Deliverable No. 6.3

Initial Standardized Cancer Hypermodels

Dissemination Level		
PU	Public	
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	**X**

COVER AND CONTROL PAGE OF DOCUMENT

Project Acronym:	**CHIC**
Project Full Name:	Computational Horizons In Cancer (CHIC): Developing Meta- and Hyper-Multiscale Models and Repositories for *In Silico* Oncology
Deliverable No.:	D6.3
Document name:	Initial Standardized Cancer Hypermodels
Nature (R, P, D, O)[1]	R
Dissemination Level (PU, PP, RE, CO)[2]	CO
Version:	1 (Final)
Actual Submission Date:	31/07/2016
Editor: Institution: E-Mail:	Georgios S. Stamatakos Institute of Communication and Computer Systems (ICCS), National Technical University of Athens (NTUA), *In Silico* Oncology and *In Silico* Medicine Group gestam@central.ntua.gr

ABSTRACT

This document outlines the initial standardized versions of the novel cancer hypermodels developed by the CHIC project concerning nephroblastoma, non small cell lung cancer (NSCLC), glioblastoma multiforme, and prostate cancer. The constituting hypomodels as well as their orchestration and links are also described. Mechanistic spatiotemporal multiscale modelling and various statistics based machine learning methods constitute the major strategies exploited for the development of hypomodels and hypermodels. Concrete clinical questions dictate the precise mathematical and computational strategy to be adopted for each case. Closely related technological procedures and components such as the process of semantically annotating hypomodels and hypermodels, the model repository, the Clinical Research Application Framework (CRAF), the Hypermodelling Editor and the procedure for the execution of hypermodels are briefly outlined in the context of the hypermodel integration process. Initial prediction results, some of which demonstrated during a number of project reviews, are also included. Comments on various aspects of the initial standardized versions of the hypermodels provide further insight into the complex process of cancer hypemordelling. The contents of the deliverable suggest that the whole process of the development of the initial standardized versions of the CHIC hypemodels as well as their handling and execution has been a successful one.

KEYWORD LIST:

cancer modelling, multiscale cancer modelling, hypermodelling, CHIC project, hypermodel, hypomodel, component model, tumour growth, angiogenesis, cancer biomechanics, cancer metabolism, molecular cancer modelling, integrated cancer model, oncosimulator, hypermodel based oncosimulator, *in silico* oncology, *in silico* medicine, systems medicine, computational oncology, computational medicine, systems oncology, finite element method, discrete event, discrete entity, partial differential equation, ordinary differential equation, semantics, model repository, clinical research application framework, CRAF, hypermodelling editor, hypermodel execution, nephroblastoma, non small cell lung cancer, glioblastoma multiforme, prostate cancer

---

[1] **R**=Report, **P**=Prototype, **D**=Demonstrator, **O**=Other

[2] **PU**=Public, **PP**=Restricted to other programme participants (including the Commission Services), **RE**=Restricted to a group specified by the consortium (including the Commission Services), **CO**=Confidential, only for members of the consortium (including the Commission Services)

MODIFICATION CONTROL			
Version	**Date**	**Status**	**Author**
1.0	2/7/2016	Draft	G. Stamatakos, ICCS-NTUA
2.0	9/7/2016	Draft	G. Stamatakos, ICCS-NTUA
3.0	15/7/2016	Draft	G. Stamatakos, ICCS-NTUA
4.0	25/7/2016	Draft	G. Stamatakos, ICCS-NTUA
5.0	30/7/2016	Revision	G. Stamatakos, ICCS-NTUA
6.0	31/7/2016	Clinical Check	N. Graf, USAAR
7.0→ 1 (Final)	31/7/2016	Revision	G. Stamatakos, ICCS-NTUA

List of contributors

G. Stamatakos, ICCS-NTUA
E. Georgiadi, ICCS-NTUA
E. Kolokotroni, ICCS-NTUA
M. Antonopoulos, ICCS-NTUA
N. Tousert, ICCS-NTUA
D. Dioynsiou, ICCS-NTUA
D. Abler, UBERN
P. Büchler, UBERN
J. A. Grogan, UOXF
H. M. Byrne, UOXF
V. Sakkalis, FORTH
E. Tzamali, FORTH
S. Sfakianakis, FORTH
I. Karatzanis, FORTH
K. Marias, FORTH
R. Radhakrishnan, UPENN
A. Ghosh, UPENN
I. Stura, UNITO
C. Guiot, UNITO
P.Grenon, UCL
S. Bnà, CINECA
N. McFarlane, BED
N. Graf, USAAR
R. Bohle, USAAR
S. Gool, KUL
L.Solie, KUL

EXECUTIVE SUMMARY

Deliverable D6.3 outlines the initial standardized versions of the novel cancer hypermodels developed by the CHIC project concerning nephroblastoma, non small cell lung cancer (NSCLC), glioblastoma multiforme, and prostate cancer. The constituting hypomodels as well as their orchestration and interaction links are also described. Mechanistic spatiotemporal multiscale modelling and various statistics based machine learning methods constitute the major strategies exploited for the development of hypomodels and hypermodels. Concrete clinical questions dictate the precise mathematical and computational strategy to be adopted for each case. Closely related technological procedures and components such as the process of semantically annotating hypomodels and hypermodels, the model repository, the clinical research application framework (CRAF), the hypermodelling editor and the procedure for the execution of hypermodels are briefly outlined in the context of the hypermodel integration process. Initial prediction results, some of which demonstrated during several project reviews, are also included. Comments on various aspects of the initial standardized versions of the hypermodels provide further insight into the complex process of cancer hypemordelling.

In more detail, the major basic science related components of the document are the following: the nephroblastoma oncosimulator hypomodel and the nephroblastoma integrated hypermodel, the non small cell lung cancer oncosimulator hypomodel and the non small cell lung cancer integrated hypermodel, the biomechanical hypomodel for nephroblastoma and non small cell lung cancer, the angiogenesis hypomodel for nephroblastoma and non small cell lung cancer, the metabolic hypomodel for nephroblastoma and non small cell lung cancer, the molecular hypomodel for nephroblastoma and non small cell lung cancer, the gross phenomenological hypermodel for nephroblastoma and non small cell lung cancer, the glioblastoma hypermodel, and the prostate hypermodel.

The content of the document suggests that the entire process of the development of the initial standardized versions of the CHIC hypemodels and their handling and execution has been a successful one. Extensions of the technological tools and services including *inter alia* the handling of missing or incomplete data and the support of the prospective hypermodeller to create a new hypermodel using the hypermodelling editor will be mainly presented in the appropriate technological deliverables.

Contents

IN. Intoduction

This document outlines the initial standardized versions of the novel cancer hypermodels developed by the CHIC project concerning nephroblastoma, non small cell lung cancer (NSCLC), glioblastoma multiforme and prostate cancer. The constituting hypomodels as well as their orchestration and interaction links are also described. Mechanistic spatiotemporal multiscale modelling and various statistics based machine learning methods constitute the major strategies exploited for the development of both the hypomodels and the hypermodels. Concrete clinical questions dictate the precise mathematical and computational strategy to be adopted for each case. Closely related technological procedures and components such as the process of semantically annotating hypomodels and hypermodels, the model repository, the Clinical Research Application Framework (CRAF), the hypermodelling editor and the procedure for the execution of hypermodels are briefly outlined in the context of the hypermodel integration process. Initial prediction results, some of which demonstrated during several project reviews, are also included. Comments on various aspects of the initial standardized versions of the hypermodels provide further insight into the complex process of cancer hypemordelling.

The document starts with the section entitled "*NO. The nephroblastoma oncosimulator hypomodel and the nephroblastoma integrated hypermodel.*" Its major components include the nephroblastoma oncosimulator component, the hypermodel integration process, the communication design and representative preliminary results. The section entitled "*LO. The non small cell lung cancer oncosimulator hypomodel and the non small cell lung cancer integrated hypermodel*" includes two approaches dictated by pertinent real clinical problems and questions: the mechanistic multiscale non small cell lung cancer oncosimulator and the machine learning based treatment response predictor. The section entitled "*BM. The biomechanical hypomodel for nephroblastoma and non small cell lung cancer (and glioblastoma multiforme)*" deals *inter alia* with the biomechanical simulator component, the mathematical model and scenario-specific adaption and patient personalisation. Section "*AN. The angiogenesis hypomodel for nephroblastoma and non small cell lung cancer*" addresses the corresponding model development, the vasculature hypomodel (nephroblastoma and lung) and high resolution angiogenesis and vascular transport models. Section "*ME. The metabolic hypomodel for nephroblastoma and non small cell lung cancer (and glioblastoma multiforme)*" addresses among other things genome-scale metabolic modelling and the integration of the cancer metabolic hypomodel with vasculature and tumour evolution components. Section "*MO. The molecular hypomodel for nephroblastoma and non small cell lung cancer*" includes the model description, results and suggestions for further work. The section entitled "*GP. the gross phenomenological hypomodel for nephroblastoma and non small cell lung cancer*" deals *inter alia* with a universal phenomenological approach to describe tumour growth and therapy response, the modelling of the response of pre-surgical chemotherapy in nephroblastoma and lung cancer and the integration with other models. Section "*GB. The glioblastoma hypermodel*" includes an outline of the corresponding background and the machine learning based hypemodel structure. Section "*PR. The prostate hypermodel*" includes predictive (hypo-)models for prostate cancer recurrence after surgery, predictive (hypo-) models for prostate cancer recurrence after radiotherapy, and (hypo-) models for evaluating resistance induction by hormonal therapies. The following closely related technological sections are included as well. Section "*SE. A Brief Outline of the Process of Semantically Annotating Hypomodels and Hypermodels*", Section "*MR. A brief outline of the model repository in the context of hypomodel integration*", Section "*CR. the clinical research application framework (CRAF) in the context of utilizing integrated hypermodels*", Section "*HE. The hypermodelling editor (HME) as a supportive technological platform for the integration of hypermodels*" and Section "*EH. Execution of hypermodels*". The document is complemented with a brief *Discussion* and *Conclusions*.

NO. The Nephroblastoma Oncosimulator Hypomodel and the Nephroblastoma Integrated Hypermodel

The development of both the Nephroblastoma (Wilms Tumour) Oncosimulator Hypomodel and the Nephroblastoma Integrated Hypermodel has been led by ICCS.

From the clinical point of view, this hypermodel aims to answer the question whether a given Wilms Tumour patient will respond to pre-operative chemotherapy through tumour shrinkage or not. Such a predictive system, following its clinical adaptation and validation, is meant to serve as a quantitative clinical decision support system fostering treatment optimization in the patient individualized context.

NO1. The Nephroblastoma (Wilms Tumour) Oncosimulator Component

The Nephroblastoma (Wilms Tumour) Oncosimulator developed by ICCS is a hypomodel (component model) simulating critical multiscale biological mechanisms involved in tumour growth and response to (chemotherapy) treatment.

The Nephroblastoma Oncosimulator also acts as the hypermodel integrator and is linked with a vasculature/angiogenesis hypomodel (of which the development has been led by UOXF), a biomechanics hypomodel (of which the development has been led by led by UBERN), a cell kill rate focusing molecular hypomodel (of which the development has been led by UPENN) and a metabolic network hypomodel (of which the development has been led by FORTH). Another parallel simplistic phenomenological model (of which the development has been led by UNITO) serves as a gross semi-quantitative alternative to the nephroblastoma multimodeller hypermodel.

However, by making use of certain plausible approximations, the ICCS nephroblastoma Oncosimulator can also function as a stand alone hypermodel constituting of just one hypomodel. This suggests that the Oncosimulator can bypass all other hypomodels in case of insufficient input data availability. Obviously the Oncosimulator predictions would generally be less accurate in such a scenario.

NO1.1 The Oncosimulator Model

NO1.1.1 Model Description – Basic Science

The Wilms Oncosimulator has been developed by the *In Silico* Oncology and *In Silico* Medicine Group (ICCS-NTUA) [Georgiadi *et al.*, 2012, Stamatakos *et al.*, 2011] and is a predominantly discrete, clinically-oriented multiscale model of tumour response to preoperative combined chemotherapy treatment of actinomycin and vincristine. It is clinical oriented and utilizes the patient's personal imaging, molecular, histological, clinical and treatment data. The following five categories (or "equivalence classes") of cancer cells are considered in the model: stem cells (cells of unlimited mitotic potential), LIMP cells (LImited Mitotic Potential or committed progenitor cells which can perform a limited number of mitoses before terminal differentiation), terminally differentiated cells, apoptotic and necrotic cells. The various cell cycle phases (G1, S, G2, M) and the dormant (G0) phase constitute subclasses in which stem or LIMP cells may reside.

The model incorporates several biological phenomena that take place in the cellular level (Fig. NO1) such as the cycling of proliferating cells through the successive cell cycle phases, the symmetric and asymmetric modes of stem cell division, the terminal differentiation of committed progenitor cells after a number of mitotic divisions, the transition of proliferating cells to the dormant phase due to inadequate supply of oxygen and nutrients, the re-entering of dormant cells into the active cell cycle

due to local restoration of oxygen and nutrient supplies, the cell death through spontaneous apoptosis, the cell death through necrosis (due to prolonged oxygen and nutrients' shortage) and the cell death due to chemotherapy-induced apoptosis. In order to simulate chemotherapy-induced cell death, lethally hit cells are assumed to enter a rudimentary cell cycle leading to apoptotic death. Cell cycle-specific, cell cycle-non specific, cell cycle phase-specific and cell cycle phase-non specific drugs can be simulated, as is detailed in [Georgiadi et al, 2012, Stamatakos et al, 2011]. "Marking" of a cell as hit by the drug is assumed to take place at the instant of drug administration. However, its actual time of death is dictated by the specific drug pharmacokinetics and pharmacodynamics. The numbers of cells hit by the drug are computed through the utilization of the cell kill probability (CKP) parameter (CKP= 1-cell survival fraction), defined as the percentage of lethally hit cells after each drug administration. A diversification of chemotherapeutic resistance between tumour stem and non-stem cells can be easily achieved through the use of different values of the corresponding CKP parameters.

For a relatively short time interval compared to the tumour lifetime (such as the duration of a simulated chemotherapeutic schedule), the various space and time-independent transition rates considered throughout the tumour region are approximately constant and reflect the means of the actual cell category/phase transition rates over the interval.
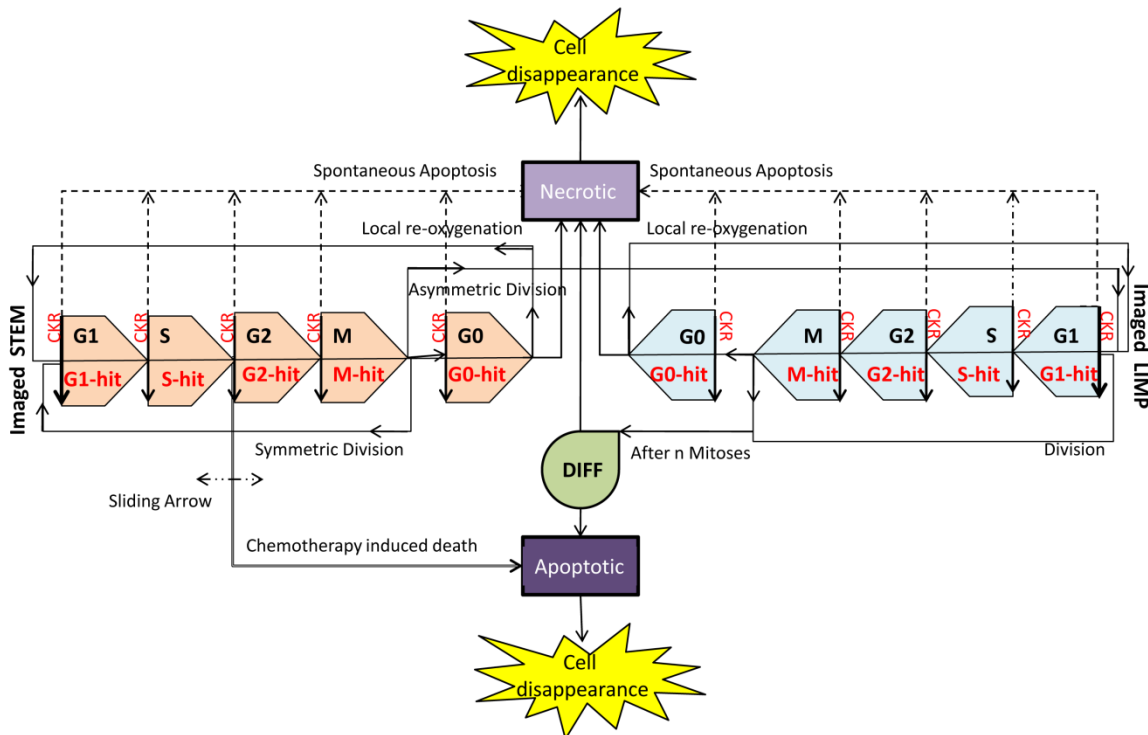
Fig. NO1. Generic cytokinetic model for tumour growth and response to chemotherapy. STEM: stem cells. LIMP: LImited Mitotic Potential cells. DIFF: terminally differentiated cells. G1: Gap 1 phase. S: DNA synthesis phase. G2: Gap 2 phase. M: Mitosis phase. G0: dormant phase. Hit: cells lethally hit by chemotherapy. The arrow indicating chemotherapy-induced death is a sliding arrow, with position dependent on drug pharmacodynamics. For a definition of the depicted model parameters see Table NO1.

NO1.1.2 Model Algorithmic Description

An algorithmic flowchart of the model is given in Fig. NO2. The basic steps are the following:

i. Initialization of the Tumour Shape and Model Parameter Initialization.

The tumour area is segmented using available patient-specific imaging data. After appropriate pre-processing this data is transformed into raw images and serves as input to the model for the spatial initialization of the virtual tumour. A three-dimensional cubic mesh discretizing the regions of interests is considered. The elementary volume of the mesh is called *geometrical cell (GC)*. Each GC belonging to tumour mass is considered occupied. The model supports the division of the tumour area into different metabolic regions (e.g. necrotic and proliferative) based on pertinent imaging data and the handling of each region separately. In this case different values of specific model parameters can be assigned to each region. If macroscopically homogeneous tumours are considered, the same model parameters values apply to all occupied GCs.

ii. Free Growth Condition:

The mass of the tumour is sustained by stem cells. Input parameter values relative to stem cells cycling define the aggressiveness of the tumour. In order to check if the parameter values set defined lead to a biologically accepted tumour, a criterion of monotonic free growth has been applied [Kolokotroni *et al.,* 2011].

iii. Spatial and Temporal Initialization of the Virtual Tumour

The occupied GCs accommodate initially a number of biological cells (NBC) which is defined based on typical solid tumour cell densities (e.g. 10^9 cells/cm³) [Steel, 2002], unless more specific information for a particular tumour is available. The cells initially residing within each GC of the mesh are distributed into the five classes and subclasses mentioned above. The initial distribution of the biological cells into the five classes (stem-LIMP-differentiated and dead) is defined by the model parameter values. The initial distribution of the proliferating cells throughout the cell cycle phases (G1, S, G2, M) is assumed proportional to the corresponding cell cycle phases durations.

In order to ensure a realistic tumour behaviour [Stamatakos *et al.* 2010; Georgiadi *et al.* 2008], an initialization technique has been implemented. Starting with a small number of stem cells occupying one GC, the evolution of this population is simulated for the set values of transition rates and fractions. Gradually, all cell categories and phases become populated and after sufficient time the relative cell categories populations reach an equilibrium state. The cell classes/phases populations are initialized using the corresponding relative populations at the state of equilibrium. The mean time spent by the biological cells of a given equivalence subclass in the same subclass is initialized using a random number generator (0-maximum time of the corresponding phase). The biological cells of the same category and at the same phase are considered synchronized if they belong to the same GC whereas if they belong to different GCs they are considered unsynchronized.

iv. Virtual Tumour Spatiotemporal Evolution:

At each time step the discretizing mesh is scanned and the basic cytokinetic, metabolic, pharmacokinetic/pharmacodynamics and mechanical rules that govern the spatiotemporal evolution of the tumour are applied. Practically, each complete scan can be viewed as consisting of two mesh scans, as described in [Stamatakos *et al.* 2010]. Briefly, the first scan aims at updating the state of each GC, by applying the rules of the cytokinetic model of Figure NO1. The second scan serves to simulate tumour expansion or shrinkage, based on the principle according to which throughout a simulation, the total population of a GC is allowed to fluctuate between a minimum and a maximum value, defined in relation to the initial typical GC cell content. At each time step, checks of each GC total population designate whether the total cell number is above/below the predefined max/min thresholds and, if necessary, specially-designed cell content shifting algorithms "create" or "delete" GCs and thereby lead to tumour expansion or shrinkage, respectively.
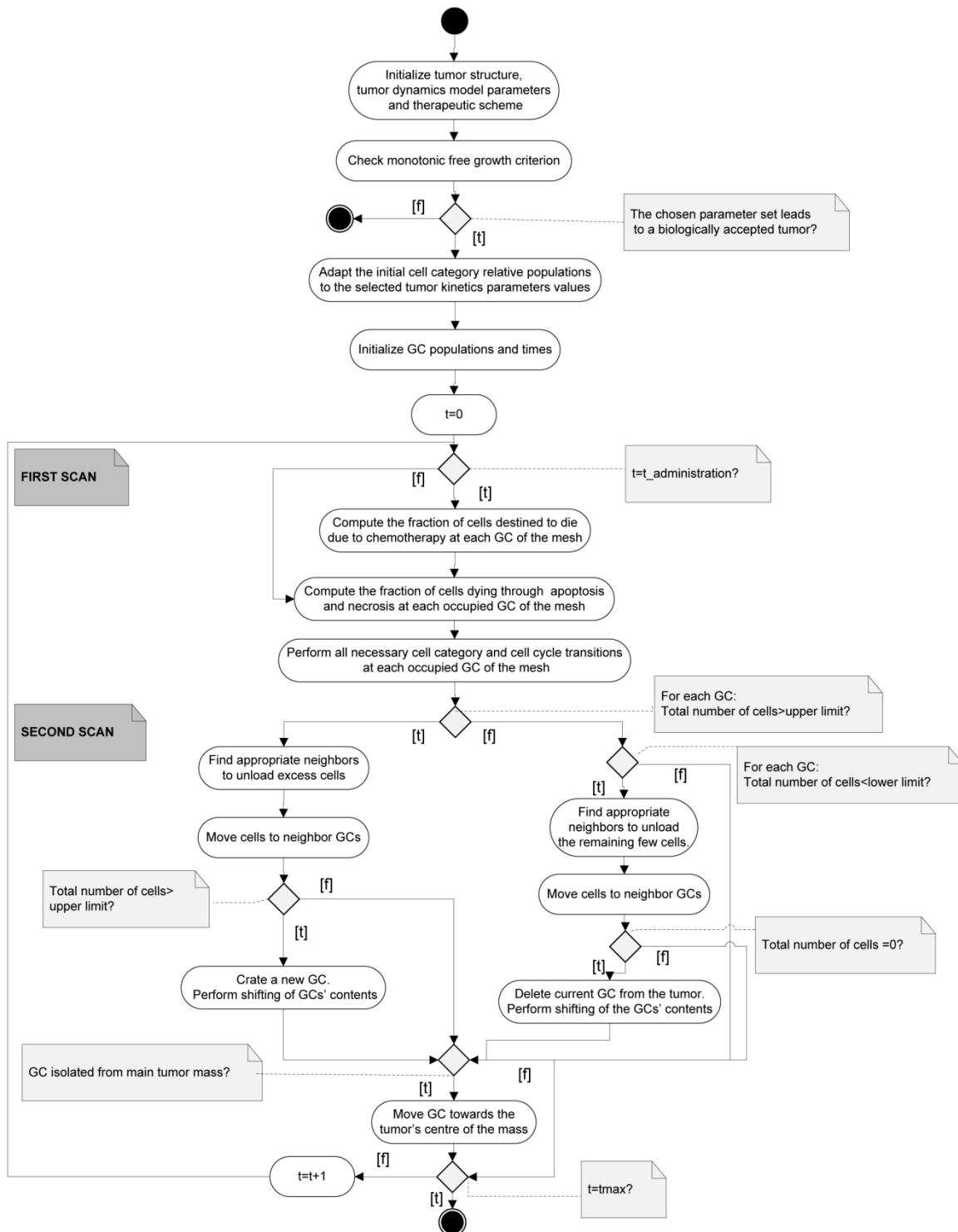
Fig. NO2: Simplified UML (Unified Modeling Language) activity diagram of the simulation procedure. GC: Geometrical Cell. [t]: True, [f]: false

NO1.1.3   Model Parameters: Input – Output

The model has been developed in C++ and is available as a standalone binary. In order to facilitate automatic iterative execution of the model as a component part of a hypermodel, the Wilms oncosimulator is available as MUSCLE [Borgdorff *et al.* 2014]-enabled component model. The input parameters are set through a command line interface.

In table NO1 the "static" input parameters of the model used to simulate the before mentioned biological phenomena are described. The parameters of the model can be categorized into:

- Cell phase durations
- Cell phases transition rates and fractions
- Miscellaneous

Table NO1: Static Input Parameters of the Wilms Tumour Oncosimulator Hypomodel

NAME	DESCRIPTION
Cell phase durations	
Cell cycle duration of stem cells	Cell cycle duration of stem cells through the phases of the active cell cycle (G1, S, G2, M- not including G0 phase)
Cell cycle duration of LIMP cells	Cell cycle duration of limited proliferating potential cells (LIMP) through the phases of the active cell cycle (G1, S, G2, M-not including G0 phase)
G0 phase duration of stem cells	Maximum G0 (dormant) phase duration before stem cells enters necrosis or re-enters G1
G0 phase duration of LIMP cells	Maximum G0 (dormant) phase duration before limp cells enters necrosis or re-enters G1
Necrosis duration in necrotic regions	Time needed for necrosis to be completed and its lysis products to be eliminated from the necrotic regions of the tumour
Necrosis duration in proliferative regions	Time needed for necrosis to be completed and its lysis products to be eliminated from the proliferating regions of the tumour
Apoptosis duration in necrotic regions	Time needed for apoptosis to be completed and its products to be eliminated from the necrotic regions of the tumour
Apoptosis duration in proliferative regions	Time needed for apoptosis to be completed and its products to be eliminated from the proliferative regions of the tumour
Cell phases transition rates and fractions	
Apoptosis rate of stem and LIMP cells	Apoptosis rate of living stem and LIMP tumour cells (fraction of non-differentiated cells dying through apoptosis per unit time).
Apoptosis rate of differentiated cells	Apoptosis rate of differentiated tumour cells per unit time
Necrosis rate of differentiated cells	Fraction of differentiated cells dying through necrosis, per unit time.
Fraction of dormant stem cells re-entering cell cycle	Fraction of the stem cells having just left dormant phase (G0) that re-enter active cell cycle at G1 phase, in the necrotic regions of the tumour.
Fraction of dormant limp cells re-entering cell cycle	Fraction of the limp cells having just left dormant phase (G0) that re-enter active cell cycle at G1 phase, in the proliferative regions of the tumour.
Fraction of cells becoming dormant in necrotic regions	Fraction of cells that enter G0 phase following mitosis in the necrotic regions of the tumour
Fraction of cells becoming dormant in proliferative regions	Fraction of cells that enter G0 phase following mitosis, in the proliferative regions of the tumour
Symmetric division fraction in necrotic regions	Fraction of the stem cells that divide symmetrically, i.e. gives birth to two stem cells, in the necrotic regions of the tumour
Symmetric division fraction in proliferative regions	Fraction of the stem cells that divide symmetrically, i.e. gives birth to two stem

	cells, in the proliferative regions of the tumour
Miscellaneous	
Image raw file	Path (including name) to the input image raw file of the tumour
Image mhd file	Path (including name) to the input image header file of the tumour
N$_{LIMP}$	Number of mitoses performed by LIMP cells before becoming differentiated
Output directory	Path (including name) to the directory where the output files are stored
Execution time	Execution stop time after initialization
Chemotherapy Parameters	
Cell kill rate of vincristine on stem cells	Fraction of stem cells lethally hit by vincristine following bolus drug administration
Cell kill rate of vincristine on limp cells	Fraction of limp cells lethally hit by vincristine following bolus drug administration
Cell kill rate of actinomycin on stem cells	Fraction of stem cells lethally hit by actinomycin following bolus drug administration
Cell kill rate of actinomycin on limp cells	Fraction of limp cells lethally hit by actinomycin following bolus drug administration
First –Sixth combination drug administration time point	Time point after initialization when the 1st -6th administration of drug combination takes place
First – Fourth single drug administration time point	Time point after initialization when the 1st-6th administration of single therapy (e.g. vincristine alone) takes place

The static output parameters of the Wilms OS component model are described in Table NO2. The output files contain all the information needed for the visualization (2D-3D) of the time and space evolution of the tumour as well as values to guide the adaptation/validation of the model.

TABLE NO2: Wilms Tumour Oncosimulator Static Outputs

NAME	DESCRIPTION
tum_evol_file.dat	A dat file containing the time evolution of the various total cell categories populations that comprise the tumour as well as its volume in mm^3, their fractions out of the total population and the number of occupied GCs for each time step. The cell categories classes reported are: stem (live+hit remaining in the tumour bulk), limp (live+hit remaining in the tumour bulk), sum of stem and limp (live+hit remaining in the tumour bulk), differentiated, dead, necrotic, apoptotic, hit cells (remaining in the tumour mass), dormant, proliferating, profiferating stem, limp, diff, dead, dormant, hit, necrotic, apoptotic
tum_doubling_time_file.dat	A dat file contaning the doubling time of the initial tumour
tum_adaptation_file.dat	A dat file listing containing characteristics of simulated initial and final tumours: Fraction of proliferating, dormant, dead and dead population out of the total tumour cell population. It also contains chemo induved relative tumour volume reduction
initial_tumour.dat	A dat file containing Number of biological cells within each voxel of the initial tumour volume
final_tumour.dat	A dat file containing Number of biological cells within each

	voxel of the final tumour volume
oncosim-parameters.log	A log file containing the values assigned to the model input parameters.
Tumour_day_*.raw	Output file of format "raw", which is an array with the size of the discretization mesh and values (255-0) that indicate the state of each GC (Tumour-normal tissue) at a specific time point. These files are used for the 3D visualization of the tumour.

NO1.1.4 Categorization of the Wilms Tumour Oncosimulator Hypomodel According to the CHIC Perspectives for Semantic Annotation

TABLE NO3: The Values of the CHIC Perspectives for the Wilms Tumour Oncosimulator Hypomodel

PERSPECTIVE CODE NUMBER	PERSPECTIVE	MODEL CATEGORY
I	TUMOUR-AFFECTED/ NORMAL TISSUE MODELLING:	✓ **tumour**
II	SPATIAL SCALE(S) OF THE MANIFESTATION OF LIFE:	✓ **cellular** ✓ **tissue** ✓ **body system**
III	TEMPORAL SCALE(S) OF THE MANIFESTATION OF LIFE:	✓ **hours** ✓ **days** ✓ **year**
IV	BIOMECHANISM(S) ADDRESSED:	✓ **cell cycling,** ✓ **apoptosis,** ✓ **necrosis,** ✓ **basic tumour biology**
V	TUMOUR TYPE(S) ADDRESSED:	✓ **Wilms cancer**
VI	TREATMENT MODALITY(-IES) ADDRESSED:	✓ **Chemotherapy**
VII	GENERIC CANCER BIOLOGY/ CLINICALLY DRIVEN CHARACTER OF THE MODELLING APPROACH:	✓ **Clinically driven**
VIII	ORDER OF ADDRESSING DIFFERENT SPATIAL SCALES:	✓ **Top-down** approach
IX	ORDER OF ADDRESSING DIFFERENT TEMPORAL SCALES:	✓ **short periods → longer periods**
X	MECHANISTIC/STATISTICAL CHARACTER OF THE MODELLING APPROACH:	✓ **explicit biological mechanism modeling (e.g. using cytokinetic diagrams)**
XI	DETERMINISTIC/STOCHASTIC CHARACTER OF THE MODELLING APPROACH:	✓ **hybrid**
XII	CONTINUOUS-FINITE-DISCRETE CHARACTER OF THE MATHEMATICS INVOLVED:	✓ **discrete**
XIII	CLOSED FORM SOLUTION / ALGORITHMIC SIMULATION MODELLING APPROACH:	✓ **algorithmic simulation**

NO2. Hypermodel Integration – Communication Design

The Wilms Oncosimulator Hypomodel developed by ICCS acts as the hyper-model integrator with the following additional models (or tools):
- [Image pre-processing tool (FORTH)]
- Molecular (UPENN) hypomodel (component model)
- Vasculature / angiogenesis hypomodel (component model) (UOXF)
- Metabolic network hypomodel (component model) (FORTH)
- Biomechanics hypomodel (component model) (UBERN)

The integration design of the complete hypermodel is depicted in Fig. NO3. Each interaction among the different hypomodels can be either "static" or "dynamic". Static interactions represent exchange of data that happens once, that is, during the initialization of the "target" (destination of data) hypomodel. Dynamic interactions on the other hand are connections that are used multiple times during the execution of both the source and the target hypomodel, i.e. they represent iterative transfer of information at runtime. The static -dynamic input and output parameters exchanged sequentially – iteratively between the hypomodels are presented in Table NO4 and explained in more detail below.
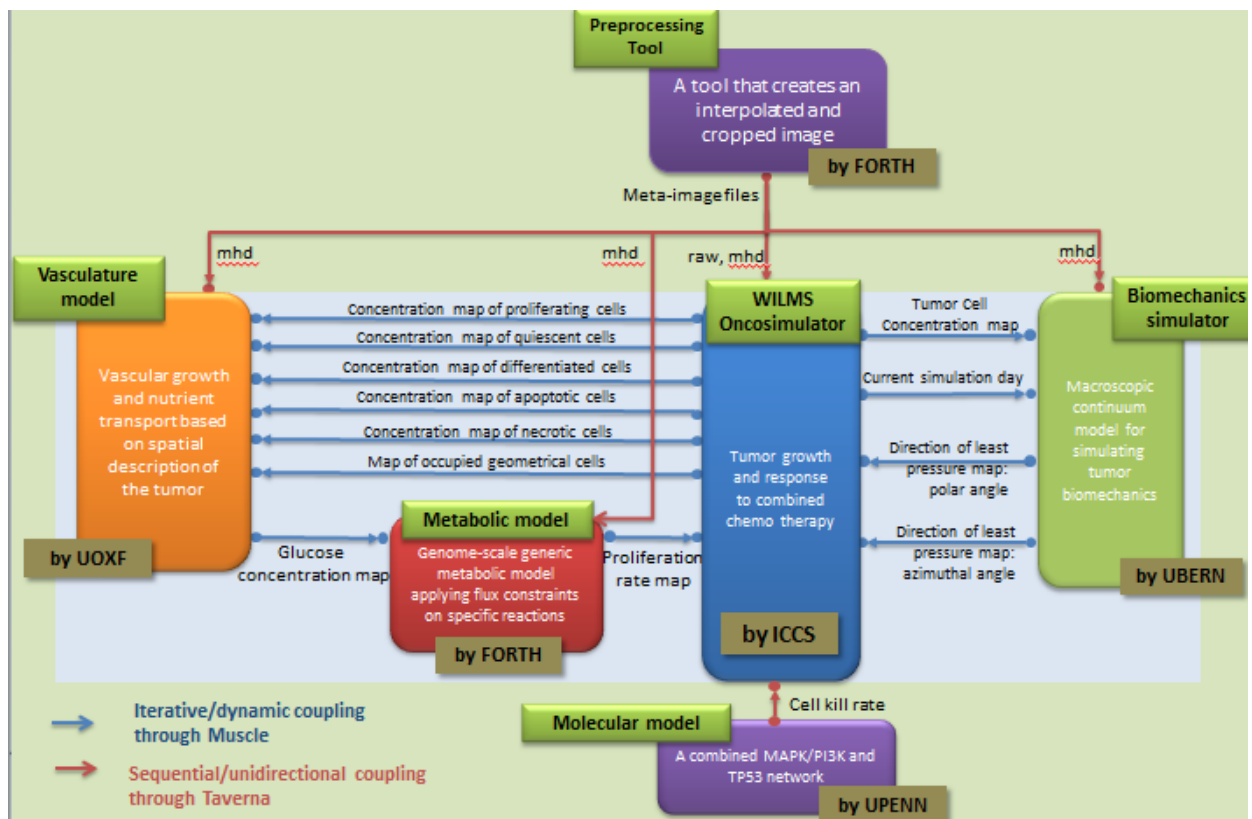

Fig. NO3: Wilms Multimodeler Hypermodel: integration scheme of Wilms oncosimulator, preprocessing tool, metabolic, molecular, vasculature and biomechanic hypomodels.

