D5.1

Generic logical model of prostate cancer

Project number:	668858
Project acronym:	PrECISE
Project title:	PrECISE: Personalized Engine for Cancer Integrative Study and Evaluation
Start date of the project:	1st January, 2016
Duration:	36 months
Programme:	H2020-PHC-02-2015

Deliverable type:	Report
Deliverable reference number:	PHC-668858 / D5.1 / V1.0
Work package contributing to the deliverable:	WP 5
Due date:	December 2016 – M12
Actual submission date:	21st December, 2016

Responsible organisation:	CI
Editor:	Pauline Traynard
Dissemination level:	PU
Revision:	V1.0

Abstract:	This deliverable details how we built a Boolean model of prostate cancer and the results of its simulations, in different conditions and for different perturbations.
Keywords:	Logical modelling, Boolean model, stochastic simulations, prostate cancer

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 668858.

This work was supported (in part) by the Swiss State Secretariat for Education, Research and Innovation (SERI) under contract number 15.0324-2. The opinions expressed and arguments employed therein do not necessarily reflect the official views of the Swiss Government.

Editor

Traynard, Pauline (CI)

Contributors (ordered according to beneficiary numbers)

Tobalina, Luis (UKAACHEN)

Guo, Tiannan (ETH)

Calzone, Laurence (CI)

Executive Summary

A network representing the pathways and the genes or proteins known to be deregulated in prostate cancer was constructed. The network encompasses signaling pathways such as the cell cycle, cell death, androgen-receptor response, DNA repair, etc. The regulatory network shows the positive and negative influences that each entity of the network has onto the others. The entities can be genes, proteins, complexes, or processes and phenotypes. A series of inputs are chosen to account for the different conditions of the microenvironment: presence of growth factors, nutrients, or hypoxia, DNA damage, TNFalfa, etc. Some outputs are explicitly added to monitor the activity of some biomarkers of phenotypes: cell proliferation, apoptosis, quiescence, etc. For example, the variable *Proliferation* gets activated when at least one of the cyclins is ON.

The network was built from knowledge extracted from publications, and databases, and completed by a tool developed by UKAACHEN, Omnipath, to query existing databases and complete semi-automatically the links and the neighbors of a gene to add in the regulatory network.

From this network, a logical model is derived describing the network dynamics in specific contexts (dependent on initial conditions or perturbations for instance). Logical models are simple, require in principle no quantitative information, and can be hence applied to large networks combining multiple pathways.

This deliverable presents some analyses made on this model. Some physiological conditions are first simulated in order to validate the model. The validation concerns both the choices made on the topology of the network (players and the interactions), and on those on the logical rules that are associated to each of the variable of the model. Then, some modifications of the wild type are explored, corresponding to mutants. The model still has to be validated on patient or clone profiles, and on data by performing some computations on the activity of pathways in the PC39 patients.

Contents

List of Figures

List of Tables

Chapter 1 Introduction

1.1 Boolean modeling

Like most cancers, prostate cancer arises from mutations on single somatic cells that induce deregulated proliferation, invasion of adjacent tissues, and metastasis. The high heterogeneity of tumor genetic and epigenetic profiles is explained by the large number of interacting proteins and the complex cross-talks between diverse cell signaling pathways that can be altered in cancer cells.

Understanding the process of tumorigenesis and tumor growth requires a systemic and dynamical description of the disease, based on mathematical modeling. At the molecular level, this can be tackled by a simplified mechanistic cell-wide model of protein interactions in the underlying pathways, dependent on external environmental signals.

Although continuous mathematical modeling has been widely used to study cellular biochemistry dynamics (e.g., ordinary differential equations), this formalism faces limits for modeling a large-scale signaling network, due to the difficulty of estimating kinetic parameter values. In contrast, the logical modeling formalism represents a convenient mean of abstraction where the causal relationships between proteins (or genes) are encoded with logic statements, and dynamical behaviours are represented by transitions between discrete states of the system. This framework is flexible, requires in principle no quantitative information, and can be hence applied to large networks combining multiple pathways. It can also provide a qualitative understanding of molecular systems lacking mechanistic detailed information.

Numerous uses of logical modeling so far have shown that this framework is able to delineate the main dynamical properties of complex biological regulatory networks (Fauré et al., 2006; Abou-jaoudé et al., 2011; Grieco et al., 2013), even with purely Boolean systems.

In particular, Boolean models have been used to describe and predict the behavior of molecular networks affected in human disease (Fumia et al., 2013; Arshad et al., 2016).

A logical model is defined by a regulatory graph, where each node represents a regulatory component, and is associated with discrete levels of activity (0, 1 and further integers when justified). Each arc represents a regulatory interaction between the source and target nodes, and is labelled with a threshold and a sign (positive or negative). The model is completed by logical rules (or functions), which assign a target value to each node for each regulator level combination. The resulting dynamics can be represented in terms of a state transition graph (STG), where the nodes denote the states of the system (i.e. vectors giving the levels of activity of all the variables) and the arcs represent state transitions (i.e. changes in variable values, according to the corresponding logical functions) (for more details, see Chaouiya *et al.*, 2012; Abou-Jaoudé *et al.*, 2016).

When concurrent variable changes are enabled at a given state, the resulting state transition depends on the chosen updating assumption. Numerous studies use the simple fully synchronous strategy where all variables are updated through a unique transition. This assumption leads to relatively simple transition graphs and deterministic dynamics. The proportion of initial states leading to given attractors is measured as the attractor landscape (Helikar et al., 2008; Fumia et al., 2013; Cho et al., 2016).

However, the synchronous updating assumption approximation often leads to spurious cyclic attractors.

On the other hand, the fully asynchronous updating assumption considers separately all possible transitions and therefore allows the consideration of alternative dynamics in the absence of kinetic data. The resulting dynamics has a branching structure which makes it more difficult to evaluate.

In this project, we consider asynchronous dynamics mixed with stochastic simulations.

The regulatory graph was constructed using GINsim software (Chaouiya et al., 2012), and then exported in a format readable by MaBoSS software (see below) in order to perform some stochastic simulations on the Boolean model.

1.2 Stochastic simulations

We use the software MaBoSS to compute Continuous Markov Chain simulations on the Boolean network. MaBoSS uses a specific language for associating transition rates to each node, enabling to account for different time scales of the processes described by the model. Given some initial conditions, MaBoSS applies Monte-Carlo kinetic algorithm (or Gillespie algorithm) to the network to produce time trajectories (Stoll et al., 2012). Time evolution of probabilities are estimated.

