



UPPSALA
UNIVERSITET

MSc BIOINF 15 004

Mathematical modeling of normal and cancer prostate signaling pathways

Sofia Stamouli

Degree project in bioinformatics, 2015

Examensarbete i bioinformatik 45 hp till masterexamen, 2015

Biology Education Centre, Uppsala University and Department of Biochemistry and Biophysics,
Science for Life Laboratory

Supervisors: Mobashir Mohammad and Niklas Schultz

Abstract

The field of systems biology has become very popular as a means to deal with cancer as well as other complex biological issues. It enables scientists to gain an insight into difficult conditions through mathematical approaches that have been developed. Prostate cancer is the second leading cause of death among men after skin cancer and its heterogeneity makes it a complex disease. In this study we focus on three pathways known to play crucial roles in the formation of prostate cancer. By using a mathematical model that combines all of them we describe the interactions taking place during signal transduction in the prostate under normal and cancer conditions.

Normal and cancer prostate signaling

Popular science summary

Sofia Stamouli

Our body is made up of cells which are the structural and functional unit of life. Normal cells can grow, divide to form new cells, or possibly die. This process is the normal cycle of cells and especially in the early stage of our life, cells grow faster allowing the person to grow. Cells undergo changes due to the actions of the molecules involved in different biological pathways. They receive specific signals from the external environment and then trigger a series of biochemical reactions which in turn transmit the information, to the nucleus for instance. Based on the nature of output response (downstream signaling molecules), cells can proliferate or die. In cancer cells, the mechanism of cell cycle regulation is altered and the cells instead of dying continue to grow continuously.

Cancer is a complex biological system and there are large number of cellular subsystems and parameters involved which are difficult to study in the laboratory. Therefore, efforts have been made to investigate the difficult conditions and understand the communication among molecules in the biological pathways by using computational approaches. In our study, we focused on prostate cancer which is the second disease, after skin cancer, causing death to men. Studies have shown that there are three main signaling pathways known to play critical role in the prostate cancer formation: the androgen receptor signaling pathway, the Ras-Raf-MEK-ERK pathway (MAPK), and the PI3K-AKT-mTOR pathway.

By using a set of ordinary differential equations we simulated these three pathways by integrating all in one. Based on the nature of output response or downstream signaling molecules (transient/sustained), cell's fate is decided. Therefore, we have developed a mathematical model to understand the differences regarding cell fate decision both in normal prostate and cancer prostate pathways and we validated the model experimentally. According to this model, we will also be able to predict new possible interactions between the signaling molecules as we already know that identifying the molecules involved in prostate cancer can be a target for drug discovery in the future.

Degree project in bioinformatics, 2015 (Examensarbete i bioinformatik 45 hp till masterexamen, 2015)

Department of Chemical Biochemistry and Biophysics, Science for Life Laboratory

Supervisor: Mobashir Mohammad and Niklas Schultz

Contents

Abstract	1
Popular science summary	3
List of figures	6
Abbreviations	7
Introduction	9
Mathematical modeling of signaling pathways	10
Prostate signaling pathways	10
Goal of this project	11
Materials and methods	12
Mathematical modeling of AR, MAPK, and mTOR pathway	12
MAPK pathway	12
Evolutionary algorithm	12
Network topology and network motifs	15
Immunohistochemistry protocol (IHC-P)	17
Results	18
Linear MAPK cascade	18
MAPK Cascade with a feedback loop	20
Normal prostate signaling	22
Cancer prostate signaling	23
Network visualization	26
Network motifs	27
Experimental validation	28
Discussion	32
Conclusions	33
Acknowledgements	34
References	35

List of figures

Figure 1. Signal transduction process	9
Figure 2. AR, MAPK, and mTOR pathways without and with interactions among them.	11
Figure 3. Steps involved in evolutionary algorithm.....	14
Figure 4. Motifs for feedback and feedforward loop	15
Figure 5. Motifs common in cancer-associated networks.....	15
Figure 6. 3-node motifs.....	16
Figure 7. Response of kinetics in MAPK linear pathway during evolution.	19
Figure 8. Fitness of MAPK linear pathway during evolution.....	20
Figure 9. Response of kinetics in MAPK pathway with feedback loop during evolution.	21
Figure 10. Fitness of MAPK pathway with feedback loop.....	22
Figure 11. Response of the model in normal prostate signaling.....	23
Figure 12. Response of the model in cancer prostate signaling because of RasGAP loss.	24
Figure 13. Response of the model in cancer prostate signaling because of PTEN loss.	25
Figure 14. Visualization of the network by using Cytoscape.	27
Figure 15. Detection of 3-node motifs for normal prostate model.	28
Figure 16. Expression of phosphorylated Erk in normal and cancer biopsies.....	29
Figure 17. Comparison of the mean fluorescent intensity of Erk and phosphorylated Erk between normal and cancer biopsies	30
Figure 18. Levels of phosphorylated Erk in normal and cancer tissue of the same biopsy.....	31
Figure 19. Comparison of the fluorescent intensity for phosphorylated Erk between normal and cancer tissue in the same biopsy	31

Abbreviations

ODEs	Ordinary Differentially Equations
AR	Androgen receptor
mTOR	Mammalian target of rapamycin
MAPK	Mitogen-activated protein kinase
RTK	Receptor tyrosine kinase
PTEN	Phosphatase and tensin homologue deleted on chromosome 10
PI3K	Phosphoinositide 3 kinase
RasGAP	Ras GTPase-activating protein
DHT	Dihydrotestosterone
IHC-P	Immunohistochemistry-Paraffin
Erk(Erk1/2)	Extracellular signal-regulated kinase
Erkp	Single phosphorylated form of Erk
Erkpp	Double phosphorylated form of Erk
MEK	MAPK/Erk kinase
Akt	AKT8 virus oncogene cellular homolog
Aktp	Single phosphorylated Akt
Raf	Rapidly accelerated fibrosarcoma
NF1	Neurofibromatosis type 1
DNA	Deoxyribonucleic acid
a.u	arbitrary unit
SE	standard error

Introduction

Signal transduction is the process of cellular communication. Three steps are involved in this process: the cell receives an extracellular signal which triggers a specific receptor either on the surface of the cell or inside it (receptor level), the signal is then transferred from outside of cell to inside through intracellular communication with a set of biochemical reactions contributing to the process creating the so-called pathways (intracellular level), and finally, the information carried by the signal is transmitted to the nucleus (effector level) [1, 2] (Figure 1). Based on the nature of the response, the cell can undergo proliferation, apoptosis, or differentiation [3]. In case of complex diseases (such as cancer, diabetes, etc.), the response nature is altered due to change in interactions between the signaling molecules. For this reason, it is critical to understand the main factors which may control the response nature. Olaf Wolkenhauer [4] describes the mechanism behind the binding of the ligand to the receptor. It is stated that an important parameter for the biochemical reactions are the kinetics which define the rate at which a reaction is taking place. Considerable attention has been paid to kinetics since changes in them result in oncogenesis.

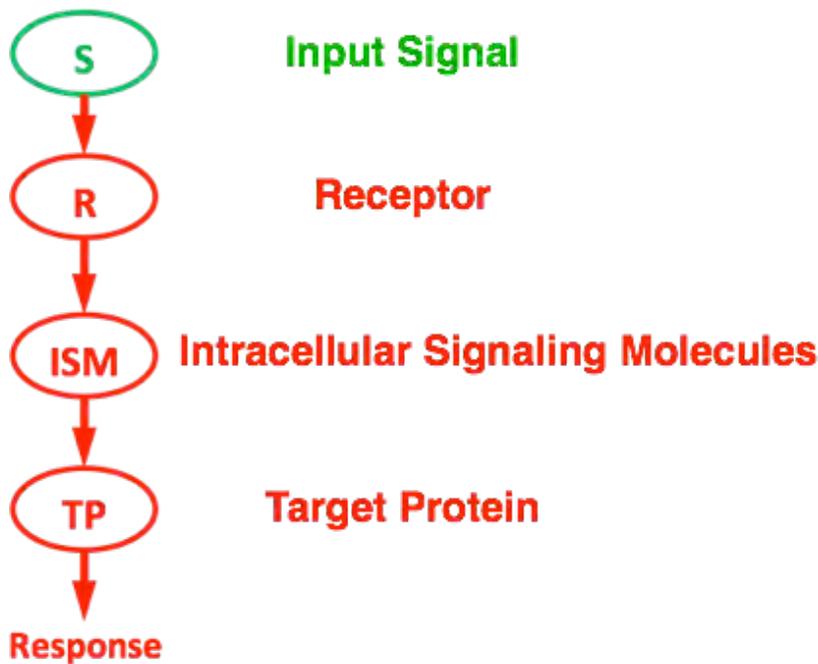


Figure 1. Signal transduction process. The receptor receives an external signal and through intracellular biochemical reactions it is transmitted to the nucleus. Based on the response nature of the target protein, the cell grows, divides or dies.

For several decades great effort has been devoted to the study of cancer. In normal cells, when DNA is damaged it is either repaired or the cell is lead to a programmed death (apoptosis). Conversely, in cancer the cells grow out of control and as reported by Douglas Hanahan and Robert A. Weinberg [5], all types of cancer have some common characteristics that control the change of

normal cells to cancer cells. Apart from their trait of growing limitless, they also invade other tissues leading to metastasis, an invasion that does not happen in normal cells. Regarding the nature of the response during cellular communication, the pathways associated with proliferation/apoptosis/growth/differentiation control are altered in such a way which favors uncontrolled proliferation in case of cancer [6]. Thus, the identification of differences between normal and cancer cells and the mutations involved has gained importance especially for the field of drug discovery.

Mathematical modeling of signaling pathways

For complex biological systems such as cancer, an approach that has been used to understand the processes of proliferation, apoptosis or differentiation in them is the establishment of mathematical models to describe signaling process and the interaction between the signaling molecules [7]. After choosing the biological pathway to be described, the biochemical reactions and their kinetics are described with the help of ordinary differential equations. (ODEs). As reported by Orton [8], while establishing the mathematical model it is often required to make some assumptions to simplify complex processes. After the assumptions being made, it is necessary to describe the kinetic parameters, the rate constants, and the initial concentration. The definition of kinetics follows either the mass-action [9] or the Michaelis Menten [10] approach. The benefit of using mathematical models is that it helps scientists to gain new insights into cases which are difficult to be addressed by experimental approaches. Pathways are unlikely to be linear. They rather have loops - feedforward and feedback- or cross-talks between them where molecules of a pathway are connected with other molecules from a different pathway. From previous studies, it is well known that signal transduction is one of the main trivial tasks which are related to many complex diseases such as cancer. Therefore, mathematical models can help to unravel possible complicated behaviors in a biological system.

Prostate signaling pathways

Prostate cancer is the most common cancer among men after skin cancer and the three biological pathways that are known to play major role are mitogen-activated protein kinase cascades (MAPK), androgen receptors (AR), and mammalian target of rapamycin (mTOR). Firstly, MAPK cascade plays a critical role in cell fate decisions [11] and therefore their function is significant in the process of transduction of extracellular signals to cellular responses. AR interacts with proteins strongly involved in the role of receptors [12] and the mammalian target of rapamycin (mTOR) is involved both in intracellular and extracellular communications having a role in cell cycle regulation [13]. Figure 2 represents the three pathways without any interactions among them (left side) and when cross-talks between AR, mTOR, and MAPK pathways are present (right side) [14].

