DOI:10.34918/87328

Jarosław GIL[1,*], Andrzej POLAŃSKI[1]

Chapter 2. MODELLING CLONAL EVOLUTION OF TUMOURS BY USING HIGH RESOLUTION STOCHASTIC SIMULATIONS

2.1. Introduction

Tumor growth, caused by accumulation of genomic somatic mutations, which disturb and eventually dysregulate cellular processes of signaling, metabolism and replication, is a complex process with various stages and modes of dynamics. Fundamental research on scenarios and mechanisms of tumor evolution, combined with observational data and biological/genetic knowledge, is of great importance for cancer research. There are many studies both theoretical and experimental devoted to characterization/observation of the process of progression of tumors. The latter ones, especially those using high throughput molecular biology assays provide quantitative background for studies on scenarios of cancer development [1-3].

Important insights to the study of dynamics of cancer evolution are given by mathematical/computational modelling tools. Mathematical models require introducing a lot of simplifying hypotheses in order to obtain interpretable conclusions. Computational stochastic simulation approaches allow for address wider range of evolution scenarios, important for explaining observational data, which are recently becoming available [3, 4]. There are several attempts to provide mathematical models able to reflect at least some properties of evolving tumor [5-7].

We present a stochastic simulation tool, based on Gillespie algorithm [8] for modelling evolution of cancer cells population. Replication process of cancer cells is affected by two types of somatic mutations occur, driver with high impact on tumor growth and passenger, assumed to have mildly deleterious effect. The elaborated algorithm allows for high resolution modelling of evolution of cancer, with

[1] Silesian University of Technology, Poland.
[*] Corresponding author: jaroslaw.gil@polsl.pl.

distinguishing between cancer cells and both types of somatic mutations occurring over time. It allows for confronting simulation results with experimental data concerning clonal structures of tumor cells and statistics of somatic mutations.

2.2. Scenario of evolution with driver and passenger mutations

Tumor cells evolution scenario, on which algorithm is based, consist of three possible events: cell death, cell division or cell division with mutation. Mutation types are considered as passenger (slightly deleterious effect) or driver (highly advantageous effect). When driver mutation occurs, its high effect on cell fitness is likely to create a new clone, which also have positive impact on cell division probability. Each cell in one clone have same driver mutation set. Passenger mutations provide differentiation between cells. Each passenger mutation causes mildly or no effect on fitness cell. When mutation is considered to have impact on cell fitness its accumulation in one cell increases death probability. Figure 1 describes all evolutionary events considered in algorithm construction.

Intensity of the cell division process depends on cell fitness and population size. When cell fitness is high it is more likely that cell will divide or/and mutate, on the contrary when cell fitness is low the probability of cell death is higher. Also, population size has impact on event probability. When population size is large more events occur and worse adapted cells are dying. In the course of evolution, the phenomenon of clonal interference takes place. Clonal interference is extinction of already existing clones, which are replaced by new ones. When a new clone with mean fitness of its cells emerges in the evolution it is likely that it displaces existing one ore one. The process of clonal interference is nonlinear, depends on fitness of neoplastic cells in clones as well as on clones' sizes.

The size of the whole population of cancer cells evolves according to the dynamics shaped by the mean fitness of all cancer cells. Depending on accumulated effects of deleterious and advantageous mutations acquired during evolution the cancer cells population can either increase in size, and eventually explode or decrease and extinct. Probabilities of these two scenarios depend on intensities of occurrences of driver and passenger mutations and on strengths (relative strengths) of their positive/negative selection effects.

