations, translocations, indels, etc., with a defined probability and dominance. These events alter the fitness and other aspects of the affected esiCell (Video 1).

esiCancer applies events to a predefined number of esiCells, each one independently subjected to four possible outcomes: no alteration; death; senescence; or cell division (Fig. 1a). If an esiCell divides, the two daughter cells receive, at random sites, a number of genetic events, defined by the user. Each event is associated with a change in the probability of division, death, mutation, and/or maximum divisions, thus impacting the population of esiCells over time. For all stochastic decisions, esiCancer uses a pseudorandom number generator initialized with a seed value. Different seeds create different evolutionary stories, which can be automatically iterated over multiple seeds to grant
high throughput simulations. A given seed will re-create the same sequence of events thus guaranteeing reproducibility (Video 2). esiCancer exports data about the cell lineages, the sequence, and frequency of events that gave rise to specific groups of esiCells, providing a complete analysis of the clonal composition of an esiTumor (Fig 1a, Video 3).

Pre-compiled Linux, Windows and MacOS GUI-based versions of esiCancer, as well as examples of esiTables, outputs, and video tutorials outlining how to use the system and analyze its output data are available at http://www.ufrgs.br/labsinal/esiCancer/. There one can also find detailed documentation about esiCancer, which includes pipelines to assist users in selecting the oncogenome and the parameters for their simulations. A guide for the production of the figures presented in this report is also provided. Source code and additional information is found at https://github.com/bernardohenz/esiCancer. esiCancer is under GNU Public License v3.0.

Randomness of the Population Fitness

Fitness in evolutionary biology is defined by the number of individuals in the n<sup>th</sup> generation (GEN<sub>n</sub>) divided by the number of individuals in the previous generation (GEN<sub>n-1</sub>). In esiCancer, fitness is directly defined by the probability of division minus the probability of death (Fig1b). If the probability of division and death are both set to 0.01, fitness value calculated with the input data (equation 1) is similar to the value calculated with the output values (equation 2), and this continues to be true after alteration in fitness produced by events. An event that affects the probability of division increases the average fitness, which is further increased by a second event. If an event increases the probability of division and decreases the probability of death, the impact of this event on the fitness reflects the impact on both division and death (Fig. 1b). As expected with exponential growth, this produces a final number of esiCells that is about 8 times higher when compared to the impact of only increasing the probability of division.

Fig. 1 Overview of esiCancer: (a) Modeling starts with normal esiCells pre-defined probabilities of division and death, and a number of maximum divisions; if an esiCell divides, it receives a number of events and if an event hits a gene, characteristics defined in the esiTable are changed. Number of esiCells, their mutations and fitness are recorded (b) Stochasticity of fitness of the esiCell population. Fitness in esiCancer is defined (1) by the number of esiCells in generation n divided by the number of cells at generation n-1 or (2) by the difference between the probability of cell division and death. Fitness using formula (1) (light grey and red line) and the number of esiCells (blue) are shown for a simulation with the conditions indicated in the boxes. Average fitness as calculated by (1) in red or (2) in green. GEN – generation. (c) Population of esiCells with 15 or 30 maximum divisions for 100 different seeds. (d) Kaplan-Meier-like plots for runs with different event rates. For these runs, 10k esiCells from an initial population of 1,000 esiCells were considered an esiTumor.
Escape from replicative senescence is another important hallmark of cancer. esiCancer allows the user to limit the number of divisions, resulting in a gradual reduction in the population since cells retain their probability of death (Fig. 1c). Events that lead to an increase in the maximum number of divisions model an escape from replicative senescence. esiCancer can also be used to generate Kaplan-Meier-like plots by plotting the number of generations required to achieve a defined threshold. Increasing the number of events per division also increases the number of simulations that reach the threshold while reducing the number of generations required to reach such condition (Fig 1d).

Survival of the fittest

In esiCancer, different simulations produce unique frequencies in gene events, but the frequency after 1100 generation of a given event on average directly correlates with its dominance (Fig 2a, i), probability (iii), and impact on the fitness (ii) as predicted by evolutionary biology. Highly dominant events will appear more frequently than events with low dominance, as the impact of a mutation on the first allele of a highly dominant gene is much stronger than on genes with low dominance values (i). Gene frequency also directly correlates with fitness (ii) and the probability of the event (iii). Therefore, these parameters will affect the probability of an event occurring and will alter the number of descendants that contain the event. A given gene can have two events, which interact allelically and, if all other conditions are the same, their frequency is higher than a gene with a single event (iv)

An event can also impact several genes, resembling copy number variation (CNV). An event affecting gene A and B, but not C, will have a frequency equal to gene A, if gene A does not receive any additional event by itself. Frequency of gene B will be the sum of the frequencies due to event AB and an additional event on gene B. Event C will not be affected by event AB (Fig. 2a, right). Lastly, the relative frequency of events at different time points indicates that the same conditions, when modeled with different seeds, can produce variable population dynamics recapitulating different models of tumorigenesis (Fig. 2b).

Gene and cell interactions in esiCancer

Cancer genes act within complex interaction networks during tumor development. A given event can affect the impact of another event, either by decreasing its impact leading to mutual exclusivity, or increasing its impact, resulting in co-occurrence (10). In a simulation containing 3 genes with equal settings and no interactions, a similar frequency of event 1 in gene 2 and 3 occurs (Fig 2c, grey). If the impact of gene 1 and 2 are mutually exclusive, gene 2 will appear less frequently altered when compared to non-interacting genes and the contrary occurs in the case of co-occurrence (Fig 2b, red and green). esiCancer also permits the modeling of interactions among cells, in which events can have impacts on the whole tumor, resulting in alterations that impact the microenvironment positively or negatively (Fig 2d).

![Fig. 2](image-url)
Conclusion
esiCancer provides a platform for simulating the genetics of tumor evolution. It was designed from the ground up to model important aspects of evolutionary biology applied to cancer using real genetic data. The unique strategy of modeling individual cells and applying single-cell decisions of division, senescence, or death reproduces key aspects of tumorigenesis. This results in the survival of the fittest, where each simulation yields a unique outcome, thereby resembling the rise of cancer in humans and capable of modeling the response to mutagens or genetic alterations. In this way, esiCancer can become an important tool to better understand the hidden aspects of tumor evolution.

Acknowledgment: This work was supported by FAPERGS/PRONEX (16-2551). All authors are or were recipients of fellowships from CNPq. We wish to thank Dr. Francisco M. Salzano (in memoriam) and Francisco Ivanio for critical reading of the manuscript and Maria Julia Oliveira for video and sound editing.

Author contributions: DCM, BH, ECFC, MMO and GL conceptualized all aspects of esiCancer. GL wrote the manuscript, with contributions from all authors. BH wrote the code under the supervision of MMO. MSO performed the simulations and produced all figures. The authors declare no potential conflicts of interest.

REFERENCES