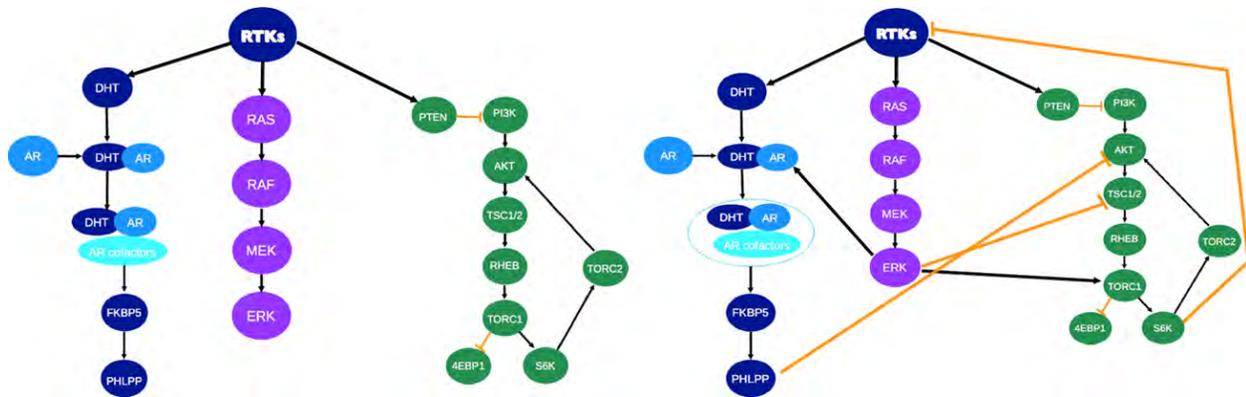


Figure 2. AR, MAPK, and mTOR pathways without and with interactions among them. On the left side, all the three independent pathways are initiated by the receptor tyrosine kinase (RTK). On the right side, the three pathways regulate each other by negative feedback describing cross-talk between the three pathways.

Rhonda L. Bitting and Andrew J. Armstrong [15] have presented how changes in signaling molecules in all of three pathways lead to prostate cancer and how each pathway affects the activity of the other two.

The literature has showed two proteins playing a crucial role in formation of prostate cancer; PTEN [16] which is considered a tumor suppressor in the PI3K pathway and the rasGTPase activating protein (rasGAP) [17]. The mutation of NF1 gene results in the loss of RasGAP which has a crucial role in the MAPK pathway.

In biological signaling, it has also been demonstrated that mTOR and MAPK are not linear pathways, yet they interact and regulate each other [18]. As far as the AR pathway is concerned, several authors [19, 20, 21, 22] have proposed the interplay of androgen receptor with the other two pathways. AR is of potential interest because of its crucial role in the development and the function of prostate.

Goal of this project

In this study, we mainly focus on how these three prostate signaling pathways control cell fate decision in normal prostate signaling and cancer prostate signaling which will help us in unravelling the possible feedforward, feedback loops (positive/negative), and cross-talks among them.

Materials and methods

Mathematical modeling of AR, MAPK, and mTOR pathway

MAPK pathway

Before combining the three main pathways in one, we first developed the MAPK pathway which has been described by several studies [11,23,24]. As already shown in Figure 2, the MAPK pathway involves the proteins Ras, Raf, MEK and Erk. The cell receives an external signal, binds to the receptor and it activates Ras converting it from GDP to GTP molecule [25]. In that state, Ras activates Raf which in turn phosphorylates and activates MEK. Finally, MEK phosphorylates Erk which is either translocated to the nucleus or binds to the DHT::AR complex of the AR pathway. During implementation, we considered two cases; MAPK pathway as linear and as a pathway with feedback loop from Erk to Raf to study the dynamics of the system [26]. Although linear pathways are not likely to be found in true biological world it was interesting to investigate the differences in the output response when loops and cross-talks are present in biological systems and when the cascade is purely linear.

For the model of MAPK pathway we used a system of differential equations based on the mass-action kinetic approach and an evolutionary algorithm has been used for the optimization of the problem [27]. The usage of differential equations helps to represent the pathways as systems of biochemical reactions. The mass-action kinetic approach states that when a system is at equilibrium, then the ratio $\frac{[C]^c [D]^d}{[A]^a [B]^b}$ is stable. A and B denote the reactants and C and D state for the products of the chemical reaction: $\alpha A + bB \rightarrow cC + dD$. The α , b , c and d are the coefficients for a balanced reaction.

Evolutionary algorithm

In many optimization problems it is not easy to find an optimal solution mostly due to heavy computations required. The use of evolutionary algorithms in problems that are difficult to solve help to perform with good approximate solutions. As A.E. Eiben reports in [28], evolutionary algorithm's concept is based on nature's concept of evolution and the steps followed are:

- Generation of population of individuals.
- Each individual is evaluated according to the concept of the survival of the fittest (selection) and a fitness function is used depending on the problem.
- The following four steps are repeated until the assumptions for the fitness function are met:
 - The fittest individuals for survival are selected for reproduction.
 - Mutation of the parameters after selection.

- Evaluation of the new individuals according to the chosen fitness function.
- Discard the unsuccessful individuals with the new ones.

Comparing genetic algorithms and evolutionary algorithms, evolutionary algorithm is based on mutation to find the possible solutions for the problem whereas genetic algorithm is based on the cross-over between the individuals [29].

Kinetics play an important role in every biochemical reaction as they describe the rate at which a reaction takes place. The role of the evolutionary algorithm is to find out the values of kinetic parameters that fit more in the output response of the model. For the optimization and the evolution of kinetic parameters we have implemented in MATLAB R2011a [30] an evolutionary algorithm enabling us to see how the populations are evolving.

Before applying the evolutionary algorithm, we have created a set of 200 generations. Each generation contained a population of 200 networks with randomly generated weak kinetic parameters so that the system was heterogeneous. For each generation, we calculated the dynamics of the populations for some input signals with predefined initial values. The kinetics were the differential equations prepared describing either production or degradation. The fitness function that we used in our study was based on the response of the networks for the six input signals with different values. Based on the response of signaling networks to the signals and by comparing it to a predefined threshold ($\frac{1}{10}$ th of the initial concentration of the target protein) we made a selection of the most successful networks ($\frac{1}{4}$ th of the total populations) and for the selected ones their kinetic parameters have been allowed to change. After the mutation we created four copies of the selected populations and we continued the process for all the 200 generations. Figure 3 describes the steps for the implementation of evolutionary algorithm as outlined in [31].

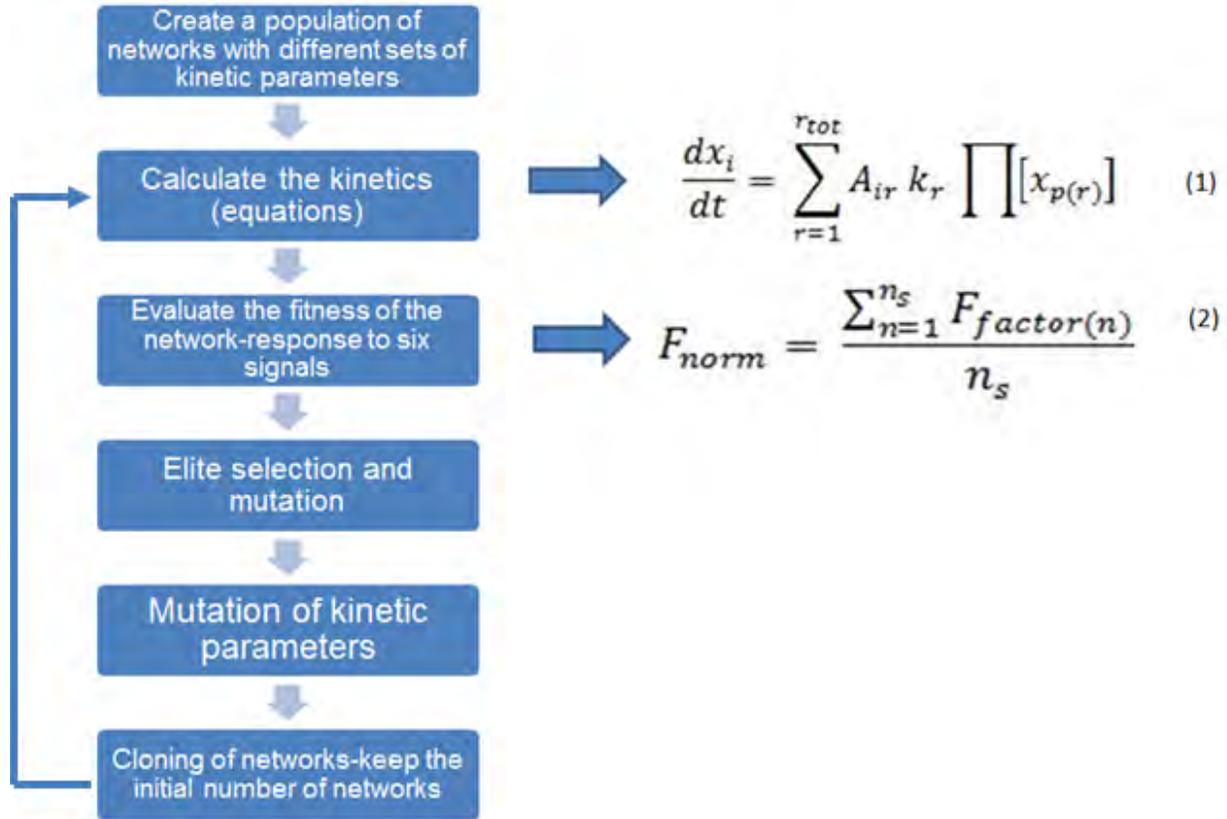


Figure 3. Steps involved in evolutionary algorithm. The kinetics is calculated according to equation 1. r_{tot} denotes the total number of reactions, A_{ir} represents the the molecules involved in each reaction, k_r describes the kinetic parameters and $\prod [x_{p(r)}]$ indicates the product arisen from the reactions. The fitness function is calculated from equation 2 where we take the mean value of each fitness factor assigned for each input signal. n_s denotes the total number of input signals.

To solve the differential equations that describe the dynamics of the system we used the solver of Matlab ode23tb: $[T, Y] = \text{ode23tb}(\text{odefun}, \text{tspan}, y0)$

- T denotes the time points.
- Y indicates the solution of the differential equations.
- odefun is the function called that contains the set of differential equations.
- tspan specifies the time space for the solution of the system.
- y0 describes the initial concentrations of the molecules involved.