Stochastic simulations allow to vary probabilities for initial states and inputs, and to measure the effect on the output probabilities. This provides a continuous intensity for each Boolean node.

In addition, fixed points are computed, as well as entropy characterizations of the whole system, which is useful to indicate oscillating systems. Indeed, stochastic trajectories that have reached a complex attractor are progressively desynchronized until the mean probabilities of each node stabilizes at its mean level over the attractor. This phenomenon is differentiated from a fixed point by a non-null local entropy.

Stochastic simulations with MaBoSS have already been successfully applied to study several logical models (Calzone et al., 2010; Remy et al., 2015; Cohen et al., 2015).

As mentioned previously, GINsim includes a functionality enabling the export of logical models into MaBoSS format.

1.3 Model perturbations

Model variants representing biologically plausible perturbations

Perturbations can be introduced in the model in order to predict the cell behaviour in altered conditions. In particular, mutations are modelled with perturbations that affect the level of a node. The amplification of a gene is translated by a forced level of 1 for the corresponding node, while a loss-of-function is translated by a forced level of 0. Testing attractor's stabilities against mutations with simulations show the effect they have on the cell behaviour, and allow to determine driver mutations that promote phenotypic transitions.

Multiple perturbations can also be considered, and can represent cell lines or patient-specific mutation profiles.

Different perturbed versions of the model can also show how cell phenotypes can evolve towards full malignancy through distinct sequences of accumulated mutations, and can predict whether the order in such sequences is important.

Finally, perturbations can also be used to model drug treatments. Thus, logic models can be used to simulate the effect of therapeutic interventions, and hence predict the expected efficacy of candidate drugs on different genomic backgrounds.

1.4 Model reduction

In this project, we begin by building a model based on a large network, encompassing all relevant nodes and pathways for prostate cancer. Stochastic simulations allow to analyse easily the dynamics in spite of the size and complexity of the system.

To perform a detailed analysis of the asynchronous dynamics while avoiding the combinatorial explosion of the number of states to consider, it is possible to reduce the number of components in a logical model while preserving the relevant dynamical properties of the master model. The dynamics of the resulting reduced model are encoded in a transition graph of smaller size, which is more amenable to a detailed dynamical analysis.

GINsim contains a function to automatically reduce a model. Only the key players can be selected and GINsim provides an equivalent dynamical model that is able to reproduce the same solutions as the full model.

Chapter 2 Model building

Building the model is done in three steps :

1- identifying signalling pathways or particular genes and proteins that are especially relevant to describe the prostate cancer tumorigenesis and tumor growth. Most of them are components that are known to be frequently altered in cancers.

2 - building a regulatory network that includes simplified representations of pathways identified as relevant for prostate cancer, as well as all individually identified genes. Each pathway is characterized by the key players that regulate it. This network takes the form of a directed graph for which positive and negative influences between components are represented.

3 - From this network, a logical model is derived describing the network dynamics in specific contexts (dependent on initial conditions or perturbations). To this end, logical rules are associated to each node of the network to indicate how it is activated or inhibited by different combinations of its regulators.

2.1 Existing logical models

In the last decade, logical modeling has successfully been used to describe the dynamics of human cellular signal transduction and gene regulation (Helikar et al., 2008; Calzone et al., 2010; Grieco et al., 2013; Flobak et al., 2015; Cho et al., 2016; Traynard et al., 2016), and their deregulation in cancer (Fumia et al., 2013; Hu et al., 2015).

Some of these already published models are available in SBML, the standard format for systems biology, in model repositories such as BioModels (Juty et al., 2015).

We combine published logical models of human signalling networks, and in particular the Boolean network model published in (Fumia et al., 2013), which is based on integrated experimental evidence of signal transduction. This model integrates major signaling pathways that have a role in regulating cell death and proliferation in many tumors. They include those involving receptor tyrosine kinase (RTKs), phosphatidylinositol 3-kinase (PI3K)/AKT, WNT/b-Catenin, transforming growth factor-b (TGF-b)/Smads, cyclins, retinoblastoma protein (Rb), hypoxia-inducible transcription factor (HIF-1), p53 and ataxia-telangiectasia mutated (ATM)/ataxia-telangiectasia and Rad3-related (ATR) protein kinases. The pathways reveal substantial cross-talks.

This initial generic network was then extended to include prostate-cancer-specific genes and proteins using several approaches presented below.

2.2 Inputs and outputs

The model aims at predicting phenotypic behaviours for healthy or cancer cells in different conditions. This amounts to assessing the reachability of phenotypes starting from ‘‘physiological’’ initial conditions. These conditions are encoded by input nodes, with no regulation but whose values are fixed for each simulation, and represent the cell’s microenvironmental characteristics.

Distinct cell phenotypes can be inferred from specific protein activities in the network attractors (for example, the activation of caspases correspond to apoptotic conditions). For

simplicity, we choose to clearly define phenotype variables as output nodes. This allows to integrate multiple phenotypic signals and obtain a 0/1 value for each phenotype. We define three main phenotypes representing the growing status of the cell: Proliferation, Apoptosis, and Quiescence. Apoptosis is activated by Caspase 8 or Caspase 9, while Proliferation is activated by either cyclin in the cell cycle. Before further refinements of the model, we define Quiescence as the absence of Proliferation and Apoptosis. This definition of phenotypes will improve in the future versions of the model. We already note that Proliferation and Apoptosis, although not directly linked, appear always mutually exclusive in simulations, which hints at a correct regulation of these nodes.

The proliferation output is sometimes described in already published models as specific stationary protein activation patterns, namely the following sequence of activation of cyclins: Cyclin D, then Cyclin E, then Cyclin A, then Cyclin B. This sequence can easily be detected in complex attractors in synchronous dynamics. However, since we prefer asynchronous dynamics for which it is more difficult to analyze complex attractors, we define Proliferation as activated by either of the four cyclins. Transient dynamics in MaBoSS simulations allow to check the correct oscillation of cyclins (see Section 3.1).

Moreover, we define several phenotypic outputs that are not mutually exclusive but merely detect the activation of some markers of cancer hallmarks: angiogenesis, epithelial-mesenchymal transition (EMT), bone metastasis, DNA repair, migration, and glycolysis.

2.3 Identification of new components based on literature search

Several studies have focused on identifying main subtypes among the heterogeneous molecular abnormalities in prostate cancer.