TABLE NO4: Parameters (static-dynamic) exchanged between hypomodels (sequentially and/or iterativelly) in the integrated Wilms Multimodeler Hypermodel.

From	To	Name	Description	Type
Pre-processing tool	**Oncosimulator**	Image raw file	Input image raw file of the tumour	static
Pre-processing tool	**Oncosimulator** **Metabolic** **Vasculature** **Biomechanical**	Image mhd file	Input image mhd file of the tumour	static
Molecular	**Oncosimulator**	Total cell kill probability	Fraction of cells lethally hit by vincristine and actinomycin following bolus drugs combined administration	static
Oncosimulator	**Vasculature**	Tumour cell population map	The total cell population at each voxel (geometrical cell) of the discretized tumour domain	dynamic
		Proliferating cell population map	Proliferating cell population at each voxel (geometrical cell) of the discretized tumour domain	
		Quiescent cell population map	Quiescent cell population at each voxel (geometrical cell) of the discretized tumour domain	
		Differentiated cell population map	Differentiated cell population at each voxel (GC) of the discretized tumour domain	
		Apoptotic cell population map	Apoptotic cell population at each voxel (GC) of the discretized tumour domain	
		Necrotic cell population map	Necrotic cell population at each voxel (GC) of the discretized tumour	

			domain	
Vasculature	**Metabolic**	Glucose concentration map	Glucose concentration at each voxel (GC) of the discretized tumour domain	dynamic
Metabolic	**Oncosimulator**	Proliferation rate map	Proliferation rate at each voxel (GC) of the discretized tumour domain	dynamic
Oncosimulator	**Biomechanical**	Tumour cell concentration map	Tumour cell concentration at each voxel (GC) of the discretized tumour domain	dynamic
		Simulation day	The simulated time in days	
Biomechanical	**Oncosimulator**	phi map	Phi spherical coordinate of the least pressure direction in each voxel (GC) of the discretized tumour domain	dynamic
		theta map	Theta spherical coordinate of the least pressure direction in each voxel (GC) of the discretized tumour domain	

NO2.1   The Wilms Tumour Oncosimulator Hypomodel and the Pre-processing Tool Integration

The Wilms Tumour OS hypomodel simulates the evolution of real patient– personalized tumours starting from the macroscopic imaging data of the patients. In the context of CHIC project, MRI images (DICOM) of nephroblastoma patients are available. The clinicians provide the segmentation of the tumour area in the form of metaimages (raw, mhd images) through the DrEye segmentation tool [Karatzanis *et al*, 2012; Skounakis *et al*, 2010; Skounakis *et al* 2009] in which the tumour area is marked with white colour (RGB-255) and the normal tissue with black colour (RGB-0). The pre-processing tool resamples the non-isotropic voxels of the metaimages in cubic voxels and crops the images around the area of interest (tumour) with a margin around it (calculated by a factor multiplied by the maximum dimension of the tumour mass) to allow tumour expansion. The Wilms Tumour OS hypomodel and the preprocessing tool communicate sequentially (Fig. NO4). The output of the pre-processing tool (raw image, mhd) is passed as static input parameters to the Wilms Tumour oncosimulator hypomodel through the TAVERNA infrastructure [Wolstencroft *et al*, 2013]. The input image header also serves as a static input to the metabolic, the vasculature and the biomechanic hypomodels.
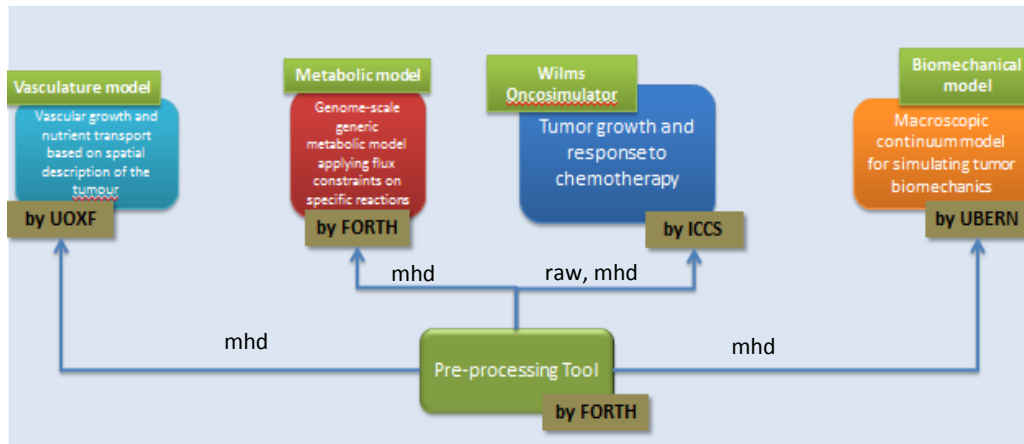
Fig. NO4: Communication scheme of pre-processing tool with the Wilms Tumour Oncosimulator, Metabolic, vasculature and biomechanic hypomodels. Parameter exchanged: metaimages of tumour area (raw, mhd). Type: Sequential coupling.

NO2.2 The Wilms Tumour Oncosimulator and the Molecular Hypomodel Integration

The molecular hypomodel using the mi-RNAs of the nephroblastoma patient estimates the total effect of chemotherapeutic drugs on the tumour cells. The total cell kill probability of combined administration of actinomycin and vincristine is calculated and forms a static output of the molecular model which is used as static input by the Wilms Tumour Oncosimulator hypomodel to simulate the effect of therapy to the tumour mass evolution (Fig. NO5). Based on literature [Groninger *et al.* 2002; Dahl *et al.* 1976; Sawada *et al.* 2005; Veal *et al.* 2005] the following relative cell kill probabilities of actinomycin and vincristine are assumed:

$$CKP_{VCR}=(3/5)*CKP_{TOTAL}$$

$$CKP_{ACT}= (2/5)*CKP_{TOTAL}$$

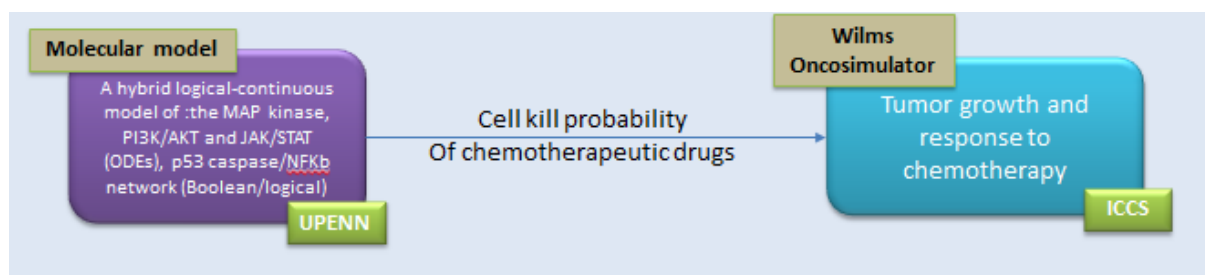The communication is sequential through the TAVERNA infrastructure [Wolstencroft *et al*, 2013].

Fig. NO5: Communication scheme of the molecular hypomodel with the Wilms Tumour Oncosimulator. Parameter exchanged: Cell kill probability of chemotherapeutic drugs. Type: Sequential coupling.

NO2.3 The Wilms Tumour Oncosimulator, the Vasculature and the Metabolic Hypomodels Integration

These three hypomodels are connected iteratively (Fig. NO6) through MUSCLE [Borgdorff *et al.* 2014]. The communication step of Oncosimulator and vasculature model is defined as MUSCLE – environment static input parameter (Table NO5).

The Wilms Tumour Oncosimulator calculates and sends per communication step to the vasculature model maps of the tumour domain containing per GC the total cell population, the number of proliferating, quiescent, differentiated, apoptotic and necrotic tumour cells.

The vasculature hypomodel which is based on tumour cell concentration calculates the normalized glucose concentration per position of the vector and send this information to the metabolic hypomodel.

The metabolic hypomodel which simulates the metabolism of cancer cells at the genome scale utilizes the current glucose concentration and calculates the local proliferation rate of the tumour cells that reside within each geometrical cell. The local conditions of nutrient supply, such as glucose concentration, primarily regulate the quiescence of the biological cells. Reduced availability of glucose will lead to increase in the percentage of dormant cells. In this framework, the Oncosimulator as a first approximation translates the proliferation rate map received by the metabolic hypomodel into the parameter P_{sleep} which regulates the fraction of newborn cells entering dormant state based on the following equation:

$$P_{sleep} = \frac{1-e^{aT_C/2}}{1-(P_{G0toG1}/T_{G0})/(a+1/T_{G0})},$$ derived from Eq. (7) in Kolokotroni *et al.* 2011

Fig. NO6: Communication scheme of the Wilms Tumour Oncosimulator, the Vasculature and the Metabolic hypomodels. Parameters exchanged:
- Oncosimulator - vasculature hypomodels: tumour cells population maps.
- Vasculature – metabolic hypomodels: glucose concentration map
- Metabolic – oncosimulator: proliferation rate map

Type: Iterative coupling

NO2.4 The Wilms Tumour Oncosimulator and the Biomechanic Hypomodels Integration

The models are connected with iterative coupling through MUSCLE (Fig NO7) [Borgdorff et al. 2014] with a communication step defined as MUSCLE environment static parameter (Table NO5).

The Wilms Tumour Oncosimulator calculates and sends a tumour cell concentration map to the biomechanical hypomodel per communication step (each simulated day). Based on this, the biomechanical hypomodel calculates the stress distribution and sends a least pressure map to the Oncosimulator. The latter utilizes these directions to define the local shrinkage or expansion direction of the tumour when necessary.
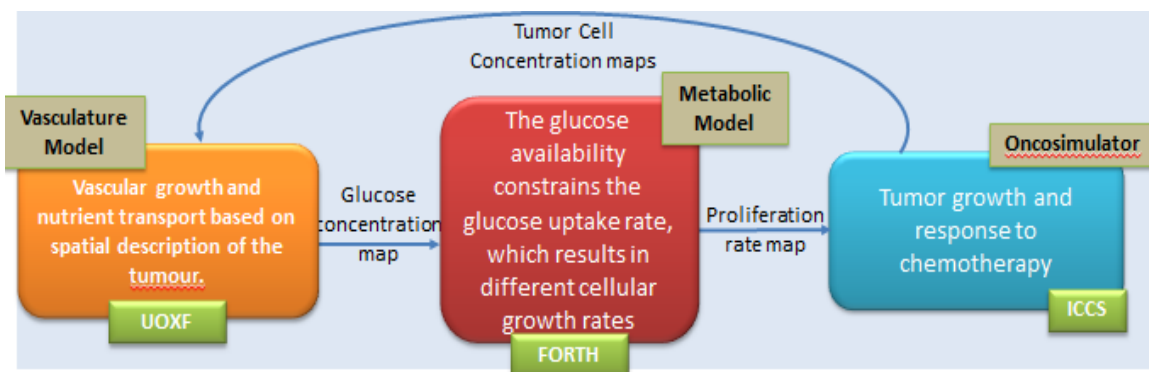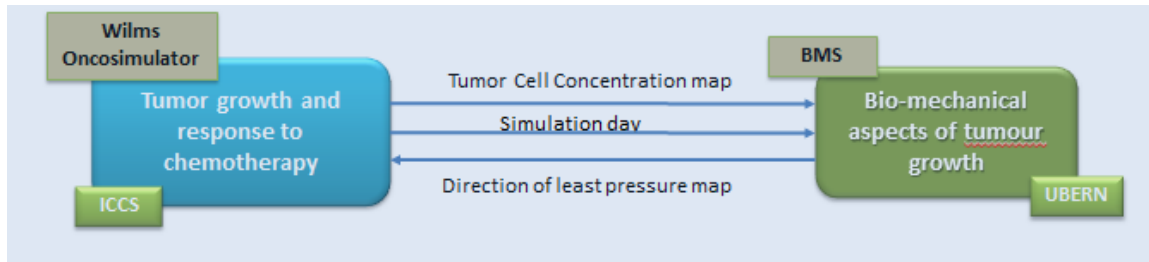
Fig. NO7: Communication scheme of the Wilms Tumour Oncosimulator and the Biomechanical hypomodels.
Parameters exchanged:
- Oncosimulator - Biomechanical hypomodels: tumour cell concentration map.
- Biomechanical– Oncosimulator hypomodels: direction of least pressure map

Type: Iterative coupling

TABLE NO5: Static Input of the Hypermodel - MUSCLE Environment Parameters

Name	Description
Communication frequency[OS-BMS]	The time interval between the subsequent interactions of the Oncosimulator with the Vasculature simulator
Communication frequency[OS-VS]	The time interval between the subsequent interactions the Oncosimulator and the Biomechanics simulator

NO3. Preliminary Results

All children with nephroblastoma receive pre-operative chemotherapy based only on imaging studies. Around 10% of patients do not respond to pre-operative chemotherapy. For these patients primary surgery would be beneficial. For this reason the clinical question formulated by clinicians and open to be addressed by the nephroblastoma hypermodel as described in detail in D2.5 is the following:

"Will a given nephroblastoma in a patient respond to pre-operative chemotherapy by tumour shrinkage, yes or no?"

The tumours of two nephroblastoma patients have been simulated with the fully integrated Wilms Tumour multimodeler hypermodel and demonstrated at project reviews. Both patients have received preoperative chemotherapy with a combination of actinomycin and vincristine according to the SIOP 2001/GPOH clinical trial (Fig. NO8) for unilateral stage I-III nephroblastoma tumours.
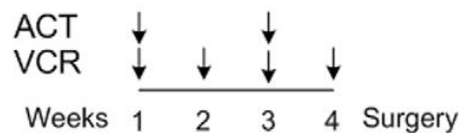
Fig. NO8: Chemotherapy treatment protocol. The simulated Wilms Tumour preoperative chemotherapy treatment protocol of the SIOP/ GPOH clinical trial.

For the two cases, two sets of MRI imaging data have been made available, one at an instant prior to chemotherapy and one at an instant after the completion of chemotherapy and before surgery. The relative dates (after pseudo-anonymization) of the two MRI acquisitions as well as the clinical tumour volumes as calculated based on the segmentation of the tumours on these imaging sets are depicted in Table NO6.

TABLE NO6: Imaging data related information of two nephroblastoma patients whose tumours have been simulated. Relative tumour shrinkage for reality and virtuality as well the deviation of virtuality from reality are also presented, It should be noted that The imaging dates are not real; they are the result of the pseudonymization process. Furthermore, the slight differences in the initial tumour volumes for reality and virtuality are mainly due to the isotropic resampling procedure which is used in virtuality for the simulation needs.

	Imaging Date	V_{CT} (cc) Clinical Tumour Volume	V_{VT} (cc) Virtual Tumour Volume	DV_{CT} (%) Relative Shrinkage (Clinical Tumour Volume)	DV_{VT} (%) Relative Shrinkage (Virtual Tumour Volume)	Deviation(%): $[(DV_{VT}$ (%)-DV_{CT} (%))/DV_{CT} (%)]X100%
Case 1	PRE: 1950/11/17	78.55	80.99	81.16	81.41	0.31
	POST: 1950/12/22	14.8	15.06			
Case 2	PRE: 1953/08/26	754.75	791.86	80.43	79.99	-0.55
	POST: 1953/10/06	147.68	158.44			

The available clinical data of the patients define the exact treatment scheme administered to each one (Fig. NO9, Fig NO10).
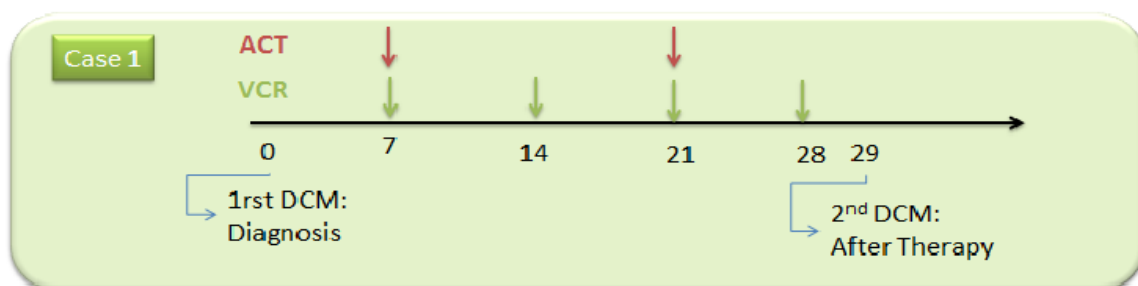
Fig. NO9: Pre-operative treatment scheme administered for the first patient (case 1).
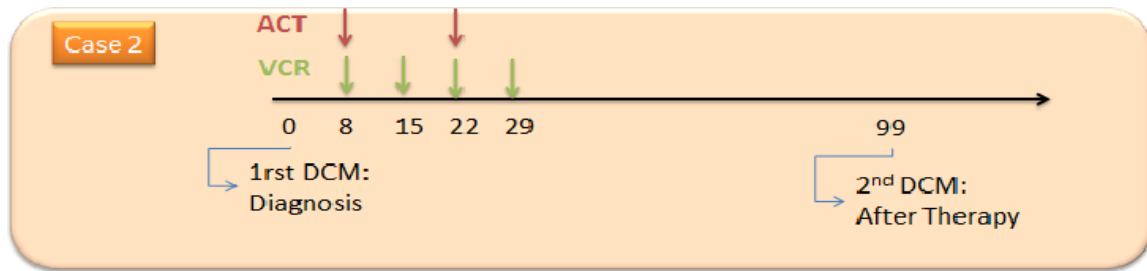
Fig. NO10: Pre-operative treatment scheme administered for the second patient (case 2).

The mi-RNA expression data of the patients is also available and serves as input for the molecular model.

The MRI data sets acquired before the beginning of therapy are segmented and serve as input to the multimodeler hypermodel. mi-RNA data sets serve also as input data. The real treatment schemes administered (Fig. NO9, Fig. NO10) which have been retrieved from the clinical data are simulated for both patients. The simulation starts at the time point of the acquisition of the MRI set before treatment whereas it finishes at the time point of the acquisition of the second available MRI imaging set after the completion of therapy. The input parameters of the Wilms Tumour Oncosimulator related to tumour cell kinetics are retrieved following adaptation of the model to real clinical, histological and imaging data (Table NO7). Information from literature has also been exploited. The resulting macroscopic characteristics of the tumours are shown in Table NO8. The real clinical tumour volume reduction, the simulated tumour volume reduction and the small deviation of virtuality from reality for both cases are shown in Table NO6.

TABLE NO7: Adaptation of parameter values related to tumour cell kinetics of the two simulated nephroblastoma patients

Parameter	Description	Value – Case 1	Value – Case 2
Tc	Cell cycle duration	6h	5
Psleep	Fraction of cells that enter G0 phase following mitosis	0.31	0.348
Psym	Fraction of stem cells that perform symmetric division	0.54	0.572
TN	Time needed for necrosis products to be removed	20	20
TA	Time needed for apoptosis products to be removed	6	6
NLIMP	Number of mitoses performed by LIMP cells before arrest	7	7
TG0	Duration of dormant phase	96	100
RA	Apoptosis rate of stem and LIMP cells	0.008	0.0047
RADiff	Apoptosis rate of differentiated cells	0.05	0.005
RNDiff	Necrosis rate of differentiated cells	0.03	0.006
PGotoG1	Fraction of G0 cells that re-enter cell cycle	0.01	0.001

TABLE NO8: Resulting Macroscopic Tumour Characteristics for the Two Simulated Cases

Cancer Characteristic	Value – Case 1	Value - Case 2
Doubling Time	10.52d	92.9 d
Fraction of Proliferating Cells	13.48%	5.48 %
Fraction of Dormant Cells	49.28%	39.13%
Fraction of DIFF Cells	12.28%	40.86%
Fraction of Dead Cells	24.97%	14.53%
Growth Fraction	17.96	6.41

The simulated time evolution of the tumours is shown in Fig. NO11. The predictions are in good agreement with clinical experience.

A.
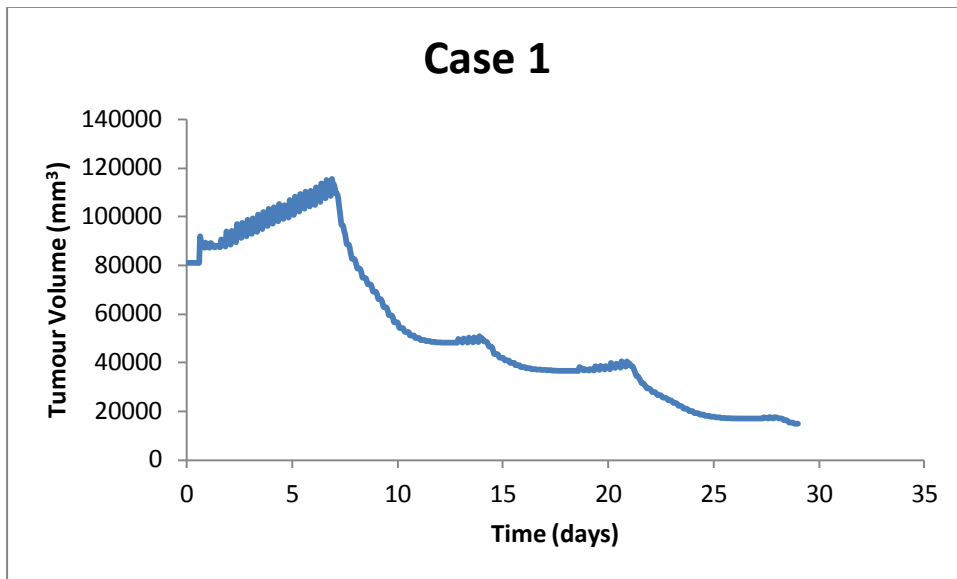
B.
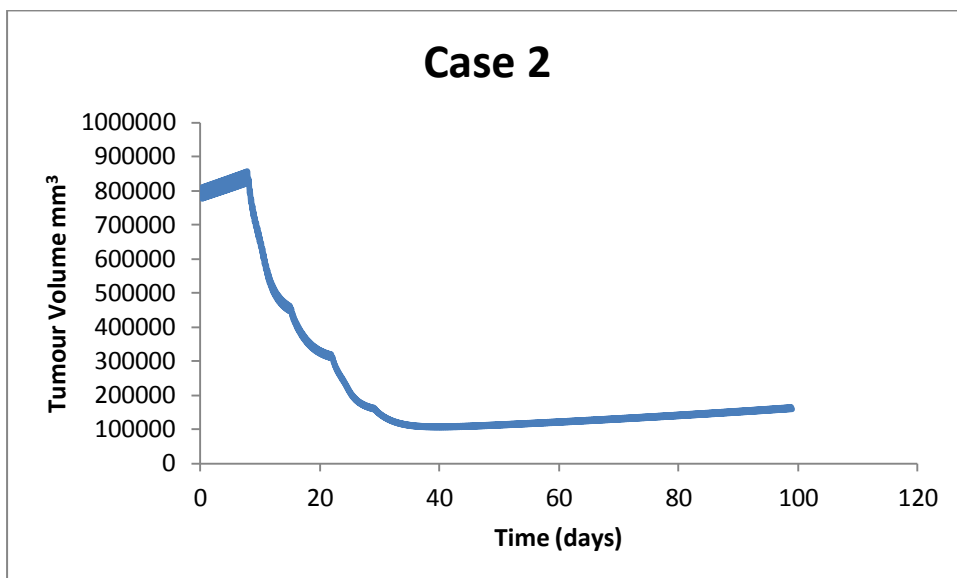
Fig. NO11: Time evolution of virtual simulated tumours of the first (A) and the second (B) case of a nephroblastoma patient.

NO4. References

J. Borgdorff, M. Mamonski, B. Bosak, K. Kurowski, M. Ben Belgacem, B. Chopard, D. Groen, P. V. Coveney, and A. G. Hoekstra, "Distributed multiscale computing with MUSCLE 2, the Multiscale Coupling Library and Environment," Journal of Computational Science, vol. 5, no. 5, pp. 719–731, Sep. 2014.

W.N. Dahl, R. Oftebro, E.O.Pettersen, T. Brustad. Inhibitory and cytotoxic effects of Oncovin (Vincristine Sulfate) on cells of human line NHIK 3025. Cancer Res. 36, 3101-3105, 1976.

E.C.Georgiadi, D.D.Dionysiou, N.Graf, G.Stamatakos, "Towards In Silico Oncology: Adapting a Four Dimensional Nephroblastoma Treatment Model to a Clinical Trial Case Based on Multi-Method Sensitivity Analysis." 2012 Nov; 42(11):1064-78. doi: 10.1016/j.compbiomed.2012.08.008.

Georgiadi EC, Stamatakos GS, Graf NM, Kolokotroni EA, Dionysiou DD, et al. (2008) Multilevel Cancer Modeling in the Clinical Environment: Simulating the Behaviour of Wilms Tumour in the Context of the SIOP 2001/GPOH Clinical Trial and the ACGT Project. In: Proceedings of the 8th IEEE International Conference on Bioinformatics and Bioengineering. 8–10 Oct 2008. Athens

E. Groninger, T. Meeuwsen-de Boer, P. Koopmans, D. Uges, W. Sluiter, A. Veerman, W. Kamps, S. de Graaf. Pharmacokinetics of Vincristine Monotherapy in Childhood Acute Lymphoblastic Leukemia. Pediatric Research 52: 113-118, 2002.

Karatzanis, I., Marias, K., & Sakkalis, V. (2015). Dr Eye. Journal of Clinical Bioinformatics, "DoctorEye platform web site," 2012, 5(1), 1–2.

Kolokotroni EA, Dionysiou DD, Uzunoglu NK, Stamatakos GS. Studying the growth kinetics of untreated clinical tumours by using an advanced discrete simulation model. Math Comput Model. 2011;54:1989-2006. doi:10.1016/j.mcm.2011.05.007

Stamatakos GS, Kolokotroni EA, Dionysiou DD, Georgiadi ECh, Desmedt C. An advanced discrete state – discrete event multiscale simulation model of the response of a solid tumour to chemotherapy. Mimicking a clinical study. J Theor Biol. 2010;266(1): 124-39. doi: 10.1016/j.jtbi.2010.05.019

S.Stamatakos, E.Ch.Georgiadi, N.Graf, E.A.Kolokotroni, and D.D.Dionysiou, "Exploiting Clinical Trial Data Drastically Narrows the Window of Possible Solutions to the Problem of Clinical Adaptation of a Multiscale Cancer Model", PLOS ONE 6(3), e17594, 2011

Stamatakos GS, Kolokotroni EA, Dionysiou DD, Georgiadi EC, Desmedt C (2010) An advanced discrete state-discrete event multiscale simulation model of the response of a solid tumour to chemotherapy: Mimicking a clinical study. J Theor Biol 266(1): 124–139.

Steel GG (2002) Basic Clinical Radiobiology. London: Arnold. pp 9–10.

K.Sawada, K.Noda, H. Nakajima, N. Shimbara, Y.Furuichi, M. Sugimoto. Differential cytotoxicity of anticancer agents in pre- and post-immortal lymphoblastoid cell lines. Biol Pharm Bull 28: 1202-1207, 2005.

E. Skounakis, V. Sakkalis, K. Marias, K. Banitsas, and N. Graf, 2009 "DoctorEye: A multifunctional open platform for fast annotation and visualization of tumours in medical images," Proc. 31st Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. Eng. Futur. Biomed. EMBC 2009, pp. 3759–3762.

E. Skounakis, C. Farmaki, V. Sakkalis, A. Roniotis, K. Banitsas, N. Graf, and K. Marias, 2010 "DoctorEye: A clinically driven multifunctional platform, for accurate processing of tumours in medical images.," Open Med. Inform. J., vol. 4, pp. 105–15.

G.J.Veal, M. Cole, J.Errington, A.Parry, J.Hale, A.D.J.Pearson, K.Howe, J.C.Chiholm, C.Beane, B.Brennan, F.Waters, A.Glaser, S.Hemsworth, H. McDowell, Y.Wright, K.ritchard- Jones, R.Pinkerton, G.Jenner, J.Nikolson, A.M.Elsworth, A.V.Boddy, and UKCCSG Pharmacology Working Groups. Pharmacokinetics of Dactinomycin in a pediatric patient population: a United Kingdom Children's Cancer Study group study. Clin Cancer Res 11(16): 5893-5899, 2005.

Katherine Wolstencroft, Robert Haines, Donal Fellows, Alan Williams, David Withers, Stuart Owen, Stian Soiland-Reyes, Ian Dunlop, Aleksandra Nenadic, Paul Fisher, Jiten Bhagat, Khalid Belhajjame, Finn Bacall, Alex Hardisty, Abraham Nieva de la Hidalga, Maria P. Balcazar Vargas, Shoaib Sufi, and Carole Goble (2013): "The Taverna workflow suite: designing and executing workflows of Web Services on the desktop, web or in the cloud", Nucleic Acids Research, 41(W1): W557-W561.

LO. The Non Small Cell Lung Cancer (NSCLC) Oncosimulator Hypomodel and the NSCLC Integrated Hypermodel

The development of both the Non Small Cell Lung Cancer Oncosimulator Hypomodel and the NSCLC Integrated Hypermodel has been led by ICCS.

LO1. Lung Oncosimulator Hypomodel

The NSCLC Oncosimulator developed by ICCS is a hypomodel (component model) simulating critical multiscale biological mechanisms involved in tumour growth and response to (radiotherapy, hemotherapy and targeted therapy) treatment.

The NSCLC Oncosimulator also acts as the hypermodel integrator and is linked with a vasculature/angiogenesis hypomodel (of which the development has been led by UOXF), a biomechanics hypomodel (of which the development has been led by led by UBERN), a cell kill rate focusing molecular hypomodel (of which the development has been led by UPENN) and a metabolic network hypomodel (of which the development has been led by FORTH). Another parallel simplistic phenomenological model (of which the development has been led by led by UNITO) serves as a gross semi-quantitative alternative to the nephroblastoma multimodeller hypermodel.

However, by making use of certain plausible approximations, the ICCS NSCLC Oncosimulator can also function as a standalone hypermodel constituting of just one hypomodel. This suggests that the Oncosimulator can bypass all other hypomodels in case of insufficient input data availability. Obviously the Oncosimulator predictions would generally be less accurate in such a scenario.

An additional statistics based machine learning hypermodel has been developed by ICCS in order to respond to particular clinical questions (see "LO2. Lung Cancer Statistical Approach").

The Lung Oncosimulator explicitly models cancer cell multiplication, cellular response to treatment and spatial expansion. It is based on the notions of cellular automata and equivalence classes and lies in the heart of the Lung Cancer Multimodeller Hypermodel. The Lung Oncosimulator, stemming from the *In Silico* Oncology and *In Silico* Medicine Group (ICCS- National Technical University of Athens) [Stamatakos et al. 2010, Kolokotroni et al. 2011], considers both radiotherapy and cisplatin–based doublet therapy. The Oncosimulator addresses clinical tumours, well beyond their initiation phase, and has been designed to incorporate patient specific data such as imaging-based, histopathological, molecular and treatment data.

The Oncosimulator is available as a standalone application or as a MUSCLE-enabled hypomodel. The MUSCLE-enabled version of the Oncosimulator explicitly addresses radiotherapy and cisplatin–vinorelbine doublet therapy. However, it can be easily adapted to other types of cisplatin-based doublet therapy.

LO1.1 Model basic notions

The model is based on the consideration of a discrete time and space stochastic cellular automaton, representing the tumour region. More specifically, the tumour region can be considered as a grid of "geometrical cells" (GCs, the elementary volume of the grid). Each GC corresponds to a cluster of heterogeneous cells found in various states. Specific rules regulate the transition between these states, as well as cell movement throughout the tumour volume; the aim is a realistic, conformal to the initial shape of the tumour, simulation of spatial evolution. A detailed description of the modelling approach can be found in D6.2. Here the basic modelling features are summarized.

Cell multiplication

Cancer cell multiplication modelling is based on the well documented hypothesis of cancer stem cell theory and incorporates the biological mechanisms of cell cycling, quiescence, recruitment, differentiation and loss via apoptosis (spontaneous) and necrosis (starvation-induced).

Stem, LIMP, DIFF, apoptotic and necrotic cells represent the distinct cell categories of the model. More specifically, tumour sustenance is attributed to the presence of a cell population that exhibits stem cell like properties. Specifically, cancer stem cells have the ability to preserve their own population, as well as give birth to cells of limited mitotic potential (LIMP cells) that follow the path towards terminal differentiation (DIFF cells). A proliferating tumour cell (stem or LIMP) passes through the successive cell cycle phases. Phases within or out of the cell cycle (G1, S, G2, M, G0) constitute different states in which cells may be found. After the completion of mitosis a fraction of newborn cells will enter the dormant phase, whereas the rest will continue to cycle. Transition to quiescence (dormant, G0, phase) and "awakening" of dormant cells are regulated by local metabolic conditions. All cell categories may die through spontaneous apoptosis. However, for dormant and differentiated cells necrosis is the main cell loss mechanism caused by inadequate nutrients' and oxygen supply.