After implementing the MAPK pathway, we followed the same approach by combining the three pathways in one. We investigated different cases for both normal and cancer prostate signaling networks by following the two systems described in Figure 2. For normal prostate signaling network we made the assumption that all inactive molecules had the same concentration and were equal to 10. We were mostly interested in PTEN and RasGAP proteins. We used different molecules as the target protein; the double phosphorylated Erk (Erkpp), the phosphorylated Akt (Akp) and the active form of the DHT::AR complex. In the first generation of the evolutionary algorithm the kinetic parameters were very weak (0.001-0.1) and during mutation the values used for kinetic parameters were ranging between 0.1 and 20.

For the signaling network of cancer prostate, we studied the response of Erkpp, Akp and DHT::AR when RasGAP and PTEN were absent respectively and therefore their initial concentration was 0.

Network topology and network motifs

Complex biological systems such as cancer can be represented as networks where the nodes denote the molecules while the edges denote the interactions within them. In a complex biological network, there are only small sets of nodes (proteins) which are known to control main function so address it, Milo R. *et. al.*, [32] introduced the term network motif i.e., patterns commonly found in a network. In biological networks these motifs have been shown to be associated with significant functions in a network such as feedforward or feedback loops. (Figure 4)

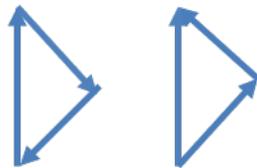


Figure 4. Motifs for feedback and feedforward loop. The first motif describes a feedback loop and the second stands for feedforward loop.

Figure 5 shows some motifs that are more likely to be present in networks that describe cancer as reported in [33].

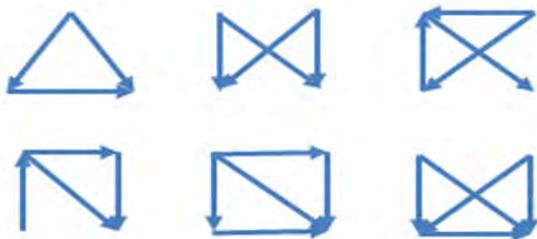


Figure 5. Motifs common in cancer-associated networks.

Visualization of complex biological systems such as cancer in order to detect frequent motifs contribute to further understanding of them by revealing potential loops which in turn can activate or deactivate a protein in a pathway.

In our study we generated a network topology from the signaling reactions in our combined model for prostate signaling. The software tools that we have used to analyze the signaling network and network motifs are Cytoscape [34] and mfinder, respectively [35].

Cytoscape visualizes the given network and unravels relationships in big data sets as well as patterns repeated within it. The user can choose among several different file formats that represent the corresponding network. The file that we used as input for the combined model is written in the following format:

- Source Target Connectivity

Source and Target represent the two molecules that are involved in each reaction. The value of the connectivity is 1 if there is activation between the molecules or -1 if blocking between the two molecules is taking place.

mfinder is a software tool that detects network motifs. In our study we focused on the 3-node motifs. There are 13 possible motifs for three nodes combinations as displayed in Figure 6.



Figure 6. 3-node motifs. There are 13 possible motifs that can appear on three nodes. We calculated the number of times each of these motifs appear in our network.

Although mfinder is able to detect up to 6-node motifs we only focused on motifs found in 3 nodes. The basic command to run mfinder is :

- mfinder <input network file > [-s motif size] [-r <num of random networks>] [-f <output file>]

input network file denotes the network to be processed. The file should be in “.txt” format and each line of the file is written as:

<source node> <target node> <edge weight>

Source node and target node represent the proteins involved in the network and the edge weight is always assigned to 1 as proposed by the manual guide of mfinder.

motif size is the number of motifs to be detected in the network. In our case the number is 3.

num of random networks describes the number of random networks to be generated. In our study we used the default number which is 100. The networks used are random in order to investigate whether the calculated motifs are a result of biological functions or are based on incidental reasons. To define their role, we tested whether their frequency is higher in a certain number of randomized networks based on the value of z-score [36]. In the section of results we will give an example of the role of z-score in the detection of the 3-node motifs.

output file defines the name of the output file.

Since all the scripts were written in MATLAB, we used the extra flag “-omat” to run mfinder. The output file contained all the information for the motifs and could be easily parsed in MATLAB.

After running the mfinder with the flags described, we obtained 100 output files in “.mat” format (100 randomized networks). Each output file contained 8 columns. The first column described the id of each one of the 13 motifs and the fifth column contained the z-score for which we were interested since it could tell us more about the frequency of the potential motifs. In order to visualize the calculated z-score for the motifs we had to parse the files to extract the z-scores.

Immunohistochemistry protocol (IHC-P)

In order to verify the validity of the mathematical model, we carried out an experiment following the immunohistochemistry analysis. IHC is a technique that identifies the proteins in tissue sections removed during biopsy and uses labelled antibodies that recognize the target protein [37]. The visualization of the interaction between the antigen and the antibody is based on fluorescence technique and a fluorophore-conjugated secondary antibody is used to detect the target protein. After applying the secondary antibody the target protein is visualized using microscopy. The antibodies are very specific and therefore they bind only to the protein that we are interested in.

In our study, we used 8 biopsies from 4 patients with prostate cancer and each biopsy contained different amount of cancer. The primary antibodies used were a mix of anti-mouse and anti-rabbit where the first recognize the proteins Erk1 and 2, while the second recognize the phosphorylated variants of these two proteins.. As secondary antibodies we used a mix of goat-anti-mouse and goat-anti-rabbit that are fluorophore-conjugated antibodies and they recognized the primary antibodies respectively. DNA was counter stained with the DNA-intercalator ToPRO3-Iodide. Our goal was to analyze the expression of Erk and the amount of phosphorylated Erk in the different biopsies.

The different biopsies were enclosed in paraffin and kept on glass slides. The first step of IHC protocol [38] was to remove the paraffin from the slides so that the antibodies could bind to the target protein. For the deparaffinization step we washed the slides with xylene and ethanol. Before the histochemical staining of the protein, we performed antigen retrieval to the samples to uncover the antigenic sites that have been masked in the previous steps. The method that we used for the

antigen retrieval was heat-induced [39]. After the deparaffinization and the antigen retrieval, we detected the target antigen with the two antibodies and used the microscope for the visualization.

Results

To understand the temporal dynamics of signaling molecules, we have investigated the output response in different generations (evolutionary period) for different sets of parameters (such as initial concentration, kinetic parameters, feedback and feedforward loops, and cross-talks between the pathways).

Linear MAPK cascade

In a linear pathway, the pathway does not have any feedback or feedforward loop. We have optimized the pathway for different range of kinetic parameters and initial concentration. We observed that in case of linear pathway, the output response is exclusively sustained as shown in Figure 7. Since, during generation of kinetic parameters before evolution, we keep the kinetic parameters as extremely low so in the first generation the kinetic parameters have weak values and therefore the system is unable to detect any input signal. During the evolutionary process kinetic parameters were mutated randomly in different ranges ($k < 10$, $k < 30$ and $k < 50$) and in all cases the output response remained sustained. We also tested the model with different initial concentrations for the inactive molecules with respect to always keeping the concentration of the target protein less than the others. The target protein of the model was Erk.

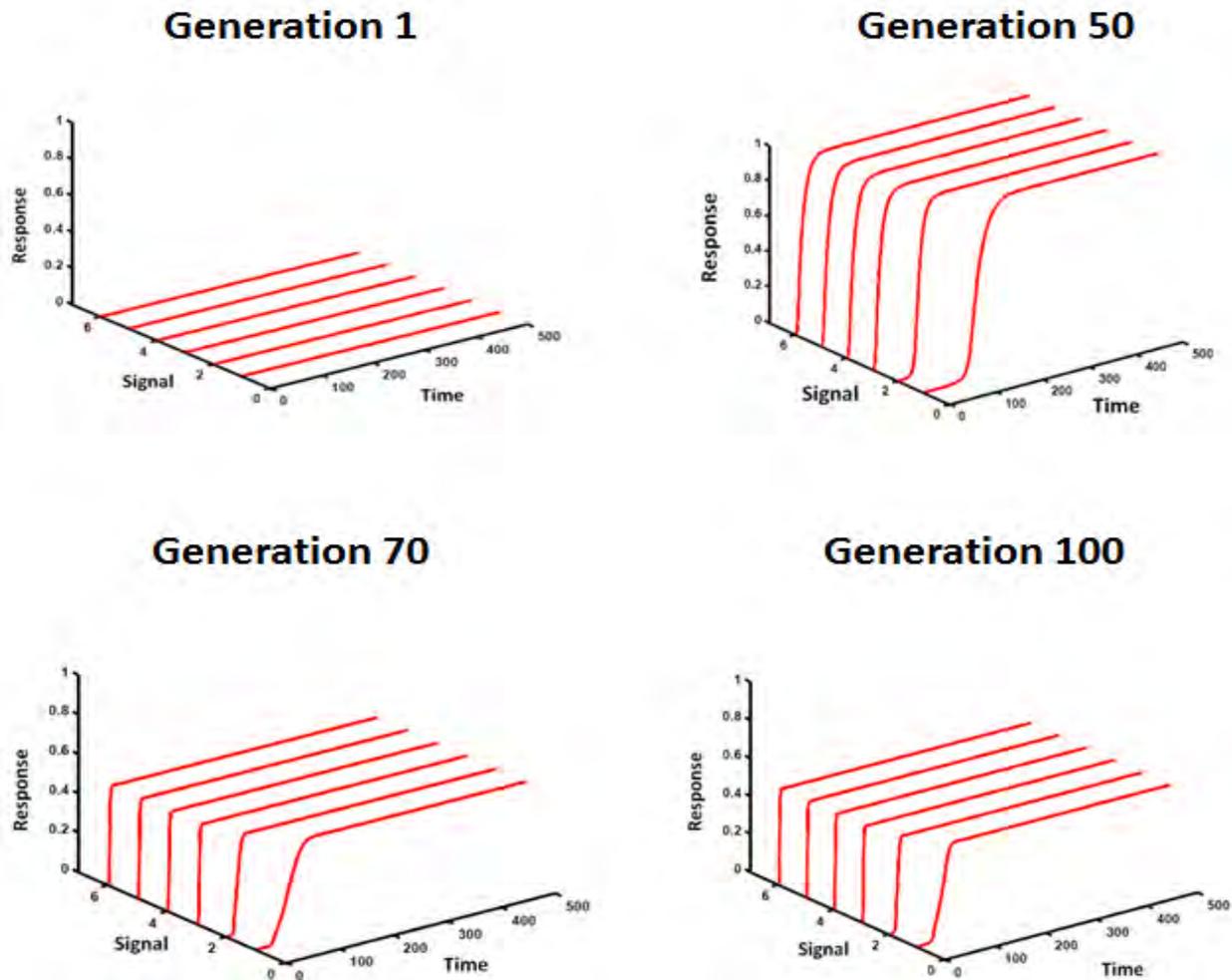


Figure 7. Response of kinetics in MAPK linear pathway during evolution. In linear MAPK pathway the output response to the six different input signals is sustained. The x-axes represent the time range that takes the system to respond to the input signals. The time varies from 0 to 500 points. The y-axes stand for the six input signals with values 0.0001, 0.001, 0.01, 0.1, 1 and 10 respectively. The z-axes show the values of system's response to the six input signals. The total number of generations is 200.