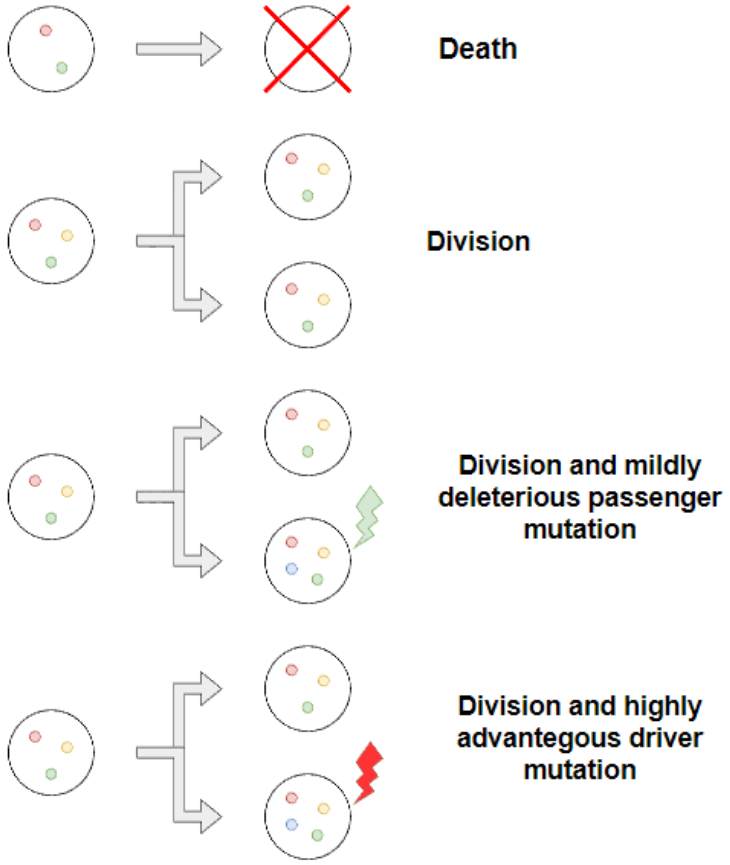

Fig. 1. Stochastic events in the evolution scenario of clonal growth of cancer
Rys. 1. Zjawiska występujące w trakcie rozwoju poopulacji komórek nowotworowych

2.2.1. Division process

Cell division probability/intensity depends on its fitness. Higher fitness provides higher chance for cell to divide. Division time can be calculated using formula (1)

$$t_{div} = random.exponential(1) \,/\, fitness \tag{1}$$

where division time of cell without mutations is calculated from exponential distribution. The "random.exponential(1)" notation in the above equation denotes a number generated by pseudo random generator function of exponential distribution with scale parameter equal to one. Scaling by fitness produces the variable on the left-hand side of the above equation with probability distribution adequately characterizing the process of division of cells of given fitness. Division time is inversely proportional to fitness value. In each simulation step event with event time as low as possible are considered.

2.2.2. Death process

Cell death is mostly dependent on population size. If number of cells is greater than population capacity it is more likely for cell to die. Death time can be calculated using formula (2)

$$t_{dth} = random.exponential(1) * \frac{N}{K} \tag{2}$$

where N denotes population size and K defines population capacity, "random.exponential(1)" has the same meaning as previously. If cells number exceeds population capacity it is more likely for cell to die than to divide.

Cell death event is considered to have bigger priority than division or division and mutation.

2.2.3. Mutation processes and its offect on fittness

Mutation can be divided into two types: passenger and driver mutation type. Passenger mutation provides cells differentiation while driver mutation provides creation of new clone. Mostly considered are events where driver mutation have mildly adventegous effect on cell fitness and passenger mutation have slightly negative effect. Cell fitness is calculated using formula (3)

$$fitness = \frac{(1+s)^l}{(1+f)^k} \tag{3}$$

where s stands for driver mutation, advantegous effect, f stand for passenger mutation deleterious effect and l and k are numbers of, respectively, driver and passenger mutations in cell.

2.2.4. Parameters of processes

In Table 1 examplary simulation parameters are presented. Driver mutation effect is considered to be quite high (about 2–5%) while passenger mutation efect should be lower (about 0.1–0.5%). Passenger mutation effect also should be considered as negative effect while driver as positive. The probabilty of mutation occurance should be respectively low for driver type and high for passenger type of mutation. As population capacity we assume the initial population size.