In particular, a recent TCGA study (Abeshouse et al., 2015) reported a comprehensive molecular analysis of 333 primary prostate carcinomas. Seven subtypes, containing 74% of these tumors, were defined by specific gene fusions (ERG, ETV1/4, and FLI1) or mutations (SPOP, FOXA1, and IDH1). Epigenetic profiles allowed to identify a methylator phenotype in the IDH1 mutant subset. SPOP and FOXA1 mutant tumors show the highest levels of AR-induced transcripts. Lesion in the PI3K or MAPK signaling pathways are observed in 25% of the prostate cancers, and DNA repair genes inactivation in 19%.

The following list of frequently mutated genes extracted from this study indicate components that could be included in the model, provided that enough information is available on their mechanistic roles:

- gene fusion: ERG, ETV1, ETV4, FLI1
- deletions: SPOP, FOXA1, IDH1, TP53, PTEN, PIK3CA, BRAF, CTNNB1, HRAS, MED12, ATM, CDKN1B, RB1, NKX3-1, AKT1, ZMYM3, KMT2C, KMT2D, ZNF770, CHD1, BRCA2, CDK12, SPINK1
- amplifications: CCND1, MYC, FGFR1, WHSC1L1.

Comparing with a recently published cohort of 150 castration-resistant metastatic prostate cancer samples (Robinson et al., 2015), the authors find a similar subtype distribution as in Abeshouse et al., with increased alteration rates in the metastatic samples, and more frequent amplification or mutation of AR, as well as DNA repair and PI3K pathway alterations.

Other studies such as (Altieri et al, 2009) focus on the role of specific pathways which play a critical role in prostate cancer maintenance, such as chaperone-mediated mitochondrial homeostasis (in particular with HSP90 found very abundant in prostate cancer), integrin-dependent cell signaling, and RUNX2-regulated gene expression in the metastatic bone microenvironment.

Notably, a set of regulatory maps of signalling pathway maps and altered circuitries of various cell biological events associated with the pathogenesis of human prostate cancer have been published recently (Datta et al., 2016). The authors manually constructed networks based on the literature. These networks constitute an important resource for retrieving information on prostate cancer specific components. Although not exhaustive, these maps are synthetic pictures of the existing knowledge on molecular events involved in prostate cancer hallmarks.

The covered hallmarks include:

(1) classical cancer hallmarks: insensitivity to anti-growth signal, self-sufficiency in growth signal, tumor promoting inflammation, genome instability, mutation and perturbation, angiogenesis, metastasis, cell death resistance, metabolic reprogramming, avoidance of immune destruction, enabling replicative immortality, tumour microenvironment; and (2) prostate cancer specific hallmarks: androgen receptor signalling, androgen independence, castration resistance.

This study points toward some candidate nodes to extend our network in order to take into account, at least in a simplified way, most pathways present in the maps. In particular, it shows that the initial network obtained through combinations of published models ignore any pathways related to inflammation, metabolism, immune evasion, or the tumor microenvironment.

However, the resource contains few mechanistic details for the interactions between its components, which are a mix of genes, proteins, molecules, processes and phenotypes.

Finally, among all these genes associated with prostate cancer, a subset has been chosen inside PrECISE for full exon sequencing: AR, PTEN, SPOP, TP53, EZH2, FOXA1, BRCA1, BRCA2, PIK3CA, AKT1, NCOA2, NCOR1, NCOR2, EP300, MYC, RB1, CHD1, CDKN1B, MED12, ZNF595, HOXB13. This list should be included in the model in priority, in order to be able to use the corresponding data.

2.4 Identification of new components based on data analysis

ROMA (Martignetti et al., 2016) is a software package written in Java for the quantification and representation of biological module activity using expression data. It uses the first principal component of a PCA analysis to summarize the coexpression of a group of genes in the gene set.

We apply ROMA analysis on the proteomic data produced on samples from 39 patients. We define gene sets as they are described in the atlas of cancer signaling networks, ACSN (www.acsn.curie.fr). ACSN is centered on signaling pathways such as DNA repair, cell death, EMT, cell adhesion, cell cycle, etc.

Using ROMA, we are able to identify some pathways significantly overdispersed over the samples, that should have relevant roles in prostate cancer and need therefore to be correctly described in the model.

The results (see Figure 1) show 21 modules that reveal a high variance of protein expression across all samples. The apoptotic pathway seems to show a progressive activation from normal to high grade tumors, so does cell adhesion pathway, whereas the MAPK, or the PI3K pathways show opposite behavior.

Therefore, ROMA provides some hints on where to extend the network to fully grab the alterations that are found in prostate cancer patients. The Hedgehog pathway was not described in the already published logical models that we used as a starting point of this model. Moreover, the DNA repair was overly simplified.

Further analysis of the proteins contributing to each module variance reveals that HSP90 (with its different forms HSP90AA1, HSP90AB1, HSP90B1) contribute highly to the variance of 6 of these 21 modules. Other proteins that contribute to several modules and could therefore have a particular relevance include YWHAB/YWHAG/YWHAQ, H2AFX, and laminins (LAMA2, LAMA4, LAMA5, LAMB1, LAMB2, LAMC1). Their role in prostate cancer regulation should be further explored before including them in the network.