We were also interested in the fitness of the system during evolution. It is indicated from Fig. 8 that the fitness of the linear cascade fluctuates in the beginning but it becomes stable quickly and before reaching the 50th generations. This suggests the evolution of robust sustained response for any kind of change in kinetic parameters in this model.

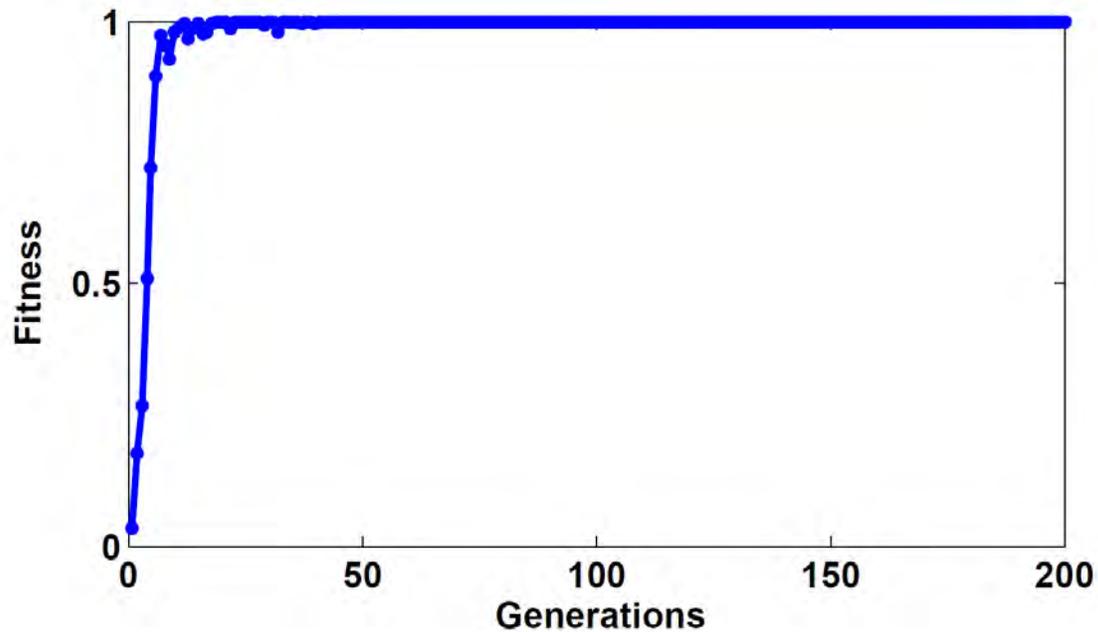


Figure 8. Fitness of MAPK linear pathway during evolution. The x-axis shows the number of generations in the model and the y-axis indicates the values that the fitness function is taking through generations.

MAPK Cascade with a feedback loop

After investigating the linear cascade, the feedback loop has been added in the linear cascade and then has been evolved for different sets of parameters and the output response has been analyzed for all successfully evolved networks. Based on the analysis, we observe that most of the evolved networks' output responses are transient (Fig. 9). However, there can also be rare case where a weak signal can be further amplified with respect to the others due to high values in kinetic parameters as illustrated in Generation 70.

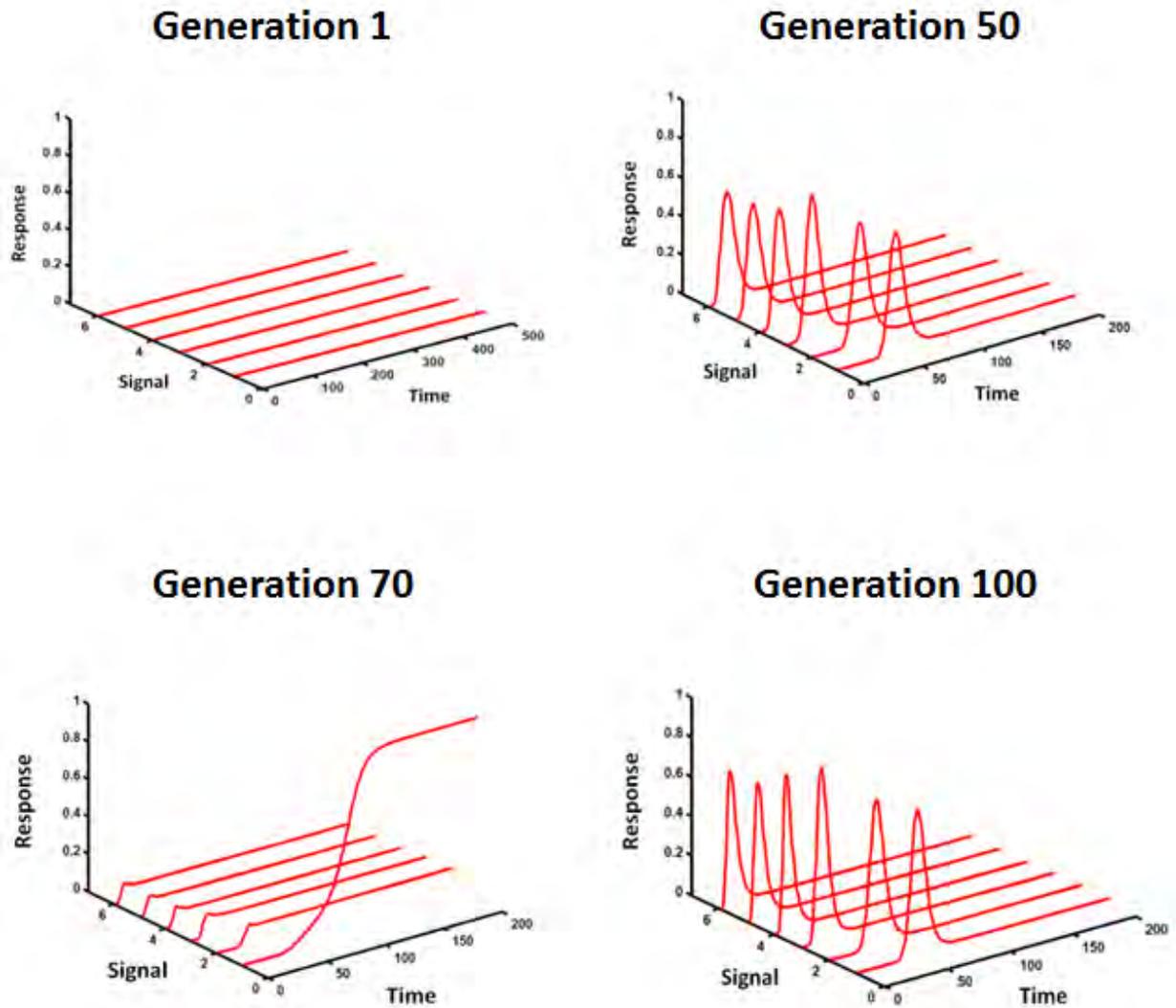


Figure 9. Response of kinetics in MAPK pathway with feedback loop during evolution. In MAPK pathway with a feedback loop, the output response to the six input signals is in most cases transient.

As far as the fitness function of this model is concerned, the fitness fluctuates for all generations suggesting transient response of the system.

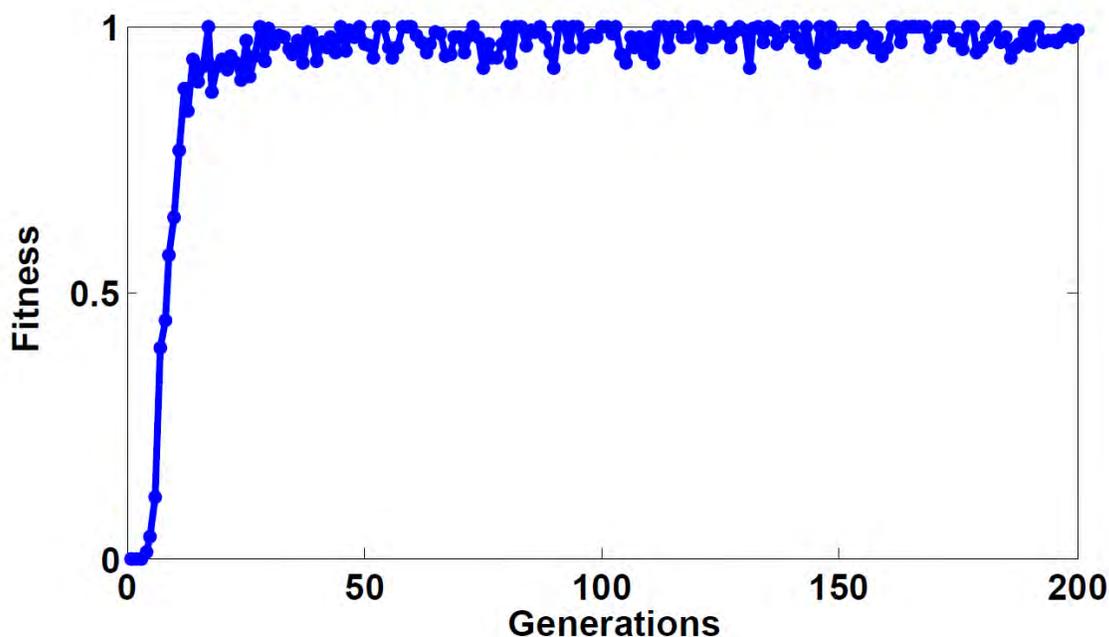


Figure 10. Fitness of MAPK pathway with feedback loop.

After analyzing the MAPK kinase pathway with and without the feedback loop, we set up a model for normal and cancer prostate signaling based on the previously described information [15-20]. This model mainly includes three pathways, AR, MAPK, and AKT pathways. Similar to above mentioned approach, this model has been optimized by using EA for different parameters sets and the kinetics of signaling molecules have been analyzed.

Normal prostate signaling

In the combined model, we have investigated the differences in normal and cancer prostate signaling by describing the model with the minimum set of interactions. Accordingly, we removed the feedback loop from Erk to Raf and we directed on the regulations among the three pathways. Figure 11 describes the evolution of the system in case of normal prostate. We focused on the presence/absence of the proteins PTEN and RasGAP. We applied the selection pressure to double phosphorylated Erk (Erkpp) and examined the evolution of the networks when Erkpp, Aktp, and the active form of DHT::AR were respectively chosen as the output response. In this model all the inactive forms of the molecules had the same initial concentration which was equal to 10 a.u. (arbitrary unit).

In the first generation, for all cases, none of the systems are able to detect the input signals due to the low values of kinetic parameters. The responses during evolution are sustained when Erkpp is chosen as the target protein. The absence of feedback loop verifies the nature of this response. In case of Aktp as output response, the system's response is transient which in general is associated

with the phase of apoptosis in the cell cycle. When the DHT::AR complex responds to the input signals the nature of the response is both transient and sustained. Fitness which is a measure of the quality of the solution in each iteration of the algorithm, reaches the maximum value and becomes stable with in 50 generations. The fitness function also displays fluctuations within 50 generations which means in this evolutionary period the all networks are not able to detect all the six input signals. And after crossing generation 50 almost all the networks are able to detect all the six input signals and the networks become robust to any further changes in the kinetic parameters.