Chemotherapy

When a tumour is chemotherapeutically treated, a fraction of cancer cells are lethally hit by the drug or its metabolites. Lethally hit cycling tumour cells enter a rudimentary cell cycle that leads to apoptotic death via a specific phase dictated by the action mechanism of the chemotherapeutic agent used. Similarly, in the case of cell cycle non-specific drugs, lethally hit dormant (G0) cells enter the G0hit phase. Marking of a cell as hit by the drug is assumed to take place at the instant of drug administration although its actual time of death is dictated by the specific pharmacokinetics and pharmacodynamics of the drug. It is pointed out however that cell cycle phase specific drugs can be readily modeled by the cytokinetic model shown in Fig. LO1 by appropriately selecting the "Chemo" induced exit from the normal cell cycle for both cases of stem and ILIMP cells.

The cytotoxic effect of cisplatin on cancer cells has primarily been attributed to the formation of cisplatin-gDNA adducts [Jordan & Carmo-Fonseca 2000, Siddik 2003, Wang et al. 2004]. In an attempt to remove these adducts and restore DNA lesions, the cells undergo a transient S arrest and a more persistent G2/M arrest [Ormerod et al.1994]. Accumulating evidence, derived from various cancer cell lines including non-small cell lung cancer, has linked the initiation of apoptotic pathway with the activity of mismatch repair proteins and the G2/M checkpoint [Sorenson et al. 1990, Jordan & Carmo-Fonseca 2000, Siddik 2003, Wang et al. 2004]. Lethally damaged cells seem to stay trapped, but viable, at G2/M phase for a few days and then proceed to death [Sorenson et al. 1990]. Cisplatin is cell cycle-non specific drug [Katzung 2001]. In our modeling approach, tumour cells are assumed to absorb Cisplatin at all cycling phases, as well as at G0 phase, whereas apoptotic death of hit cells takes place at the end of G2 phase.

Vinorelbine, a member of the *Vinca* alkaloid class, is a microtubule-targeting agent. Microtubules are polymers involved in the formation of the mitotic spindle that pulls apart the sister chromatids during cell division. At relatively low, but clinically relevant, concentrations, vinorelbine can suppress microtubule dynamics (dynamic instability and treadmilling,), whereas at higher concentrations it can prevent the polymerization of tubulin into microtubules, cause microtubules depolymerization or induce formation of tubulin paracrystals. In all cases, the subsequent disruption of mitotic splinde formation and function and, thus, the inhibition of chromosomal segregation leads to a prolonged arrest during mitosis, and, eventually, apoptosis through the intrinsic mitochondrial apoptotic pathway [Jordan and Wilson 2004; Esteve et al. 2007, Chiu et al, Bourgarel-Rey et al. 2009]. Vinorelbine is considered cell cycle specific [Katzung 2001]. In our modeling approach, tumour cells

are assumed to absorb Vinorelbine at all cycling phases, whereas apoptotic death of hit cells takes place at the end of M phase.

Radiotherapy

In the case of radiation therapy lethally damaged cells die through a radiation-induced necrotic mechanism. These cells enter a rudimentary cell cycle and die after undergoing a few mitotic divisions. The probability of cells to be hit by irradiation depends primarily on the phase they reside. Cell killing by irradiation is described by the Linear Quadratic or LQ Model, which is widely used in the pertinent literature [Steel, 2002]:

$$S(D) = \exp[-(\alpha D + \beta D^2)]$$

where $S(D)$ is the surviving fraction after a (uniform) dose D (Gy) of radiation to a population of cells. The parameters α (alpha) (Gy^{-1}) and β (beta) (Gy^{-2}) are called the radiosensitivity parameters of the LQ model. Cell radiosensitivity varies considerably throughout the cell cycle [Steel 2002, Perez and Brady 1998]. The S phase is regarded as the most resistant. Cells in any proliferating phase are more radiosensitive than hypoxic cells residing in G0. Based on these observations the model currently uses different values for the radiosensitivity parameters of the LQ model for the S phase (α_s, β_s), the remaining proliferating phases G1, G2, M (α_p, β_p), and the G0 phase (α_{G0}, β_{G0}).

The values of α_s, β_s and α_{G0}, β_{G0} can be derived as perturbations of the (α_p, β_p) values, as for example in: $\alpha_{G0} = \alpha_p / OER$, $\beta_{G0} = \beta_p / OER^2$, $\alpha_s = 0.6 \alpha_p + 0.4\alpha_{G0}$, $\beta_s = 0.6\beta_p + 0.4\beta_{G0}$, where OER is the Oxygen Enhancement Ratio (Perez and Brady 1998) Typical values of OER found in literature vary between 2.5-3 [Murray et al., 2003; Laperriere et al., 2002; Perez and Brady 1998].
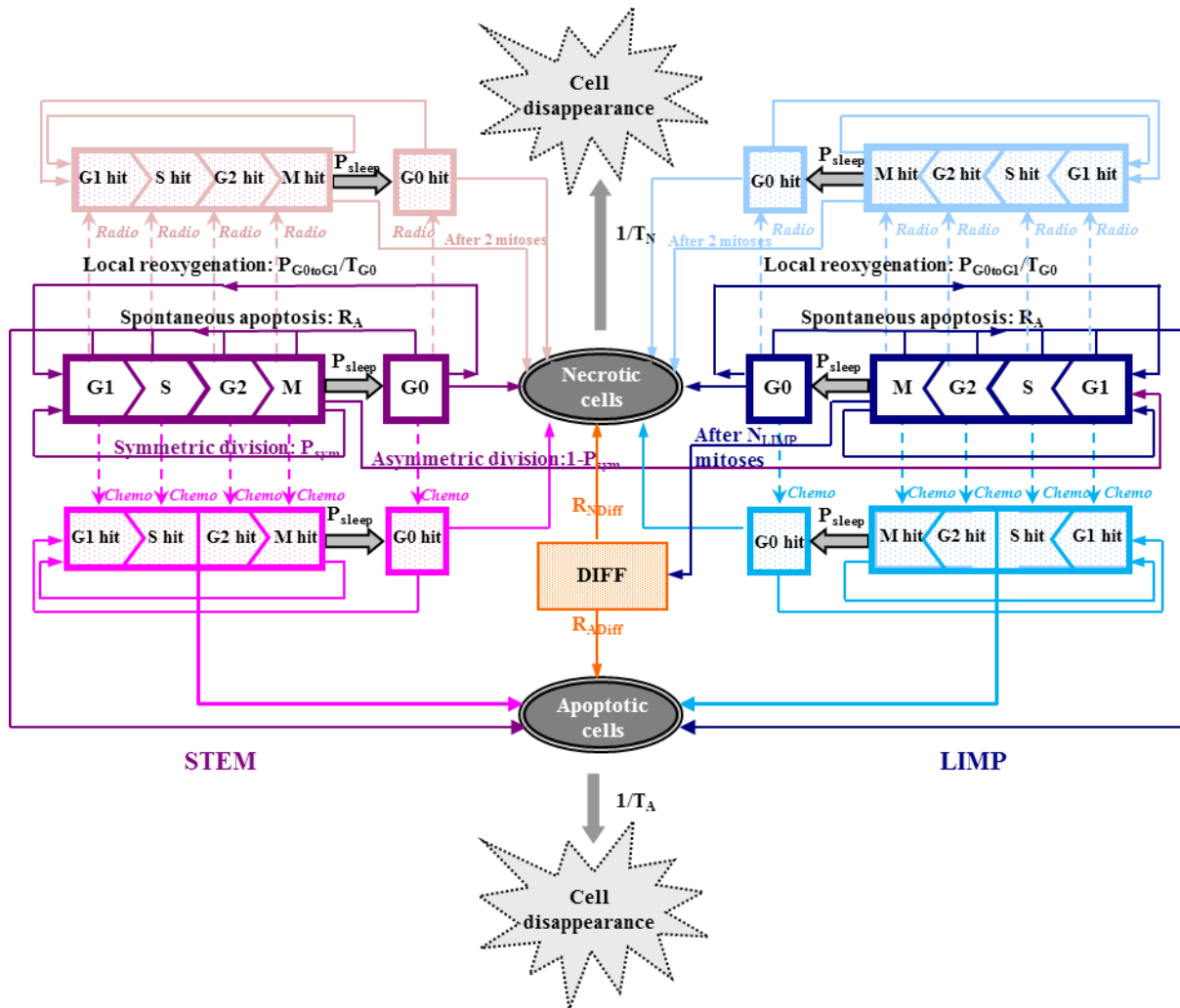
Fig. LO1. General cytokinetic model for tumour response to irradiation and chemotherapy.
STEM: stem cells. LIMP: Limited proliferative potential cells. DIFF: terminally differentiated cells. G1: Gap 1 phase. S: DNA synthesis phase. G2: Gap 2 phase. M: Mitosis phase. G0: Dormant, resting phase. Radio: Radioation treatment. Chemo: Chemotherapy Hit: Cells lethally hit by irradiation or chemotherapy. P_{G0toG1}: fraction of dormant cells that reenter the cell cycle, R_A: spontaneous apoptosis rate, P_{sleep}: fraction of newborn cells that enter G0, P_{sym}: fraction of stem cells that perform symmetric division, T_{G0}: duration of dormant phase, T_A: duration of apoptosis, T_N: duration of necrosis, R_{NDiff}: necrosis rate of differentiated cells, R_{ADiff}: apoptosis rate of differentiated cells.

LO1.2 Categorization According to CHIC Perspectives for Semantic Annotation

PERSPECTIVE CODE NUMBER	PERSPECTIVE	MODEL CATEGORY
I	TUMOUR-AFFECTED/ NORMAL TISSUE MODELLING:	✓ **tumour**
II	SPATIAL SCALE(S) OF THE MANIFESTATION OF LIFE:	✓ **cellular** ✓ **tissue** ✓ **body system**
III	TEMPORAL SCALE(S) OF THE MANIFESTATION OF LIFE:	✓ **hours** ✓ **days** ✓ **year**
IV	BIOMECHANISM(S) ADDRESSED:	✓ **cell cycling,** ✓ **apoptosis,** ✓ **necrosis,** ✓ **basic tumour biology**
V	TUMOUR TYPE(S) ADDRESSED:	✓ **lung cancer**
VI	TREATMENT MODALITY(-IES) ADDRESSED:	✓ **Chemotherapy** ✓ **Radiotherapy**
VII	GENERIC CANCER BIOLOGY/ CLINICALLY DRIVEN CHARACTER OF THE MODELLING APPROACH:	✓ **Clinically driven**
VIII	ORDER OF ADDRESSING DIFFERENT SPATIAL SCALES:	✓ **Top-down** approach
IX	ORDER OF ADDRESSING DIFFERENT TEMPORAL SCALES:	✓ **short periods → longer periods**
X	MECHANISTIC/STATISTICAL CHARACTER OF THE MODELLING APPROACH:	✓ **explicit biological mechanism modeling (e.g. using cytokinetic diagrams)**
XI	DETERMINISTIC/STOCHASTIC CHARACTER OF THE MODELLING APPROACH:	✓ **hybrid**
XII	CONTINUOUS-FINITE-DISCRETE CHARACTER OF THE MATHEMATICS INVOLVED:	✓ **discrete**
XIII	CLOSED FORM SOLUTION / ALGORITHMIC SIMULATION MODELLING APPROACH:	✓ **algorithmic simulation**

LO1.3 Integration - Hypermodel Communication Scheme

Oncosimulator – Preprocessing Tool

The tumour morphology needs to be defined at a preprocessing step and provided to the Oncosimulator in the form of a raw/mhd metaimage. In this file, different subregions of interest should be noted with distinct color numbers (e.g. white-255 for tumour area, black -0 for normal tissue) and the voxels should be cubic. Voxels of 1 or 2 mm acne (resolution) are commonly used. Furthermore, in order to reduce computational resources, the requested three-dimensional representation should not correspond to the entire image scan, but to a 'region of interest' centered

on the tumour, with available space around it to permit growth simulations. The amount of this extra space can be defined based on the maximum dimension of the tumour.

The transformation of the segmented metaimages to a raw file format compatible to the Oncosimulator input specifications, is implemented by the preprocessing tool. Since, voxels in the initial imaging data sets are usually non-isotropic, i.e. their dimensions differ in the x-, y- and z-direction, the segmented metaimages need resampling. The tool also performs cropping to reduce the size of the initial image at a smaller region of interest centered on the tumour.

The communication between the oncosimulator and the preprocessing tool is sequential. The cropped and interpolated metaimages of the tumour domain (in the form of mhd and raw files) produced by the preprocessing tool are passed as a static input to the Oncosimulator via TAVERNA.

Oncosimulator – Molecular Model

Cellular intrinsic sensitivity or resistance to treatment is a determinant of treatment outcome. The cell kill probability of a tumour cell is explicitly computed by the molecular model based on the molecular profile (e.g. EGFR mutations, miRNA expression data etc.) of the patient.

The communication between the Oncosimulator and the molecular model is sequential. The probability of a tumour cell to be lethally hit by a given therapy regimen is passed as a static input to the Oncosimulator via TAVERNA. In the case of chemotherapy the value is exploited directly by the Oncosimulator. In the case of radiotherapy the value is translated to the α (alpha) (Gy^{-1}) and β (beta) (Gy^{-2}) radiosensitivity parameters of the LQ model by applying the formula:

$$a = -\frac{\ln(1 - CKR)}{D + \frac{D^2}{\alpha/\beta}}$$

$$\beta = \frac{\alpha}{\alpha/\beta}$$

where α/β the alpha to beta ratio, CKR the cell kill rate provided by the molecular model and D (Gy) the radiation dose.

The cell radiosensitivity and the cell kill probability is then calculated throughout the cell cycle, as explain in the paragraph 'Radiotherapy'.

Oncosimulator – Vasculature – Metabolic models

The hyper-modelling scenario dictates an iterative communication scheme between the three hypomodels (Fig. LO2). This dynamic coupling is implemented via MUSCLE [Borgdorff et al. 2014]. The Oncosimulator computes the updated tumour domain geometry and the number of proliferating, quiescent, differentiated, apoptotic and necrotic tumour cells at each geometrical cell/voxel of the grid and send them to the vasculature model. Based on tumour cell concentration, the vasculature model passes the normalized concentration of glucose in each geometrical cell to the metabolic model. The metabolic model, which describes the aberrant metabolism of cancer cells at the genome scale based on the current glucose concentration, returns to the Oncosimulator the local proliferation rate of the tumour cells that reside within each geometrical cell. The local conditions of nutrient supply, such as glucose concentration, primarily regulate the withdrawal of tumour cells in a quiescent state, in an attempt by the tumour to sustain viability under conditions of reduced nutrient supply. Hence, a reasonable first approximation, is to translate the proliferation rate, a, to the fraction of newborn cells entering quiescent state, P_{sleep}, using the following formula:

$$P_{sleep} = \frac{1-e^{aT_C/2}}{1-(P_{G0toG1}/T_{G0})/(a+1/T_{G0})}, \quad \text{derived from Eq. (7) in Kolokotroni } et\ al.\ 2011$$

where T_C the cell cycle duration, T_{G0} the residence time of tumour cells in quiescent state and P_{G0toG1} the fraction of quiescent cells re-entering cell cycle.

Oncosimulator – Biomechanics simulator

The hyper-modelling scenario dictates an iterative communication scheme between the two simulators (Fig. LO2). This dynamic coupling is implemented via MUSCLE [Borgdorff et al. 2014]. The Oncosimulator computes cell proliferation in the case of free growth or cell loss in the case of treatment and sends the updated cell concentration information to the biomechanical solver. Based on this information the Biomechanics simulator calculates the stress distribution and passes the preferred direction of cell movement and, hence, of tumour growth/shrinkage, to the Oncosimulator. This direction corresponds to the direction against which the cells sense a minimum pressure.
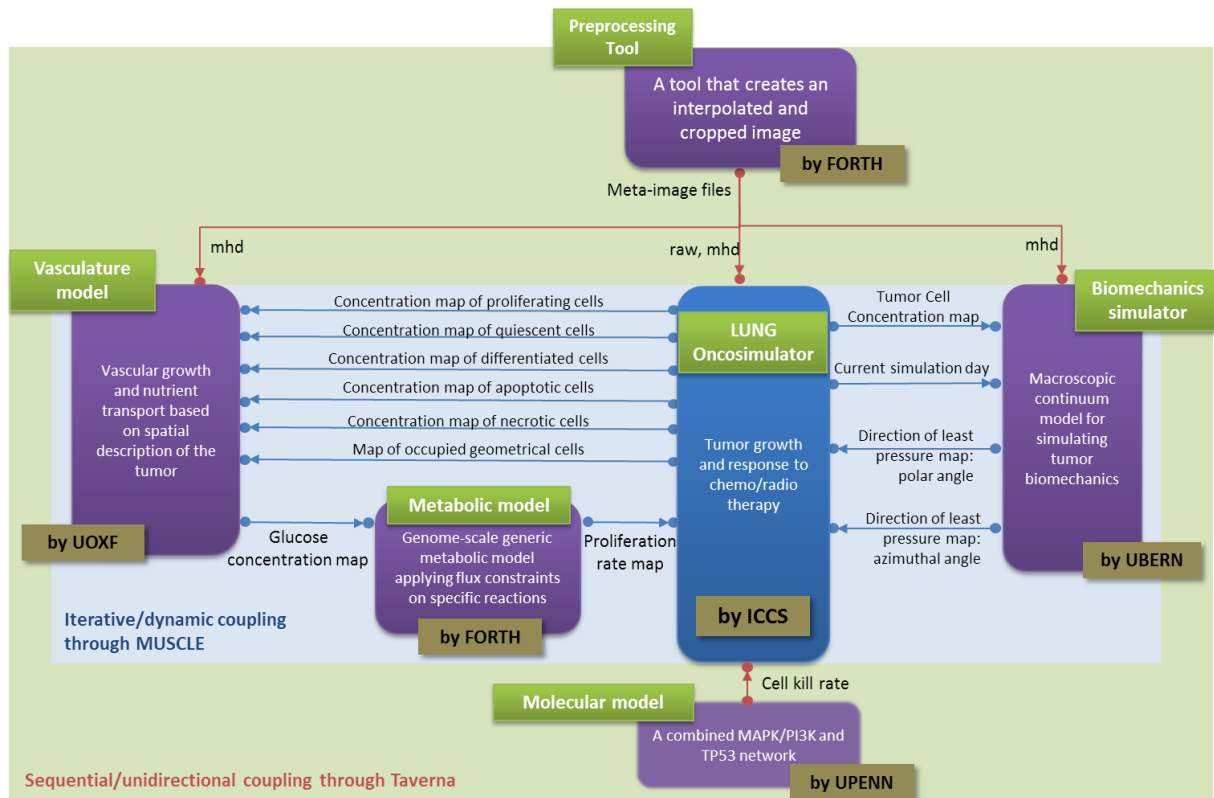
Fig. LO2. Lung multimodeler hypermodel communication scheme

LO1.4 Cancer-Specific Adaptation

With reference to lung cancer paradigm, the CHIC project primarily deals with surgically treated adenocarcinoma (ADC). The comprehensive literature survey presented in D6.2 can guide the selection of the most appropriate values of characteristic proliferation kinetics for this histological subtype, in the case of unavailability of patient-specific data. The less aggressive nature of ADC in respect to squamous cell carcinoma can be been reflected on a relatively slow T_d, a low proliferation, i.e. GF and stem cell frequency, and a low necrosis. The following assumptions/constraints can be imposed based on literature considering a typical lung adenocarcinoma case:

- A T_d ~ 300d, corresponding to the average of mean values reported in the reviewed literature for ADC (D6.2)

- A fraction of stem cells (out of living cells) <10^{-4} in line with the frequencies reported for highly permissive xenotransplantation conditions (D6.2)

- A GF ~ 0.35 approximately in the middle of the maximum range reported in the reviewed literature (D6.2)

- A fraction of apoptotic cells (out of total cells) ~ 0.01 corresponding to the average of median values reported in the reviewed literature (D6.2)

- A fraction of necrotic cells (out of total cells) ~ 0.1 near the lower boundary of the values reported in the reviewed literature (D6.2)

Table LO1 lists the value range of the Oncosimulator parameter values adapted for the case of lung adenocarcinoma. The development of more sophisticated strategies for optimizing the model adaptation process, as requested by the reviewers during the 5th project review, is under investigation and the outcome will be presented in deliverable D6.4: "Clinical adaptation and partial validation of hypermodels".

TABLE LO1. Typical model parameter values for lung adenocarcinoma.

Parameter symbol	Description	Units	Value range	References
CELL PHASE DURATIONS				
T_c[class*]	Cell cycle duration	hours	18-134	Lieber et al., 1976; Kimura et al. 1979; Loh et al. 1984; Olsson et al. 1984; Brower et al. 1986; Masuda et al. 1991; Campling et al. 1992 Liu & Tsao, 1993; Giaccone et al. 1992; Li et al. 2012
T_{G0}[class*]	G0 (dormant phase) duration i.e. time interval before a dormant cell re-enters cell cycle or dies through necrosis	hours	90-1200	extension of Durand & Sham 1998
T_A[region‡]	Time needed for both apoptosis to be completed and its products to be removed from the tumour	hours	1-25	Kerr et al. 1972, Gavrieli et al., 1992; Bursch et al. 1990
T_N[region‡]	Time needed for both necrosis to be completed and its lysis products to be removed from the tumour	hours	1-200	Ginsberg T., 1996, estimation based on fraction of necrotic cells
CELL CATEGORY/PHASE TRANSITION RATES AND FRACTIONS[i]				
R_A	Apoptosis rate of living stem and limp tumour cells, i.e. fraction of cells dying through apoptosis per unit time	hours^{-1}	0-0.001	estimation
R_{ADiff}	Apoptosis rate of differentiated tumour cells	hours^{-1}	0.0001-0.02	extension of Rawlins & Hogan, 2006, 2008; Lippmann 2000; Flindt 2006;
R_{NDiff}	Necrosis rate of differentiated tumour cells	hours^{-1}	0-0.02	Assumption, estimated to achieve the given GF
P_{sym}[region‡]	Fraction of stem cells that perform symmetric division	-	0-0.4	estimation based on Pine et al. 2010; Morrison et al. 2012
P_{sleep}[region‡]	Fraction of cells entering the G0 phase	-	0-1	Estimated to achieve the

	following mitosis			given T_d
P_{G0toG1}[class*] [region‡]	Fraction of dormant (stem and LIMP) cells that re-enter cell cycle	-	0-1	-
MISCELLANEOUS PARAMETERS				
N_{LIMP}	Number of mitoses performed by LIMP cells before becoming differentiated	-	8-24	Estimation based on fraction of stem cells
CHEMOTHERAPY PARAMETERS				
$T_{chemo, adm}$ [n][drug†]	Time point of nth drug administration, n=1,…	days	-	clinical data
Cell kill rate[drug†]	The numbers of biological cells lethally hit by the drug at each administration	-	0-1	-
CKF [class*]	Cell kill factor i.e. factor adapting cell killing probability to stem or LIMP cells	-	0-1 for stem, 1 for LIMP	-
RADIOTHERAPY PARAMETERS				
α/β	alpha to beta ratio	Gy	10	Jin et al. 2010; Lindblom et al. 2015
OER	Enhancement of therapeutic or detrimental effect of ionizing radiation due to the presence of oxygen	-	2.5 - 3	Perez and Brady 1998; Murray et al. 2003; Laperriere et al. 2002
D	Dose of radiation to a population of cells	Gy	-	Depends on scheme
$T_{radio, adm}$ [n]	Time point of nth radio administration, n=1,…	days	-	clinical data
Cell kill rate [radio]	The numbers of biological cells lethally hit by the radio at each administration	-	0-1	-

*Defined separately for **stem** and **LIMP** cells (class: {stem, limp})
‡Defined separately for **proliferating** and **necrotic** region (region: {proliferating, necrotic})
†Defined separately for each drug administered
iThe parameters included under this term express fractions and, therefore, can theoretically take any value between zero and unity. Whenever possible this range has been narrowed based on logical assumptions supported by literature or basic science.

LO1.5 Preliminary Results

The hypermodel execution has been successfully demonstrated on a recurrent lung adenocarcinoma treated with irradiation. The clinical question addressed is:

How much will be the imageable volume reduction of the tumour one year following the completion of tumour irradiation?

The patient received four fractions of 15 Gy. The radiation schedule is depicted in Fig LO3. The patient specific data that have been exploited by the Oncosimulator are the applied chemotherapeutic scheme (drugs, administration instants) and the 3D image of the tumour as reconstructed from CT imaging data. The sets of imaging data were provided for two time instants before and after the completion of the treatment (Fig LO3). There is no visible tumour approximately one year after treatment. Due to the non-availability of data related to any distinct internal metabolic regions, the virtual tumour implemented is homogeneous with a shape compliant to the reconstructed tumour image.

Table LO2 lists the parameter values of Oncosimulator related to tumour cell kinetics, following adaptation of the hypermodel to the histological subtype. Fig LO4 shows the simulated time course of the tumour volume for the clinical case considered. The onset (time 0) and the finishing time point of simulation correspond to the time instances of the initial and the final tumour CT image acquisition. The onset of therapy is evident from the rapid decrease in tumour volume. The hypermodel predicts

a very small nodule of 5mm in diameter at the time point of the final CT acquisition. Additional adaptation work has been planned for the last year of the project to explain such discrepancies and improve the accuracy of the hypermodel. The clinical case was also simulated assuming lower radiation doses. The hypermodel successfully demonstrated a worse treatment response and a faster tumour cell repopulation with decreasing radiation dose.
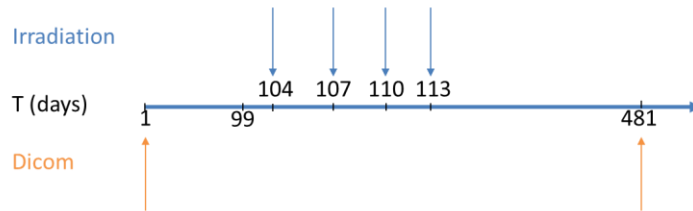
Fig. LO3. Treatment schedule and volumetric measurements

TABLE LO2. Adaptation Of Parameter Values Related To Tumour Cell Kinetics And Resulting Cancer Characteristics

Parameter	Description	Value
T_c	Cell cycle duration	40 h
T_{G0}	Duration of dormant phase	168 h
T_N	Time needed for necrosis products to be removed	23 h
T_A	Time needed for apoptosis products to be removed	4 h
N_{LIMP}	Number of mitoses performed by LIMP cells before arrest	22
R_A	Apoptosis rate of stem and LIMP cells	0.0001 h^{-1}
$R_{A,DIFF}$	Apoptosis rate of differentiated cells	0.017 h^{-1}
$R_{N,DIFF}$	Necrosis rate of differentiated cells	0.025 h^{-1}
P_{G0toG1}	Fraction of G0 cells that re-enter cell cycle	0.1
P_{sleep}	Fraction of cells that enter G0 phase following mitosis	0.263
P_{sym}	Fraction of stem cells that perform symmetric division	0.322
Cancer Characteristic		**Value**
Doubling Time		293 d
Fraction of Stem Cells‡		0.00009
Fraction of LIMP Cells‡		0.856
Growth Fraction‡		0.356
Fraction of Dormant Cells ‡		0.500
Fraction of DIFF Cells‡		0.144
Fraction of Apoptotic Cells*		0.007
Fraction of Necrotic Cells*		0.121

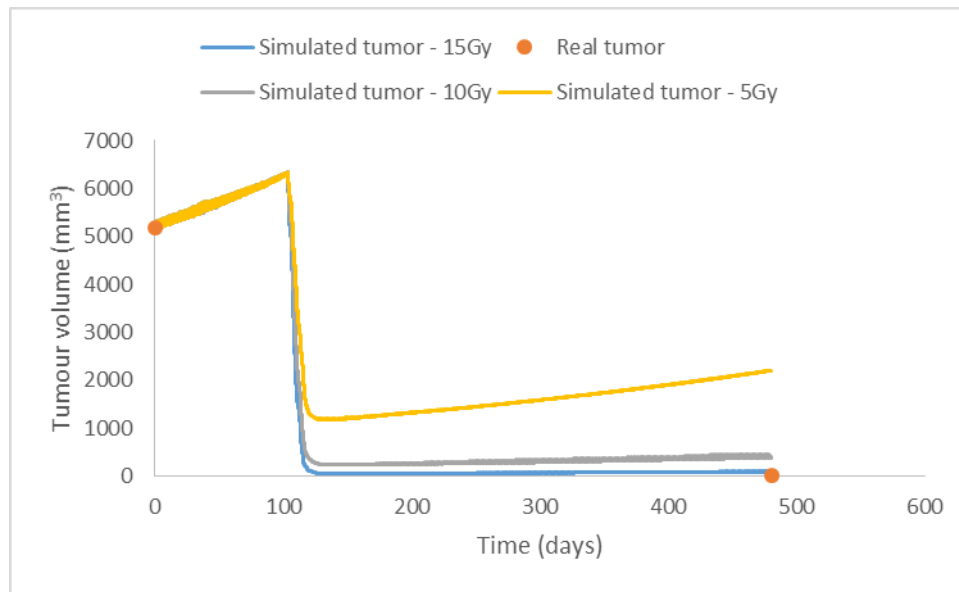‡out of living cells
*out of total cells

Fig. LO4. Simulated time course of the tumour volume for the clinical case considered.

LO2. Lung Cancer Statistical Approach

Regarding NSCLC adenocarcinoma, a machine learning approach is under development. The aim is to identify group of patients with high risk of recurrence based on their miRNA data and/or EGFR/BRAF/ALK mutation and clinical profile. A preliminary Naive Bayes classifier based on miRNA data only has been developed. A standalone application utilizing the above classifier, compliant to the CHIC specifications with respect to model deployment and packaging, has been released and successfully tested on the CHIC platform.

LO2.1 Methodological Aspects

At the molecular level, statistical models exploiting gene or pathway expression data have traditionally been recruited to serve as prognostic tools for relapse-free survival. The selection of the most appropriate technique is a fundamental step in data analysis. An important principle of model selection is to balance bias, referring to the errors occurring when the model is not complex enough to describe the true generating process, and variance, referring to the errors when the model has more degrees of freedom than the true generating process (overfitting). The tradeoff of bias and variance is generally resolved with proper penalization in the models.

The goal of feature selection is to identify a minimal subset of features that optimize prediction accuracy. The benefits are manifold, the most important ones being: (a) to avoid overfitting and improve model performance, i.e. prediction performance in the case of supervised classification and better cluster detection in the case of clustering, (b) to provide faster and more cost-effective models and (c) to gain a deeper insight into the underlying processes that generated the data [Saeys et al., 2007]. Three major categories of feature selection techniques can be recognized: filter methods, wrapper methods and embedded methods. Filters (e.g. Euclidean distance, t-test, Information gain, gain ratio, correlation-based feature selection etc.) estimate an index for each feature that expresses how relevant a feature is to the target. They are computationally simple and fast, and they are independent of the classification algorithm. On the other hand, wrappers rank feature subsets by the prediction performance of a classifier on the given subset. Greedy search strategies are generally preferred, such as sequential forward selection or backward elimination. Embedded methods select features based on criterions that are generated during the learning process of a specific classifier. When the training sample size is small, filters are preferred, whereas

the more advanced wrappers and embedded methods will perform better when the training sample size increases, due to their ability to detect mutual information between features. Guyon and Elisseeff (2003) suggested the following steps in solving a feature selection problem: 1. Feature constructions 2. Feature ranking with filter methods 3. Excluding outliers 4. Comparisons between different feature selection methods 5. Assessing stable solution.

A common problem in statistical analysis is the handling of missing values. The simplest way of dealing with it is to discard the observations that contain them. However, this method is applicable only when there are a few missing values. Rather than removing the observations with missing data, another approach is to fill in or "impute" the missing values. A variety of imputation approaches exists ranging from extremely simple (e.g replace each missing value with the mean) to rather complex (e.g. Maximum Likelihood, Multiple imputation) [Pigott 2001]. In the case of gene expression data the simplest imputation algorithms consist of replacing the missing values by zero or by the corresponding row/column average [de Souto et al. 2015]. More complex algorithms that employ gene correlations include, the weighted k-nearest neighbor (WKNN), local least squares, expectation maximization approach and Bayesian principal component analysis [de Souto et al. 2015].

In all cases, the choice of the most appropriate technique is problem specific and related to the type of data (binary, integer or continuous), their size and any prior information about the data distribution, according to their definitions and properties.

LO2.2 Categorization According to the CHIC Perspectives for Semantic Annotation

PERSPECTIVE CODE NUMBER	PERSPECTIVE	MODEL CATEGORY
I	TUMOUR-AFFECTED/ NORMAL TISSUE MODELLING:	✓ tumour
II	SPATIAL SCALE(S) OF THE MANIFESTATION OF LIFE:	-
III	TEMPORAL SCALE(S) OF THE MANIFESTATION OF LIFE:	-
IV	BIOMECHANISM(S) ADDRESSED:	-
V	TUMOUR TYPE(S) ADDRESSED:	✓ lung cancer
VI	TREATMENT MODALITY(-IES) ADDRESSED:	-
VII	GENERIC CANCER BIOLOGY/ CLINICALLY DRIVEN CHARACTER OF THE MODELLING APPROACH:	✓ clinically driven
VIII	ORDER OF ADDRESSING DIFFERENT SPATIAL SCALES:	-
IX	ORDER OF ADDRESSING DIFFERENT TEMPORAL SCALES:	-
X	MECHANISTIC/STATISTICAL CHARACTER OF THE MODELLING APPROACH:	✓ based on "black box" statistical modeling
XI	DETERMINISTIC/STOCHASTIC CHARACTER OF THE MODELLING APPROACH:	-
XII	CONTINUOUS-FINITE-DISCRETE CHARACTER OF THE MATHEMATICS INVOLVED:	-
XIII	CLOSED FORM SOLUTION / ALGORITHMIC SIMULATION MODELLING APPROACH:	-

LO2.3 Preliminary Results

The classifier aims to predict the clinical evolution of a surgical treated patient with NSCLC adenocarcinoma. The classifier was trained based on the miRNA data from 10 patients with known clinical evolution. Two classes were considered: progression or complete remission. The fold change in miRNA expression was considered. There were no missing values.

The workflow for the development and validation of the classifier consists of the following steps:
- *Feature selection*: A t-test statistic was chosen due to the small size of the sample (10) compared to the considerable large size of the feature space (2550). The 10 most significant features based on t statistics (i.e. the ones with the lowest p-value) were selected.