Table 1

Example simulation parameters
Model parameters

Mutation effect	High for driver mutation: f.e. 2%, low for passenger mutation: f.e. 0.2%
Initial population	10’000
Population capacity	10’000
Mutation probability	Low for driver mutation: f.e. 1%, high for passenger mutation: f.e. 50%
Simulation where provided till population gained about 1’000’000 cells	

Driver and passenger mutation effects and probabilities are assumed for scenario when only one kind of both types occurs. It is possible to implement more than one effect for both types to differentiate mutation kinds.

2.3. Gillespie simulation algorithm

Basing on described scenario we have elaborated appropriate version of Gillespie algorithm. We introduce modification which provides a possibility to track every mutation in population in reasonable time stamp. Modification is based on division of the whole population into clones what enables parallel computing. Each clone is composed of clone parameters: size, previous clone, driver mutation list and passenger mutation matrix. Each cell in one clone have same set of driver mutation so it is not needed to include them into cell mutation profile. The main element in clone is mutation matrix which enables tracking each mutation in every cell. Matrix rows represent clone cells and columns represents mutations. First column describes each cell fitness which is computed directly after mutation event but only for new cell. Each other column represents events of mutations existing in cell. Including mutation matrix provide possibility to copy only cells which divide or mutate without interacting with whole matrix in every simulation loop.

Each clone is updated simultaneously one cycle after another. Firstly, based on cells fitness, random variables representing time of events in the evolution, are generated according to exponential distribution. Mutation events occur only when division event happen. Mutation probabilities differ between drivers and passengers. It is possible to

implement few driver and passenger mutation types to differentiate mutation probability and effects. For each cell in clone event time is compared to simulation step – tau step. Every cell which death time is smaller is simply deleted from clone. When division time is smaller mutation flag is checked. If cell mutate type of mutation have to be considered – when driver mutation occurs new cell will create new clone, when passenger mutation occurs simply new mutation is added to mutation matrix. If cell only divide copy of that cell with no new mutation is added to mutation matrix. Block diagram of that Gillespie algorithm modification is presented on Figure 2.

After every clone update simulation time is extended by tau step and new loop is considered. Simulation end when ending condition is completed – maximum size of population is reached or simulation time exceeds assumed time.