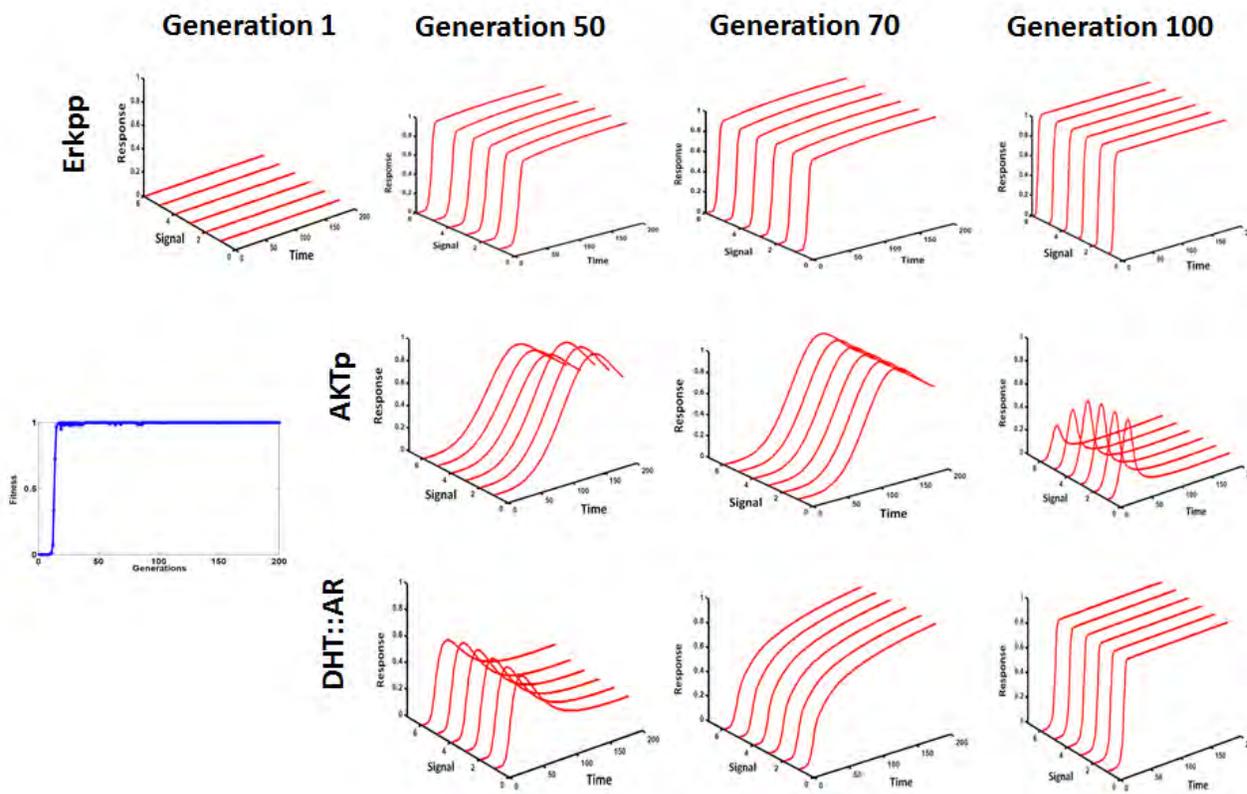


Figure 11. Response of the model in normal prostate signaling. Evolution of the system in normal prostate when different proteins are chosen as the output response and when all the inactive forms of molecules including PTEN and RasGAP have the same concentration. The fitness of the model is also presented.

Cancer prostate signaling

In case of cancer prostate, Figure 12 and 13 describe the two conditions examined. Figure 12 represents the evolution of the network and the fitness when the concentration of RasGAP is 0 a.u. The role of RasGAP protein is to deactivate Ras in MAPK pathway. Therefore its loss results in a continuous activation of Ras [40]. As already mentioned in the introduction the loss of RasGAP is due to a mutation in NF1 gene. Neurofibromatosis type 1 (NF1) is responsible for the synthesis of neurofibromin protein which acts as tumor suppressor protein [41]. Klose A. *et. al.* [42] studied the Ras-GTPase-activating role of neurofibromin and showed that a mutation in NF1 gene functions as

an off signal for RasGAP activity. This results in its absence and eventually to the excessive activation of Ras leading to tumor formation.

Figure 12 describes the evolution of the system during cancer prostate. The RasGAP was absent and therefore the concentration was 0. All the other molecules had the same concentration which was equal to 10. In this case also we applied the selection pressure to double phosphorylated Erk (Erkpp) and examined the evolution of the networks when Erkpp, Aktp, and the active form of DHT::AR were respectively chosen as the output response. The kinetic parameters were randomly ranging between 0 and 20.

In the first generation, for all cases, none of the systems are able to detect the input signals due to the low values of kinetic parameters. The responses during evolution are sustained when Erkpp is chosen as the target protein explained by the removed feedback loop from Erk to Raf. Both in cases of Aktp and DHT::AR as outputs the responses are sustained which is an indication of continuous growth as it is happening in real cancer biological systems. The fitness function oscillates in the first generations but becomes stable after 70th proposing sustained response.

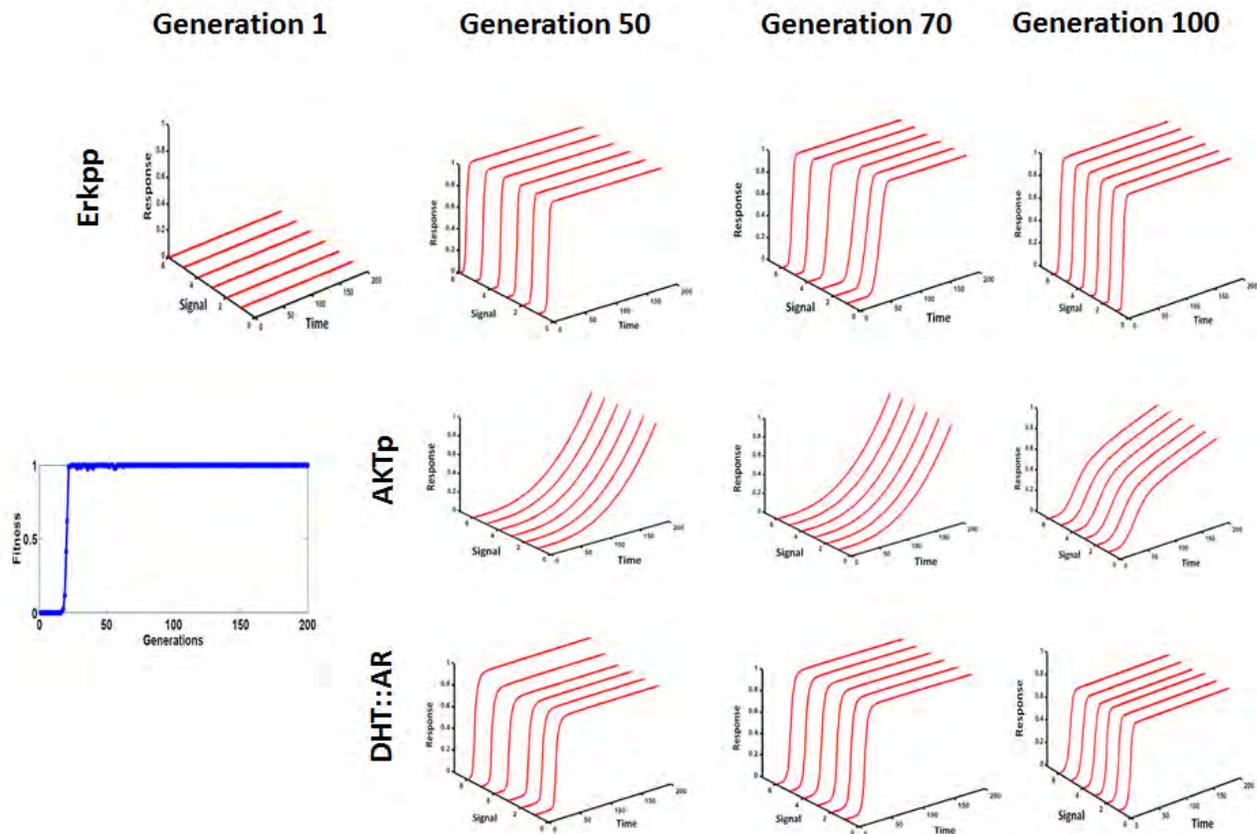


Figure 12. Response of the model in cancer prostate signaling because of RasGAP loss. Evolution of the network when RasGAP is absent for different proteins chosen as the output response

The second condition tested is the loss of PTEN gene. PTEN is a tumor suppressor gene that is either deleted or mutated in prostate cancer. It negatively regulates the PI3K-AKT-mTOR pathway controlling AKT and is responsible for growth and proliferation in cells. Accordingly, the loss of PTEN results in the increased activity of AKT in the pathway affecting also the other molecules in the pathway like mTOR and S6K[43, 44]

In this condition we used the same conditions as when RasGAP was absent, with the difference that PTEN's concentration was 0 and all the other molecules (including RasGAP) had concentrations which were equal to 10. The kinetics were also ranging between 0 and 20.

After analyzing this model for 200 generations, all the responses are sustained when Erkpp is the target protein. Both sustained and transient responses are observed in both Aktp and DHT::AR complex as target proteins. The presence of transient response points out that Aktp is not affected by the loss of PTEN in the same way as is affected by the loss of RasGAP. The complex of DHT::AR is not influenced in generation 50 but in the other two generations it demonstrates sustained response (Fig. 13).