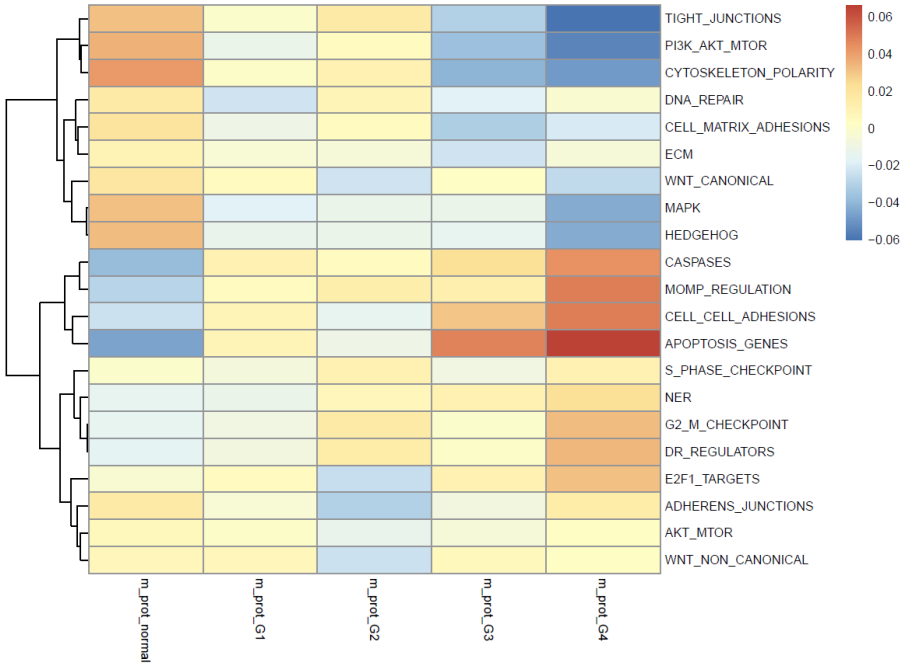

Figure 1: Mean activities by samples subgroups for gene modules defined from pathways described in ACSN and significantly overdispersed over all samples. Blue indicates low pathway activity, red indicates high pathway activity.

2.5 Model extension with Omnipath via pypath

OmniPath is developed by one of the partners of PrECISE. It is a comprehensive collection of high confidence, literature curated, human signaling pathways. It is accompanied and developed together with Pypath, a Python module for cellular signaling pathways analysis.

Pypath is a python module used to query the content of Omnipath in order to retrieve components and interactions in the human protein-protein signaling network associated with annotations, especially sources, literature references, direction, effect signs (stimulation/inhibition) and enzyme-substrate interactions.

The development of pypath allows to build sophisticated and personalised queries. For instance, existing interaction paths between a protein of interest and a list of user-defined proteins can be found, with a given size for the paths. We use this in the extension process of our network to automatically find new interactions between a new gene and the genes already included in the network. We filter the interactions found to select the ones for which the direction and sign are known.

For example, when extending the network with the chaperone protein HSP90AA1, we generate the graph displayed in Figure 2, which shows all signed directed interactions linking HSP90AA1 to the network. The associated references given as annotations are useful to check the mechanism behind each interaction and manually infer a logical rule.

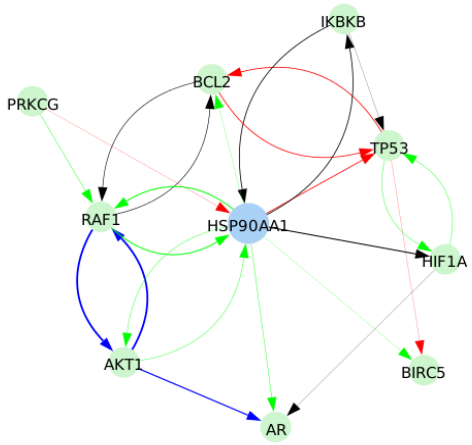

Figure 2: Signed directed interactions between HSP90AA1 and nodes already taken into account in the model.

2.6 Model extension with the literature

PPI (protein-protein interactions) and signaling databases are useful to find quickly established interactions between genes and proteins. However, they are not exhaustive and in particular they often lack recent findings. It is therefore necessary to rely on manual literature search to find information on specific prostate cancer components.

The roles of the fusion gene TMPRSS2:ERG and the tumor suppressor NKX3-1 are examples where the information from databases retrieved from Omnipath or PPI databases is lacking, and for which we found additional information from the literature.

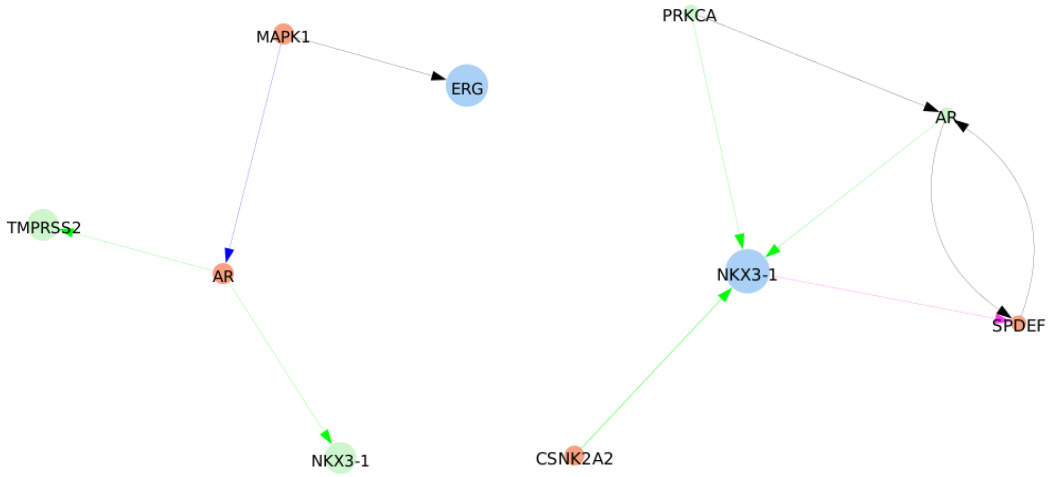

Figure 3: shortest paths found between ERG and TMPRSS2 or NKX3-1 by Pypath: no direct interaction is found.

Fusion genes are frequently found in human prostate cancer, and have been identified as a specific subtype marker (Abeshouse, 2015). The most frequent is TMPRSS2:ERG. It involves the transcription factor ERG, which leads to cell-cycle progression. ERG fuses with the AR-regulated TMPRSS2 gene promoter to form an oncogenic fusion gene that is especially common in hormone-refractory prostate cancer, conferring androgen responsiveness to ERG. This fusion is not found with Pypath, nor is any target of ERG (see Figure 3). However, literature search reveals that ERG directly regulates EZH2, oncogene c-Myc and tumor suppressor NKX3-1 and many other targets in prostate cancer (Kunderfranco et al., 2010).

We model the gene fusion with an activation of ERG by the logical product AR & TMPRSS2. In the wild-type case, TMPRSS2 is fixed to the level 0. The occurrence of the gene fusion is represented with the model perturbation where TMPRSS2 is fixed to 1.