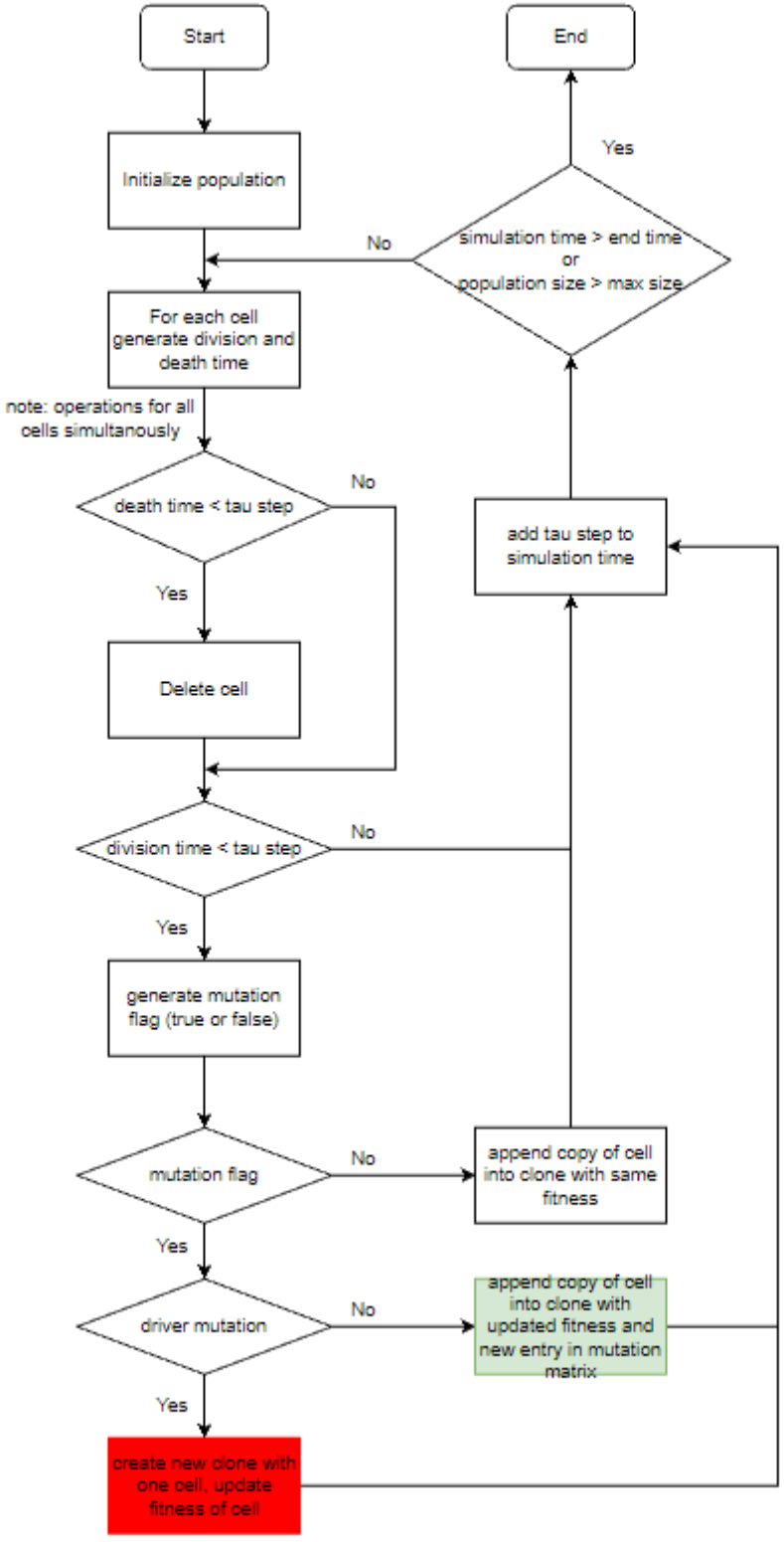

Fig. 2. Block diagram of the Gillespie simulation algorithm using parallel computing and mutation matrix

Rys. 2. Diagram blokowy algorytmu Gillespiego z wykorzystaniem obliczeń równoległych i struktury macierzy mutacji

2.4. Implementation of the mutation matrix

Table 2

Fragment of mutation matrix structure

Fitness	Pass. Mut. ID	Pass. Mut. ID	Pass. Mut. ID	Pass. Mut. ID	Pass. Mut. ID
	0	1	0	0	0
	0	0	0	1	1
	0	0	1	1	0
	1	0	0	0	0

Mutation matrix represents, in the form of binominal values, mutations presence or absence in cells. One mutation can be assigned to multiple cells and one cell can be assigned to multiple mutation. If cell contains mutation, value 1 is present in appropriate column. Because of matrix structure it can have big dimensions. To reduce memory usage that structure can be saved as sparse matrix (compressed form like map but with every ability of matrix). In python environment we are using scipy library with sparse matrix structures.

First column describes the cumulative effect of all mutation. Each driver mutation (same for every cell in clone, not present in mutation matrix) have positive effect on cell fitness. The highest fitness in clone always have cell from which clone starts progression. Each passenger mutation (all mutations in matrix) have negative effect on cell what is calculated only when parent cell mutates.

Mutation matrix application provides possibility to easily handle evolutionary events. When cell dies simply row is extracted from matrix and deleted. When cell divides, hard copy of row is appended to matrix. If cell also mutates (passenger mutation type) besides row copy, one more column is appended with value 1 in last row.