- *Training* a Naïve Bayes classifier with the selected features.

- *Cross-Validation*: Due to the small size of the sample, the classifier was validated by applying the leave-one-out method.

The accuracy of the Naïve Bayes classifier estimated based on the leave-one-out cross-validation method is 100% in the sample considered (Table 3). Even though the results are promising, no definite conclusions can be drawn yet, due to the small size of the sample.
'

TABLE LO3. Confusion matrix of Naïve Bayes performance

Predicted class

		complete remission	recurrence	Recognition
Actual class	**complete remission**	4	0	100%
	recurrence	0	6	100%
	Precision	100%	100%	

LO3. References

J. Borgdorff, M. Mamonski, B. Bosak, K. Kurowski, M. Ben Belgacem, B. Chopard, D. Groen, P. V. Coveney, and A. G. Hoekstra, "Distributed multiscale computing with MUSCLE 2, the Multiscale Coupling Library and Environment," Journal of Computational Science, vol. 5, no. 5, pp. 719–731, Sep. 2014.

Bourgarel-Rey V, Savry A, Hua G, Carré M, Bressin C, Chacon C, Imbert J, Braguer D, Barra Y. Transcriptional down-regulation of Bcl-2 by vinorelbine: identification of a novel binding site of p53 on Bcl-2 promoter.

Biochem Pharmacol. 2009 Nov 1;78(9):1148-56. doi: 10.1016/j.bcp.2009.06.025. Epub 2009 Jun 23.

Brouckaert G, Kalai M, Krysko DV, Saelens X, Vercammen D, Ndlovu MN, et al. Phagocytosis of necrotic cells by macrophages is phosphatidylserine dependent and does not induce inflammatory cytokine production. Mol Biol Cell. 2004 Mar;15(3):1089-100.

Brower M, Carney DN, Oie HK, Gazdar AF, Minna JD. Growth of cell lines and clinical specimens of human non-small cell lung cancer in a serum-free defined medium. Cancer Res. 1986 Feb;46(2):798-806.

Bursch W, Paffe S, Putz B, Barthel G, Schulte-Hermann R. Determination of the length of the histological stages of apoptosis in normal liver and in altered hepatic foci of rats. Carcinogenesis. 1990 May;11(5):847-53.

Campling BG, Haworth AC, Baker HM, Greer DL, Holden JJ, Bradley EC, et al. Establishment and characterization of a panel of human lung cancer cell lines. Cancer. 1992 Apr 15;69(8):2064-74.

Chiu WH, Luo SJ, Chen CL, Cheng JH, Hsieh CY, Wang CY, Huang WC, Su WC, Lin CF. Vinca alkaloids cause aberrant ROS-mediated JNK activation, Mcl-1 downregulation, DNA damage, mitochondrial dysfunction, and apoptosis in lung adenocarcinoma cells. Biochem Pharmacol. 2012 May 1;83(9):1159-71. doi: 10.1016/j.bcp.2012.01.016. Epub 2012 Jan 20.

Durand RE, Sham E. The lifetime of hypoxic human tumour cells. Int J Radiat Oncol Biol Phys. 1998;42(4):711–5.

Emely Lindblom, Alexandru Dasu & Iuliana Toma-Dasu (2015) Optimal fractionation in radiotherapy for non-small cell lung cancer – a modelling approach, Acta Oncologica, 54:9, 1592-1598, DOI: 10.3109/0284186X.2015.1061207

Estève M-A, Carré M, Braguer D. Microtubules in apoptosis induction: are they necessary? Curr Cancer Drug Targets. 2007;7:713–729

Flindt, R. Amazing Numbers in Biology. Berlin: Springer -Verlag; 2006.

Gavrieli Y, Sherman Y, Ben-Sasson SA. Identification of programmed cell death in situ via specific labeling of nuclear DNA fragmentation. J Cell Biol. 1992 Nov;119(3):493-501.

Giaccone G, Gazdar AF, Beck H, Zunino F, Capranico G. Multidrug sensitivity phenotype of human lung cancer cells associated with topoisomerase II expression. Cancer Res. 1992 Apr 1;52(7):1666-74.

Ginsberg, T., 1996. Modellierung und Simulation der Proliferationsregulation und Strahlentherapie normaler und maligner Gewebe, Fortschritt-Berichte, VDI Verlag: Duesseldorf, Reihe 17: Biotechnik, Nr 140, 103-107.

Guyon I. and A. Elisseeff,.An Introduction to Variable and Feature Selection Journal of Machine Learning Research 3, 1157-1182 2003

Jin JY, Kong FM, Chetty IJ, Ajlouni M, Ryu S, Ten Haken R, Movsas B. Impact of fraction size on lung radiation toxicity: hypofractionation may be beneficial in dose escalation of radiotherapy for lung cancers. Int J Radiat Oncol Biol Phys. 2010 Mar 1;76(3):782-8. doi: 10.1016/j.ijrobp.2009.02.079.

Jordan P, Carmo-Fonseca M. Molecular mechanisms involved in cisplatin cytotoxicity. Cell Mol Life Sci. 2000;57(8-9):1229–35.

M.A. Jordan, L. Wilson Microtubules as a target for anticancer drugs Nat. Rev. Cancer, 4 (2004), pp. 253–265

Katzung BG, editor. Basic and Clinical Pharmacology. 8th ed. United States of America: McGraw-Hill; 2001.

Kerr JF, Wyllie AH, Currie AR. Apoptosis: a basic biological phenomenon with wide-ranging implications in tissue kinetics. Br J Cancer. 1972;26(4):239-57.

Kimura N, Shibuya T, Niho Y, Nakamura H, Matsuo S, Imamura T, et al. Human lung cancer cell line (KSNY) producing colony-stimulating activity which affects both human and mouse marrow cells. Gan. 1979 Dec;70(6):807-10.

Kolokotroni EA, Dionysiou DD, Uzunoglu NK, Stamatakos GS. Studying the growth kinetics of untreated clinical tumours by using an advanced discrete simulation model. Math Comput Model. 2011;54:1989-2006. doi:10.1016/j.mcm.2011.05.007

N. Laperriere, L. Zuraw, G. Cairncross. "The cancer Care Ontario Practice Guidilines Initiative Neuro-oncology Disease Site Group "Radiotherapy for newly diagnosed malignant glioma in adults: a systematic review"." Rad Oncol., vol. 64, pp. 259-273, 2002.

Li J, Yang H, Chen L, Li Y, Zhu Y, Dai Y, et al. Establishment and characterization of human non-small cell lung cancer cell lines. Mol Med Rep. 2012 Jan;5(1):114-7. doi: 10.3892/mmr.2011.613

Lieber M, Smith B, Szakal A, Nelson-Rees W, Todaro G. A continuous tumour-cell line from a human lung carcinoma Lieber with properties of type II alveolar epithelial cells. Int J Cancer. 1976 Jan 15;17(1):62-70.

Liu C, Tsao MS. Proto-oncogene and growth factor/receptor expression in the establishment of primary human non-small cell lung carcinoma cell lines. Am J Pathol. 1993 Feb;142(2):413-23.

Loh PM, Clamon GH, Robinson RA, White ML, Hukku B, Rossi NP, et al. Establishment and characterization of four new human non-small cell lung cancer cell lines. Cancer Res. 1984 Aug;44(8):3561-9.

Masuda N, Fukuoka M, Takada M, Kudoh S, Kusunoki Y. Establishment and characterization of 20 human non-small cell lung cancer cell lines in a serum-free defined medium (ACL-4). Chest. 1991 Aug;100(2):429-38.

Morrison BJ, Steel JC, Morris JC. Sphere Culture of Murine Lung Cancer Cell Lines Are Enriched with Cancer Initiating Cells. PLoS One. 2012;7(11):e49752. doi:10.1371/journal.pone.0049752

Morton Lippmann. Environmental Toxicants: Human Exposures and Their Health Effects. 3rd ed. New York: John Wiley & Sons, Inc.; 2000.

D. Murray, R. Mirzayans, A.L. Scott, J.M. Allalunis-Turner. "Influence of oxygen on the radiosensitivity of human glioma cell lines." Am J Clin Oncol.., vol. 26: e169-e170, 2003.

Olsson L, Sorensen HR, Behnke O. Intratumoural phenotypic diversity of cloned human lung tumour cell lines and consequences for analyses with monoclonal antibodies. Cancer. 1984 Nov 1;54(9):1757-65.

Ormerod MG, Orr RM, Peacock JH. The role of apoptosis in cell killing by cisplatin: a flow cytometric study. Br J Cancer. 1994 Jan;69(1):93-100.

C. Perez, L. Brady L. Principles and Practice of Radiation Oncology. Philadelphia: Lippincott-Raven, 1998.
Pigott TD. A Review of Methods for Missing Data. Educational Research and Evaluation 2001, Vol. 7, No. 4, pp. 353±383

Pine SR, Ryan BM, Varticovski L, Robles AI, Harris CC. Microenvironmental modulation of asymmetric cell division in human lung cancer cells. Proc Natl Acad Sci U S A. 2010 Feb 2;107(5):2195-200. doi: 10.1073/pnas.0909390107

Rawlins EL, Hogan BL. Ciliated epithelial cell lifespan in the mouse trachea and lung. Am J Physiol Lung Cell Mol Physiol. 2008 Jul;295(1):L231-4. doi: 10.1152/ajplung.90209.2008

Rawlins EL, Hogan BL. Epithelial stem cells of the lung: privileged few or opportunities for many? Development. 2006 Jul;133(13):2455-65.

Saeys Y, Inza I, Larrañaga P. A review of feature selection techniques in bioinformatics. Bioinformatics. 2007 Oct 1;23(19):2507-17.

Siddik, Zahid H. Cisplatin: mode of cytotoxic action and molecular basis of resistance. Oncogene. 2003;22(47):7265–79.

Sorenson CM, Barry MA and Eastman A. Analysis of Events Associated With Cell Cycle Arrest at G2 Phase and Cell Death Induced by Cisplatin. J Natl Cancer Inst. 1990 May 2;82(9):749-55

de Souto MC, Jaskowiak PA, Costa IG. Impact of missing data imputation methods on gene expression clustering and classification. BMC Bioinformatics. 2015 Feb 26;16:64. doi: 10.1186/s12859-015-0494-3.

Stamatakos GS, Kolokotroni EA, Dionysiou DD, Georgiadi ECh, Desmedt C. An advanced discrete state – discrete event multiscale simulation model of the response of a solid tumour to chemotherapy. Mimicking a clinical study. J Theor Biol. 2010;266(1): 124-39. doi: 10.1016/j.jtbi.2010.05.019

G. Steel. Basic Clinical Radiobiology. London: Arnold, 2002.

Wang G, Reed E, Li QQ. Molecular basis of cellular response to cisplatin chemotherapy in non-small cell lung cancer (Review). Oncol Rep. 2004 Nov;12(5):955-65.

BM. The Biomechanical Hypomodel for Nephroblastoma and Non Small Cell Lung Cancer (and Glioblastoma Multiforme)

The development of the Biomechanical Hypomodel for Nephroblastoma and Non Small Cell Lung Cancer (and Glioblastoma Multiforme) has been led by UBERN. ICCS has ensured the compatibility of the hypomodel with the rest of the interacting hypomodels as well as with the overarching topology of the corresponding hypermodels.

BM1. The Biomechanical Simulator Component

The Biomechanical Simulator (BMS) is a hypomodel or component model for the simulation of bio-mechanical aspects of macroscopic tumour growth. It relies on the Finite Element Method (FEM) to compute mechanical stresses and strains resulting from tumour growth or shrinkage in a patient-specific anatomy. The resulting information (3D stress, strain, displacement and pressure maps) can be used as input to other hypo-models.

We currently focus on the interaction with OncoSimulator (OS), where a 'direction of most-likely tumour growth' (for OS) is derived from pressure information computed by BMS. Further Interaction with other hypo-models can also be envisaged for the future. For example, the pressure map resulting from biomechanical simulations could be used to compute pressure-induced changes in blood perfusion and thus in the local availability of nutrients.

BM2. The Mathematical Model

The rationale and mathematical model underlying BMS has been introduced in Deliverable 6.2 and will therefore be summarized only briefly here. Two phenomena of tumour growth are considered: The mechanical effect of volumetric changes in a nodal tumour, the so-called mass effect, and healthy tissue invasion by a diffusively growing tumour.

Tumour Mass Effect

Simulation of tumour mass effect is based on a hyper-elastic material model as detailed in deliverable 6.2, section B7. Changes in cell concentrations relative to an initial reference concentration value c_{init} are linked to volumetric increase or decrease of the respective geometrical cell by a growth-induced strain ε_{growth} which, together with the mechanical strain ε_{mech}, contributes to the total strain in the tissue:

$$\varepsilon_{total} = \varepsilon_{mech} - \varepsilon_{growth}$$

Healthy Tissue Invasion

A theoretical framework for the modelling of healthy tissue invasion for brain tumour simulations was introduced by [BM1, BM2, BM3] based on a reaction-diffusion equation. This approach describes active tumour cell migration as a random walk-process which is mathematically expressed as a passive diffusion process from regions of higher cell concentration to regions of lower cell concentration. The evolution of the tumour cell population in space and time thus depends on the motility of tumour cells and the generation of new cells:

$$\frac{\partial c}{\partial t} = D \, \Delta c + \rho \, c \, (1 - c)$$

The isotropic and locally constant diffusion D represents the migration rate of cells. Logistic growth with growth rate ρ is assumed as reaction term, and $c = c(\boldsymbol{r}, t)$ is the normalised concentration of tumour. This formulation, known as Fisher's equation [BM4], leads to the creation of a non-proliferating zone in regions of high tumour cell concentration, in agreement with clinical observation of the formation of a necrotic core.

BM3 Hypo-model Simulator & Interfaces

The Biomechanical Simulator (BMS) focuses on the simulation of solid nodal tumours such as Nephroblastoma and NSCLS, and does not take into account aspects of healthy tissue invasion at this moment. Simulation of diffusive tumour growth and tissue invasion is of importance mainly for one of the CHIC scenarios, namely high-grade glioma growth. Modelling for this scenario is explored in a separate simulation framework which is introduced in the section "Modelling healthy-tissue invasion in GBM".

BMS is available as standalone (file-based) or MUSCLE [5]-enabled component model. Three types of static modelling parameters can be distinguished, all of which are configurable via a single configuration file in xml format:

- "Technical parameters", related to file-management and the FEM backend
- Scenario / body-site specific parameters, such as mechanical material properties of tumour and healthy tissues, as well as mechanical boundary conditions.
- Patient-specific parameters, namely the precise morphology of patient anatomy in the region of interest.

To facilitate automatic iterative execution of the model in a coupled workflow, a command-line interface is provided for a selected subset of parameters. This allows model configuration via a scenario-specific configuration file and patient-specific command-line parameters.

Dynamic inputs and outputs consist of snapshots of 3D maps of tumour cell concentration and "direction of least pressure" at given time points. Internally, as well as for file-based data exchange, BMS uses the VTK data format for representing parameter maps in 3D space. For MUSCLE-based model communication, outgoing parameter maps are transformed into byte-streams and, likewise, incoming parameter maps are serialised into VTK representations based on shared information about the domain size.

Figure BM1 illustrates the interfaces of BMS, as well as its dynamic in- and output ports. 'Tumour Domain' indicates the spatial map of cell concentrations provided by OS, 'BMS Simulation Domain' corresponds to the domain model on which biomechanical simulations are performed. Results of FEM simulation and post-processing are returned in both BMS and OS simulation domains. Details of these domains and the parameter exchange process are further explained in section "Modelling healthy-tissue invasion in GBM".
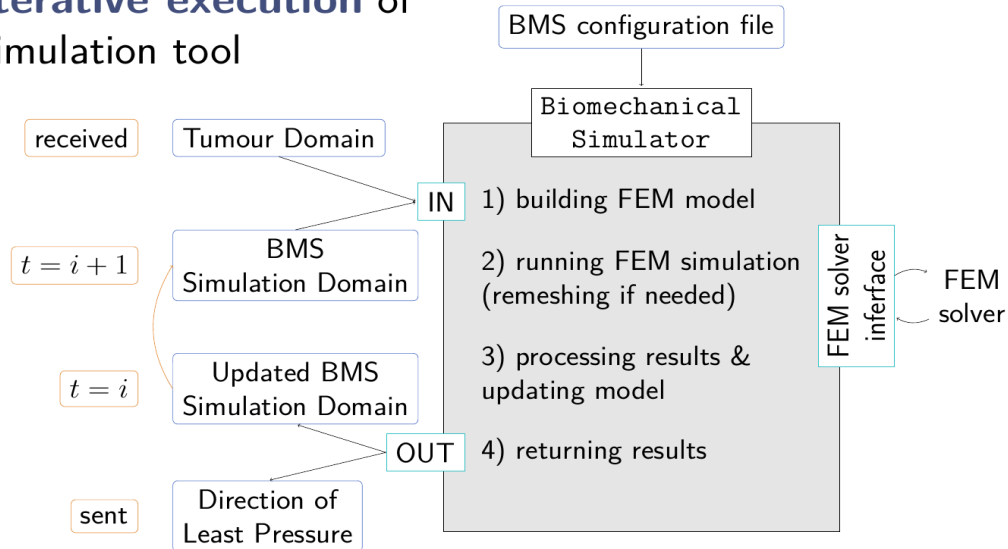
Fig. BM1 Static, dynamic and computational interfaces of the Biomechanical Simulator (BMS)

BM4. Scenario-Specific Adaptation and Patient Personalisation

The BMS simulator component can be adapted to different body-sites and be personalised to patient specific anatomies by means of model parametrization.

BM4.1 Scenario-Specific Parameters

Depending on its location, the tumour is surrounded by different tissues. We only take into account those tissues explicitly that are expected to make a distinct contribution to the tumour's mechanical landscape, either because of their immediate vicinity to the developing tumour or because of their distinctly different mechanical properties. Average bulk values are assumed for other tissues. Also, boundary conditions are specified per body-site/scenario, not for the individual patient.

The following paragraphs briefly detail the tissue types considered for the different scenarios in CHIC, together with estimates for their mechanical material properties, and boundary conditions.

Nephroblastoma

Fully constrained surface nodes are assumed as mechanical boundary conditions. Table BM1 presents tissue types and their mechanical material properties. Fig. BM2 illustrates a patient-specific FEM model with those tissues.

TABLE BM1: Tissues and mechanical tissue parameters for Nephroblastoma scenario

Tissue Type	E [Pa]	Poisson ratio
Healthy kidney	$5.3 \cdot 10^3$	0.40
Bone	$1.0 \cdot 10^9$	0.30
Other tissues	$5.0 \cdot 10^3$	0.40
Tumour	$20.0 \cdot 10^3$	0.40
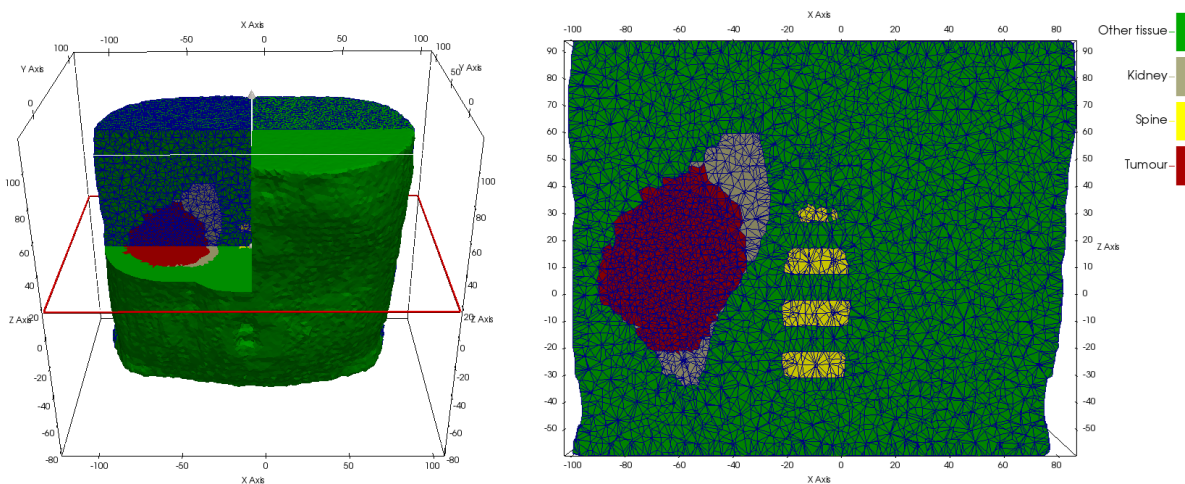
Fig. BM2: Personalised FEM model for Nephroblastoma scenario, derived from patient anatomy.

Non-Small Cell Lung Cancer (NSCLC)

Fully constrained surface nodes are assumed as mechanical boundary conditions. Table BM2 presents tissue types and their mechanical material properties. Fig. BM3 illustrates a patient-specific FEM model with those tissues.

TABLE BM2: Tissues and mechanical tissue parameters for NSCLC scenario

Tissue Type	E [Pa]	Poisson ratio
Lung tissue	$5.0 \cdot 10^3$	0.40
Bone	$1.0 \cdot 10^9$	0.30
Other tissues	$5.0 \cdot 10^3$	0.40
Inner organs	$5.0 \cdot 10^3$	0.40
Bronchi	$5.0 \cdot 10^3$	0.40
Tumour	$10.0 \cdot 10^3$	0.40
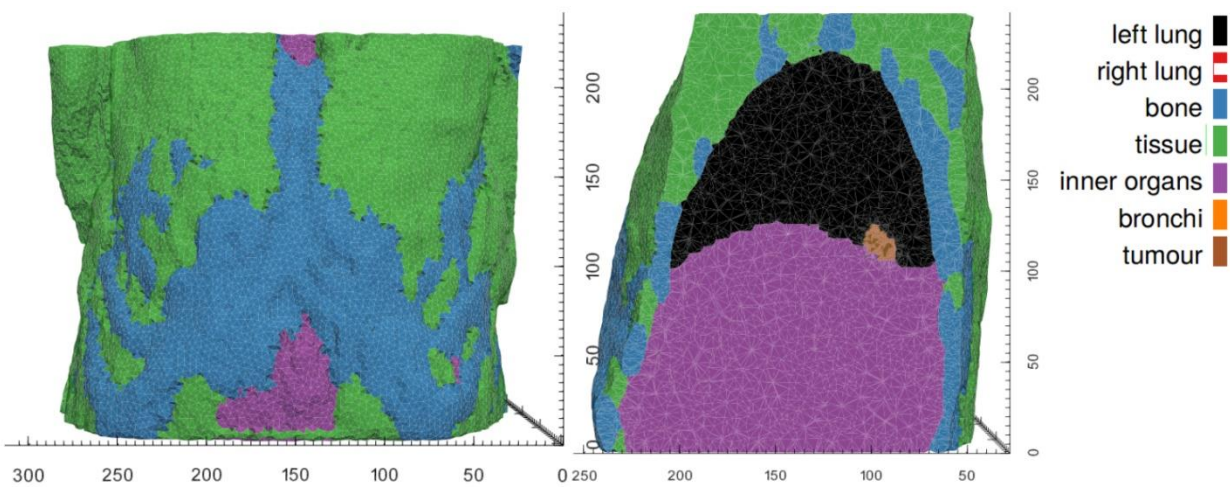
Fig. BM3: Personalised FEM model for NSCLC scenario, derived from patient anatomy.

Glioblastoma

Fully constrained surface nodes are assumed as mechanical boundary conditions. Table BM3 presents tissue types and their mechanical material properties. Figure 3 illustrates a patient-specific FEM model with those tissues and additional tumour subregions (edema as manifestation of microscopic presence of tumour cells).

TABLE BM3: Tissues and mechanical tissue parameters for GBM scenario

Tissue Type	E [Pa]	Poisson ratio
White Matter	$3.0 \cdot 10^3$	0.45
Grey Matter	$3.0 \cdot 10^3$	0.45
Tumour	$6.0 \cdot 10^3$	0.45
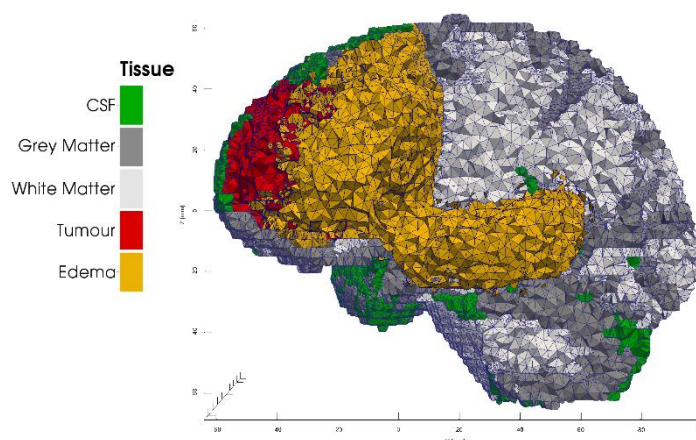
Fig. BM4. Personalised FEM model for GBM scenario, derived from patient anatomy.

BM4.2 Patient-specific parameters

While material parameters and boundary conditions are assumed to be applicable across patients within a scenario, anatomical models are patient-specific. For each patient, a segmented image of the region and tissues of interest is required to build the patient's personalised BMS simulation domain, as illustrated in figures Fig. BM2-Fig. BM4.

BM5. Data Preparation

The model configuration process comprises multiple steps that are both error-prone and time-consuming when performed manually. Therefore, a pre-processing pipeline has been developed to automate model creation, site-specific adaptation and personalisation steps as sketched in Fig. BM5:

- **Creation of tissue label maps of region of interest from medical images**

  In Nephroblastoma and NSCLC this remains a manual process that has to be completed before simulation. When automatic healthy tissue segmentation tools become available, this step can be automated. For example for Glioblastoma, automatic segmentation tools (e.g. BraTumIA [BM6]) are available and able to identify all tissues of interest.

- **Creation of patient-specific anatomical model: Meshing**

  For large simulation domains, a fixed spatial resolution as provided by the image voxels results in conflicting requirements for short simulation time (coarse mesh) and high accuracy (fine mesh) in specific regions of interest within the domain. As solution to this problem, a meshing routine for tetrahedral meshes has been implemented, which allows for variable mesh granularity in different regions of the simulation domain.

- **Parametrisation of patient-specific anatomical model with scenario-specific tissue parameters**

  All distinct tissues present in the anatomical segmentation are annotated with the material parameters and boundary conditions defined for the specific cancer scenario, see Tables BM1-BM3.

The resulting model provides the initial simulation domain on which growth-related biomechanical effects are simulated. All information is stored as VTK 'unstructured grid' and converted into the appropriate input file format for the chosen FEM backend. FeBio [BM7] is currently supported as simulation backend and further converters for Abaqus (Simulia) are being integrated.
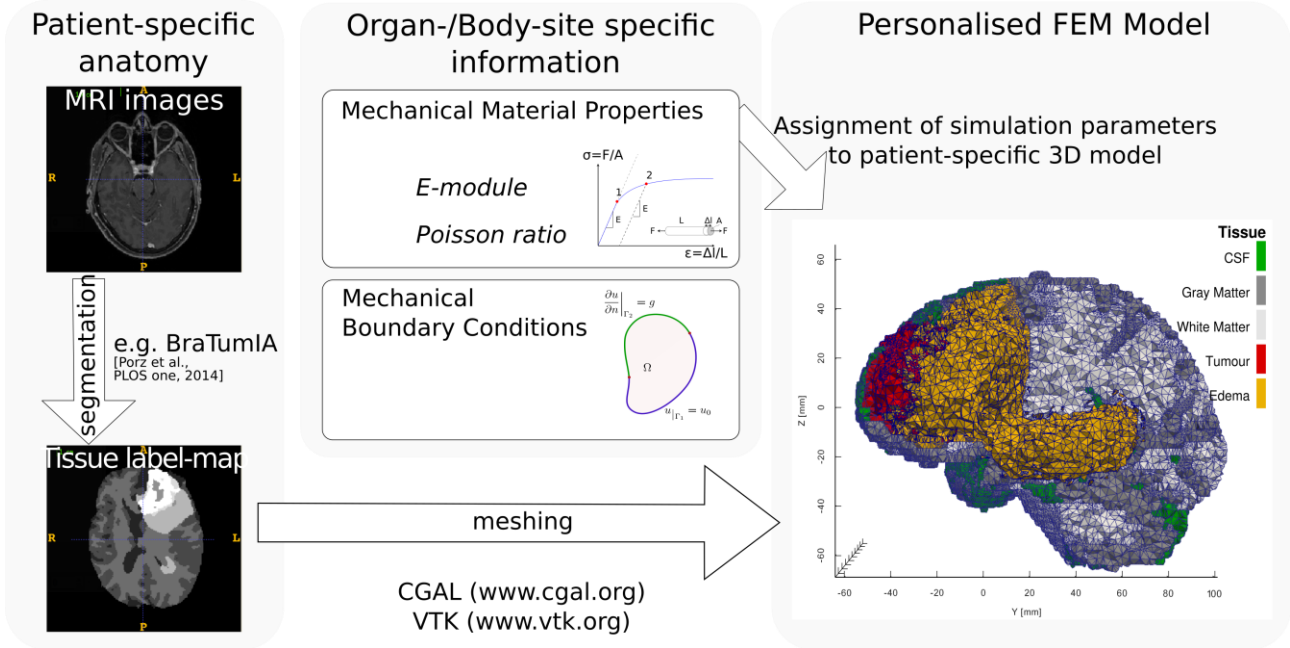
Fig. BM5: Data preparation pipeline for model generation.

BM6. Hyper-Model Execution

Once the simulation domain is generated, BMS can be used in iterative coupled execution with other hypo-models. Fig. BM6 illustrates this process on the example of BMS-OS interaction. Starting from an initial spatial map of cancer cell concentrations, in each time step, BMS receives an (updated) tumour cell concentration map from OS (1). Since both simulators operate on distinct domain discretisation (OS: tumour only, regular grid; BMS: tumour & healthy tissue, unstructured mesh), OS cell concentrations are mapped into the BMS domain (2). From this information, spatial maps of growth, stress and pressure are computed (3-5) as described in section BM2. The Mathematical Model". The spatial pressure distribution is then mapped back from the BMS simulation domain into the regular grid of the OS (6) where the pressure gradient is computed. Based on this, a map of the "direction of least-pressure" for every grid cell is returned to OS (7) where it informs the movement and redistribution of tumour cells. This process is repeated until the hyper-model execution loop terminates.

BM6.1 Additional BMS Functionalities

As described in the foregoing section, personalised FEM simulations on a larger domain demand the creation of patient-specific meshes with adaptable element sizes. For this purpose, BMS integrates a meshing tool (using CGAL and VTK libraries) that generates custom meshes based on a segmentation of the domain of interest and a set of meshing parameters.

One challenge for the biomechanical simulations consists in handling large deformations due to tumour growth. These deformations lead to distortion of finite elements in the given domain discretisation, which may result in non-convergence of the computational problem. A 're-meshing' mechanism has been implemented that creates a new high-quality mesh (with the same mesh properties as original undeformed one) from the current simulation domain. It uses above mentioned

meshing tool, and is automatically invoked when the FEM backend fails to converge in the first solution attempt. This approach is promising for cases in which deformation and element distortion accumulates over multiple simulations, however, at this moment it does not address excessive deformations that occur within a single BMS execution. BMS as well as the integrated re-meshing mechanism do not yet support "pre-straining"/"pre-stressing" of the updated (deformed) simulation domain for the next time-step.

Parameter exchange between BMS and OS, as well as the re-meshing and future pre-straining/stressing procedure, require spatial parameter distributions to be mapped between domain discretisations: between structured grids (OS) and unstructured meshes (BMS), or between different unstructured meshes of (parts of) the same domain. A mechanism for this mapping is currently implemented as intrinsic part of the BMS simulator using standard VTK filters.

BM6.2 Common Functionalities and Recommendations

Both functionalities, mesh creation from image segmentations and mapping of 3D parameters distributions between domains, are a commonly used in Finite Element or Finite Difference based simulations. Although not part of the actual (mathematical) model, these 'convenience functions' are crucial for functioning simulator components and their implementation is time-consuming. Furthermore, their reliability and accuracy are of utmost importance to ensure correctness of the individual hypo-model as well as the composed hyper-model simulation.

We believe that each of these functionalities could be well encapsulated in a standalone hypo-model in the future to be made available on the shared computational platform. This would not only greatly facilitate the creation of new personalised FEM models and the parameter exchange between other component simulators; it would also ensure consistent handling of these critical simulation and communication aspects across the platform.
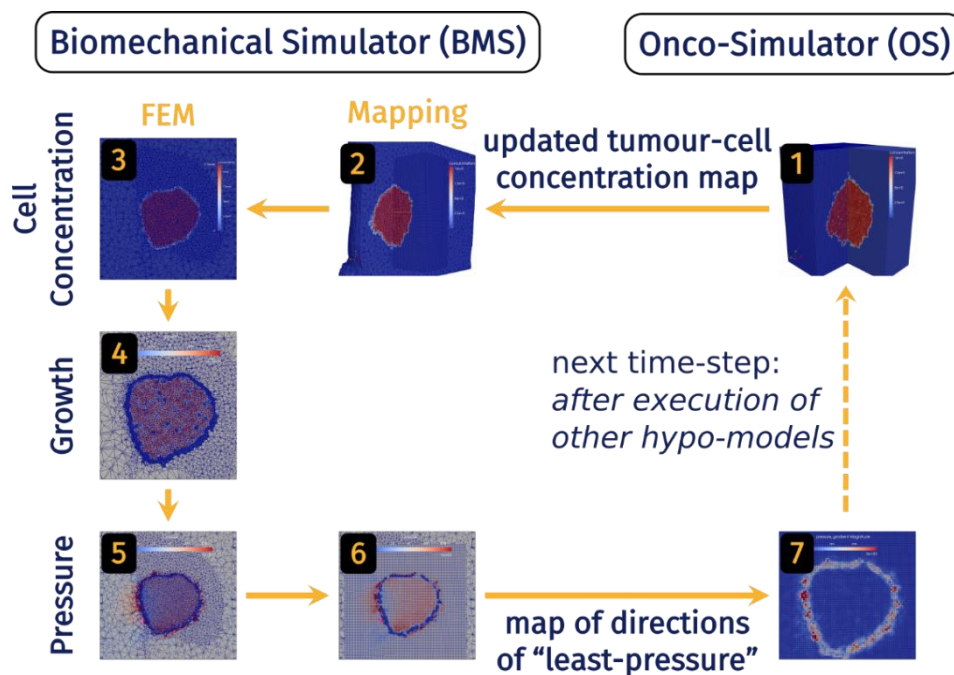
Fig. BM6: Data exchange and Computation in BMS-OS coupled execution.