Moreover, it turns out that ERG expression has a major impact on cell invasion and epithelial–mesenchymal transition (EMT) through the upregulation of the *FZD4* gene, a member of the frizzled family of receptors. In our model, we choose for simplicity to consider ERG as a marker of EMT, with a direct activation of the output node EMT by ERG. (Adamo et al., 2016)

NKX3-1 has been identified as a tumor suppressor for prostate cancer. Since it is frequently mutated, it should be included in the model. Some of its regulations can be found with Pypath (see Figure 4), in particular its activation by AR and PKC. However, its role is not identified. The literature search highlighted its role in accelerating the DNA repair response and in particular in avoiding the gene fusion TMPRSS2:ERG. NKX3-1 binds to AR at the ERG gene breakpoint and inhibits both the juxtaposition of the TMPRSS2 and ERG gene loci and also their recombination, by influencing the recruitment of proteins that promote homology-directed DNA repair. Thus, loss of NKX3-1 favors recruitment to the ERG gene breakpoint of proteins that promote error-prone non-homologous end-joining (Bowen et al., 2015).

We therefore add the absence of the node NKX3-1 as a new requirement for the activation of ERG by AR and TMPRSS2 in the model. The effect of the gene fusion can be seen in combination with the perturbation that maintains NKX3-1 to the null level.

In contrast with these examples where some knowledge can be retrieved from the literature, some new nodes can not be included in the model in a satisfactory manner, because of missing information about their regulation or role. High-throughput studies have allowed to identify genes with mutations or expression associated with prostate cancer progression or prognosis. But for many of them, the precise mechanisms behind this association remains to be elucidated.

For example, IDH1 (isocitrate dehydrogenase 1) exhibits a recurrent mutation in 1% of primary prostate cancers that defines a specific subtype (Abeshouse et al., 2015). This mutant status is associated with a DNA hypermethylation phenotype. Despite a lack of detailed mechanisms linking this gene to the regulation network, we can still reflect this association in the model by including IDH1 as a non-regulated gene, whose absence (level 0) induces the activation of a new output node Hypermethylation. The regulation of both new nodes IDH1 and Hypermethylation should be refined when new knowledge is found.

In some cases, we cannot provide any link for a new node, either to an existing node or to a phenotypic output, even qualitatively. For example, ZNF595 has been linked to prostate cancer progression and is therefore going to be sequenced in PrECISE. However, this gene encodes a protein belonging to the Cys2His2 zinc finger protein family, whose members function as transcription factors that can regulate a broad variety of developmental and cellular processes. This knowledge is not detailed enough to add this node in the model yet. However, future mutation data from prostate cancer samples, associated with clinical data, will allow to test several hypothesis.

2.7 Extended model

The extended model contains 130 nodes and 357 interactions. The network is displayed in Figure 4. The model is provided as .bnet file where each row contains a target node and its associated logical rule.

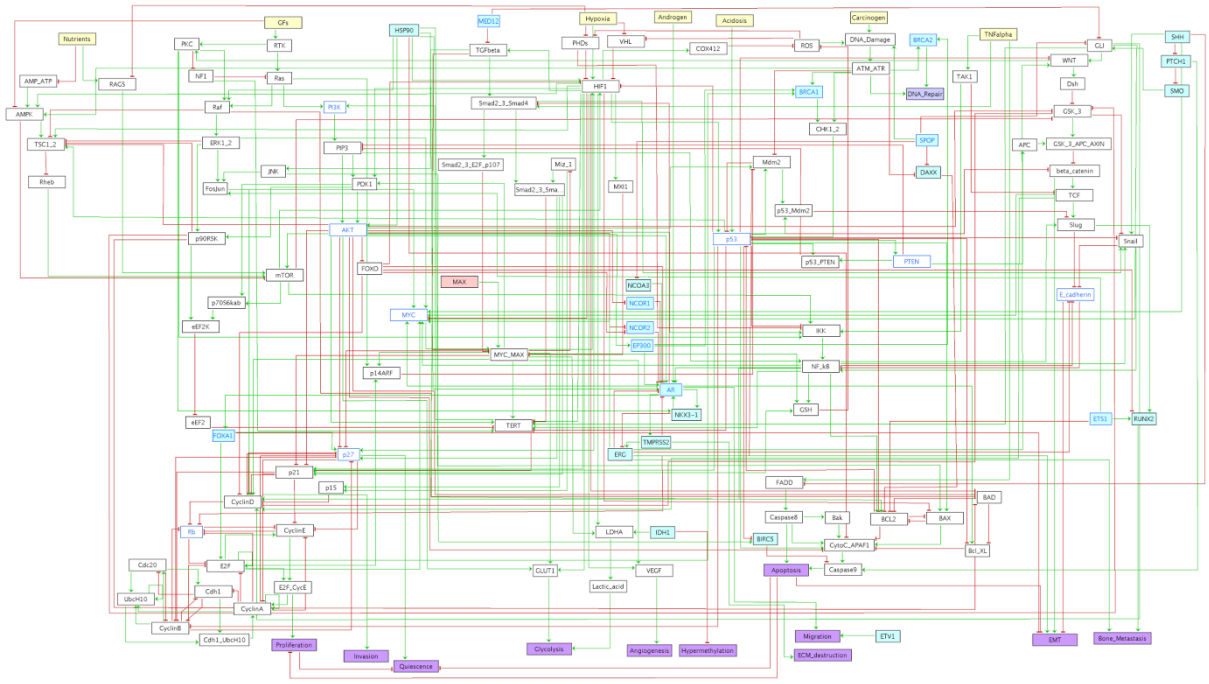

Figure 4: Network of pathways altered in prostate cancer, with input nodes in yellow, output nodes in purple, prostate-cancer-specific nodes in blue.

Chapter 3 Model analysis

3.1 Simulations of healthy cells in different conditions

The wild-type model comprises all pathways involved in prostate cancer, in unaltered state. It can be understood as a model of healthy prostate cells. These cells should exhibit quiescence in absence of inputs. The activation of proliferation should be dependent on the presence of nutrients and growth factors. Cell death factors should trigger full apoptosis, while hypoxia and carcinogen should trigger partial apoptosis. Androgen should help the activation of proliferation. We check that these observations are indeed verified by stochastic simulations in each condition.

Figure 5 shows the state distribution at the end of the simulation in absence of any input, as well as the mean transient probabilities, for the three main outputs (Proliferation, Quiescence, Apoptosis). The results in different conditions are summarized in Table 1. Note that as said in Section 2.2, some outputs are not mutually exclusive, therefore the sum of all output probabilities can be higher than 1.

As mentioned in Chapter 1, in proliferating conditions, transient mean probabilities of the cyclins can be used to check that the order of activations of these nodes in the paths leading to the cyclic attractor is consistent with a correct progression of the cell cycle (see Figure 6).