2.5. Results

High resolution stochastic simulation, implementing the idea of mutation matrix, enables tracking every mutation both separately in clones and in the whole population. Our method is able to provide information about VAF of every mutation. Mutation distribution can be viewed as VAF histogram for whole population with or without division into clones, or for one clone.

2.5.1. Clone diagrams of the evolution of cancer cells population

Basic information about clone – size and parent clone, can be used for plotting clone diagram. In Figure 3 clone size distribution along population progression has been shown. Different clones are marked by using different colors. Clone diagram provides information about dominant clone in each generation, when occurs driver mutation which have huge impact on population and how fast population is rising. The simulation end condition was about 10 times enlargement of initial population.

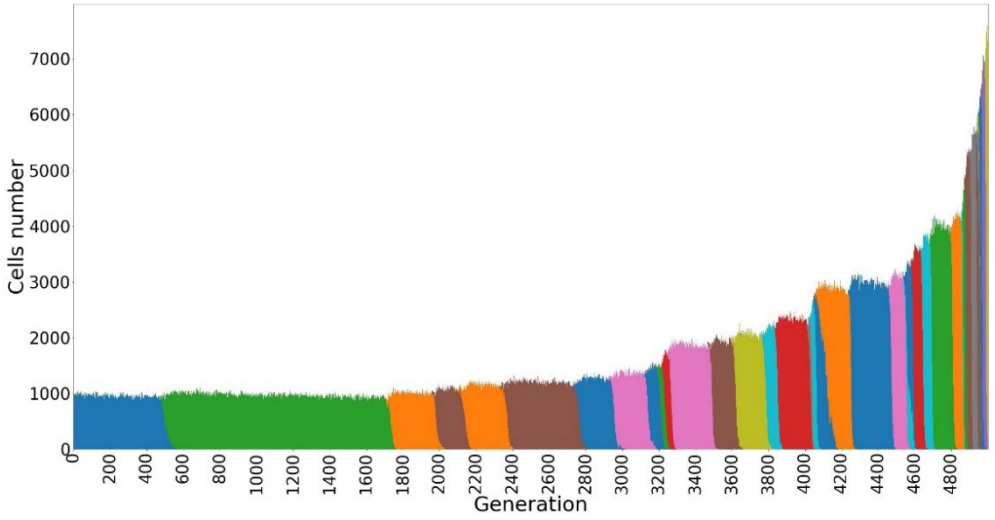

Fig. 3. Clone diagram of the evolution of cancer cells population
Rys. 3. Wykres zmian struktury klonów w trakcie ewolucji populacji

2.5.2. Evolution of variant allele frequencies (VAFs)

Due to emergence and extinction of clones, VAFs change dynamically over time. Frequencies are calculated basing on maximum cells number of interpreted groups. VAF histograms provide appropriate information to track cells differentiation and clone evolution.

2.5.3. Evolution of variant allele frequencies inside clones

In Figure 4 below we show two examples of VAF frequecies computed, separately in clones, in the course of evolution of cancer cell population. It is seen that VAFs inside

both clones are bimodal. Driver mutation, whose occurrence lead to new clones are present in all cell of the clone. All cells of the clone also contain passenger muations from the ancestral cell of the clone. In the process of evolution of the clone new passenger mutations are also generated. These passenger muations appear only in small subsets of clonal cells and therefore they correspond to low values of VAFs, concentrated on the left hand side of the chart. The above described mechanism leads to bimodality of VAF distribution shown in both charts in Figure 4.