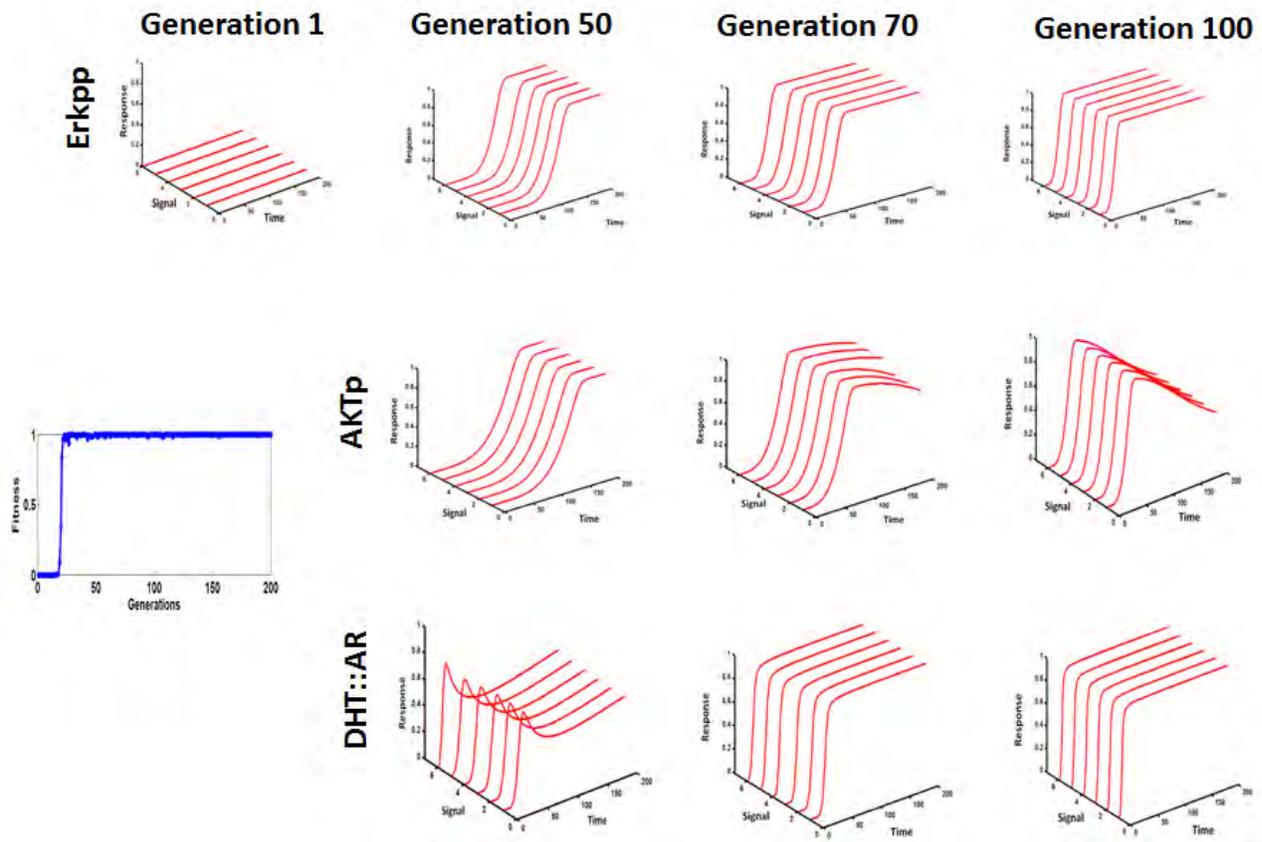


Figure 13. Response of the model in cancer prostate signaling because of PTEN loss. Evolution of the networks due to loss of PTEN and when different proteins are chosen as the output response.

Some experiments that were also carried out, included Erkpp, Aktp, and the DHT::AR as output response with different initial concentrations of the inactive forms of Erk, Akt and AR. The concentrations were taking values 0.1 a.u, 1 a.u, 5 a.u and 10 a.u. We ran also the simulations with different kinetic parameters ($k < 10$, $k < 20$ and $k < 50$) each time. The results gave mostly sustained responses.

Network visualization

After analyzing the kinetic parameters, we were interested in visualizing the network. As mentioned earlier the visualization of complex biological systems help to find patterns that are hidden because of the complexity. Figure 14 shows how the integration of the three pathways have been captured by using Cytoscape. In a delimited text file we prepared the network by writing the biochemical reactions in the form of nodes and edges. The nodes represent the molecules and the edges denote if a molecules activates or inhibits the other.

Here is two lines example from the cytoscape file describing activation and inhibition respectively:

Signal_PTEN	PTEN	1
PTENa	PI3Ka	-1

By looking at the topology of the network we identify the molecules with high connectivity like the single and double phosphorylated forms of Akt, Erk and Mek and the single phosphorylated form of Raf. The other molecules have mostly similar topology.

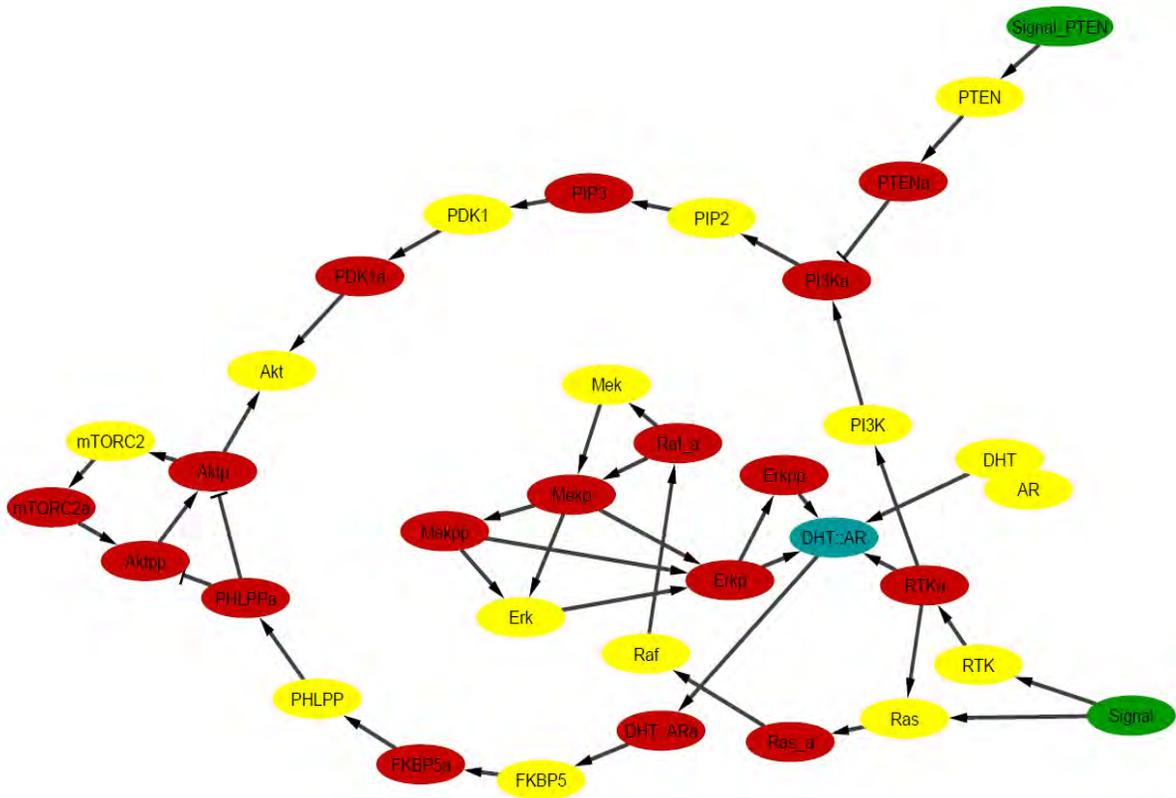


Figure 14. Visualization of the network by using Cytoscape. Nodes in green denote the signals, nodes in yellow define the inactive molecules, nodes in red represent the phosphorylated forms of molecules(single and double) and the node in teal describes the DHT::AR complex.

Network motifs

The main 3-node sub-functional networks are shown in Figure 15 as they have been arisen from mfinder. In order to create the file to use for mfinder we were based on the file used in Cytoscape with the only difference that instead of names we numbered the molecules and we used 1s for the last column of the file.

The z-score of the motifs is based on the number of times each motif appears in the randomly generated networks as described in the section of methods because it tells us if each motif in our network is significant compared to the motifs in the random networks. The higher the value of z-score, the most significant is the motif in the network. In our network the fifth motif is the most significant one, representing a feedforward loop

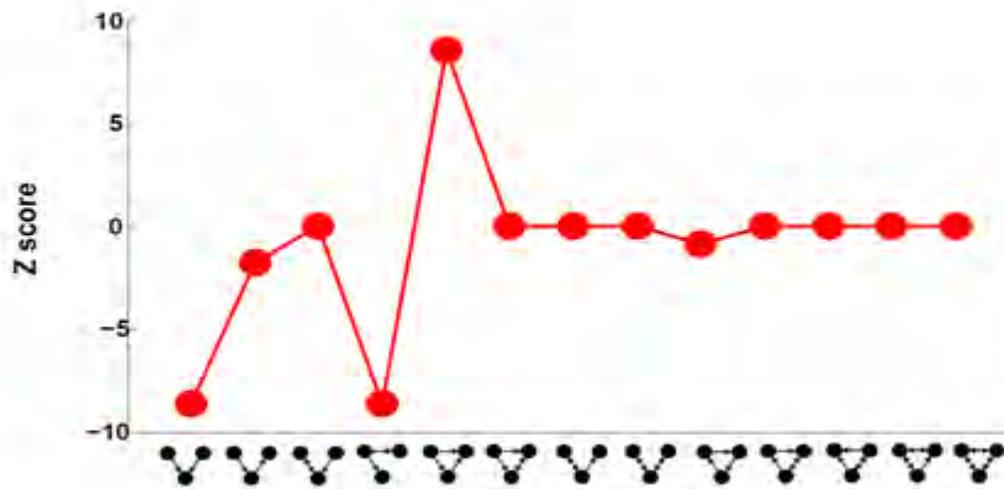


Figure 15. Detection of 3-node motifs for normal prostate model. The mean value of z-score for all the 100 randomized networks is displayed. The fifth motif is more likely to be present. The higher the positive value of z-score, the higher the significance.

Experimental validation

The mathematical model of the three pathways was validated experimentally using the method of immunohistochemistry. We were interested in the expression of phosphorylated Erk in normal and cancer prostate.

In Figure 6, 4 biopsies from 2 different patients are displayed. The upper left quadrant of the figure shows 2 different biopsies from a normal tissue and the upper right part of the figure displays different parts of the same biopsies. Similarly, the down left and right parts of the figure describe biopsies from patients with cancer. For each biopsy there are three pictures. The first demonstrates how phosphorylated Erk is expressed, the second picture includes both phosphorylated Erk and DNA and the third one contains the two forms of Erk (inactive and active) as well as the DNA.

DNA is used as a marker of nuclei and is illustrated with blue color. The Erk protein is denoted by the green color and its phosphorylated form by red color. The main difference between the biopsies without and with cancer is that in normal biopsies there is structure in cells while in cancer biopsies the cells do not have any architecture. An indication of severe cancer is seen in biopsy 4 where no structure is observed especially in the left part.

The aim of this experiment is to compare the intensity of phosphorylated Erk between the different biopsies. The signal of phosphorylated Erk is strong in biopsies with cancer especially in biopsy 3

indicating continuous proliferation in cells. In contrast, there is no high signal of phosphorylated Erk in normal biopsies.