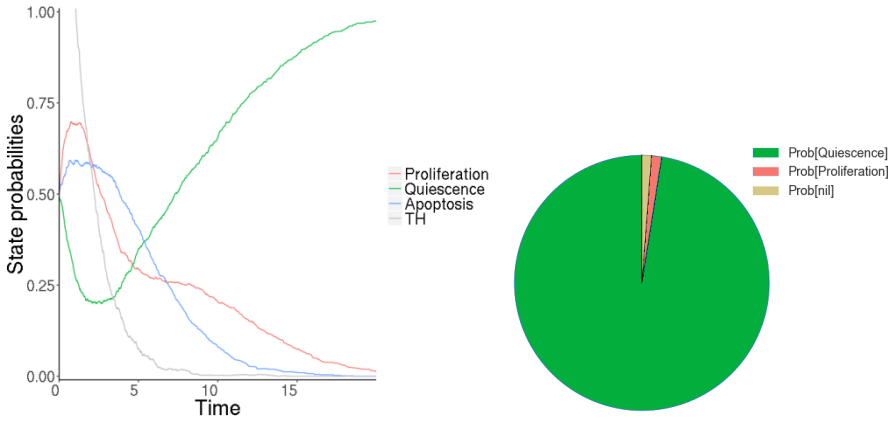

Figure 5: Stochastic simulation in the condition with no input and random initial states (probabilities of 0.5). Measured variables are the levels of the nodes Proliferation, Quiescence, Apoptosis. Left: transient probabilities. Right: state distribution at the end of the simulation.

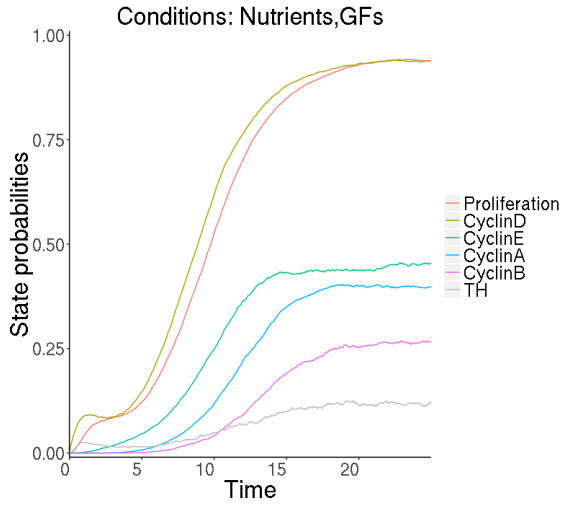

Figure 6: Mean probabilities of the nodes characterizing the cyclins and proliferation, with nutrients and growth factors as inputs. We choose initial states for the nodes involved in the cell cycle that correspond to quiescence (cyclins OFF, cell cycle inhibitors Rb and p27 ON), in order to visualize the order of activation of the cyclins: first Cyclin D, then Cyclin E, Cyclin A and finally Cyclin B. The mean probabilities reach asymptotic levels because of the desynchronization of stochastic trajectories in the population. The non-null transition entropy TH indicates a cyclic attractor.

Inputs	**Proliferation**	**Quiescence**	**Apoptosis**	**EMT**	**Bone metastasis**	**Angiogenesis**	**Glycolysis**	**Migration**
No input	0.01	0.98	0	0	0	0	0	0
GFs	0.43	0.57	0	0	0	0	0	0
Nutrients	0.05	0.93	0	0	0	0	0	0
Androgen	0.21	0.77	0	0	0	0	0	0
GFs,Androgen	0.99	0.02	0	0	0	0	0	0
Nutrients,Androgen	0.16	0.82	0	0	0	0	0	0
Nutrients,GFs	0.93	0.08	0	0	0	0.08	0.05	0
Nutrients,GFs, Androgen	0.81	0.19	0	0	0	0.24	0.14	0
Nutrients,GFs, Carcinogen	0.63	0.20	0.20	0	0	0.03	0.04	0
Nutrients,GFs, Carcinogen,Androgen	0.39	0.37	0.29	0	0	0.10	0.07	0
Nutrients,GFs, Hypoxia	0	0.50	0.49	0	0	0	0	0
Nutrients,GFs, Hypoxia,Androgen	0	0.46	0.54	0	0	0	0	0
Nutrients,GFs, TNFalpha	0	0.37	0.63	0	0	0	0	0

Nutrients,GFs, TNFalpha,Androgen	0	0.99	0	0	0	0	0	0

Table 1: Output probabilities for the wild-type model in different input conditions. Partial proliferation is induced by GFs and Androgen, while full Proliferation requires Nutrients and GFs. Carcinogen, Hypoxia and TNFalpha induce partial Apoptosis. Androgen induces low probabilities of Angiogenesis and Glycolysis in some conditions. Values below 0.01 are written as 0.

3.2 Mutation simulations

A mutant in the logical framework is simulated by setting the node corresponding to the gene mutated to 0 in the case of loss of function and to 1 in the case of gain of function.

The effect of a mutation is assessed by comparing the probabilities for reaching a phenotype in the wild type to the probabilities in the mutant conditions.

Therefore, mutations should either decrease or increase the phenotypes: Apoptosis, Quiescence, Proliferation, EMT, angiogenesis, inflammation, Invasion, Migration, etc.

Single mutations

The single mutations of some of the main nodes of the network show some changes in the probabilities of reaching the phenotypes when compared to wild type conditions.

The examples on Figure 7 show that a loss-of-function mutation of FOXA1 in proliferative conditions (nutrients and growth factors) results in the activation of EMT (epithelial-mesenchymal transition). A loss-of-function mutation of TP53 in the same condition with the addition of carcinogen results in the loss of the apoptosis induced by DNA damage.