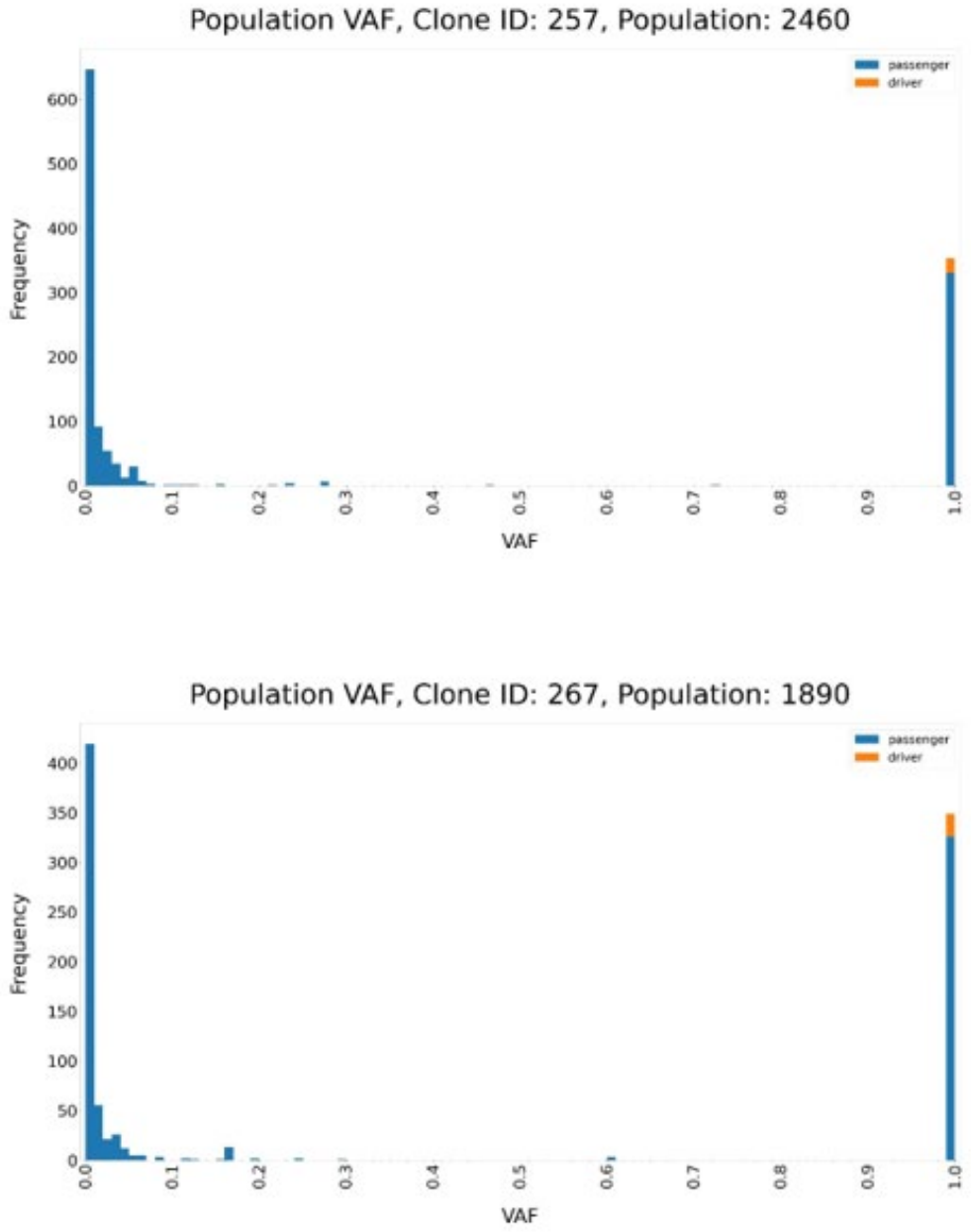

Fig. 4. VAF histograms for single clone in one generation of whole population. On chart mutation types are colored in different colors. Mutation frequency is calculated basing on only one clone cells

Rys. 4. Histogram VAF dla jednej generacji pojedynczego klonu. Na wykresie wyróżnione zostały rodzaje mutacji poprzez zastosowanie różnych kolorów. Częstotliwość mutacji obliczana jest na podstawie komórek pojedynczego klonu

2.5.4. Evolution of variant allele frequencies for the whole population

Simulated data enable also showing mutation frequencies in the whole population, as presented in Figure 4. Here the whole population of cancer cells contains two clones. Combination of all mutations in one plot results in modification of frequencies of mutations. VAFs of mutations in the whole population depend on sizes of clones and on mutations frequencies inside clones. This is seen in the plot in Figure 5. Two colors are used for showing mutations coming from two different clones. Combining VAFs of mutations of all clones in one plot results now in multimodal distribution.