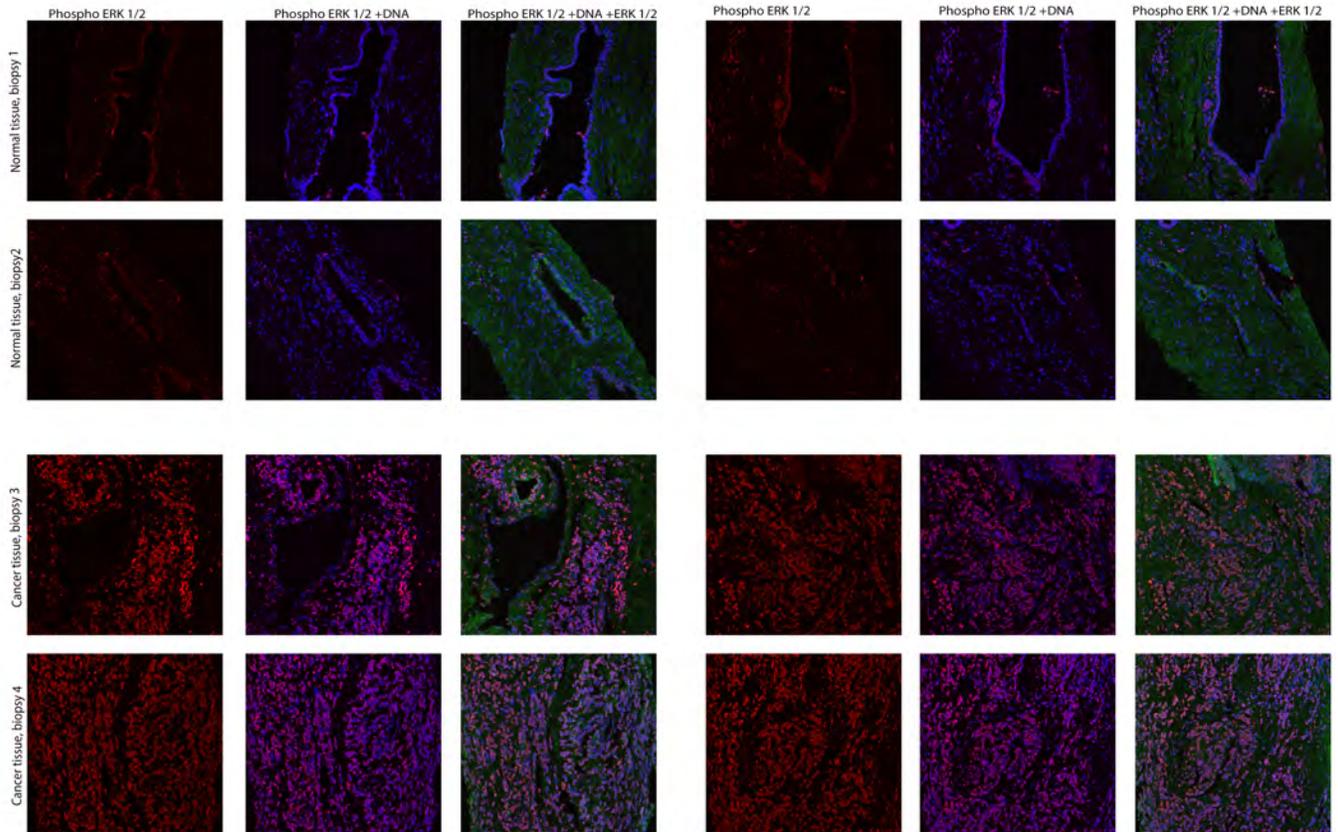


Figure 16. Expression of phosphorylated Erk in normal and cancer biopsies. Red color is used to show the phosphorylated Erk, green is used for its inactive form and blue for DNA which is used as a mask. Cells in cancer biopsies do not have any structure and high levels of phosphorylated Erk are found.

The diagram in Fig. 17 compares the mean fluorescent intensity of Erk and phosphorylated Erk among the biopsies. The difference in phosphorylated Erk between normal and cancer biopsies is significant ($p < 0.001$, two sided t-test, more than 3000 cells measured) while there is no huge difference concerning the intensity of Erk.

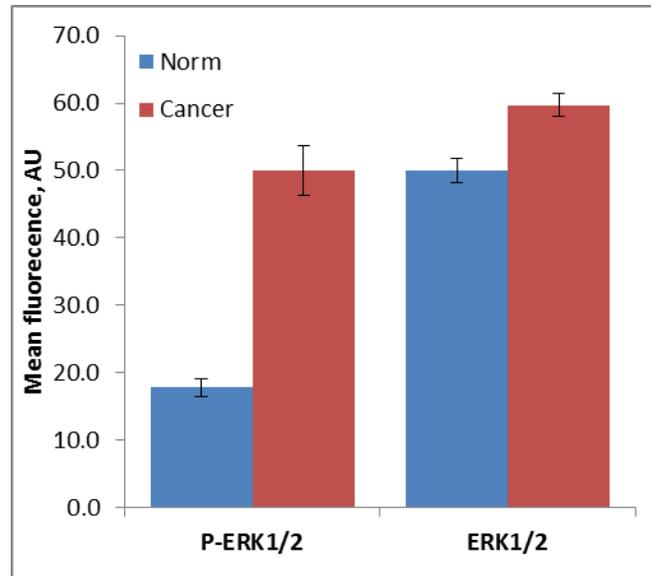


Figure 17. Comparison of the mean fluorescent intensity of Erk and phosphorylated Erk between normal and cancer biopsies. Error bars shows SE.

The microscopic examination of a normal and cancer tissue in the same biopsy is presented in Figure 18. The images show a lower degree of structure in the cancer tissue compared with normal tissue. However, compared with the previous cancer biopsies, this cancer is more structured and thus not so advanced and the levels of phosphorylated Erk are lower.

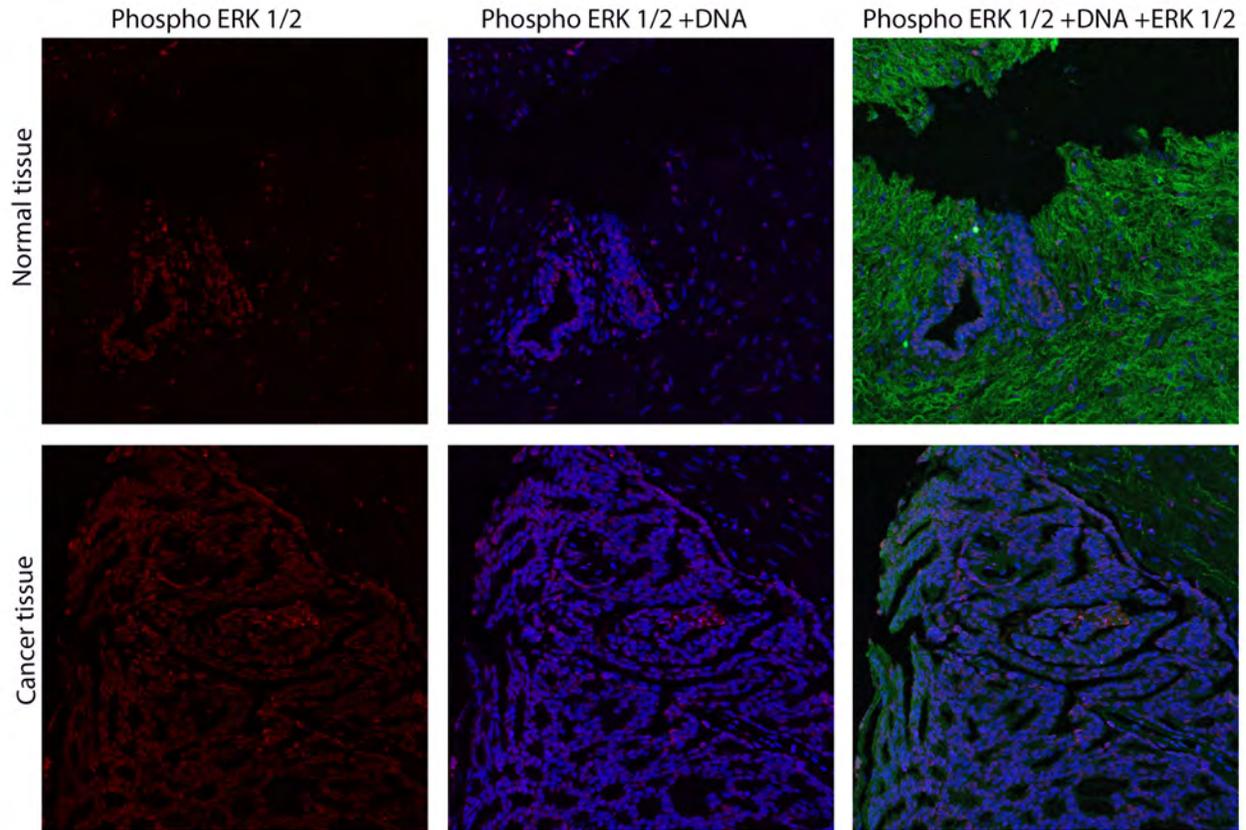


Figure 18. Levels of phosphorylated Erk in normal and cancer tissue of the same biopsy.

Figure 19 summarizes the fluorescence intensity of phosphorylated Erk from the biopsy of Fig.18. It is observed that there is a small but significant difference ($p < 0.001$, two-sided t-test, more than 1500 cells measured) between cancer and normal tissue concerning the levels of phosphorylated Erk.

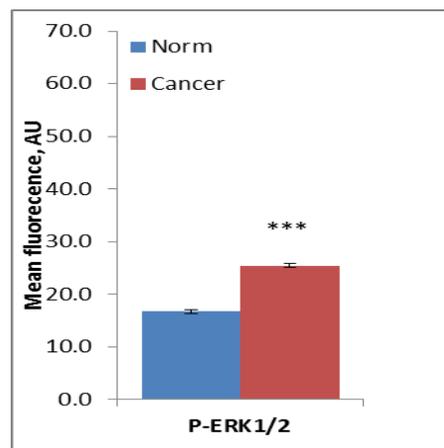


Figure 19. Comparison of the fluorescent intensity for phosphorylated Erk between normal and cancer tissue in the same biopsy. Error bars show SE.

Discussion

Complex diseases such as cancer cannot always be studied in broad perspectives in laboratory because of its complexity at multiple level and the parameters associated with it [45, 46]. A deeper understanding of the pathways involved can be achieved through the field of systems biology and the use of mathematical modeling for their simulation. In this study, we have established a mathematical model for the three most known pathways involved in prostate cancer and investigated the change in the output response nature (sustained or transient) under different conditions with an interest to the loss of PTEN gene and RasGAP protein. Sustained responses were identified in most cases in the model describing prostate cancer. The sustained response found is associated with the continuous proliferation in cancer cells instead of apoptosis that is taking place in normal cells. Besides, we identified the molecules with high connectivity in the topology of our network as well as the most common pattern by using software tools for visualization. Although kinetics play a major role in the biochemical reactions, we were rather focused on the minimum possible set of interactions that reassemble prostate cancer in real biological world.

The experimental validation of the theoretical model verified the high concentration of phosphorylated Erk in prostate cancer. In the theoretical model the double phosphorylated Erk was used as the target protein as we considered its double phosphorylated form to be the fully activated form of the protein. The single phosphorylated form represents the partially activated Erk.

Heterogeneity is one of the main reasons that make cancer a complex disease. Heterogeneity which is the variation both among tumors (inter-tumor heterogeneity) and within tumors (intra-tumor heterogeneity) exist both in genetic and in epigenetics level and affects significant pathways in all types of cancer [47]. Since, heterogeneity is a dominant trait during cancer progression, it leads to different mutations within genes influencing their expression. Therefore, it is possible that the mathematical model for the three pathways will result in different nature responses for different cancer cells as they can show different changes in their proliferation and their morphology

As future prospects, the already established model for prostate cancer can be further improved by integrating the intra-tumor heterogeneity and therefore, be able to study the different levels of gene expression taking also into account the role of kinetics. The analysis of gene expression would give a better insight in the components crucial for the formation of prostate cancer.