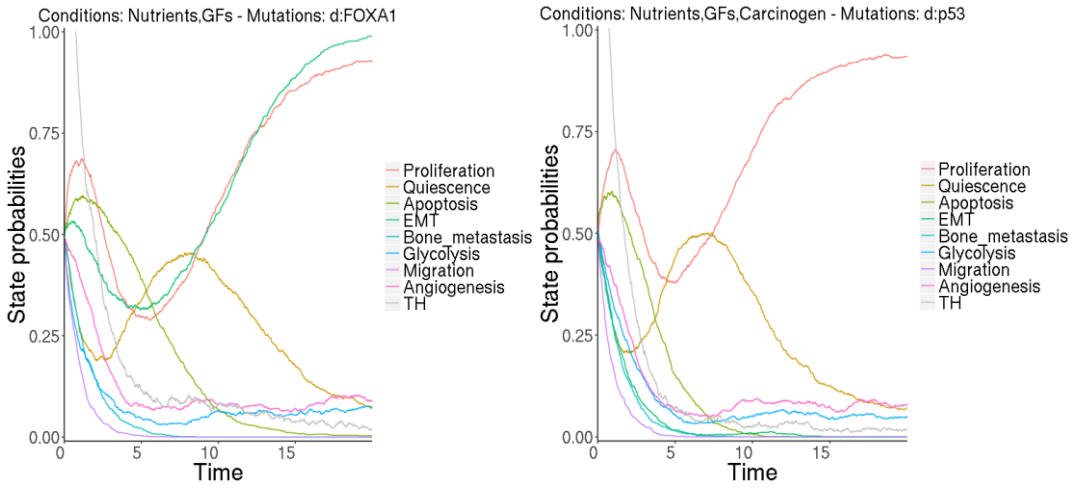

Figure 7: Mean probabilities in simulations of mutated models. Left: loss-of-function mutation of FOXA1. Right: loss-of-function mutation of TP53.

We provide in Table 2 a selection of simulations for nodes contained in the model that have been reported to be frequently mutated in prostate cancer, with their effects on the model outputs in different conditions.

Mutation	Proliferation	Quiescence	Apoptosis	EMT	Bone metastasis	Glycolysis	Migration	Angiogenesis
d:E-cadherin	0.04	0.21	0.76	0.05	0	0	0	0
u:GLI	0.11	0.17	0.71	0	0	0	0	0
u:ETV1	0.05	0.21	0.75	0	0	0	0.58	0
d:FOXA1	0.05	0.22	0.74	0.05	0	0	0	0
u:AKT	0.17	0.26	0.67	0	0	0	0	0.01
u:HSP90	0.18	0.62	0.19	0	0	0	0	0.01
u:ETS1	0.07	0.49	0.44	0	0	0	0	0
d:p53	0.12	0.67	0.21	0	0	0.01	0	0.01

Table 2: Output probabilities for a subset of single perturbations of the model, with random input conditions. “d:” corresponds to inhibition and “u:” to overactivation of the node. Values below 0.01 are written as 0.

Multiple mutations

Cancer progression is characterized by the accumulation of genetic alterations that affect multiple pathways in the signaling network. The logical model allows to easily simulate all possible combinations of mutations and study the potential redundancy or synergy of alteration effects and the importance of order. An example of double mutation is shown in Figure 8, where the combination of the gene fusion TMPRSS2:ERG and the loss-of-function of NKX3-1 activates Bone Metastasis signals in proliferative conditions with androgen induction.

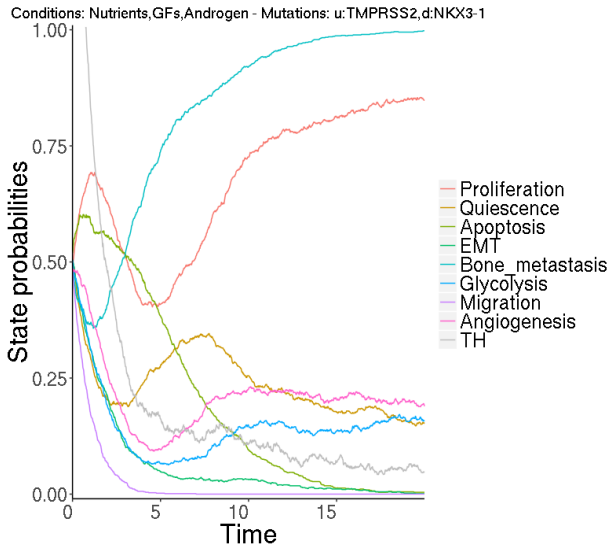

Figure 8: Mean probabilities in simulations of the model with a multiple simulation: the gene fusion TMPRSS2:ERG and a loss-of-function of NKX3-1.

The model allows to study easily all possible associations of mutations to assess synergies or redundancies. It can also reproduce sets of mutations observed in tumors. Different sequences of possible acquired mutations can be simulated and compared to what is already known about patients harboring these mutations.

We can see on Figure 9 an example of a sequence of mutations, showing the relative importance of each new alteration in a genetic profile regarding the probability of each phenotype.

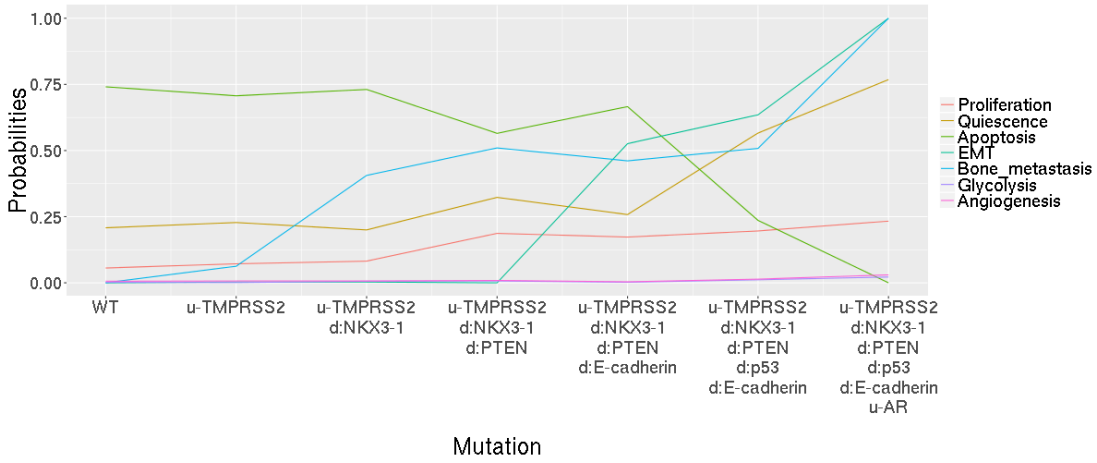

Figure 9: Variation of output probabilities in all conditions along a mutation sequence. “d:” corresponds to inhibition and “u:” to overactivation of the node. The mutation u:TMPRSS2 corresponds to the gene fusion TMPRSS2:ER

Chapter 4 Summary and Conclusion

In conclusion, we have built a logical model of prostate cancer, involving the main signaling pathways altered in prostate cancer cells. Stochastic simulations allow us to obtain quantitative output probabilities summarizing complex asynchronous dynamics. This model can predict a set of phenotypic behaviors for a prostate cell given its microenvironmental condition. The effect of different perturbations of the model on these phenotypes show how genomic alterations initiate and contribute to the different hallmarks composing cancer progression.