BM7. Modelling Healthy-Tissue Invasion in GBM

Aggressive tumours are characterised by fast proliferation and their ability to invade healthy tissue. This characteristic is a prominent feature of Glioblastoma multiforme (GBM), the most frequent and most malignant subtype of glioma.

A mechanically coupled reaction-diffusion model has been set up to investigate this effect. It follows the mathematical model outlined in section "BM2. The Mathematical Model" "and has been implemented in Abaqus (Simulia) as (mathematically equivalent) heat-transport problem with thermo-elastic coupling. Integration of this model into BMS requires further study and discussions among the involved partners. The remainder of this section provides an overview of parameters choices and first simulation results.

Grey and White Matter are assigned isotropic diffusion coefficients DWM / GM of different magnitudes (D_{WM} = 5 D_{GM} [BM8]) representing the motility of GBM cells in the respective tissue. Parameter choices for diffusivity and proliferation are based on literature values estimated from clinical observations in glioblastoma patients. The model uses normalised cell concentration values. Mechanical coupling is achieved via a linear expansion coefficient λ=0.15 corresponding to a maximum tumour-related local strain of 15% [BM9].

Fig. BM7 shows a rendering of the concentration 80% iso-contour, corresponding to the bulk tumour visible on T1Gd MRI images, at different time points during tumour development. The second panel shows the distribution of cell concentration and clearly indicates the presence of tumour cells beyond the boundaries of the visible tumour.

Fig. BM7: Growing tumour from mechanically-coupled reaction diffusion model at different time steps. Upper panel: tissue types and tumour; lower panel: cancer cell concentration. The rendered tumour surface corresponds to relative cell concentrations > 0.8.

Fig. BM8 illustrates the mass-effect caused by the growing tumour, leading to displacements up to 4mm in the simulated configuration.
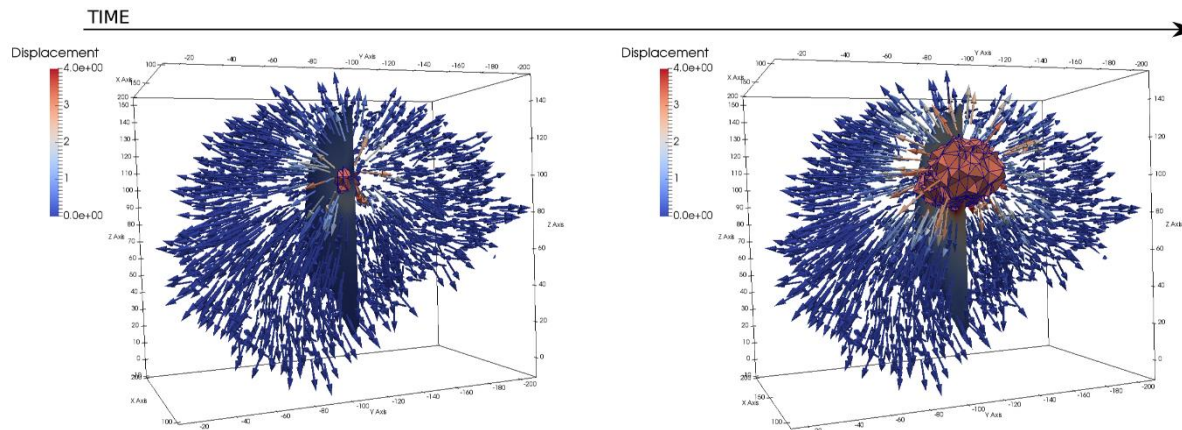
Fig. BM8: Displacement fields due to tumour mass-effect at different stages of tumour growth.
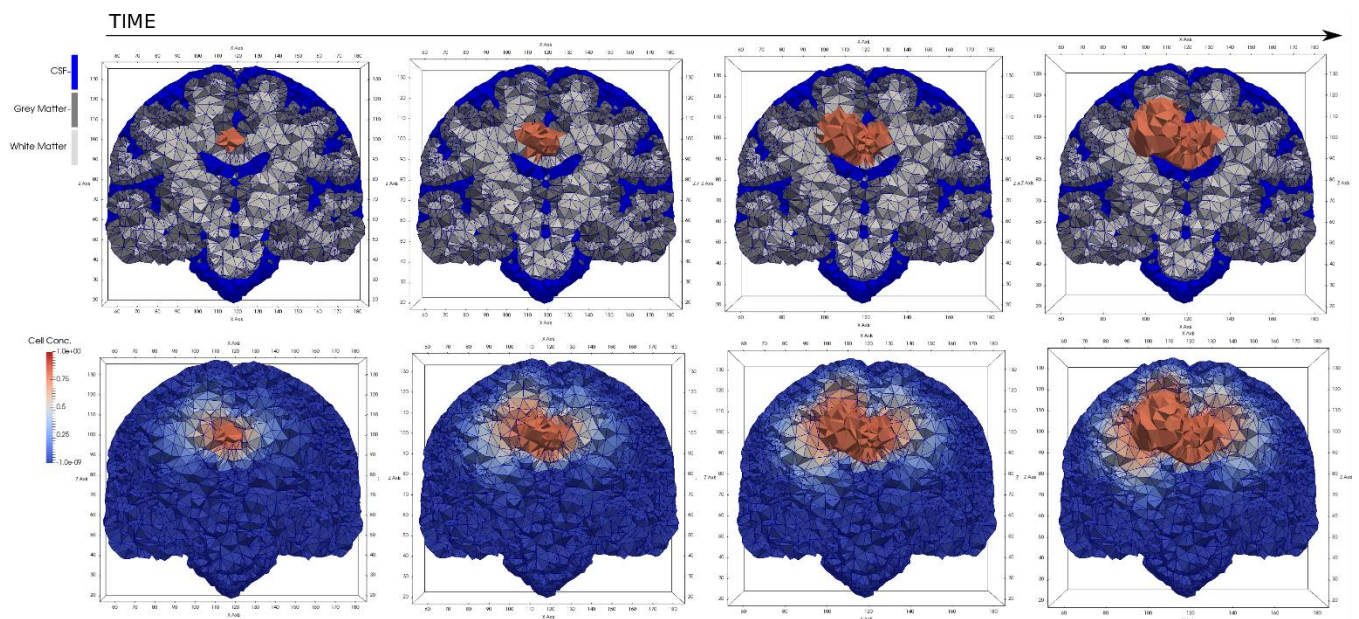
BM8. References

BM1. P. Tracqui, "From passive diffusion to active cellular migration in mathematical models of tumour invasion," Acta Biotheor., vol. 43, no. 4, pp. 443–464, Dec. 1995.

BM2. D. E. Woodward, J. Cook, P. Tracqui, G. C. Cruywagen, J. D. Murray, and E. C. Alvord, "A mathematical model of glioma growth: the effect of extent of surgical resection," Cell Proliferation, vol. 29, no. 6, pp. 269–288, Jun. 1996.

BM3. P. K. Burgess, P. M. Kulesa, J. D. Murray, and E. C. Alvord, "The Interaction of Growth Rates and Diffusion Coefficients in a Three-dimensional Mathematical Model of Gliomas:," Journal of Neuropathology and Experimental Neurology, vol. 56, no. 6, pp. 704–713, Jun. 1997.

BM4. R. A. Fisher, "THE WAVE OF ADVANCE OF ADVANTAGEOUS GENES," Annals of Eugenics, vol. 7, no. 4, pp. 355–369, Jun. 1937.

BM5. J. Borgdorff, M. Mamonski, B. Bosak, K. Kurowski, M. Ben Belgacem, B. Chopard, D. Groen, P. V. Coveney, and A. G. Hoekstra, "Distributed multiscale computing with MUSCLE 2, the Multiscale Coupling Library and Environment," Journal of Computational Science, vol. 5, no. 5, pp. 719–731, Sep. 2014.

BM6. E. Rios Velazquez, R. Meier, W. D. Dunn Jr, B. Alexander, R. Wiest, S. Bauer, D. A. Gutman, M. Reyes, and H. J. W. L. Aerts, "Fully automatic GBM segmentation in the TCGA-GBM dataset: Prognosis and correlation with VASARI features," Scientific Reports, vol. 5, p. 16822, Nov. 2015.

BM7. S. A. Maas, B. J. Ellis, G. A. Ateshian, and J. A. Weiss, "FEBio: finite elements for biomechanics," J Biomech Eng, vol. 134, no. 1, p. 11005, Jan. 2012.

BM8. K. R. Swanson, E. C. Alvord, and J. D. Murray, "A quantitative model for differential motility of gliomas in grey and white matter," Cell Prolif., vol. 33, no. 5, pp. 317–329, Oct. 2000.

BM9. O. Clatz, M. Sermesant, P.-Y. Bondiau, H. Delingette, S. K. Warfield, G. Malandain, and N. Ayache, "Realistic simulation of the 3-D growth of brain tumours in MR images coupling diffusion with biomechanical deformation," IEEE Trans Med Imaging, vol. 24, no. 10, pp. 1334–1346, Oct. 2005.

AN. The Angiogenesis Hypomodel for Nephroblastoma and Non Small Cell Lung Cancer

The development of the Angiogenesis Hypomodel for Nephroblastoma and Non Small Cell Lung Cancer has been led by UOXF. ICCS has ensured the compatibility of the hypomodel with the rest of the interacting hypomodels as well as with the overarching topology of the corresponding hypermodels.

AN1. Overview

The Angiogenesis Hypomodel describes the transport of nutrients in the tumour at the tissue-scale. In addition, 'capillary–scale' models of nutrient transport and blood flow in tumour micro-vessels have been developed, informed by experimental measurements of mouse tumour micro-vascular structure. These models can complement and inform the tissue-scale transport model and may be used as part of the CHIC modelling framework for the purpose of scientific research.

AN2. Model Development

- Tissue-scale spatial model of tumour nutrient transport developed. The model has run successfully as part of the nephroblastoma and lung demonstrator hypermodels and is available on the model repository.

- Development of high-resolution, capillary-scale models of oxygen transport and blood flow in tumour micro-vessels. Micro-vessel structures and micro-fluidics measurements have been provided by experimental collaborators for model parameterization. Simulation code will be made available to the public later this year.

- Development of a capillary-scale model that simulates vascular tumour response to radiotherapy. Code will be made available to the public later this year.

AN3. Vasculature Hypomodel (Nephroblastoma and Lung)

AN3.1 Background

The vasculature plays a vital role in the transport of nutrients and therapeutics to tumours, with tumour size limited by its ability to co-opt and maintain a vessel network. Many approaches have been adopted for modelling the effect of vascularization on tumour growth. The simplest models account for the effects of the vasculature in a phenomenological manner, by describing its ability to support tumour growth through a 'carrying capacity' term. A well-known example is the model of Hahnfeldt [AN1], which describes the rate of change of tumour volume V as a function of carrying capacity K:

$$V' = -\lambda V log\left(\frac{V}{K}\right)$$

where λ is a growth rate parameter. In this model, the carrying capacity is a phenomenological, non-spatial, variable that represents the ability of the vasculature to supply nutrient to the tumour. The adopted rate of change of the carrying capacity term:

$$K' = -\lambda_2 K + bV - dKV^{\frac{2}{3}} - eKg(t)$$

where λ_2, b, d, e are constants and $g(t)$ is the time dependent concentration of anti-angiogenic drug, assumes that the tumour is spherical. Advantages of models of this type are that they have few parameters and are simple to implement, verify and validate. Disadvantages are that they are non-spatial, assuming a spherically-symmetric tumour, and contain system parameters such as carrying capacity do not have a clear physical meaning. The latter makes it challenging to include information from histology (micro-vessel density) or functional imaging (vessel volume fraction, permeability, perfusion) within the model.

Detailed spatial models of nutrient transport, vessel growth and vessel occlusion have been developed at the tissue-scale [AN2]. These models allow tumour growth to be described in a more mechanistic fashion, relaxing the assumption of spherical geometry and allowing quantities such as vessel density and permeability to be explicitly described. In theory, such models can be parameterised with data related to vessel volume fractions and permeability from functional imaging; however this has not yet been performed. A disadvantage of such models, is that they contain many more parameters than their non-spatial counterparts. In addition, their practical implementation is more challenging and computationally expensive. This makes testing, verification and validation more difficult. Also, vessels are described as a scalar density or volume fraction field. In reality the tumour vessel network is branching structure exhibiting spatial heterogeneity and multi-fractal characteristics. It is not clear whether transport in, and temporal evolution of, such a structure can be described using such tissue-scale descriptions.

A thrid approach for simulating vascularization and nutrient transport is to model at the 'capillary-scale' [AN3]. At this size-scale transport of nutrients and blood, angiogenesis and vessel regression can be described in a mechanistic manner. However, the application of such models demands spatial resolution far greater than that available in clinical imaging to resolve small scale structures such as individual blood vessels. In addition, many more parameters and more complex software are required to describe and implement such models.

Evidently, the alternative approaches for modelling tumour vascularization and nutrient transport have distinct advantages and disadvantages. An ideal approach for modelling these processes is to integrate all of the modelling approaches over the multiple biological size-scales involved in tumour growth. This is the goal of current modelling efforts by UOXF. Ultimately, a model used for clinical application needs to be computationally efficient and have a small number of parameters to ensure it can be readily tested, verified and validated. This precludes the direct use of the more detailed models. However, they can be used to inform the rules used in more simple implementations.

AN3.2 Model Requirements

This section describes the requirements for the vasculature hypomodel used in the multi-modeller hypermodels for nephroblastoma and lung cancer. These requirements were the basis for the final form of the model.

- Coupling: Hypermodels are built by 'connecting' individual hypomodels and 'piping' the output of one model into the input of another. Care is needed in the design of the model itself, and associated software, to allow for such coupling. In particular, handling spatial data and time-stepping need care.

- Exploitation of Clinical Data: The model should exploit available clinical data. In particular, for the vasculature component, three-dimensional tumour volumes from imaging and micro-vessel destiny from histology are available.

- **Verification:** Due to the target application of models in clinical decision making, and the coupled nature of hypermodels, it is important that models and software can be suitably tested. This has implications for model run-times, which should be relatively short. In addition, models should be sufficiently simple so that the entire range of behaviours of the hypermodel can be explored in tests for mathematical/numerical and software stability.

- **Validation:** It is important that the hypomodels do not have an excessive number of parameters, so that hypermodel predictions can be reliably validated with respect to clinical data.

With these requirements in mind, a simple model of nutrient transport was chosen for inclusion in the nephroblastoma and lung-cancer hypermodels. In particular, a steady-state tissue-scale transport problem with static vasculature was used to avoid challenges with time stepping when coupling models and to minimize the number of model parameters. A spatial description of the tumour was used to exploit available clinical imaging data, relaxing the assumption of a spherical tumour in the model of Hahnfeldt [AN1]. A regular grid based finite difference method was used to solve the transport equations. Lattice size is identical to that of the input model in the hyper-modelling pipeline, avoiding spatial interpolation problems.

The model is flexible regarding the use of parameters to facilitate the 'plug-and-play' nature of the hypermodelling framework. A generic nutrient (e.g. oxygen, glucose) can be modelled with suitable parameter choices, vessels can be described in terms of vessel volume fraction, surface area density or line density and cell populations can be described in absolute numbers or volume fractions. The model is described in detail in the next section.

AN3.3 Model Overview: Theory

The vasculature hypomodel assumes steady-state, diffusion-limited transport of nutrient with concentration c, which is supplied by the vasculature at a rate dependent on vessel amount (volume fraction or density) V, and is consumed by tumour tissue at a rate proportional to the number or volume fraction of viable cells P. The tissue is assumed to comprise a tumour region and a non-tumour region, as shown in Fig. AN1.
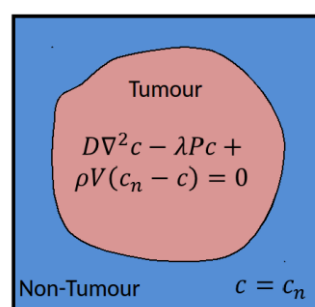
Fig. AN1. The simulation domain for the transport problem. Distinct tumour and non-tumour regions are assumed.

In practice these regions are determined based on segmentations of clinical images, performed in a pre-processing step external to the vasculature component in the hypermodel execution. In the non-tumour region it is assumed that the tissue is well vascularized and nutrient concentrations are set to a reference value c_n . In the tumour region nutrient transport is described according to:

$$D\nabla^2 c - \lambda Pc + \rho V(c_n - c) = 0$$

where D is the effective nutrient diffusion coefficient in the tumour tissue, λ (redefined) is the rate of nutrient consumption and ρ is rate of nutrient delivery by vessels. In this simple model the nutrient concentration in the tumour will approach the value in the surrounding healthy tissue as the vessel volume fraction increases, or as the rate of consumption by cells decreases.

Although the model has a physical basis, determination and physical interpretation of parameter values (D, λ, ρ, V, c_n) is challenging. In practice, these parameters should be treated in a phenomenological manner and used in fitting model predictions to observed clinical tumour growth (or shrinkage) rates. Discrimination of the relative influence of delivery by the vasculature and consumption by cells may be aided by additional observations of tumour micro-vessel density, vascular function through functional imaging or cell numbers by functional imaging. However, this has not been performed to date.

In order to obtain reasonable estimates for these parameters for model evaluation, literature values based on observations in tumour spheroids are adopted [AN4]. Within the hyerpmodelling framework dependent components use glucose concentrations to predict cell proliferation rates in the tumour. Although the mechanism of glucose consumption by cells also depends on oxygen availability [AN4], to preserve the simplicity of the model it is assumed that the diffusing nutrient is glucose and glucose consumption is independent of oxygen concentration. Parameter values are shown in Table AN1.

TABLE AN1. Assumed parameter values for the vasculature hypomodel for nephroblastoma and lung.

Description	Symbol	Unit	Value	Source
Glucose Diffusivity	D	mm^2.hr^{-1}	0.396	[4]
Glucose Consumption Rate	λ	(Num cells)$^{-1}$.hr^{-1}	7.6e-10	Modified from [4]
Glucose Concentration in Non Tumour Regions	c_n	Kg.m^3	0.9	Value used in metabolic hypomodel
Vascular Delivery Efficiency	ρV	hr^{-1}	0.25	User chosen/fit to data

The described transport model is simple, with considerable scope for improved biological realism and detail. However, it is deemed prudent that a simple model such as this is used first in the development, verification and validation of nephroblastoma and lung multi-modeller hypermodels. More detailed models, described subsequently, can be readily incorporated in the framework if justified by available clinical data.

AN3.4 Model Overview: Software and Integration

The nutrient transport problem is solved on a regular finite difference grid in 3D. This method was chosen for computational efficiency and to avoid interpolation when used with the grid-based descriptions of cell growth used in other hypomodels. In the demonstrator models the cell population P is received from the ICCS tumour growth hypomodel at each grid point. The vascular hypomodel output is the normalized glucose field ($\frac{c}{c_n}$) at each point on the grid. This is passed to the FORTH metabolic component, where it as averaged for the whole simulation domain. A sample model output for one of the nephroblastoma demonstrator cases is shown in Fig. AN2.
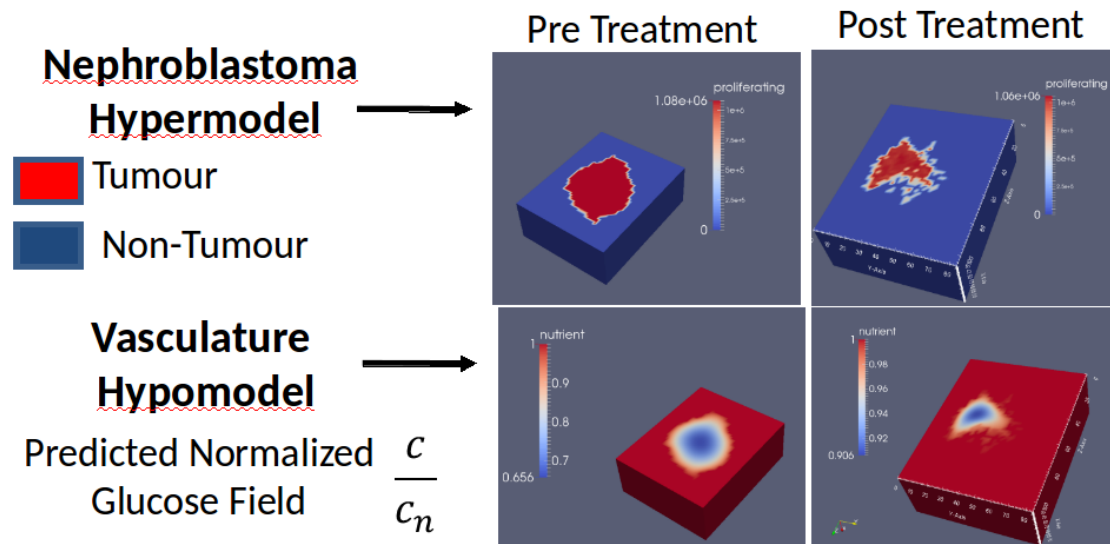
Fig. AN2. Input tumour cell populations and output normalized glucose fields pre-and post- treatment for the nephroblastoma case.

The software can output 3D renderings of the tumour and nutrient regions using VTK, as shown in Fig. AN2, although results are only used for debugging and are not passed along the hypermodel execution workflow. Models can be executed using either the CHIC TAVERNA or MUSCLE environments, or can run in standalone mode for testing purposes.

The model is implemented in the Chaste [AN5] open-source C++ framework for soft tissue modelling. As part of developments in this work package (WP6) a custom Chaste build was developed to allow the incorporation of MUSCLE libraries for run-time coupling of hypomodels and also packaging as a standalone executable. The latter was important for uploading software to the CHIC model repository and deploying to sandboxes on remote systems. A consequence is that many features of the Chaste library, including discrete cell based modelling, soft tissue mechanics and coupled PDE system solvers [AN5] can be used to rapidly build hypomodels for use with the CHIC based TAVERNA and MUSCLE execution workflows.

AN3.5 Limitations and Potential Extensions

The transport model used in in the vasculature hypomodel has several limitations, due to its generic treatment of a single 'nutrient' field. In reality, when used to model glucose transport, oxygen availability should also be accounted for, as per [AN4]. This coupling could readily be introduced to the model; however it results in the addition of extra parameters and makes validation, which is yet to be attempted for the simple case, more challenging.

A limitation of the current adoption of the vasculature hypomodel in the demonstrator hypermodels is that the metabolic hypomodel (FORTH) uses an independent model of glucose consumption. In theory, the rate of glucose consumption from the metabolic model could be passed back to the vasculature component and used to update glucose concentrations. However, this presents practical difficulties, as it would lead to an iterative spatial/non-spatial coupling between models, which is not accounted for in the current workflow design.

The vasculature is assumed to be 'static' in the current model, in that it does not evolve in time. Previous iterations of the model included temporal evolution of the vasculature (refer to previous WP6 deliverable reports, e.g. D6.2). These features were removed from the model as it was difficult to justify their inclusion from available clinical data, they increased computational expense and they

introduced difficulties managing time stepping with other hypomodels. If justified by available data they can be re-activated.

In addition to the tissue-scale transport model used in the clinical demonstrators, a range of more spatially resolved models of transport in tumour micro-vessels have been developed using Chaste [AN5]. These models can be used to inform the hypomodel used in the demonstrators, or may be useful as scientific tools in their own right.

AN4. High Resolution Angiogenesis and Vascular Transport Models

This section briefly overviews other model developments, which are not directly used in demonstrator hypermodels, but may be of scientific interest run used as individual hypomodels.

AN4.1 High Resolution Transport Models

Models of nutrient and drug transport in realistic tumour micro-vessel geometries have been developed, and are being parameterized based on collaborations with experimentalists in the Dept. of Radiation Oncology, University of Oxford and Centre De Recerca Matematica, Universitat Autonoma de Barcelona. Models include: i) finite element and network models of fluid flow in vessels structures, ii) finite element, greens function and finite difference models of advection-diffusion-reaction of solutes inside vessels and surrounding tissues and iii) discontinuous Galerkin finite element methods for transport through semi-permeable vessel walls. A typical work-flow from an intrivatal image of a tumour micro-vessel through to a simulation of nutrient convection and consumption is shown in Fig. AN3.
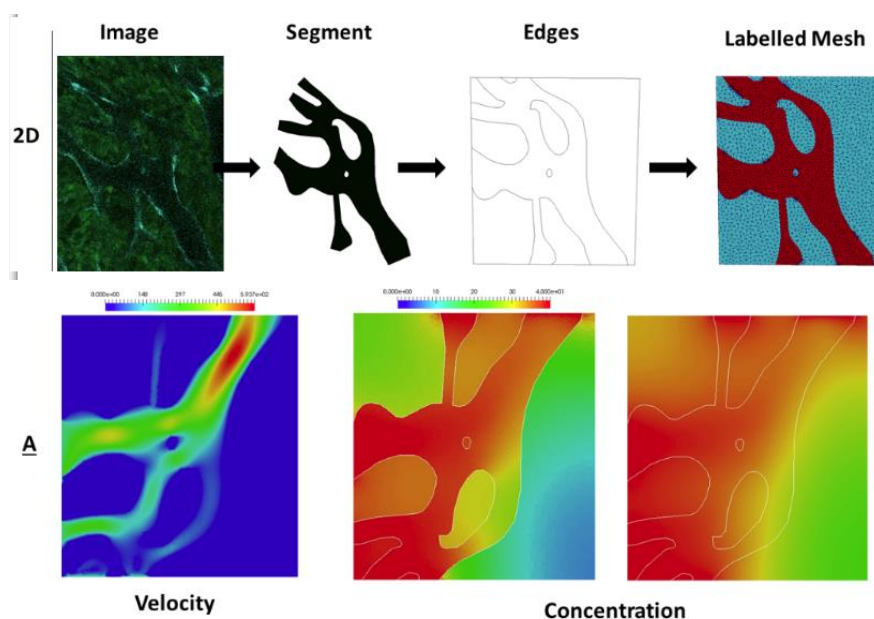
Fig. AN3. Pipepline from segmentation of a tumour micro-vessel through to simulation of blood flow and oxygen transport and consumption.

AN4.2 Effect of Vessel Network Structure on Predicted Radiotherapy Outcome

A 3D model of radiation therapy based on discrete representations of vessels and tumour cells has been developed and used to investigate whether different spatial descriptions of the micro-vessel network affects predicted radiotherapy outcome. Of particular interest is that it is predicted that models based on point-wise descriptions of the vessel network, which may be obtain from histology,

can overestimate the effect of spatial heterogeneity on radiation response. An example of the predicted tumour cell death at different dosing times is shown in Fig. AN4.
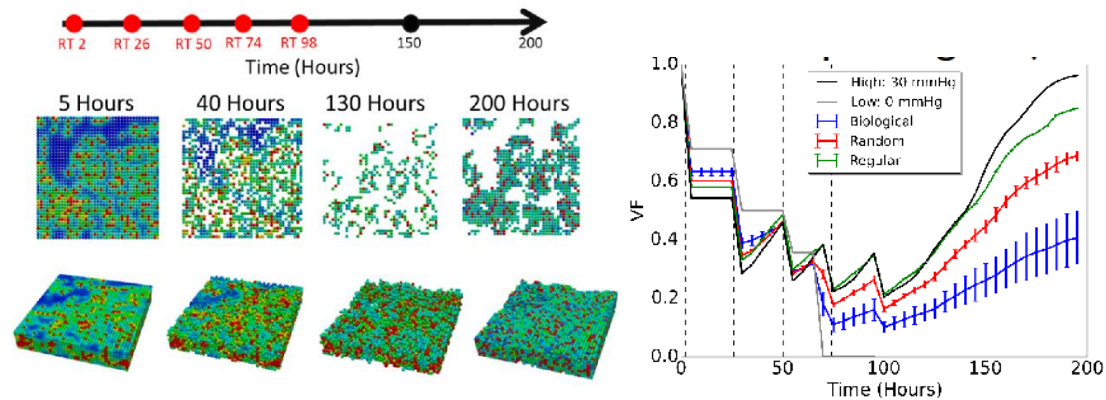
Fig. AN4. Left, the change in cell populations in 2D (above) and 3D (below) simulations of vascular tumour response to radiotherapy. Administration times for a 2 Gy dose are also shown. Right, predicted cell viable fractions for different methods of spatially representing vessels networks in 2D and 3D.

AN5. References

[AN1] P. Hahnfeldt, D Panigraphy, J. Folkman, L. Hatky, "Tumour Development under Angiogeneic Signalling: A Dynamical Theory of Tumuor Growth, Treatment Response and Postvascular Dormancy", Cancer Research 59, pp4470-4775, 1999.

[AN2] M.E. Hubbard, H.M. Byrne, "Multiphase Modelling of Vascular Tumour Growth in Two Spatial Dimensions, Journal of Theoretical Biology, 316, pp70-89, 2013.

[AN3] M.R. Owen, T. Alarcon, P.K. Maini, H. M. Byrne, "Angiogenesis and Vascular Remodelling in Normal and Cancerous Tissues", Journal of Mathematical Biology, 58, pp689-721, 2009.

[AN4] T. Roose, S.J. Chapman, P.K. Maini, "Mathematical Models of Avascular Tumour Growth", SIAM Review, 49(2), PP 179-208, 2007.

[AN5] G.R. Mirams et al., "Chaste: An Open Source C++ Library for Computational Physiology and Biology," PLoS Computational Biology, vol. 9, no.3, p. e1002970, 2013.

ME. The Metabolic Hypomodel for Nephroblastoma and Non Small Cell Lung Cancer (and Glioblastoma Multiforme)

The development of the Metabolic Hypomodel for Nephroblastoma and Non Small Cell Lung Cancer (and Glioblastoma Multiforme) has been led by FORTH. ICCS has ensured the compatibility of the hypomodel with the rest of the interacting hypomodels as well as with the overarching topology of the corresponding hypermodels.

ME1.Genome-scale Metabolic Modeling

Available high-throughput data allow the description of cellular metabolism at high resolution [ME1-ME3]. The reconstructed network, which considers the set of all biochemical reactions that take place within the cell is converted into a mathematical model and commonly analyzed under the constraint-based framework, since genome-scale kinetic information regarding reaction rate constants, enzyme concentrations and metabolite concentration, is largely missing. These methods have been widely used in the past to describe the metabolic capabilities of microorganisms. Constraint-based metabolic approaches utilize the genome-scale metabolic network reconstructions and describe the metabolic activity of the chemical reactions at flux level [ME3-ME5]. The core assumption of constraint-based models is that the cellular system constrained by its stoichiometry, reaches a steady state (intracellular flux balancing) that satisfies the physiochemical constraints under a given environmental condition. Flux Balance Analysis (FBA) is a constrained-based method that further assumes that a cell follows an optimization strategy in order to accomplish cellular tasks. The most commonly applied objective is the maximization of growth rate reflected in biomass production. Thermodynamic constraints, enzymatic capacity constraints and substrate availability can also be included to place limits on the range of possible reactions. The optimization problem is mathematically formulated to a linear programming problem, which can be solved efficiently and exactly providing the optimal flux distribution of the system.

ME1.1 Cancer Metabolic Modeling

ME1.1.1 Cancer Metabolism

Normal mammalian cells are exposed to a continuous supply of oxygen, glucose and other nutrients in circulating blood. Glucose is taken up by specific transporters and is converted to pyruvate in the cytoplasm through glycolysis generating 2 moles of ATP per glucose. In the presence of sufficient oxygen, pyruvate is then completely oxidized in mitochondria through oxidative phosphorylation generating additional 36 moles of ATP per glucose. However, when oxygen is insufficient, pyruvate is redirected away from mitochondrial oxidation and is converted to the waste product lactate. In contrast to normal cell metabolism, Warburg's observations [ME6] showed that cancer cells produce a substantial amount of energy inefficiently metabolizing glucose to lactate, independent on oxygen availability, a phenomenon termed as Warburg effect or aerobic glycolysis. The exact regulatory mechanisms of tumour metabolism are far from complete. The tumour microenvironment significantly affects the metabolic activity and rewiring. It alters metabolite transporters and glycolytic enzymes, while signaling pathways involving a number of oncogenes and tumour suppressor genes have been found to be implicated in the altered metabolism [ME7-ME8].

ME1.1.2. Growth Maximization Strategy

In order to model the metabolic adaptations of highly proliferating cancer human cells, Shlomi et al. [ME9] utilized a genome-scale human metabolic network accounting for 1496 ORFs, 3742 reactions and 2766 metabolites [ME1] and assumed that cancer cells are under a selective pressure to increase their proliferation rate. As commonly approached, they introduced the metabolic demands for biomass synthesis required for high proliferation rates and used the flux through biomass as the

objective function. In addition to that, they accounted for solvent capacity constraints to further limit the fluxes of the metabolic reactions and showed that this incorporation was capable of reproducing several metabolic characteristics observed experimentally during cancer development.

ME1.1.3 Lactate Maximization Strategy

We extend the work of Shlomi et al. [ME9] and apply a metabolic strategy that allows near optimal growth solution, while maximizing lactate secretion in order to describe the high-flux mechanisms that lead to a substantial increase of lactate production that is observed in tumour cells. Sub-optimal solutions have been observed to describe the metabolic capabilities of microorganisms under environmental stress and in the absence of sufficient evolutionary pressure [ME10] indicating that it is not unexpected for biological systems including cancer to show variability around optimal growth solutions. Furthermore, the lactate maximization strategy is built based on the experimental observations that i) support high lactate production by tumour cells, and ii) lactate secretion levels and proliferation rates do not always scale up. In particular, a negative correlation has been observed between proliferation and lactate production across the entire NCI-60 collection [ME11].

The lactate maximization strategy is mathematically described as a two-step optimization problem, similarly to the Flux Variability Analysis (FVA) method [ME12] which has been used to identify alternate optimal and sub-optimal metabolic states. The lactate strategy uses however an iterative procedure in an attempt to identify the minimal possible compromise in growth rate that achieves lactate production. More specifically, the first step solves the optimization problem as described in [ME9]. Cells are assumed to maximize their growth rate subject to flux balancing constraints, uptake bounds in the substrate reactions and the solvent capacity constraint. Under these specific constraints, the first problem identifies the maximum growth rate.

The second optimization problem aims to maximize the lactate production subject to flux balancing, uptake bounds and the solvent capacity constraint but also the constraint that the growth rate is not less than a given percentage, k, of the optimal growth rate determined in the first problem. As long as lactate secretion rate is less than a value of tolerance (0.01umol/mgDW/h), the second step is repeated for smaller values of k until a solution is found. As lactate rate is conversely related with cellular growth rate, varying k from maximum to lower values, the model provides a solution that is closer to optimal growth. The model has shown to capture several metabolic phenotypes observed experimentally in cancer and that slight deviations around the optimal growth rate (90-99%) were sufficient for adequate lactate production with increasing deviations to be observed at lower glucose uptake bounds.