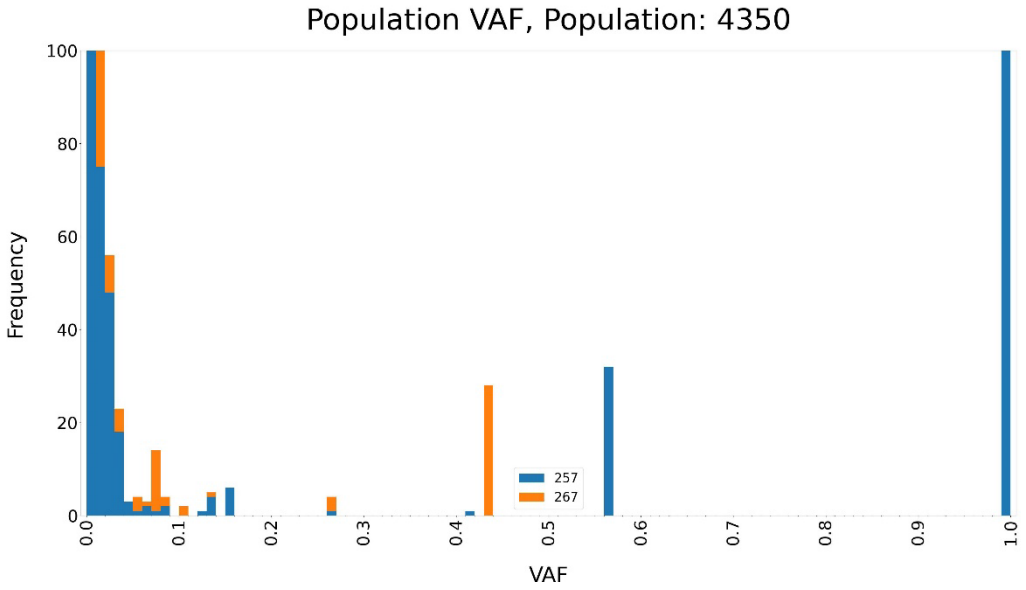

Fig. 5. VAF histogram for one generation of whole population. Each clone is presented as different color. VAF is calculated with respect to mutation frequency in all cells

Rys. 5. Histogram VAF dla jednej generacji całej populacji. Każdy klon został zaznaczony za pomocą odrębnego koloru. VAF obliczany jest na podstawie częstotliwości mutacji w odniesieniu do całej populacji

2.6. Summary and conclusion

Clonal composition of tumor is a very important determinant of its potential for malignancy transition to methstasis and predicted response to therapies. The research devoted to studying clonal structures of tumors and their relations to biological background of cancer development and clinical observations is

a wide area in computational oncology [9–13]. This scientific research area is also supported by mathematical and computational models of cancer clonal development.

We presented a high – resolution simulation system for tracing clonal evolution of tumor with interacting driver and passenger mutations. Evolution of clones when cancer cells population progresses and increases in time leads to dynamic changes of frequencies of mutations. During the evolution of separate clones distributions of VAFs (allelic frequencies) of passenger muations are biomodal, as shown in Figure 4. When distributions of VAFs are drawn for the whole cancer cells populations, as shown in Figure 5, they display multimodal shapes. These results are consistent with many observations of VAFs obtained in experimental studies [4].