Conclusions

Summing up the results, it can be concluded that in MAPK pathway without any feedback loop the double phosphorylated Erk as the target protein produces a sustained response. The findings for MAPK pathway having a feedback loop from Erk to Raf suggest transient response. In normal prostate model that simulates the three pathways, the double phosphorylated Erk as output produces sustained response because we have not applied the feedback loop from Erk to Raf. Nevertheless, when Aktp and the complex DHT::AR are tested as the target protein they produce mostly transient responses. Finally regarding the cancer prostate model, we applied two known facts involved in formation of prostate cancer including the loss of PTEN and the loss of RasGAP. The absence of RasGAP led to sustained responses in all cases whereas the loss of PTEN led to a mixture of sustained and transient responses. The increased levels of phosphorylated Erk from the cancer biopsies verified the sustained responses of Erkpp in the mathematical model.

Acknowledgements

I would like to thank Mobashir Mohammad and Niklas Schultz for their help throughout the project, for being positive to explain things that I could not understand and for the interesting discussions that we had.

I would also like to thank Prof. Thomas Helleday that gave me the chance to be part of his group and learn so many new things in the field.

Last but not least, a special thank you to my friends and family for their love and support all these years and especially these last two that I am far away from them.

References:

- [1] Berg, J.M., Tymoczko, J.L., Stryer, L., 2002. *Biochemistry*. 5th edition. New York: W H Freeman. Chapter 15, Signal-Transduction Pathways: An Introduction to Information Metabolism.
- [2] Pharmaceuticals A, Pharmaceuticals K., 2000. Signal transduction as a drug-discovery platform Mapping the key signaling molecules in biochemical pathways will be central. *Nature Biotechnol.* 18(3):37–39.
- [3] Mohammad, M., Schraven, B., Beyer, T., 2012. Simulated evolution of signal transduction networks. *PLoS ONE* 7:e50905.
- [4] Wolkenhauer, O., 2004. *Systems Biology - Dynamic Pathway Modelling*.
- [5] Hanahan, D., Weinberg, R.A., 2011. Hallmarks of cancer: the next generation. *Cell* 144: 646–674.
- [6] Evan, G.I., Vousden, K.H., 2001. Proliferation, cell cycle and apoptosis in cancer. *Nature* 411(6835):342–348.
- [7] Bachmann, J., Raue, A., Schilling, M., Becker, V., Timmer, J., Klingmüller, U., 2012. Predictive mathematical models of cancer signalling pathways. *J Intern Med.* 271: 155–165.
- [8] Orton, R.J., Sturm, O.E., Vyshemirsky, V., Calder, M., Gilbert, D.R., Kolch, W., 2005. Computational modelling of the receptor-tyrosine-kinase-activated MAPK pathway. *Biochem J.* 392(Pt 2):249–261.
- [9] Steven, B., Bricker, C., 1969. The law of mass action. *J Chem Educ.* 46 (8): 499.
- [10] Johnson, K.A., Goody, R.S., 2011. The Original Michaelis Constant: Translation of the 1913 Michaelis–Menten, 50 (39): 8264–8269.
- [11] Zhang, W., Liu, H.T., 2002. MAPK signal pathways in the regulation of cell proliferation in mammalian cells. *Cell Res.* 12(1):9–18.
- [12] Brooke, G., Bevan, C., 2009. The Role of Androgen Receptor Mutations in Prostate Cancer Progression. *Curr Genomics* 10(1):18–25.
- [13] Cuyas, E., Corominas-Faja, B., Joven, J., Menendez, J.A., 2014. Cell cycle regulation by the nutrient-sensing mammalian target of rapamycin (mTOR) pathway. *Methods Mol Biol.* 1170:113–144.
- [14] Barlow, L.J., Shen, M.L., 2013. SnapShot: Prostate Cancer. *Cancer Cell* 24(3): 400.

- [15] Bitting, R.L., Armstrong, A.J., 2013. Targeting the PI3K/Akt/mTOR pathway in castration-resistant prostate cancer. *Endocr Relat Cancer* 20(3):R83–99.
- [16] Keniry, M., Parsons, R., 2008. The role of PTEN signaling perturbations in cancer and in targeted therapy. *Oncogene* 27:5477–5485.
- [17] Ratner, N., Miller, S.J., 2015. A RASopathy gene commonly mutated in cancer: the neurofibromatosis type 1 tumour suppressor. *Nat Rev Cancer* 25: 290–301.
- [18] Mendoza, M.C., Er, E.E., Blenis, J., 2011. The Ras-ERK and PI3K-mTOR Pathways: Crosstalk and Compensation. *Trends Biochem Sci.* 36(6):320–328.
- [19] Lee, S.H., Johnson, D., Luong, R., Sun, Z., 2015. Crosstalking between androgen and PI3K/AKT signaling pathways in prostate cancer cells. *J Biol Chem.* 290(5): 2759–2768.
- [20] Kaarbø, M., Klok, T.I., Saatcioglu, F., 2007. Androgen signaling and its interactions with other signaling pathways in prostate cancer. *Bioessays* 29(12):1227–1238.
- [21] Zhu, M.L., Kyprianou, N., 2008 Androgen receptor and growth factor signaling cross-talk in prostate cancer cells. *Endocr Relat Cancer* 15(4):841–849.
- [22] Lonergan, P.E., Tindall, D.J., 2011. Androgen receptor signaling in prostate cancer development and progression. *J Carcinog.* 10:20.
- [23] Schoeberl, B., Eichler-Jonsson, C., Gilles, E.D., Müller, G., 2002. Computational modeling of the dynamics of the MAP kinase cascade activated by surface and internalized EGF receptors. *Nat Biotech.* 20:370–375.
- [24] Tian, T., Song, J., 2012. Mathematical Modelling of the MAP Kinase Pathway Using Proteomic Datasets. *PLoS ONE* 7(8):e42230.
- [25] Filchtinski, D., Sharabi, O., Rüppel, A., Vetter, I.R., Herrmann, C., Shifman, J.M., 2010. What makes Ras an efficient molecular switch: a computational, biophysical, and structural study of Ras-GDP interactions with mutants of Raf. *J Mol Biol.* 399(3):422–435.
- [26] Fritsche-Guenther, R., Witzel, F., Sieber, A., Herr, R., Schmidt, N., Braun, S., Brummer, T., Sers, C., Blüthgen, N., 2011. Strong negative feedback from Erk to Raf confers robustness to MAPK signaling. *Mol Syst Biol.* 7:489.
- [27] Kaneko, K., 2007. Evolution of robustness to noise and mutation in gene expression dynamics. *PLoS ONE* 2(5): e434.
- [28] Eiben, A.E., Smith, J.E., 2003. *Introduction to Evolutionary Computing*. 1st ed. Springer.
- [29] Simon, D., 2013. *Evolutionary Optimization Algorithms*. 1st ed. WILEY.

- [30] MATLAB and Statistics Toolbox Release 2012b, The MathWorks, Inc., Natick, Massachusetts, United States.
- [31] Mohammad, M., Madhusudhan, T., Isermann, B., Beyer, I., Beyer, T., Schraven, B., 2014. Negative interactions and feedback regulations are required for transient response. *Sci Rep.* 4:3718.
- [32] Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., Alon, U., 2002. Network motifs: simple, building blocks of complex networks. *Science* 298:824–827.
- [33] Awan, A., Bari, H., Yan, F., Moksong, S., Yang, S., Chowdhury, S., Cui, Q., Yu, Z., Purisima, E.O., Wang, E., 2007. Regulatory network motifs and hotspots of cancer genes in a mammalian cellular signaling network. *IET Syst Biol.* 1(5):292–297.
- [34] Saito, R., Smoot, M.E., Ono, K., Ruscheinski, J., Wang, P.L., Lotia, S., Pico, A.R., Bader, G.D., Ideker, T., 2012. A travel guide to Cytoscape plugins. *Nat Methods* 9(11):1069–1076.
- [35] Kashtan, N., Itzkovitz, S., Milo, R., Alon, U., 2004. Efficient sampling algorithm for estimating subgraph concentrations and detecting network motifs. *Bioinformatics* 20(11): 1746–1758.
- [36] Weizman Institute of Science, 2015. Mfinder Tool Guide. Web. <https://www.weizmann.ac.il/mcb/UriAlon/NetworkMotifsSW/mfinder/MfinderManual.pdf>. Date visited 18 May 2015.
- [37] Gounder, S.K., Chang, V.T., Hoover, D., Gonzalez, M.L., Ahmed, S., Finch-Cruz, C., Duque, L., Zhong, F., Toomey, K., Kasimis, B.S., 2008. Prostate cancer immunohistochemical (IHC) stains and survival in stage D3 patients (pts). *J Clin Oncol.* 26:22187.
- [38] Life Technologies, 2015. Overview of Immunohistochemistry. Web. <https://www.lifetechnologies.com/se/en/home/life-science/protein-biology/protein-biology-learning-center/protein-biology-resource-library/pierce-protein-methods/overview-immunohistochemistry.html>. Date visited 25 May 2015.
- [39] Abcam, 2015. IHC antigen retrieval protocol. Web. <http://www.abcam.com/protocols/ihc-antigen-retrieval-protocol>. Day visited 25 May 2015.
- [40] Maartens, O., Cichowski, K., 2014. An expanding role for RAS GTPase activating proteins (RAS GAPs) in cancer. *Adv Biol Regul.* 55:1–14.
- [41] Carroll, S.L., Stonecypher, M.S., 2004. Tumor suppressor mutations and growth factor signaling in the pathogenesis of NF1-associated peripheral nerve sheath tumors. I. The role of tumor suppressor mutations. *J Neuropathol Exp Neurol.* 63(11):1115–1123.

- [42] Klose, A., Ahmadian, M.R., Schuelke, M., Scheffzek, K., Hoffneyer, S., Gewiwa, A., Schmitz, F., Kaufmann, D., Peters, H., Wittinghofer, A., Nürnberg, P 1998. Selective disactivation of neurofibromin GAP activity in neurofibromatosis type 1 (NF1). *Hum Mol Gen.* 7(8): 1261–1268.
- [43] Ruscetti, M., Hong, W., 2013. PTEN and Prostate Cancer, In D.J. Tindall's Book "Prostate Cancer: Biochemistry, Molecular Biology, and Genetics, Protein Reviews", (Chapter 4): 87–137.
- [44] Hollander, M.C., Blumenthal, G.M., Dennis P.A., 2011. PTEN loss in the continuum of common cancers, rare syndromes and mouse models. *Nat Rev Cancer* 11(4): 289–301.
- [45] Boyd, L.K., Mao, X., Lu, Y.J., 2012. The complexity of prostate cancer: genomic alterations and heterogeneity. *Nat Rev Urol.* 9:652– 664.
- [46] Marte, B., 2013. Tumour heterogeneity. *Nature* 501: 327.
- [47] Burrell, R.A., McGranahan, N., Bartek, J., Swanton, C., 2013. The causes and consequences of genetic heterogeneity in cancer evolution. *Nature* 501:338–345.