ME1.2 Context-specific Metabolic Modeling

The genome-scale human metabolic network [ME1] is a global reconstruction superimposing the metabolic functions found in a variety of human cells and tissue types. Considering that the inefficient glycolytic phenotype of cancer cells, which is mainly characterized by highly proliferating rates as well as the significantly increased glucose uptake and lactate production comprise a common characteristic of most tumours, the global genome-scale human metabolic network is plausibly a network to begin with and apply constraint-based metabolic strategies that can accommodate these observations while taking into account the interconnectivity of the metabolic reactions. Nevertheless, high-throughput analytical methods have shown increased variability in metabolism among different cancer types as well as extensive intra-tumour heterogeneity within the same tumour [ME13]. Thus, the glycolytic phenotype is an average behaviour where extensive heterogeneity and stochasticity in the metabolic reactions and capabilities can be also present. In order to demonstrate how tumours with different metabolic capabilities might evolve, reported differentially expressed metabolic proteins are used.

The development of context-specific models that integrate several types of data from transcriptomic to proteomic and metabolomics in order to construct the active tissue/cancer specific metabolic model that allows individualized medicine and therapy planning is anticipated. Although several approaches have been proposed [ME14-ME18], the problem remains challenging.

In order to construct a tumour-specific metabolic model in a simplified manner, we included constraints in the metabolic reactions of the model, which are associated with bibliographically reported differentially expressed metabolic genes/proteins in these tumours. mRNA levels cannot accurately determine enzyme concentrations as inaccuracies in experiments, post-translational modifications and other effects might occur. However, they can determine an upper bound on the amount of available enzyme concentrations. In particular, enzyme levels, E_i, bound the fluxes of the corresponding metabolic reactions v_i through $v_i = Kcat_i E_i$, where $Kcat_i$ corresponds to the enzyme's turnover number. However, in the absence of quantitative information, metabolic reactions catalyzed by up-regulated metabolic proteins/enzymes, are constrained to carry non-zero fluxes via a lower bound, which is set equal to 0.1umol/mgDW/h unless stated otherwise, for all the involved reactions. Different bounds have also been tested. It is important to mention that the level of flux bound substantially alters the metabolic capabilities of the cells. Downregulated genes constrain the corresponding reactions via an upper bound, which is usually set equal to zero unless stated otherwise.

ME1.2.1 Non-Small Cell Lung Cancer (NSCLC)

NSCLC is a leading cause of cancer mortality worldwide (> 900000 deaths/year). NSCLC is a highly glycolytic lung cancer accounting for more than 85% of all lung cancers. Two major NSCLC subtypes have been reported: adenocarcinoma (LADC) and squamous cell carcinoma (LSCC). It has been shown that aerobic glycolysis in NSCLC is promoted through oncogenic mutations in two critical proteins, K-RAS and EGFR [ME13, ME19]. Ras-driven cancer cells display increased glucose uptake and aerobic glycolysis that support both nucleotide biosynthesis and protein glycosylation for growth signaling. However, it should be noted that high heterogeneity in metabolism proteome has been observed i)compared to normal lung tissue, ii)between lung subtypes and iii) between primary and metastatic lung cancer.

Recently, an extensive omics analysis [ME13] integrating DNA, RNA and proteomics data from normal lung, patient primary tumours and primary tumour-derived xenograft tumours revealed sets of proteins that are consistently up- or downregulated across tumours, recapitulated in xenograft tumours and their associated genes map into regions of focal amplification or deletion respectively. This DNA->RNA->protein association indicates a response to selective pressure driving cancer phenotype. From the reported metabolism proteome clusters in [ME13] , we used specific clusters of proteins consistently upregulated in LADC (cluster index: C15) and LSCC (cluster index: C10) to constrain the corresponding metabolic fluxes of the genome-scale metabolic network. It is also important to mention that individual proteome clusters have been correlated with overall survival in cancers other than NSCLC.

 Interestingly, for some clusters, the patient's outcome –better or worse- did not coincide among the different cancer types. For example, cluster C10 was associated with better outcome in lung SCC, but with worse outcome in head and neck SCC, indicating that the whole proteome signature associated with a specific cancer type and not individual clusters is important for predicting a patient's outcome.

ME1.2.2 Glioblastoma Multiforme

Gliomas are one of the most common tumours that originate in the Central Nervous System (CNS). Glioblastoma multiforme (GBM) is the most aggressive type of glioma, classified as grade IV by the World Health Organization. GBMs are highly therapeutically resistant with median survival of 12-16

months despite surgery, and advanced radiation and chemotherapeutic protocols. The main characteristics of GBM include cellular polymorphism, brisk mitotic activity, increased microvascular proliferation, extensive necrosis with regions of pseudopalisading perinecrotic cells, high degree of invasiveness and infiltrative edema. Furthermore, GBMs are highly glycolytic upregulating glycolysis more than three times that of normal brain tissue producing increased amounts of lactate. Numerous differentially up- and downregulated metabolic genes have been identified in relation to normal brain tissue [ME20]. These genes have been mainly related to 1) glucose metabolism, 2) fatty acid metabolism, 3) nucleotide metabolism and 4) glutamine metabolism. The set of these genes has been used to constrain the generic metabolic network and produce a GBM-specific metabolic model [ME21].

ME1.2.3 Nephroblastoma

Wilms tumour (WT), also known as nephroblastoma, is the most common malignant pediatric kidney tumour. WTs are believed to arise from the malignant transformation of renal stem cells that abnormally persist after embryogenesis and maintain embryonic differentiation capacity. Although there are a few studies that have shown metabolic alteration related to glycolytic phenotype in WT, unfortunately, there are no thorough studies currently available, which have investigated the metabolism of WT in detail. There is only indirect evidence from the genes altered in WT that there are alterations of cell metabolism. Thus, in the absence of bibliographic or other data, we use the generic cancer metabolic model to describe WT metabolism, which can be supported by the fact that Wilms' tumour cells are believed to derive from pluripotent embryonic renal precursor cells.

ME2. Integration of the Cancer Metabolic Hypomodel with the Vasculature and the Tumour Evolution Components

As tumour grows well-vascularized regions providing sufficient nutrients to cancer cells can coexist with nutrient-limited regions within the tumour mass. In this work, glucose is assumed to be the only limiting resource, although oxygen can also be incorporated as well as glutamine. As described in many studies [ME22-ME23], the dependence of glucose uptake on glucose concentration, C, can be modeled using Michaelis-Menten kinetics.

A relatively slow varying environment is assumed where cancer cells can operate at optimal or near optimal growth rates constrained by the current nutrient availability. As illustrated in Fig. ME1, at each time interval, the glucose concentration is estimated at every position in the computational grid. The spatiotemporal-dependent inflow of glucose flux is constrained by the Michaelis-Menten kinetics model and an instantaneous optimization problem is solved for each cell position and time point. During that time interval, the fluxes of the metabolic model are assumed constant. The metabolic model provides information regarding the uptake fluxes (e.g. glucose), intake fluxes (e.g. lactate) and proliferation rate given the available glucose that is then used from the tumour evolution hypomodel to update its state. For computational efficacy, the cellular metabolic capabilities have been pre-calculated for the different glucose concentrations and a lookup table has been used instead of repeatedly solving the optimization problem at each position and time interval.
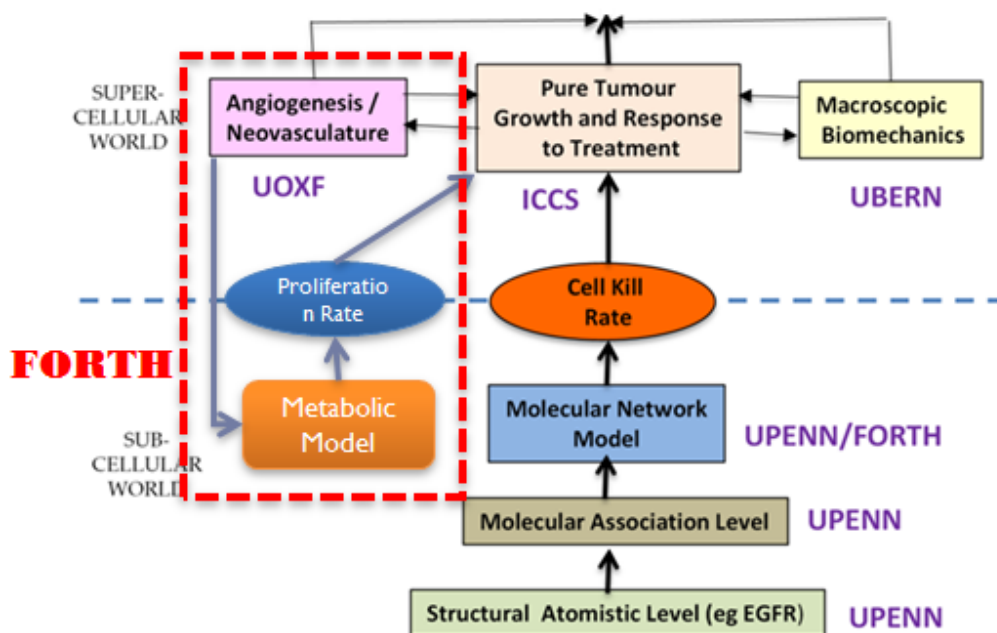
Fig. ME1. Schematic of the multi-modeller hypermodel framework with the metabolic hypomodel (FORTH) highlighted. The project patterns that are directly involved with the metabolic model of FORTH include the University of Oxford (UOXF) and the Institute of Communications and Computer Systems (ICCS) Greece. The University of Bern (UBERN) and the University of Pennsylvania (UPENN) are indirectly involved in the metabolic hypomodel through the tumour growth model and therapy response of ICCS.

ME3. References

[ME1]   N. C. Duarte, S. A. Becker, N. Jamshidi, I. Thiele, M. L. Mo, T. D. Vo, R. Srivas, and B. O. Palsson, "Global reconstruction of the human metabolic network based on genomic and bibliomic data," Proc Natl Acad Sci U S A, vol. 104, pp. 1777-82, Feb 6 2007.

[ME2]   I. Thiele, N. Swainston, R. M. Fleming, A. Hoppe, S. Sahoo, M. K. Aurich, H. Haraldsdottir, M. L. Mo, O. Rolfsson, M. D. Stobbe, S. G. Thorleifsson, R. Agren, C. Bolling, S. Bordel, A. K. Chavali, P. Dobson, W. B. Dunn, L. Endler, D. Hala, M. Hucka, D. Hull, D. Jameson, N. Jamshidi, J. J. Jonsson, N. Juty, S. Keating, I. Nookaew, N. Le Novere, N. Malys, A. Mazein, J. A. Papin, N. D. Price, E. Selkov, Sr., M. I. Sigurdsson, E. Simeonidis, N. Sonnenschein, K. Smallbone, A. Sorokin, J. H. van Beek, D. Weichart, I. Goryanin, J. Nielsen, H. V. Westerhoff, D. B. Kell, P. Mendes, and B. O. Palsson, "A community-driven global reconstruction of human metabolism," Nat Biotechnol, vol. 31, pp. 419-25, May 2013.

[ME3]   M. A. Oberhardt, B. O. Palsson, and J. A. Papin, "Applications of genome-scale metabolic reconstructions," Mol Syst Biol, vol. 5, p. 320, 2009.

[ME4]   A. M. Feist and B. O. Palsson, "The growing scope of applications of genome-scale metabolic reconstructions using Escherichia coli," Nat Biotechnol, vol. 26, pp. 659-67, Jun 2008.

[ME5]   K. J. Kauffman, P. Prakash, and J. S. Edwards, "Advances in flux balance analysis," Curr Opin Biotechnol, vol. 14, pp. 491-6, Oct 2003.

[ME6]   O. Warburg, "On the origin of cancer cells," Science, vol. 123, pp. 309-14, Feb 24 1956.

[ME7]   R. A. Cairns, I. S. Harris, and T. W. Mak, "Regulation of cancer cell metabolism," Nat Rev Cancer, vol. 11, pp. 85-95, Feb 2011.

[ME8]   M. G. Vander Heiden, L. C. Cantley, and C. B. Thompson, "Understanding the Warburg Effect: The Metabolic Requirements of Cell Proliferation," Science, vol. 324, pp. 1029-1033, May 22 2009.

[ME9]   T. Shlomi, T. Benyamini, E. Gottlieb, R. Sharan, and E. Ruppin, "Genome-scale metabolic modeling elucidates the role of proliferative adaptation in causing the warburg effect," PLoS Comput Biol, vol. 7, Mar 2011.

[ME10] R. U. Ibarra, J. S. Edwards, and B. O. Palsson, "Escherichia coli K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth," Nature, vol. 420, pp. 186-9, Nov 14 2002.

[ME11] M. Jain, R. Nilsson, S. Sharma, N. Madhusudhan, T. Kitami, A. L. Souza, R. Kafri, M. W. Kirschner, C. B. Clish, and V. K. Mootha, "Metabolite Profiling Identifies a Key Role for Glycine in Rapid Cancer Cell Proliferation," Science, vol. 336, pp. 1040-1044, 2012.

[ME12] R. Mahadevan and C. H. Schilling, "The effects of alternate optimal solutions in constraint-based genome-scale metabolic models," Metab Eng, vol. 5, pp. 264-76, Oct 2003.

[ME13] L. Li, Y. Wei, C. To, C. Q. Zhu, J. Tong, N. A. Pham, P. Taylor, V. Ignatchenko, A. Ignatchenko, W. Zhang, D. Wang, N. Yanagawa, M. Li, M. Pintilie, G. Liu, L. Muthuswamy, F. A. Shepherd, M. S. Tsao, T. Kislinger, and M. F. Moran, "Integrated omic analysis of lung cancer reveals metabolism proteome signatures with prognostic impact," Nat Commun, vol. 5, p. 5469, 2014.

[ME14] T. Shlomi, M. N. Cabili, M. J. Herrgard, B. O. Palsson, and E. Ruppin, "Network-based prediction of human tissue-specific metabolism," Nat Biotechnol, vol. 26, pp. 1003-10, Sep 2008.

[ME15] R. Agren, S. Bordel, A. Mardinoglu, N. Pornputtapong, I. Nookaew, and J. Nielsen, "Reconstruction of genome-scale active metabolic networks for 69 human cell types and 16 cancer types using INIT," PLoS Comput Biol, vol. 8, p. e1002518, 2012.

[ME16] S. A. Becker and B. O. Palsson, "Context-specific metabolic networks are consistent with experiments," PLoS Comput Biol, vol. 4, p. e1000082, May 2008.

[ME17] N. Vlassis, M. P. Pacheco, and T. Sauter, "Fast reconstruction of compact context-specific metabolic network models," PLoS Comput Biol, vol. 10, p. e1003424, Jan 2014.

[ME18] Y. Wang, J. A. Eddy, and N. D. Price, "Reconstruction of genome-scale metabolic models for 126 human tissues using mCADRE," BMC Syst Biol, vol. 6, p. 153, 2012.

[ME19] S. M. Davidson, T. Papagiannakopoulos, B. A. Olenchock, J. E. Heyman, M. A. Keibler, A. Luengo, M. R. Bauer, A. K. Jha, J. P. O'Brien, K. A. Pierce, D. Y. Gui, L. B. Sullivan, T. M. Wasylenko, L. Subbaraj, C. R. Chin, G. Stephanopolous, B. T. Mott, T. Jacks, C. B. Clish, and M. G. Vander Heiden, "Environment Impacts the Metabolic Dependencies of Ras-Driven Non-Small Cell Lung Cancer," Cell Metab, vol. 23, pp. 517-28, Mar 8 2016.

[ME20] A. Wolf, S. Agnihotri, and A. Guha, "Targeting metabolic remodeling in glioblastoma multiforme," Oncotarget, vol. 1, pp. 552-62, Nov 2010.

[21]   E. Tzamali, V. Sakkalis, and K. Marias, "The effects of near optimal growth solutions in genome-scale humanME cancer metabolic model," presented at the Bioinformatics & Bioengineering (BIBE), 2012 IEEE 12th International Conference, Larnaca, Cyprus, 2012.

[ME22] G. Cheng, P. Markenscoff, and K. Zygourakis, "A 3D hybrid model for tissue growth: the interplay between cell population and mass transport dynamics," Biophys J, vol. 97, pp. 401-14, Jul 22 2009.

[ME23] R. Chignola and E. Milotti, "A phenomenological approach to the simulation of metabolism and proliferation dynamics of large tumour cell populations," Phys Biol, vol. 2, pp. 8-22, Mar 2005.

MO. The Molecular Hypomodel for Nephroblastoma and Non Small Cell Lung Cancer

The development of the Molecular Hypomodel for Nephroblastoma and Non Small Cell Lung Cancer has been led by UPENN. ICCS has ensured the compatibility of the hypomodel with the rest of the interacting hypomodels as well as with the overarching topology of the corresponding hypermodels.

MO1. Introduction

The ErbB family of receptor tyrosine kinases and the signaling networks they influence have been implicated in a wide variety of cancers. The wild type and mutant forms of this receptor have critical role in determining tumour cell fates (death and proliferation) and resistance to various forms of targeted and systemic therapies. The ErbB family of receptors form multiple homo and hetero dimers and are activated by various ligands like epidermal growth factor (EGF), neuregulins (NRG). The activated dimers can turn on multiple signaling pathways downstream like mitogenic Ras-MAPK cascade and PI3K/AKT pathway. These pathways and their interaction with ErbB receptors have been extensively studied and modeled in literature using various computational paradigms like mechanistic compartmental/spatial models as well as statistical machine learning based models. These models can be of great clinical value in determining drug effectiveness, dosage and duration as well as investigating development of resistance to drugs and effect of intra and inter-tumoural heterogeneity.

The ErbB receptor mediated signaling pathways, important as they are, are however one of many other important determinant of tumour cell fate and drug resistance. Of these, tumour suppressor TP53 mediated signaling pathways are particularly important and they interact in important ways with the ErbB receptor mediated signaling pathways to determine ultimate cell fates. The p53 mediated signaling pathways are particularly important in determining tumour cell response to DNA damage as induced by chemotherapeutic drugs like doxorubicin and vincristine as well as radiation therapy. Due to its importance this signaling pathway has also been studied in detail. However due to its complexity and incomplete knowledge of the network, detailed mechanistic models are very few and not as predictive. These pathways have mainly been modeled using discrete two or multi-state network models. Although these models are simplistic, a variety of useful information regarding cell fate in response to DNA damage can be extracted from them and these predictions can be easily verified with experiments leading to further refinement of the models.

Among the important signaling modules mentioned above – ErbB receptor mediated Ras-MAPK and PI3K-AKT pathways and the TP53 mediated DNA damage response pathways have been modeled individually with different modeling paradigms to varying degree of success. However, there have been very few attempts to integrate these models into a combined cellular model. Such an integrated model will be of great clinical value due to its scope and its ability to test a great variety of situations to give results that can be directly useful to a clinician or can be used as an input to phenomenological models of tumour growth in response to chemo and radiation therapies. However, due to large differences in the relevant time and spatial scales of the two modules and due to the different nature of modeling paradigms which have been successfully applied to them, such an integration is a challenging task. In this work, we have successfully integrated these modules by identifying various interfaces from the literature and by developing a simple framework for information exchange keeping in mind the disparate time scales involved. By modeling the TP53 mediated signaling modules using a discrete network model we refine its predictions by running a continuum ordinary differential equations (ODE) based model of the ErbB receptor mediated signaling and passing information across the identified interfaces in both directions. In order to consider the effect of molecular profiling data from clinical subjects, in this work, we have also incorporated the miRNA expression data for various patients to re-normalize the initial expression levels of corresponding mRNAs to a given patient. This makes our model predictions patient-specific.

In doing so, however, we have also taken into account the heterogeneity of the microenvironment and have adopted an ensemble of models approach by averaging over multiple conditions of receptor expression, growth factor availability, and nature of the memory coupling signaling and transcriptional modules. This modeling framework has been successfully applied to lung cancer and nephroblastoma demonstrators where the model predictions in the form of a Cell Kill Rate (CKR) has been used as an input to phenomenological tumour growth models.

MO2. Model Description

As mentioned in the previous section, the model is composed of two main signaling modules – ErbB receptor mediated Ras-MAPK and PI3K/AKT signaling module and TP53 mediated DNA damage response module. The main features of model are summarized below:

Input:

1. Patient microRNA from tissue or serum; oncogene mutation
2. Treatment conditions- Radiation: dose and regimen

Model:

1. Tumour microenvironment- growth factor level, temporal growth factor availability, receptor expression, heterogeneity of the tumour microenvironment
2. Integration of cell cycle, DNA damage response, mitogenic and survival signals, drugs

Output:

Patient-specific cell kill probability, cell growth probability, and cell senescence probability in the presence and absence of specific drug/combinations, radiation treatment

The individual model components and the interfaces are described below:

A) ErbB Receptor Mediated Ras/Raf/MAPK and PI3K/AKT Pathways

Fig MO1: A schematic diagram of the ErbB receptor mediated Ras-MAPK and PI3K/AKT pathway. The black arrows indicate normal and red arrows indicate feedback interactions. From Schoeberl et. al. [MO1]

This part of the model has been adapted from Schoeberl et. al. [MO1] after suitable modifications. This is a multi-compartment model with 504 distinct species, 827 elementary mass action type reactions, 252 parameters. It consists of all the ErbB family of receptors and a subset of the homo and heterodimers. Though the model includes both EGF and HRG as the growth factors for simplicity we only consider the effect of EGF here. This model incorporates the effect of receptor internalization and recycling and the cross talks involved between the Ras-MAPK cascade and PI3K/AKT cascade. This has also been extended to take into account various mutant forms of EGFR receptor like L858R and deletion mutants which can be constitutively active.

This module gives the steady state concentrations of the various key downstream markers of the Ras-MAPK and PI3K/AKT cascade like activated ERK and AKT. The typical time to achieve steady state for this model is about 4-6 hours.

B) TP53 Mediated DNA Damage Response Module

This part of the model has been adapted and modified from Choi et. Al [MO2]. This consists of two main submodules corresponding to cell cycle progression and apoptosis. The first module consists of cyclin-CDK mediated cell cycle progression which affects the G1 to S phase entry of the cell cycle. The cell death module is the intrinsic apoptotic pathway which is mediated by key proteins like Bax and Bcl-2 which regulates the cell death protein Caspase 9. Damage to DNA activates ATM kinase

which in turn activates TP53 through kinases Chk1 and Chk2. Activated TP53 in turn activates DNA damage repair pathways or cell death pathways depending on the extent of damage. The ultimate cell fate will depend on the combined interactions of all the various components of the network and their initial activation state. This module consists of 16 nodes with 160 negative and 218 positive feedbacks.
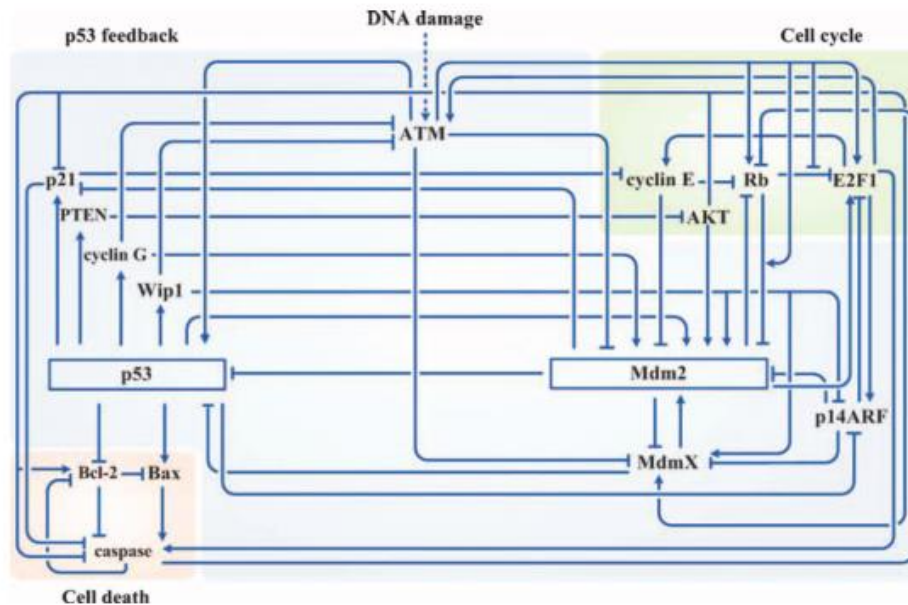

Fig MO2: A schematic diagram of the TP53 mediated DNA damage response pathway. From Choi et. al. [MO2]. The cell proliferation and apoptosis modules are shaded in green and red respectively. Arrows indicate activating interactions and dashes represent repression

Unlike the continuous ODE based ErbB signaling pathway model the TP53 module is modeled using a discrete Boolean model. Depending on set thresholds, each node of the network can have two possible states – ON or OFF. The interaction between the nodes is also a discrete number which can be both positive and negative depending on whether it activates or represses the downstream node. These kind of discrete models can give two possible outcomes:

i) **Point Attractor**: a single steady state where the activation state of all the nodes in the network do not change over successive time steps
ii) **Cyclic Attractor**: a sequence of repeating states (cycle)

For the current model there were three possible final outcomes which were identified with three different cell fates:

i) **Cell cycle progression**: (Markers- point attractor with high cyclin-G and low p53 activity)
ii) **Apoptosis**: (Markers – point attractor high p53 and high caspase activity)
iii) **Cell Senescence**: (Markers – cyclic attractor with oscillations in p53 and Mdm2)

TABLE MO1

Table 1. Stable states of the p53 network of a normal cell in the presence or absence of DNA damage. The cellular states are cell proliferation (P), cell cycle arrest (A), and cell death (D). White or dark blue boxes represent protein state as "off" or "on," respectively. The presence or absence of DNA damage is represented by "1" or "0," respectively, in column 1.

DNA damage _condition	Type	ATM	p53	Mdm2	Mdmx	Wip1	cyclin G	PTEN	p21	AKT	cyclin E	Rb	E2F1	p14ARF	Bcl-2	Bax	caspase	Basin size (ratio)	Cell state
0_normal	Point		■	■					■	■		■	■					65,536 (1)	P
1_normal	Cyclic	■		■					■	■		■	■					65,536 (1)	A
1_normal Wip1 inhibition	Point	■	■			■			■						■	■		65,536 (1)	D

C) Module Interfaces and Hybrid Simulator Algorithm

As mentioned in the introduction, in order to integrate the two above modules, we need to identify the specific interfaces between the modules. In this case we searched the literature to identify the various interfaces between the modules. These are given below:
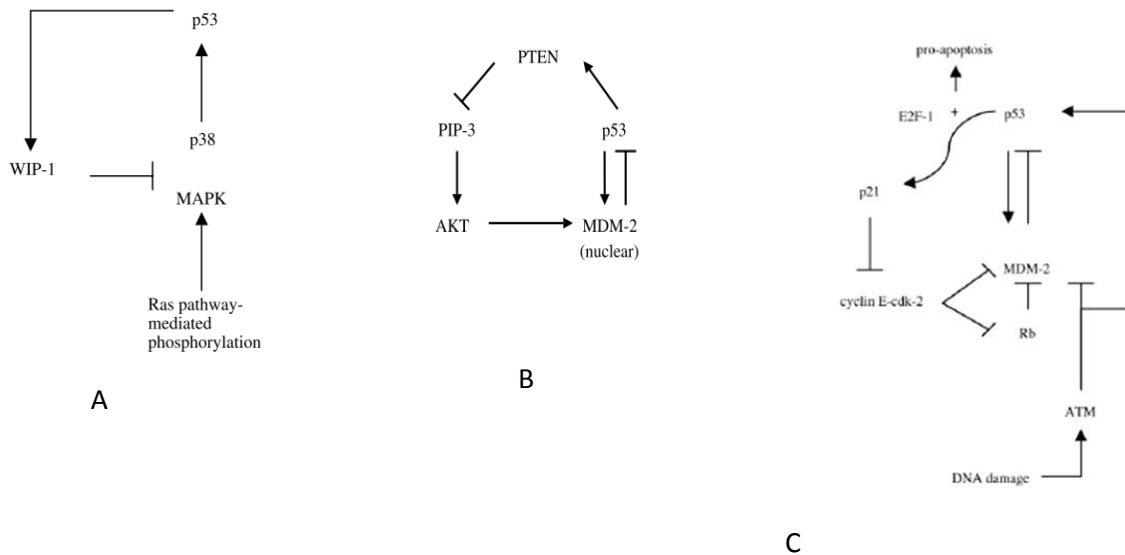
Fig. MO3: Various interfaces between the ErbB and p53 modules as identified from the literature. A – Interaction between p53 and MAPK pathway through Wip1 [MO3] B – The double negative interaction of p53 and Akt through PTEN and Mdm2 [MO4]. C- Interface between DNA damage and the nodes of the pathway.

The above interfaces form the conduit for information flow between the two modules. As seen above there are some species like Erk, Akt and PTEN which are common to both signaling modules. In this framework we run the modules sequentially where the final states of the interface nodes

obtained from each module is fed to the other module. Since the characteristic time scale of the two modules are quite different we assume that the faster ODE model will be at steady state during the start of each new time step of the slower Boolean model. The overall algorithm of the hybrid simulator is given below:
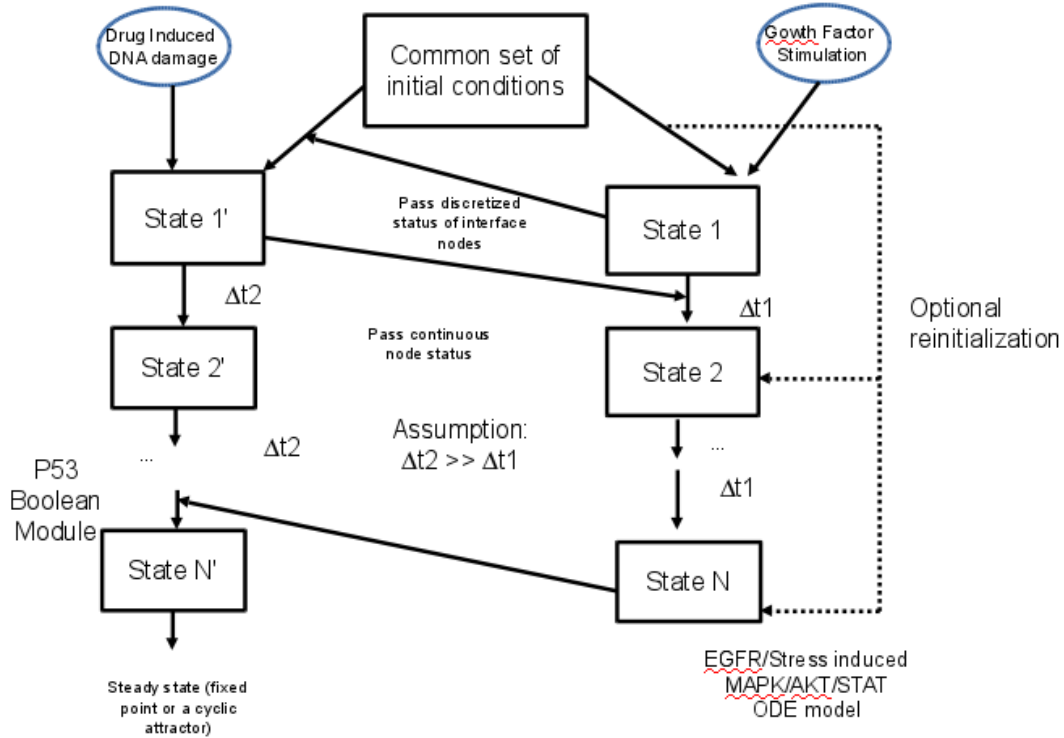
Fig MO4: Algorithm of the hybrid simulator. The Boolean and ODE modules are run sequentially with information regarding the states of interface species passed at the end of each run. The main assumption is that the time scale to steady state for the Boolean model is much larger than that of the ODE which allows us to pass only steady state information

Integration of the Boolean framework with the ODE framework occurs is performed using a hybrid framework as depicted below (Fig. MO5).

Piecewise linear DE for interface species (Boolean Network)

$$\frac{d\widehat{S}_i}{dt} = B_i\left(S_{i_1}, S_{i_2}, \ldots, S_{i_{k_i}}\right) - \gamma_i \widehat{S}_i$$

$$S_i = \begin{cases} 0, & C_i < \theta_i \\ 1, & C_i \geq \theta_i \end{cases}$$

Model Integration through interface species (Species common to both ODE and Boolean Network

$$\frac{dC_i}{dt} = f(C_{i_1}, \ldots, C_{i_{k_i}}, \widehat{S}_i C_i; kf_i, kr_i)$$

Continuous time DE for interface species (ODE Network)

Output: Network node fraction open vs. time

Output: Network node concentration/particle number vs. time

Fig. MO5: (Top) key nodes connecting the Boolean model with the ODE model. (Bottom, left) flow chart of interaction of Boolean model with ODE model. (Bottom, right) Information exchange between the Boolean model and the ODE model.