Acknowledgements

This publication was supported by the Department of Graphics, Computer Vision and Digital Systems, under research project for young scientists and statue project (RAu6, 2022), Silesian University of Technology (Gliwice, Poland).

Bibliography

1. Nowell P. (1976). The clonal evolution of tumor cell populations, Science 194(4260):23–28.
2. Greaves M., Maley C.C. “Clonal evolution in cancer”. Nature 481.7381 (2012): 306–313.
3. Kumar S. et al. 2020, Passenger Mutations in More Than 2,500 Cancer Genomes: Overall Molecular Functional Impact and Consequences, Cell, Volume 180, Issue 5, 5 March 2020, Pages 915–927.e16 https://doi.org/10.1016/j.cell. 2020.01.032.
4. Campbell P.J., Getz G., Korbel J.O., Stuart J.M., Stein L.D. The ICGC/ TCGA Pan-Cancer Analysis of Whole Genomes Consortium (2020). Pan-cancer analysis of whole genomes. Nature 578, 82–93.
5. Bozic I. et al. “Accumulation of driver and passenger mutations during tumor progression”. Proceedings of the National Academy of Sciences 107.43 (2010): 18545–18550.
6. McFarland C.D. et al. “Impact of deleterious passenger mutations on cancer progression”. Proceedings of the National Academy of Sciences 110.8 (2013): 2910–2915.

7. McFarland C.D., Mirny L.A., Korolev K.S. “Tug-of-war between driver and passenger mutations in cancer and other adaptive processes”. Proceedings of the National Academy of Sciences 111.42 (2014): 15138–15143.
8. Gillespie D.T. (2007). Stochastic simulation of chemical kinetics. Annu. Rev. Phys. Chem., 58, 35–55.
9. Martincorena I., Campbell P.J. (2015). Somatic mutations in cancer and normal cells, Science 25 Sep 2015: Vol. 349, Issue 6255, pp. 1483–1489, DOI: 10.1126/science.aab4082.
10. Strino F. et al. (2013). TrAp: a tree approach for fingerprinting subclonal tumor composition. Nucleic Acids Res. 41(17):e165.
11. Andor N. et al. (2014). EXPANDS: expanding ploidy and allele frequency on nested subpopulations, Bioinformatics, 30(1):50–60.
12. Roth A. et al. (2014). PyClone: statistical inference of clonal population structure in cancer. Nat Methods 11(4):396–398.
13. Miller C.A. et al. (2014). SciClone: inferring clonal architecture and tracking the spatial and temporal patterns of tumor evolution. PLoS Comput Biol 10(8):e1003665.

MODELLING CLONAL EVOLUTION OF TUMOURS BY USING HIGH RESOLUTION STOCHASTIC SIMULATIONS

Abstract

Tumor growth, caused by accumulation of genomic somatic mutations, which disturb and eventually dysregulate cellular processes of signaling, metabolism and replication, is a complex process with various stages and modes of dynamics. Fundamental research on scenarios and mechanisms of tumor evolution, combined with observational data and biological/genetic knowledge, is of great importance for cancer research.

Important insights to the study of dynamics of cancer evolution are given by mathematical/computational modelling tools. Mathematical models require introducing a lot of simplifying hypotheses in order to obtain interpretable conclusions. Computational stochastic simulation approaches allow for address wider range of evolution scenarios, important for explaining observational data, which are recently becoming available.

We present a stochastic simulation tool, based on Gillespie algorithm for modelling evolution of cancer cells population. Replication process of cancer cells is affected by two types of somatic mutations, driver with high impact on tumor growth and passenger, assumed to have mildly deleterious effect. The elaborated algorithm allows for high resolution modelling of evolution of cancer, with distinguishing between cancer cells and both types of somatic mutations occurring over time. It allows for confronting simulation results with experimental data concerning clonal structures of tumor cells and statistics of somatic mutations.

Keywords: Stochastic simulation, Clonal evolution, VAF histograms, Gillespie algorithm