We validate the model by verifying the consistency between predicted phenotypes and expected behaviors of different cancer cells.

This generic model constitutes a starting point for the study of prostate cancer inside PrECISE, and will be updated regularly throughout the project. In particular, further extension of the regulatory network will include interactions and pathways identified in other work packages with methods of network inference and data analysis, such as the prostate cancer interactome described in D4.1.

Perspectives for the model have been planned in the project, with the instantiation of data-based patient-specific versions of the model. In addition to patient mutation profiles, normalized protein or mRNA expression from patients will be used as initial probabilities for the nodes present in the model, in order to obtain patient-specific outcome probabilities. Pathway analysis methods such as ROMA can be used to set values for unobserved nodes.

Finally, the Boolean network model will be employed to evaluate the outcome of molecularly targeted cancer therapies. This will allow to orient the selection of ligand and inhibitor targets to generate phosphoprotein data that will be the most instructive. This data will ultimately allow to validate further the model, or correct it to account better for this new data.

Chapter 5 Bibliography

Abeshouse, A., Ahn, J., Akbani, R., Ally, A., Amin, S., Andry, C. D., … Zmuda, E. (2015). The Molecular Taxonomy of Primary Prostate Cancer. *Cell*.

Abou-jaoude, W., Chaves, M., & Gouzé, J.-L. (2011). A Theoretical Exploration of Birhythmicity in the p53-Mdm2 Network. PLoS ONE, 6(2). http://doi.org/10.1371/Citation

Abou-Jaoudé, W., Traynard, P., Monteiro, P. T., Saez-Rodriguez, J., Helikar, T., Thieffry, D., & Chaouiya, C. (2016). Logical Modeling and Dynamical Analysis of Cellular Networks. Frontiers in Genetics, 7, 94.

Adamo, P, Ladomery, MR (2016). The oncogene ERG: a key factor in prostate cancer. *Oncogene*, 35, 4:403-14.

Arshad, O. A., Venkatasubramani, P. S., Datta, A., & Venkatraj, J. (2016). Using Boolean Logic Modeling of Gene Regulatory Networks to Exploit the Links Between Cancer and Metabolism for Therapeutic Purposes. IEEE Journal of Biomedical and Health Informatics, 20(1), 399–407.

Bowen, C, Zheng, T, Gelmann, EP (2015). NKX3.1 Suppresses TMPRSS2-ERG Gene Rearrangement and Mediates Repair of Androgen Receptor-Induced DNA Damage. *Cancer Res.*, 75, 13:2686-98.

Calzone, L., Tournier, L., Fourquet, S., Thieffry, D., Zhivotovsky, B., Barillot, E., & Zinovyev, A. (2010). Mathematical modelling of cell-fate decision in response to death receptor engagement. *PLoS Computational Biology*, *6*(3), e1000702.

Chaouiya, C, Naldi, A, Thieffry, D (2012). Logical modelling of gene regulatory networks with GINsim. *Methods Mol. Biol.*, 804:463-79.

Cho, S.-H., Park, S.-M., Lee, H.-S., Lee, H.-Y., Cho, K.-H., Kolch, W., … Lukas, J. (2016). Attractor landscape analysis of colorectal tumorigenesis and its reversion. *BMC Systems Biology*, *10*(1), 96.

Cohen, D. P. A., Martignetti, L., Robine, S., Barillot, E., Zinovyev, A., & Calzone, L. (2015). Mathematical Modelling of Molecular Pathways Enabling Tumour Cell Invasion and Migration. *PLoS Computational Biology*, 11(11), e1004571.

Datta, D, Aftabuddin, M, Gupta, DK, Raha, S, Sen, P (2016). Human Prostate Cancer Hallmarks Map. *Sci Rep*, 6:30691.

Flobak, Å., Baudot, A., Remy, E., Thommesen, L., Thieffry, D., Kuiper, M., & Lægreid, A. (2015). Discovery of Drug Synergies in Gastric Cancer Cells Predicted by Logical Modeling. *PLOS Computational Biology*, *11*(8), e1004426.

Fumia, H. F., & Martins, M. L. (2013). Boolean Network Model for Cancer Pathways: Predicting Carcinogenesis and Targeted Therapy Outcomes. *PLoS ONE*, *8*(7).

Grieco, L., Calzone, L., Bernard-Pierrot, I., Radvanyi, F., Kahn-Perlès, B., & Thieffry, D. (2013). Integrative modelling of the influence of MAPK network on cancer cell fate decision. *PLoS Computational Biology*, *9*(10), e1003286.

Helikar, T., Konvalina, J., Heidel, J., & Rogers, J. A. (2008). Emergent decision-making in biological signal transduction networks. *Proceedings of the National Academy of Sciences*, *105*(6), 705088105.

Juty, N, Ali, R, Glont, M, Keating, S, Rodriguez, N, Swat, MJ, Wimalaratne, SM, Hermjakob, H, Le Novère, N, Laibe, C, Chelliah, V (2015). BioModels: Content, Features, Functionality, and Use. *CPT Pharmacometrics Syst Pharmacol*, 4, 2:e3.

Kunderfranco, P., et al., (2010) ETS transcription factors control transcription of EZH2 and epigenetic silencing of the tumor suppressor gene Nkx3.1 in prostate cancer. *PLoS One*,5 (5): p. e10547

Martignetti, L., Calzone, L., Bonnet, E., Barillot, E., & Zinovyev, A. (2016). ROMA: Representation and Quantification of Module Activity from Target Expression Data. Frontiers in Genetics, 7, 18.

Remy, E., Rebouissou, S., Chaouiya, C., Zinovyev, A., Radvanyi, F., Calzone, L., … Network, A. of C. S. (2015). A Modeling Approach to Explain Mutually Exclusive and Co-Occurring Genetic Alterations in Bladder Tumorigenesis. *Cancer Research*, 75(19), 4042–52.

Stoll, G., Viara, E., Barillot, E., & Calzone, L. (2012). Continuous time Boolean modeling for biological signaling: application of Gillespie algorithm. *BMC Systems Biology*, *6*(1), 116.