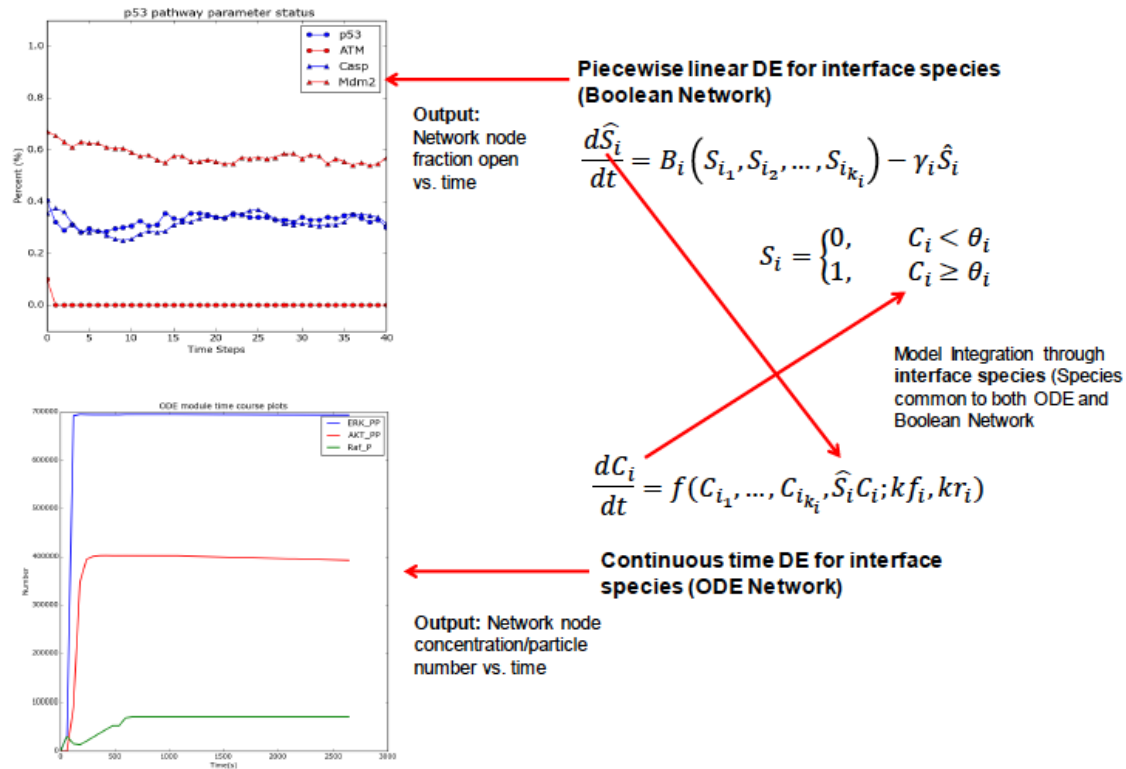In Fig. MO5, the MAPK, Akt modules are described by continuous time ODEs and p53 mediated apoptosis module is described by a Boolean system. The two modules communicate through the states of nodes, which are common to both modules; in this case is Erk, Akt and Raf. The states are described by a continuous time concentration function and a discrete time Boolean function. We assume that the reactions in the ODE module are much faster than the p53 module. This enables us to partially uncouple them by assuming that within the short time step of the ODE module the state of p53 module is invariant. Hence the Boolean module contribution to the state of the interface species is also constant and the only varying contribution is coming from the ODE module. The Boolean module will evolve with its own time scale but its behaviour will be modified by the information about the interface species it receives from ODE module. In Figs. MO4, MO5, the algorithm itself can be described as:

1. Run ODE for time steps tODE up to time tBool and get the time course data
2. Get the discrete states of the interface nodes (Erk, Akt and Raf) by applying appropriate thresholds
3. Run Boolean for time tBool while modifying the status of interface species at each tODE interval using the information received from ODE run
4. Get the activation fraction for the interface nodes at the end of the time step and pass it to the ODE model in the next run (Fig. MO5)
5. Go back to step 1 and repeat until steady state is achieved

The cellular states provided by the model are provided in Table MO1. The fingerprint of 0s and 1s of the proteins listed in the columns define microstates. The cellular states P (proliferation), A (cell

cycle arrest), and D(death) are mapped onto the microstates as given in Table MO2. The mapping is based on cell cycle progression (Markers- point attractor with high cyclin-G and low p53 activity); Apoptosis: (Markers – point attractor high p53 and high caspase activity); Cell Senescence: (Markers – cyclic attractor with oscillations in p53 and Mdm2).

D) Using miRNA Expression Data to Generate Patient Specific Predictions

Micro-RNAs are short non-coding RNAs that regulate gene expression post-transcriptionally by either inhibiting translation or promoting mRNA degradation. Hence not surprisingly these miRNAs have been found to play critical role in various forms of cancer. For CHIC lung and nephroblastoma demo normalized tissue and serum miRNA expression data are available in minml format. We were able to utilize this expression data by identifying the specific targets of top 20 miRNAs using miRTarBase [MO5]. For those mRNAs which are present in our network we adjusted the initial expression level accordingly before each model run. Hence the final outcomes were tailored to the particular expression profile of the patients to generate clinically useful outcomes. An example table is shown below indicating some top expressed miRNA and their corresponding mRNA targets obtained from miRTarBase [MO6].

The molecular hypermodel takes as an input, the miRNA profiles of a given patient from the CHIC data repository and then uses a database mapping (miRTARbase) to map the enriched miRNA in either tissue or serum of a patient to the corresponding mRNA, see Fig. MO6. The mRNA information is then used to constrain the nodes of our network in order to capture the molecular effect; similarly, nodes are constrained based on the drug interactions. In total, we consider doxorubicin, vincristine, and actinomycin as the three drugs for nephroblastoma. The model is run based on the input miRNA and drug treatment and averages over several tissue conditions such as growth factor levels and receptor expression. An average as well as a distribution of cell kill, cell senescence, and cell growth probabilities are obtained for a given patient, which are then passed on to the multi-modeler framework.

miRNA	WT vs. control*	Targets in Combined Network†
miR-223-3p	9.63	STAT5A, ATM, E2F1, TP53, MDM2
miR-143-3p	6.87	KRAS,MAPK,HRAS,AKT,MDM2,BCL2,MAP3K,CASP5
miR-130b-3p	6.47	RB,PTP,E2F,CASP,PTEN,MAP3K,WIP,CDK4
miR-17-5p	6.2	
miR-100-5p	5.29	ATM,MTOR,MAPK(ERK),BCL,PTP,GRB2,RB,AKT
miR-24-3p	4.37	E2F1, MDM2, MAP3K, MAP2K, MAPK, CDK, ATF, PTP, BCl2
miR-3651	3.58	
let-7f-5p	3.35	
miR-18a-5p	2.91	PTEN,ATM,BCL2,ARF6,TP53
miR-181-b-5p	2.86	BCL2,MAP3K,E2F1,PTEN,MTOR,ATM,ARF6,MAP2K,PDK3
let-7a-5p	2.75	
miR-335-5p	2.29	
miR-320a	1.71	
miR-126-3p	1.68	
miR-195-5p	1.67	
miR-101-3p	1.31	
miR-638	1.15	
miR-572	-1.29	
miR-1825	-2.55	

*miRNA Expression data from Ludwig, Nicole, et al. "Circulating serum miRNAs as potential biomarkers for nephroblastoma." Pediatric blood & cancer (2015).
†Target genes obtained from miRTarBase: (http://mirtarbase.mbc.nctu.edu.tw)

How will the miRNA data be combined with network model?
- Identify top 5-10 miRNA which are overexpressed in particular patient data pre and post CT
- Use miRTarBase to obtain a list of top target proteins (can be automated using a simple Perl/Python script)
- Run combined model with lower levels of activity or concentration (determined based on the expression levels of miRNA with respect to control) for the target proteins
- Report observed cell fate/cell kill rate

Fig. MO6: Algorithm for miRNA implementation in the molecular model.

MO3. Results

A) Application of the Model to Lung Cancer Demonstration

We have completed the implementation of the molecular model for lung cancer. In addition to the example above, the molecular modelling of the EGFR mutants has yielded the construction of networks specific to mutants, which encode dynamics and logic that are different from the wildtype. These, results have been combined with the Boolean model to yield a comprehensive molecular model for lung cancer. The model takes input from a patient-specific miRNA profile and renormalizes the nodes of the network. In addition, the model incorporates the sequencing information by considering the mutational status of EGFR, KRAS, BRAF, and AML/ALK in order to produce a customized network to match the patient's molecular profiles. We have further implemented the effect of radiation dosage in the model through the linear-quadratic model (LQ) for radiation treatment. In brief, the linear-coefficient is used to constrain the node that regulates the p53 activation is a radiation dose and regimen dependent fashion. The cell kill probability output by the model is then a function of the radiation exposure as well as the miRNA-mediated signalling. This property is then returned to the multi-modeler (described later in Fig. MO11) for further processing of the clinical outcome.

The above general framework was adapted to the lung cancer demonstrator. For this case the selected patient underwent a treatment regimen as described in Fig. MO7. As indicated, only the radiation treatment was considered for the review.

Imaging and treatment data overview:	Data	Date	Observations
	Dicom	1/5/1951	TUMOUR EVIDENT, SURGERY AND COMPLETE REMISSION
	Dicom	6/12/51	
	Dicom	5/2/53	
The radiation treatment (with 4 fractions in total) has been considered for the review	Surgery	4/3/1953	FIRST RECURRENCE
	Dicom	30/10/1953	TUMOUR NOT EVIDENT
	Chemo 1st session start	15/1/1954	
	Chemo 4th session end	26/3/1954	
	Dicom	25/5/1956	TUMOUR EVIDENT
	Dicom	31/8/1956	TUMOUR PROGRESSION
	Radiation start	6/9/1956	
	Radiation end	15/9/1956	
	Dicom	17/9/1957	NO TUMOR, RADIATION NECROSIS

Fig. MO7: Summary of treatment regimen for a lung cancer patient selected from the CHIC data repository.

Radiation dosage introduces DNA single and double strand breaks which activates the p53 mediated DNA repair and apoptotic pathways. The effect of radiation dosage on cell kill is modeled almost universally using the phenomenological Linear-Quadratic model [MO7] which gives the fraction of surviving cell as a function of radiation dosage (Gy) as below where D is the radiation dosage in Gy and α, β are empirical constants which correspond to dose dependent and independent parts of radiation response. G is the generalized Lea-Catcheside time factor.

$$S = exp(-\alpha D + \beta G D^2)$$

We chose to replace the empirical constants α and β by using information obtained from our detailed mechanistic model. However due to the discrete nature of our p53 mediate DNA damage model, we are only able to account for the dose independent part (the linear term involving α).

Future version will aim to account for the dose dependent part as well which requires a more detailed model which takes into account the various double strand DNA repair mechanisms. We activate the ATM kinase levels according to a probability exp(-α *D) to obtain the cell survival. This survival is modified by the quadratic term involving b which is obtained from the literature. The overall scheme is shown below in Fig. MO8.

Fig. MO8: Overall scheme for the lung cancer hypomodel. The input to the model consists of the usual miRNA and mutation data for the specific patient. Also we activate the DNA damage by calculating a probability based on the radiation dosage D (Gy-1). The cell survival fraction obtained from the model is modified using the empirically calculated quadratic term of the LQ model.

The specific radiation dosage data and values of the empirical constants are as given below in Table MO2:

Table MO2: LQ Model Parameters

LQ Model Parameters	
alpha (Gy-1)	0.35
beta (Gy-2)	0.035
Pat JGUP5Z Radiation Dosage	
No. of fractions (n)	4
G (approx)	0.25
Total Dosage D (Gy)	20

B) Predicted Cell Kill Rate for Demo Patient for the Lung Cancer Scenario

The model as adapted above was run using the available miRNA data and radiation dosage information to obtain cell kill fraction for different radiation dosages. The predicted cell kill rate from the molecular model was compared with that obtained from the empirical LQ model. We observe that the predicted cell kill rates converge at higher radiation dosages but molecular model prediction are lower compared to the LQ model at lower dosage fractions. The results obtained from the molecular model are averages over various growth factor and time scale considerations. The individual results and their distribution are also shown in Fig. MO9.

A

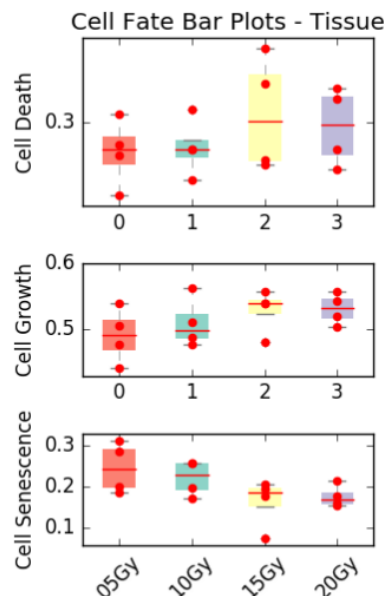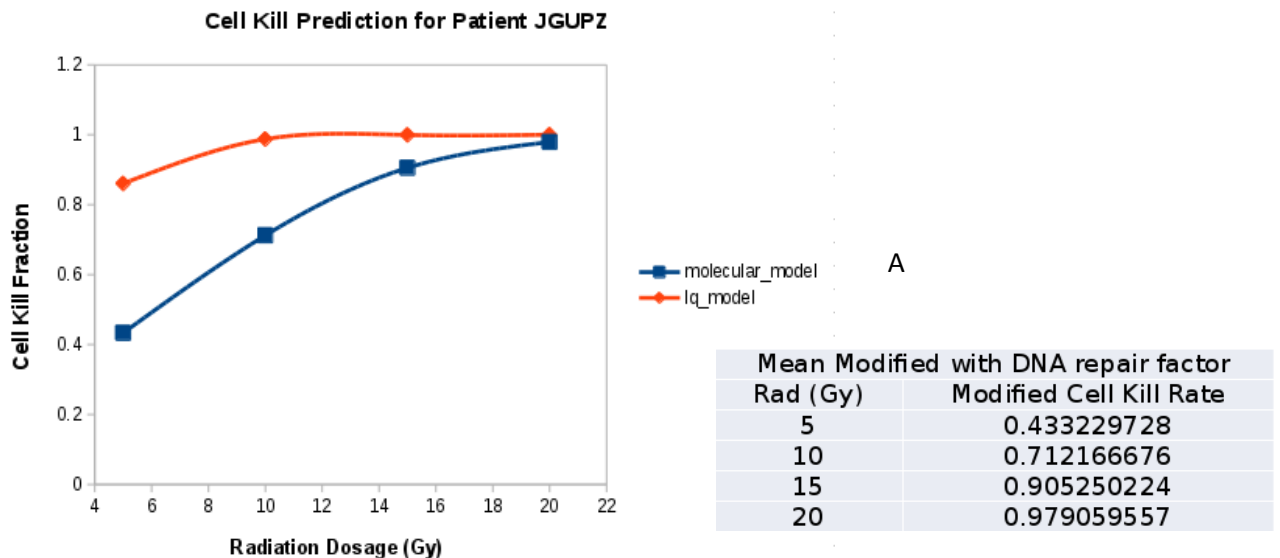| Mean Modified with DNA repair factor	
Rad (Gy)	Modified Cell Kill Rate
5	0.433229728
10	0.712166676
15	0.905250224
20	0.979059557

Fig. MO9: Main results of the lung cancer hypomodel. A – Cell kill fraction averaged over various growth factor concentration and cell cycle times as obtained from molecular model compared with the results obtained from the LQ model; B – Boxplots showing the distribution of the individual cases for different radiation dosages.

C) Application to the Nephroblastoma Scenario

In Fig. MO10, we describe the results of our model for a given patient in the nephroblastoma scenario selected from the CHIC data repository. The Table in Fig. MO10 shows the cell death probabilities predicted by the model under the action of different chemotherapeutic drugs (A: doxorubicin; V: vincristine; A: actinomycin). The panels below show the cell state transitions and the time evolution of key transcriptional activators.

Results Summary for Patient 5XIHQGQZ2GDYMITS5KON

Table 1: Cell Fate Probabilities

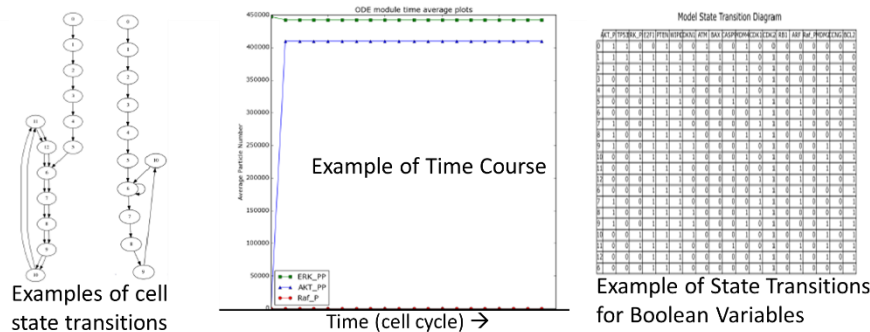| mirna-source	Drug(D,V,A)†	Cell-Death($\mu \pm \sigma$)	Cell-Growth($\mu \pm \sigma$)	Cell-Senescence($\mu \pm \sigma$)
tissue	(D-,V-,A-)	0.18($\pm$0.03)	0.51($\pm$0.15)	0.32($\pm$0.13)
tissue	(D+,V+,A+)	0.63($\pm$0.08)	0.08($\pm$0.07)	0.29($\pm$0.14)
tissue	(D-,V+,A+)	0.53($\pm$0.18)	0.12($\pm$0.15)	0.35($\pm$0.1)

† D-Doxorubicin, V-Vincristine, A-Actinomycin

Fig. MO10: Demonstration of the molecular model for a patient in the nephroblastoma scenario.

MO4. Conclusions and Future Work

The present work we describe an integrated cellular framework to model key cell signaling pathways operating at different time scales – a well-recognized challenge in the field. Here we model the ErbB receptor mediated Ras-MAPK and PI3K/AKT pathway and integrate it with p53 mediated DNA damage response pathway to obtain a cell kill rate under specific drug dosing and patient specific miRNA expression levels. The obtained cell kill rate was directly used as an input to phenomenological tumour growth models. The aim of such integrated molecular model is to provide a mechanistic foundation to the more empirical models used in the field to obtain cell kill and growth rates under particular dosage conditions. The integration of the models was accomplished by identifying model interfaces and passing information between runs of the two models. This framework has been tested for the lung cancer and the nephroblastoma scenarios and the future work will be focused on performing a detailed sensitivity analysis to simulate the inherent tumour heterogeneity and also the effect of various mutations and subject the framework to clinical validation. The integration of the UPENN molecular model with the multimodeler framework of CHIC is shown in Fig. MO11.

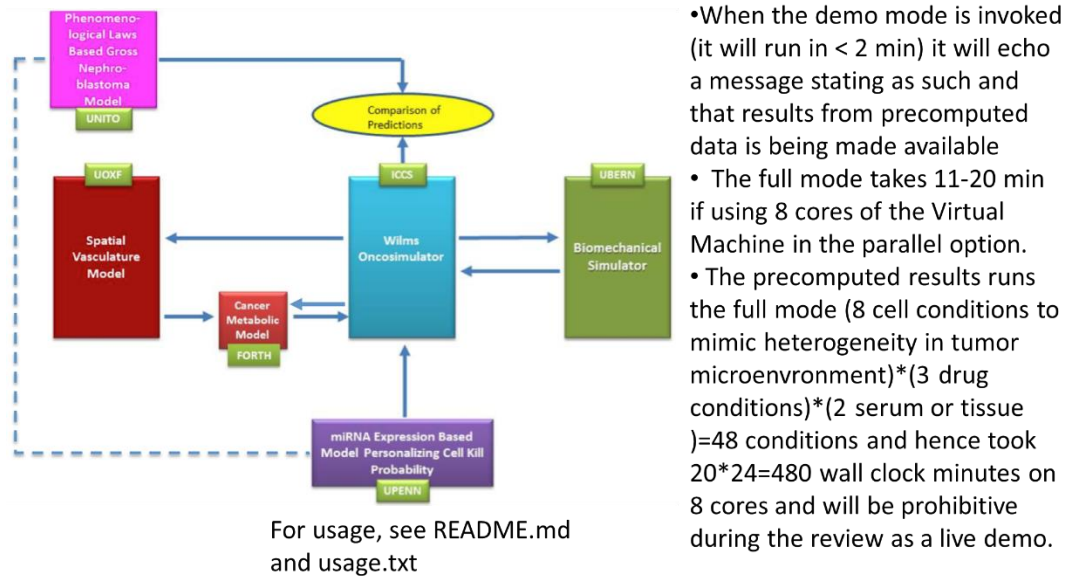# Integration of UPENN Model in the CHIC Multimodeler

•When the demo mode is invoked (it will run in < 2 min) it will echo a message stating as such and that results from precomputed data is being made available
• The full mode takes 11-20 min if using 8 cores of the Virtual Machine in the parallel option.
• The precomputed results runs the full mode (8 cell conditions to mimic heterogeneity in tumor microenvronment)*(3 drug conditions)*(2 serum or tissue)=48 conditions and hence took 20*24=480 wall clock minutes on 8 cores and will be prohibitive during the review as a live demo.

For usage, see README.md and usage.txt

Fig. MO11: Flowchart showing the passing of information between the hypermodels including the molecular hypermodel.

MO5. References

MO1. Chen, W. W., Schoeberl, B., Jasper, P. J., Niepel, M., Nielsen, U. B., Lauffenburger, D. a, & Sorger, P. K. (2009). Input-output behaviour of ErbB signaling pathways as revealed by a mass action model trained against dynamic data. Molecular Systems Biology, 5(239), 239.

MO2. Choi, M., Shi, J., Jung, S. H., Chen, X., & Cho, K.-H. (2012). Attractor landscape analysis reveals feedback loops in the p53 network that control the cellular response to DNA damage. Science Signaling, 5(251), ra83.

MO3. Cuadrado, A., & Nebreda, A. R. (2010). Mechanisms and functions of p38 MAPK signalling. The Biochemical Journal, 429(3), 403–417

MO4. Gottlieb, T. M., Leal, J. F. M., Seger, R., Taya, Y., & Oren, M. (2002). Cross-talk between Akt, p53 and Mdm2: possible implications for the regulation of apoptosis. Oncogene, 21(8), 1299–1303.

MO5. Chou, C. H., Chang, N. W., Shrestha, S., Hsu, S. D., Lin, Y. L., Lee, W. H.,Tsai, T. R. (2016). miRTarBase 2016: updates to the experimentally validated miRNA-target interactions database. Nucleic acids research, 44(D1), D239-D247

MO6. Ludwig, N., Nourkami-Tutdibi, N., Backes, C., Lenhof, H.-P., Graf, N., Keller, A., … Background. (1996). Circulating Serum miRNAs as Potential Biomarkers for Nephroblastoma. Journal of Clinical Oncology, 14(5), 1526–1531.

MO7. Brenner, D. J. (2008). The Linear-Quadratic Model Is an Appropriate Methodology for Determining Isoeffective Doses at Large Doses Per Fraction. Seminars in Radiation Oncology, 18(4), 234–239

GP. The Gross Phenomenological Hypermodel for Nephroblastoma and Non Small Cell Lung Cancer

The development of the Gross Phenomenological Hypermodel for Nephroblastoma and Non Small Cell Lung Cancer has been led by UNITO. ICCS has ensured the compatibility of the hypomodel with the rest of the interacting hypomodels as well as with the overarching topology of the corresponding hypermodels.

GP1. The Phenomenological Universalities (PUN) Approach to Describe Tumour Growth and Therapy

The PUN approach describes tumour macroscopic growth according to an unknown function which can be approximated by its series expansion. The first term of such expansion are the exponential function (N=0), the Gompertz function (N=1) and the West function (N=2). [GP1,GP2]

Apart from the exponential function, which describe an unlimited proliferation, both the Gompertz and the West approaches combine growth with some tumour-control term, which can be modelled to account for both environmental and/or therapeutic effects.

For Gompertz eqn:

$$\frac{1}{N}\frac{dN}{dt} = \alpha_0 \exp(-\beta t)$$

It is possible to assume that N=N1 + N2 being N1 the cells duplicating at constant rate and N2 those dying at the same time:

$$\frac{dN_1}{dt} = \alpha_0 \ N_1$$

$$\frac{dN_2}{dt} = -\alpha_0 \exp(-\beta t)N_2 - \alpha_0(1-\exp(-\beta t))N_1$$

$$\alpha_0 = \frac{\ln 2}{T}$$

being the growth rate, related to the duplication time T, and

$$\beta = \alpha_0 \ln(\frac{N_\infty}{N_0})$$

a parameter related to the carrying capacity N_∞

In case of therapy one or more 'kill rate' parameters can be added to the equation.

For West [GP3] eqn.:

$$\frac{dN}{dt} = aN^{3/4} - bN$$

$$\frac{a}{b} = N_\infty^{1/4} \rightarrow b = \frac{a}{N_\infty^{1/4}}$$

where the first term accounts for proliferation and the second one for death.
When therapy targets the growth rate, then a becomes $a' < a$, when it targets tumour bulk then N_∞ becomes $N'_\infty \ll N_\infty$

GP2. Modelling the response of pre-surgical chemotherapy in nephroblastoma and lung cancer

In NSC lung / nephroblastoma cancer pre-surgery chemotherapy is commonly performed to reduce tumour volume [GP4].

It is therefore necessary both to reduce tumour proliferation AND impact on its bulk volume.
Normally, at least two drugs should be combined in order to *reduce the bulk volume*:

- **ACTINOMYCIN-D** which affects both quiescent and proliferating (**BLACK DRUG**)

- **VINCRISTINE** which **delays the cell proliferation** (**WHITE DRUG**)
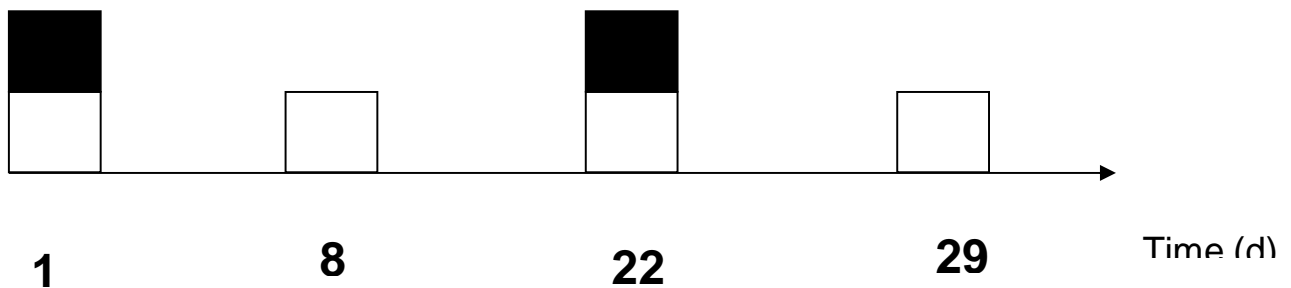
This treatment can be modelled easily with the PUN approach, inserting two kill rates (one for each drug) acting in the days of the treatment.
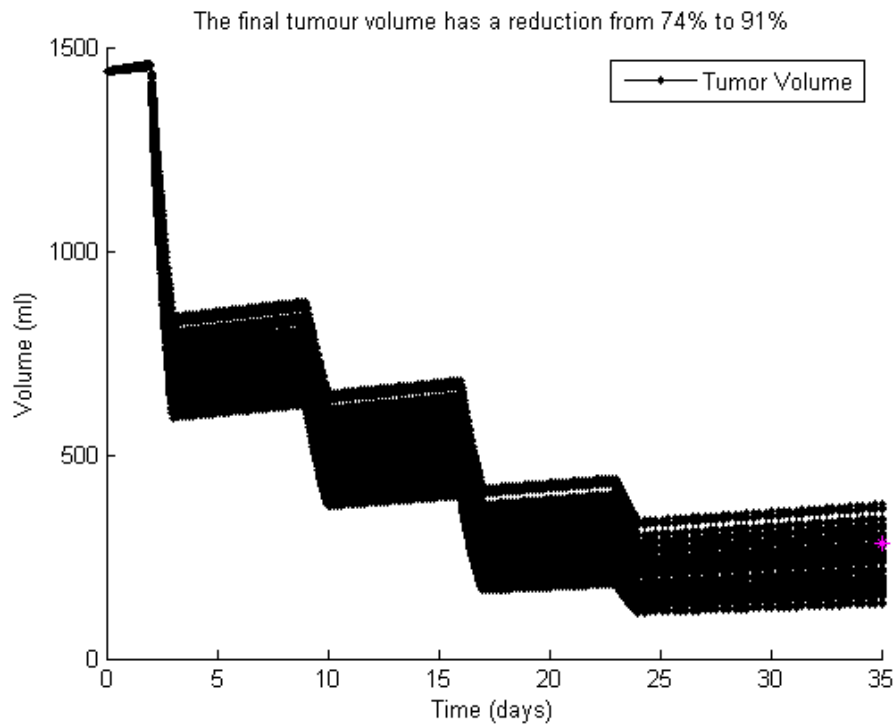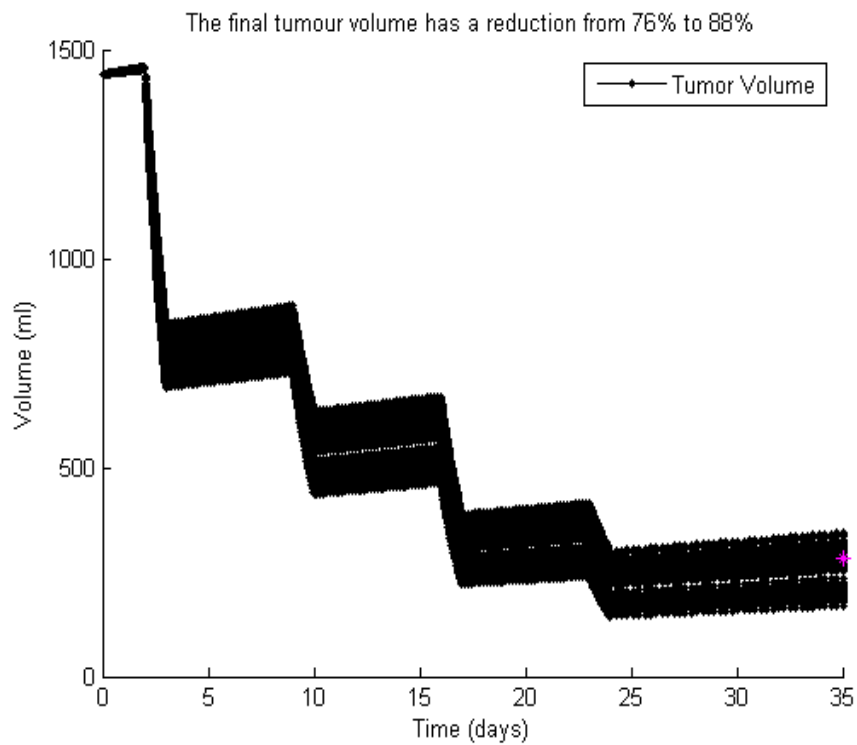
Fig. GP1a: Output with standard kill rates

Fig. GP1b: Output with kill rates from UPENN model

The response to treatment, however, is not the same for each patient. Using the provided patients, an interval confidence of the kill rates where found. The output of the model (see Fig GP1a) are underlined different simulations with underlined different kill rates chosen in the confidence interval. For example, the expected tumour reduction for the selected patient in Fig. GP1a is between 74%. and 91%. The real tumour reduction is 80%.

Moreover, the accuracy of the model can be increased using the molecular UPENN model. In fact, the UPENN model provides the real kill rate of the patient with a certain confidence interval, based on molecular analysis. These kill rates and confidence interval are substituted to the standard ones in our model, predicting the final tumour reduction. Considering the same patient of Fig GP1a, for example, the predicted shrinkage using the UPENN kill rates is between 76%. and 88%.

GP3. Integration with other modellers: the 'two pillars' scheme

The phenomenological model can be seen as a 'validator' of the main 'pillar', constituted by the interconnection of different models.
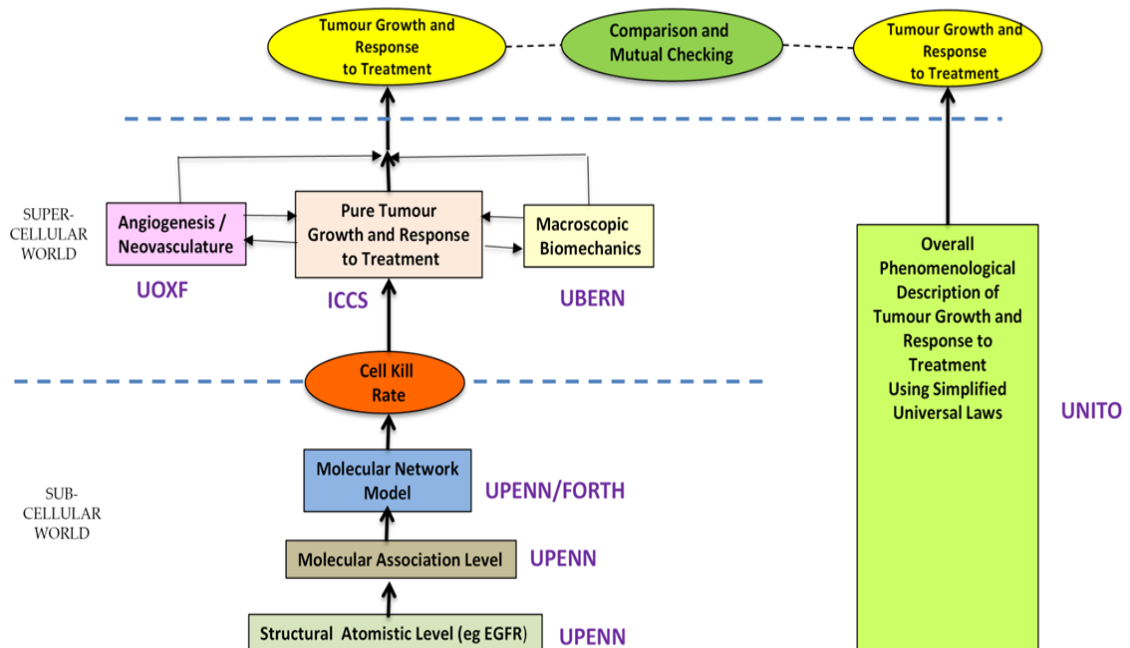
CHIC WP6 MASTER TOPOLOGY OF THE LUNG CANCER MULTI-MODELLER HYPERMODEL

Conversely, as mentioned before, the single models of the main pillar can improve the result of the UNITO model. In fact, the Molecular model by UPENN can provide the kill rate values for the specific patient, in order to personalize the prediction.

GP4. Modelling Prostate Cancer: Similarities and Differences with the Other Tumours

	PROSTATE	NEPHRO/LUNG
Tumour stage	Recurrent after surgery/RT	primary
Measured parameter value	PSA	Tumour volume
Expected response to therapy	Absence of recurrence	Tumour shrinkage
Main secondary therapy	ADT	Surgery
Number of patients considered For model validation	Some thousands	Some units
Number of patients considerable for CRAF implementation	5 (Eureka1) + 5 (Eureka2)	Some units

The basic idea is the same both for lung, nephroblastoma and prostate cancer: the PUN approach can model correctly the tumour growth.

However, in prostate cancer scenario we focused our attention on prostatectomized patients who had a relapse in the next ten years [GP5, GP6], in particular on how to estimate the free tumour growth after surgery. The other types of cancer considered in CHIC have a very fast growth and are treated after few days of their discovery. Conversely, in prostatectomized patients we have a very long period of observation, in which a free growth of the tumour can be monitored without therapies. This protocol allowed us to create a model which is able to predict the future development of the tumour considering the next 2 years after surgery.

Although the most used treatment is the Androgen Deprivation Therapy (ADT), including different drugs for hormone therapy, chemotherapy and radiotherapy can be considered as secondary 'salvage' therapies. The molecular UPENN model could provide the kill rate of the ADT by analyzing the miRNA sequences that are available for some patients.

GP5 References

[GP1] Guiot,C. and Degiorgis, P.G. and Delsanto,PP. and Gabriele,P. and Deisboeck, TS., *Does Tumour growth follow a 'universal law'?*, Journal of Theoretical Biology (2003), 225, 147-151

[GP2] Stura,I. And Venturino,E. and Guiot,C. *A two-clones tumour model: Spontaneous growth and response to treatment*, Mathematical Bioscience (2015), 271, 19-28

[GP3] West,G.B. and Brown,J.H. and Enquist,B.J., *A general model for ontogenetic growth*, Nature (2001), 413, 628-631

[GP4] Stamatakos,G. S. and Kolokotroni,E. and Dionysiou,D. and Veith,C. and Kim,Y. -. and Franz,A. and Marias,K. and Sabczynski,J. and Bohle,R. and Graf,N. *In silico oncology: Exploiting clinical studies to clinically adapt and validate multiscale oncosimulators*, Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS (2013), p. 5545-5549

[GP5] Stura,I. and Gabriele,D. and Guiot,C., *Modeling prostate cancer within CHIC*, Minerva Urologica and Nefrologica, 1:1. 97-98

[GP6] Stura,I. and Gabriele,D. and Guiot,C., *A Simple PSA-Based Computational Approach Predicts the Timing of Cancer Relapse in Prostatectomized Patients*, Cancer Research, to appear (in August)

GB. The Glioblastoma Hypermodel

The development of the Glioblastoma Multiforme (GBM) Hypermodel has been led by ICCS.

GB1. Background

As depicted in Fig. GB1, both the adaptive and the innate immune system interact with GBM, through a variety of cell types and molecular mechanisms.
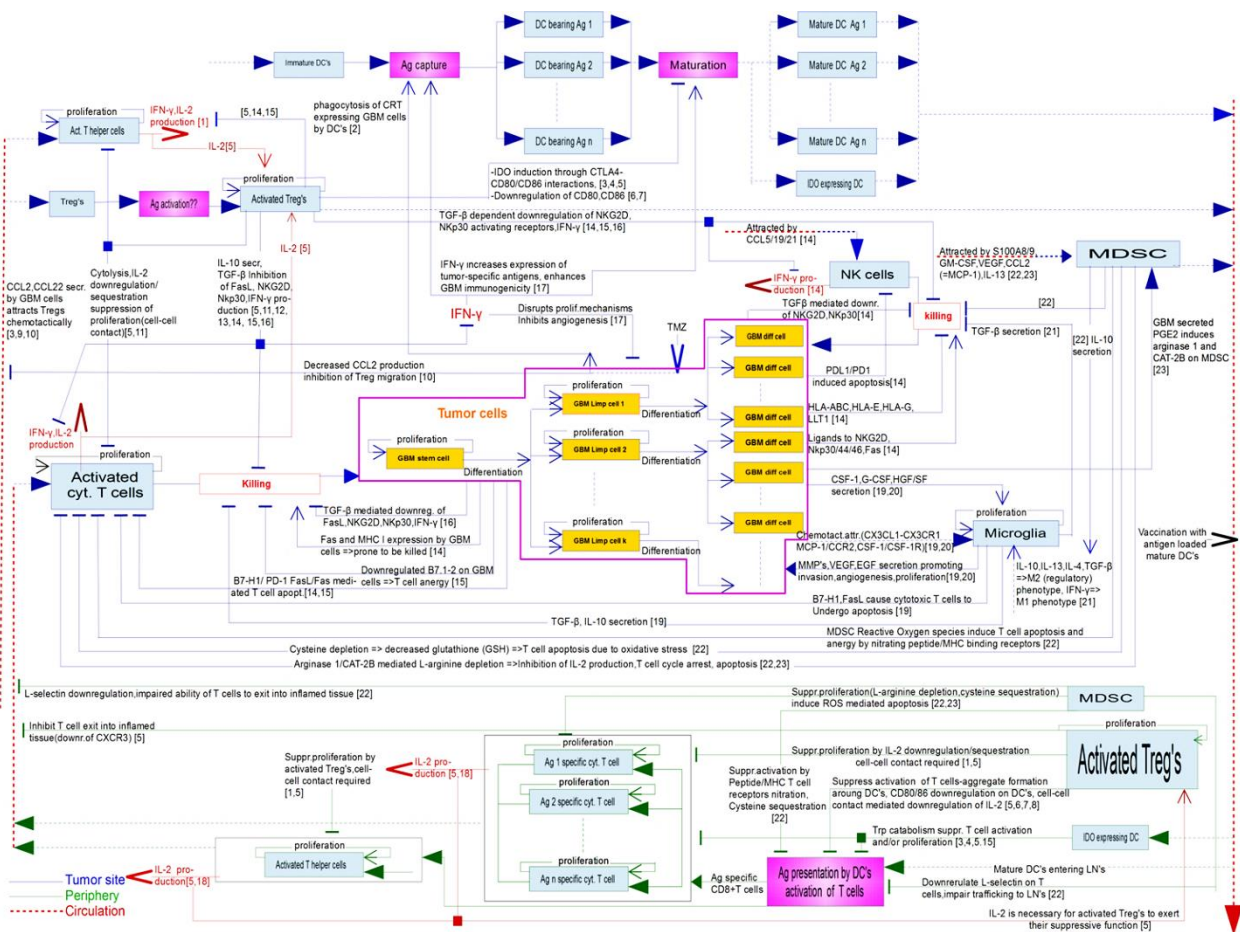

Fig. GB1.Cellular and molecular mechanisms involved in the interaction of GBM and the immune system
(References are included in the diagram. They appear in the end of the GB section preceded by the code GB)

Most commonly, the immune system of a patient suffering from GBM is found to be completely paralyzed. Concerted action of immunosuppressive cells like regulatory T cells, myeloid derived suppressor cells and microglia suppress a variety of immune reactions against the tumour; Regulatory T cells impair the antigen-presenting functions and maturation of dendritic cells and thereby suppress the production of antigen-specific, activated cytotoxic T cells in the lymph nodes. Furthermore, the killing potential of cytotoxic T cells and natural killer (NK) cells is impaired by the immunosuppressive action of myeloid derived suppressor cells and microglia. These mechanisms contribute to the induction of tolerance against the tumour[Refs GB1-GB34, included in the diagram]

Dendritic cell vaccination is a method which aims at circumventing these immunosuppressive mechanisms and reshape the patient's immune system so that it reacts effectively against the tumour.

White blood cells drawn from the patient's blood are differentiated to dendritic cells. These are furtherly cultivated with lysates from the resected tumour, aiming to produce antigen-loaded mature dendritic cells which are then injected into the patient's body in specific vaccination schedules (Fig. GB2). The rational is that these in-vitro maturated antigen-bearing dendritic cells have the capacity to present their antigens to naïve T cells in the lymph nodes, and induce their activation and proliferation. These activated cytotoxic T cells are expected to react against the tumour, by killing cancer cells. This process is further expected to induce immunological memory against GBM. [GB35, GB36, GB37, GB38, GB39, GB40]
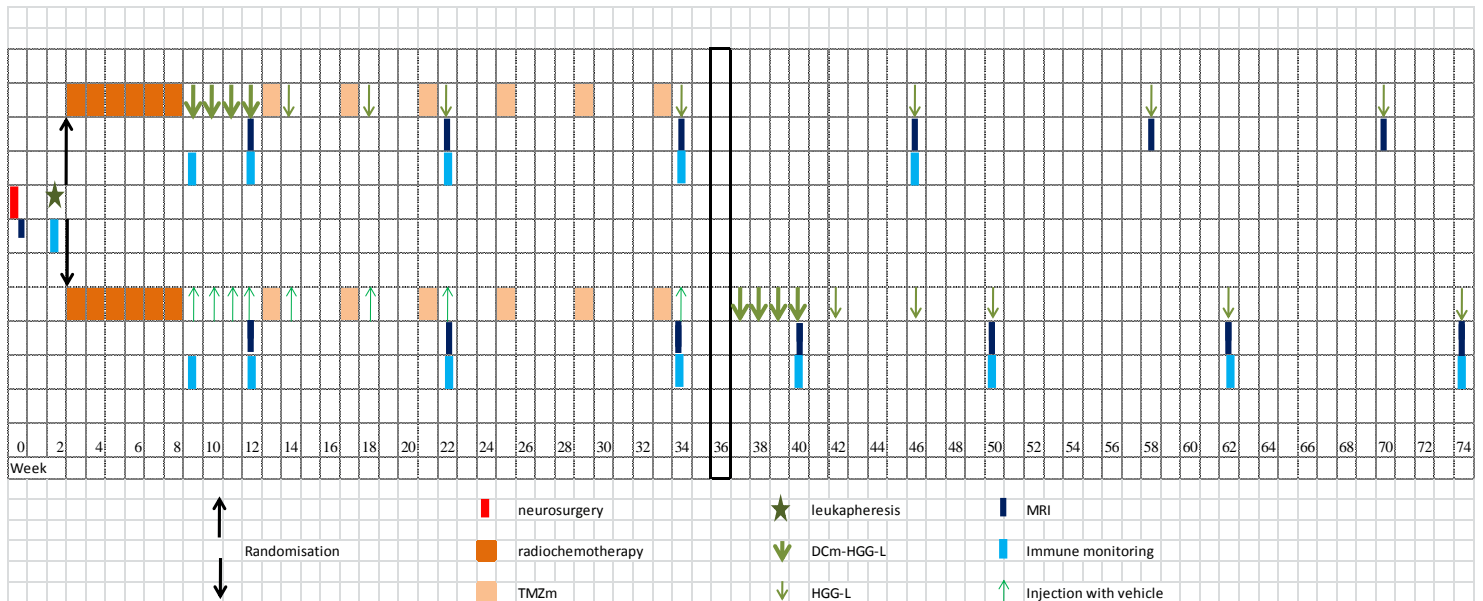

Fig. GB2. Dendritic cell vaccination and overall treatment schedule. There are two possible vaccination schedules; Just after radiochemotherapy or after 6 cycles of maintainance chemotherapy. Patients are randomly assigned to one of these schedules.
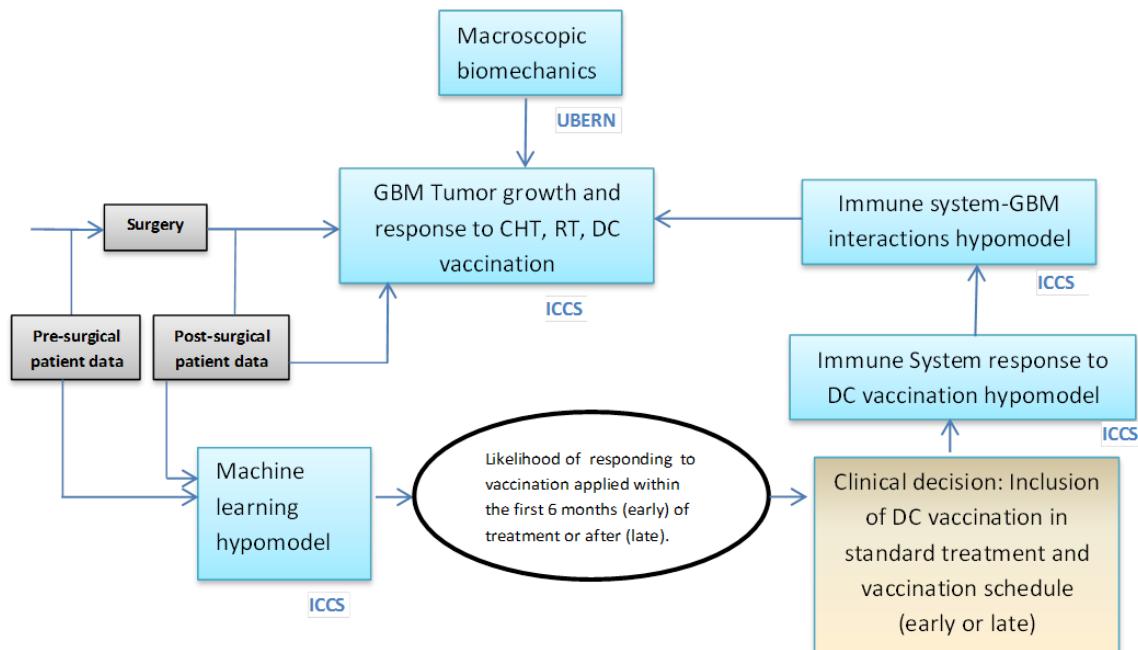
GB2. Hypermodel Structure

Fig. GB3: Master topology of the planned GBM hypermodel.

The master topology of the GBM hypermodel is depicted in Fig GB3. It consists of the following modules/ hypomodels.

Machine Learning Hypomodel: This hypomodel will classify patients to (probable) "responders" or "non-responders" to Dendritic Cell (DC) vaccination. It takes input from both pre-surgical and post-surgical data and vaccination schedule.

Immune System response to DC vaccination hypomodel: This hypomodel calculates the evolution of the population of various immune system cells during and after DC and tumour lysate vaccination. This hypomodel is a mechanistic one, exploiting data concerning the actual applied vaccines (prepared using the particular patient's lymphocytes and tumour lysates) and immune monitoring measurements.

Immune system-GBM interactions hypomodel: This hypomodel correlates the reshaped immune system (in terms of immune cells populations) with the immune response to GBM.

Up to now, available medical knowledge and clinical data do not allow building the last two hypomodels in a sensible mechanistic way. The effect of DC vaccination to the tumour will be included in the hypermodel in a simple phenomenological way, i.e. a GBM cell kill rate due to immune system response will be introduced to the oncosimulator. If pertinent data enabling the design of these last two hypomodels become available, this GBM cell kill rate could be estimated from them.

GB3. Progress on the GBM Machine Learning Model

The basic clinical questions posed by the clinicians are

1)   Will a specific GBM patient become a long term survivor (overall survival > 24 months) or suffer from early relapse and death (overall survival < 12 months)? → based on patient-, pathology- and immune profile

2)   Will dendritic cell vaccination help to reach long term survival?

3)   If yes, should vaccination be given early or late?

A summary of the available data is depicted in Fig. GB4. They regard patient characteristics for 82 cases.

Name	Type	Statistics	Range	Missings
RPA_EORTC	polynominal	mode = RPA_4 (60), least = RPA_5 (7)	RPA_4 (60), RPA_3 (14), RPA_5 (7)	1
MMSE_at_inclusion	binominal	mode = normal(>=27) (73), least = Abnormal(<27) (5)	normal(>=27) (73), Abnormal(<27) (5)	4
KPS_at_inclusion	binominal	mode = >=70 (77), least = <70 (3)	>=70 (77), <70 (3)	2
Height	real	avg = 1.734 +/- 0.079	[1.540 ; 1.900]	3
Weight	real	avg = 75.800 +/- 13.608	[44.000 ; 105.000]	3
Body_Mass_Index	real	avg = 25.187 +/- 4.038	[17.070 ; 34.810]	3
Sex	binominal	mode = Male (52), least = Female (30)	Male (52), Female (30)	0
Hemiparesis	binominal	mode = NO (71), least = YES (10)	NO (71), YES (10)	1
Behavioral_Changes	binominal	mode = NO (66), least = YES (15)	NO (66), YES (15)	1
Confusion	binominal	mode = NO (69), least = YES (12)	NO (69), YES (12)	1
Memory_Disturbance	binominal	mode = NO (68), least = YES (13)	NO (68), YES (13)	1
Aphasia	binominal	mode = NO (60), least = YES (21)	NO (60), YES (21)	1
Cranial_Nerve_Deficit	binominal	mode = NO (71), least = YES (10)	NO (71), YES (10)	1
Hemianopsia	binominal	mode = NO (70), least = YES (11)	NO (70), YES (11)	1
Epilepsy	binominal	mode = NO (57), least = YES (24)	YES (24), NO (57)	1
Motoric_Symptoms	binominal	mode = NO (62), least = YES (19)	NO (62), YES (19)	1
Cognitive_Deterioration	binominal	mode = NO (46), least = YES (35)	NO (46), YES (35)	1
Age_at_diagnosis	real	avg = 57.086 +/- 8.898	[35.000 ; 79.000]	1
Number_of_locations	binominal	mode = 1.0 (77), least = 2_or_3 (5)	1.0 (77), 2_or_3 (5)	0
Main_location_of_tumor	polynominal	mode = Frontal_l (30), least = Corp_Call (2)	Parietal_l (14), Temporal_l (29), Frontal_l (30), Occipital_l (7), Corp_Call (2)	0
Hemisphere_of_tumor_location	polynominal	mode = Right (41), least = Both (2)	Left (39), Right (41), Both (2)	0
Preoperative_Diffusion_chatacteristics	binominal	mode = ADC<1x10-³mm²/sec (51), least = ADC>1x10-³mm²/sec (3)	ADC<1x10-³mm²/sec (51), ADC>1x10-³mm²/sec (3)	28
Preoperative_Perfusion_chatacteristics	binominal	mode = rrCBV>1.75 (26), least = rrCBV>1.75 (26)	rrCBV>1.75 (26)	56
Preoperative_Oedema_characteristics	polynominal	mode = Distinct (29), least = No (4)	Moderate (23), Minimal (12), Distinct (29), No (4)	14
Preoperative_Contrast_enhancement	binominal	mode = All_c.e.t. (44), least = Partly_c.e.t. (24)	All_c.e.t. (44), Partly_c.e.t. (24)	14
Postoperative_Diffusion_chatacteristics	binominal	mode = Dis_of_all_c.e. (25), least = ADC<1x10-³mm²/sec (19)	Dis_of_all_c.e. (25), ADC<1x10-³mm²/sec (19)	38
Postoperative_Oedema_characteristics	polynominal	mode = Moderate (28), least = No (4)	Minimal (23), No (4), Moderate (28), Distinct (11)	16
Postoperative_Perfusion_characteristics	binominal	mode = Dis_of_all_c.e. (25), least = rrCBV>1.75 (10)	Dis_of_all_c.e. (25), rrCBV>1.75 (10)	47
Residual_Tumor_Volume	polynominal	mode = RTV=0 (39), least = RTV>2cm² (13)	RTV<2cm² (27), RTV=0 (39), RTV>2cm² (13)	3
Early/Late_vaccination	binominal	mode = late (41), least = late (41)	late (41), early (41)	0
RBC (10^12 / L)- aph	binominal	mode = normal (44), least = low (38)	low (38), normal (44)	0
RBC (10^12 / L)- P1/V1	polynominal	mode = normal (48), least = relapse_before_P1/V1 (12)	normal (48), relapse_before_P1/V1 (12), low (18)	4
RBC_difference	polynominal	mode = steady (43), least = drop (5)	rise (17), steady (43), relapse_before_P1/V1 (12), drop (5)	5
WBC(10^9/L)-aph	polynominal	mode = normal (70), least = low (6)	normal (70), low (6), high (6)	0
WBC(10^9/L)-P1/V1	polynominal	mode = normal (52), least = high (1)	normal (52), relapse_before_P1/V1 (12), low (13), high (1)	4
WBC_difference	polynominal	mode = steady (48), least = rise (2)	steady (48), relapse_before_P1/V1 (12), drop (15), rise (2)	5
Lymphocytes(10^9/L)-aph	binominal	mode = normal (59), least = low (23)	low (23), normal (59)	0
Lymphocytes(10^9/L)-P1/V1	polynominal	mode = low (47), least = relapse_before_P1/V1 (12)	low (47), normal (19), relapse_before_P1/V1 (12)	4
Lymphocytes_difference	polynominal	mode = steady (33), least = rise (3)	steady (33), drop (30), relapse_before_P1/V1 (12), rise (3)	4

Monocytes(10^9/L)-aph	Lymphocytes_difference	mode = normal (70), least = low (1)	normal (70), low (1), high (11)	0
Monocytes(10^9/L)-P1/V1	polynominal	mode = normal (60), least = low (2)	normal (60), relapse_before_P1/V1 (12), low (2), high (4)	4
Monocytes_difference	polynominal	mode = steady (55), least = rise (3)	steady (55), relapse_before_P1/V1 (12), rise (3), drop (8)	4
Neutrophils(10^9 / L)-aph	polynominal	mode = normal (70), least = high (5)	normal (70), low (7), high (5)	0
Neutrophils(10^9 / L)-P1/V1	polynominal	mode = normal (56), least = high (1)	normal (56), relapse_before_P1/V1 (12), low (9), high (1)	4
Neutrophils_difference	polynominal	mode = steady (51), least = rise (4)	steady (51), rise (4), relapse_before_P1/V1 (12), drop (11)	4
Eosinophils(10^9 / L)-aph	binominal	mode = normal (78), least = high (2)	normal (78), high (2)	2
Eosinophils(10^9 / L)-P1/V1	polynominal	mode = normal (61), least = high (5)	normal (61), relapse_before_P1/V1 (12), high (5)	4
Eosinophils_difference	polynominal	mode = steady (57), least = drop (2)	steady (57), relapse_before_P1/V1 (12), rise (5), drop (2)	6
Basophils(10^9 / L)-aph	binominal	mode = normal (76), least = high (3)	normal (76), high (3)	3
Basophils(10^9 / L)-P1/V1	binominal	mode = normal (66), least = relapse_before_P1/V1 (12)	normal (66), relapse_before_P1/V1 (12)	4
Basophils_difference	polynominal	mode = steady (61), least = drop (3)	steady (61), relapse_before_P1/V1 (12), drop (3)	6
Blood_Platelets (10^9 / L)-aph	polynominal	mode = normal (74), least = low (3)	normal (74), high (5), low (3)	0
Blood_Platelets (10^9 / L)-P1/V1	polynominal	mode = normal (52), least = relapse_before_P1/V1 (12)	normal (52), relapse_before_P1/V1 (12), low (14)	4
Blood_Platelets_difference	polynominal	mode = steady (48), least = rise (1)	steady (48), relapse_before_P1/V1 (12), drop (15), rise (1)	6
Immature platelet fraction (%)-P1/V1	polynominal	mode = normal (51), least = relapse_before_P1/V1 (12)	normal (51), relapse_before_P1/V1 (12), low (12)	7
Hemoglobin(g/dL)-aph	binominal	mode = low (47), least = normal (35)	low (47), normal (35)	0
Hemoglobin(g/dL)-P1/V1	polynominal	mode = normal (37), least = relapse_before_P1/V1 (12)	normal (37), low (29), relapse_before_P1/V1 (12)	4
Hemoglobin_difference	polynominal	mode = steady (44), least = drop (6)	rise (16), steady (44), relapse_before_P1/V1 (12), drop (6)	4
Hematocrite-aph	binominal	mode = low (44), least = normal (38)	normal (38), low (44)	0
Hematocrite-P1/V1	polynominal	mode = normal (49), least = relapse_before_P1/V1 (12)	normal (49), relapse_before_P1/V1 (12), low (17)	4
Hematocrite_difference	polynominal	mode = steady (36), least = drop (5)	steady (36), rise (25), relapse_before_P1/V1 (12), drop (5)	4
Reticulocytes(10^9 / L)-P1/V1	polynominal	mode = normal (59), least = high (1)	normal (59), low (3), relapse_before_P1/V1 (12), high (1)	7
Immature reticulocyte fraction(%)-P1/V1	polynominal	mode = normal (36), least = relapse_before_P1/V1 (12)	low (27), normal (36), relapse_before_P1/V1 (12)	7
PFS_after_6_cycles_TMZm?	binominal	mode = Progression (48), least = No_progression (34)	No_progression (34), Progression (48)	0
OS_months	integer	avg = 18.573 +/- 8.992	[4.000 ; 42.000]	0
PFS_months	integer	avg = 10.400 +/- 8.838	[1.000 ; 32.000]	2

Fig. GB4. Summary of available data. Further explanations are included in main text.

Known prognostic factors for GBM include RPA_EORTC classification and residual tumour volume after resection [GB36]. Including the vaccination schedule as an input feature, visualization of the data set based on known prognostic factors and patient stratification in early and late vaccine patients results in the trees depicted in Figures GB5 and GB6.
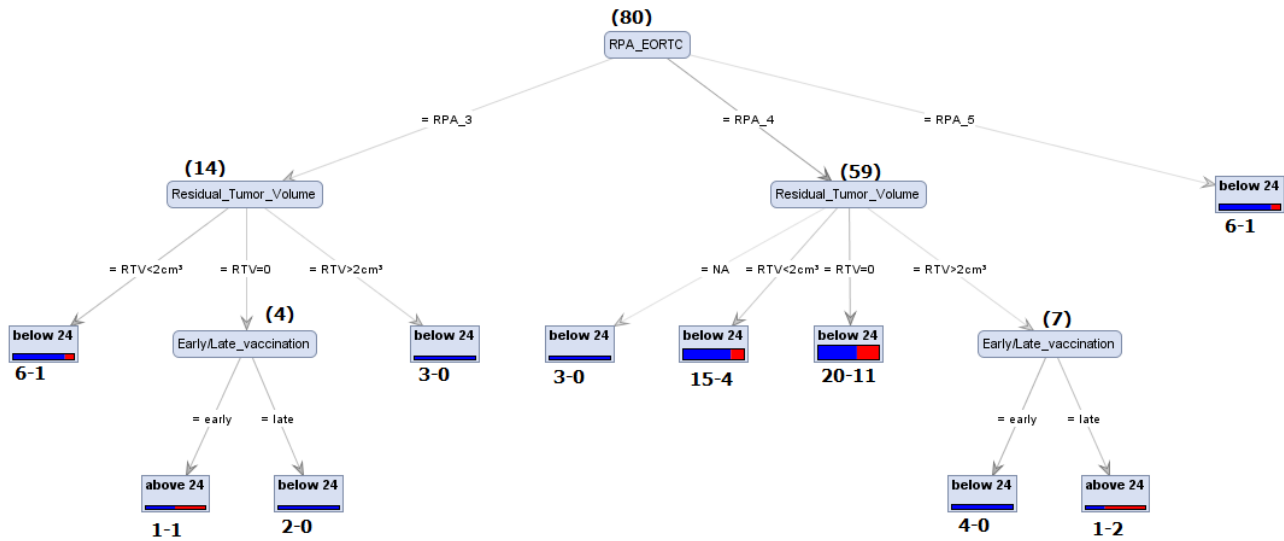

Fig. GB5. Visualization of the data set based on known prognostic factors and patient stratification in early and late vaccine patients
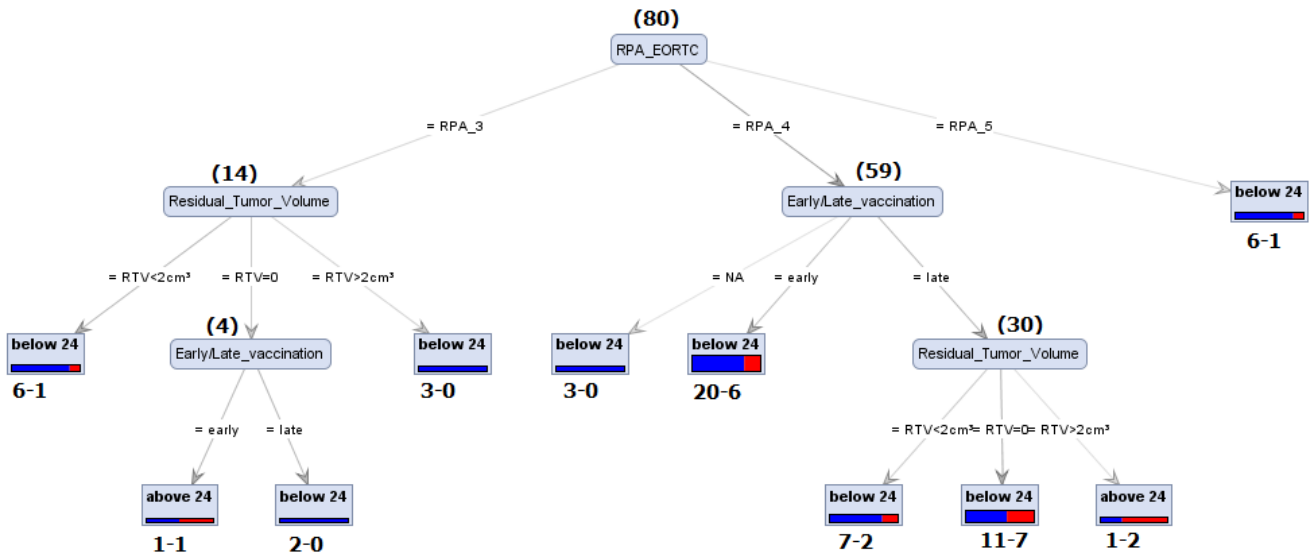
Gig. GB6 Visualization of the data set based on known prognostic factors and patient stratification in early and late vaccine patients

Despite providing visualization, the classification performance of these two trees is of limited value. The size of almost all ending nodes (leaves) is too small to be interpreted as actual probabilities. However, one can detect slightly better outcomes for patients for which total resection (residual tumour volume=0) was feasible.

Posed as classification problems, answering questions 1 and 3 consists in finding a) which of the patient features available to medical doctors up to the P1/V1 point (first vaccine, active or placebo) of the trial can predict the class (OS>24 months, 24 months>OS>12 months ,OS<12 months) and b) to construct a classifier which predicts the class of a patient based on these features. Since the vast majority of available data had categorical values, to answer these questions we evaluated the performance of a variety of tools, able to handle categorical data. These included [GB41, GB42]

- rule-based based classification methods
- decision trees
- random forests
- frequent pattern based classification methods
- TAN, K2, Hill Climbing, Genetic Search and Simulated annealing derived topologies of probabilistic graphical networks [GB43, GB44, GB45]
- ensemble classification methodologies.

 Feature selection included various heuristic selection methods, including forward/backward elimination and was further optimized by use of genetic algorithms and particle swarm attribute weighting.

All of the preceding classifier structures were outperformed by the naive Bayes classifier. In fact, naïve Bayes was the only classifier structure that provided acceptable results. In all cases, ensemble classification methods like AdaBoost [GB41, GB42] did not improve performance. The use of two patient classes (OS>24 and OS<24) provided much better results, and since the main class of interest is the OS>24 one, it was decided to use these two classes, instead of three, as was proposed in research question 1.

Additional features where created for each blood count performed before and after radiochemotherapy. These are denoted by e.g. Lymphocytes_difference and take values "rise", "drop" and "steady" depending on the change of level of Lymphocytes before and after radiochemotherapy. As shown below, feature selection methods indicated that some of these features if combined with others had predictive value.

One detail is that, in order to make use of as many patients as possible, for the patients who had missing P1/V1 blood counts and relapsed before the end of radiochemotherapy or P1/V1, their P1/V1 blood data was filled with the categorical value "early_relapse_before_P1/V1". This is an artificially introduced value, which however, makes a clear medical sense. Introducing this value means of course that for consistency, the same had to be done for every patient with relapse before P1/V1 and no missing P1/V1 blood values. There were a few patients (5-6) whose relapse (PFS date) was detected after at most 1 week (7 days) after P1/V1. In these cases, P1/V1 blood counts were also replaced with the "early_relapse_before_P1/V1" value.

As was expected, heuristic methods for feature selection provided more than one feature subset able to classify the patients in the two classes of interest. However, this subsets do not in general differ that much, at least qualitatively. They all consist of various combinations of features regarding sex, presenting symptoms, preoperative radiology (main location of tumour), blood counts before and after radiochemotherapy and vaccination schedule.

A total of 12 classifiers, each one utilizing a set of 70-75 patients, and each with different input features have been designed and tested in 5-7-10 fold and leave-one-out cross validation schemes. Many random initializations of the cross validation folds were performed to test robustness of the classifiers, and for the vast majority of them precision and recall for the OS>24 class varied from 70 to 90%.

Input features and leave-one-out cross validation results for each classifier are depicted in the following figures.

ExampleSet (73 examples, 1 special attribute, 20 regular attributes)

Role	Name	Type	Statistics	Range
label	OS_months	nominal	mode = below 24 (52), least = above 24 (21)	below 24 (52), above 24 (21)
regular	Sex	binominal	mode = Male (48), least = Female (25)	Male (48), Female (25)
regular	Confusion	binominal	mode = NO (61), least = YES (12)	NO (61), YES (12)
regular	Memory_Disturbance	binominal	mode = NO (60), least = YES (13)	NO (60), YES (13)
regular	Motoric_Symptoms	binominal	mode = NO (56), least = YES (17)	NO (56), YES (17)
regular	Cognitive_Deterioration	binominal	mode = NO (38), least = YES (35)	NO (38), YES (35)
regular	Number_of_locations	[Motoric_Symptoms]bminal	mode = 1.0 (68), least = 2_or_3 (5)	1.0 (68), 2_or_3 (5)
regular	Main_location_of_tumor	polynomial	mode = Temporal_l (29), least = Corp_Call (1)	Parietal_l (11), Temporal_l (29), Frontal_l (25), Occipital_l (7), Corp_Call (1)
regular	Early/Late_vaccination	binominal	mode = late (37), least = early (36)	late (37), early (36)
regular	WBC(10^9/L)-aph	polynomial	mode = normal (63), least = low (5)	normal (63), low (5), high (5)
regular	WBC(10^9/L)-P1/V1	polynomial	mode = normal (47), least = high (1)	normal (47), relapse_before_P1/V1 (12), low (13), high (1)
regular	Lymphocytes_difference	polynomial	mode = steady (30), least = rise (3)	steady (30), drop (28), relapse_before_P1/V1 (12), rise (3)
regular	Monocytes(10^9/L)-aph	polynomial	mode = normal (63), least = low (0)	normal (63), low (0), high (10)
regular	Monocytes(10^9/L)-P1/V1	polynomial	mode = normal (56), least = low (2)	normal (56), relapse_before_P1/V1 (12), low (2), high (3)
regular	Neutrophils(10^9 / L)-aph	polynomial	mode = normal (62), least = high (4)	normal (62), low (7), high (4)
regular	Neutrophils_difference	polynomial	mode = steady (47), least = rise (4)	steady (47), rise (4), relapse_before_P1/V1 (12), drop (10)
regular	Eosinophils(10^9 / L)-P1/V1	polynomial	mode = normal (57), least = high (4)	normal (57), relapse_before_P1/V1 (12), high (4)
regular	Eosinophils_difference	polynomial	mode = steady (55), least = drop (2)	steady (55), relapse_before_P1/V1 (12), rise (4), drop (2)
regular	Immature platelet fraction (%)-P1/V1	polynomial	mode = normal (49), least = relapse_before_P1/V1 (12)	normal (49), relapse_before_P1/V1 (12), low (12)
regular	Hemoglobin(g/dL)-aph	binominal	mode = low (42), least = normal (31)	low (42), normal (31)
regular	Immature reticulocyte fraction(%)-P1/V1	polynomial	mode = normal (34), least = relapse_before_P1/V1 (12)	low (27), normal (34), relapse_before_P1/V1 (12)

Fig. GB7a. Input features, class and patient utilization (73/82) for classifier 1.

	true below 24	true above 24	class precision
pred. below 24	45	1	97.83%
pred. above 24	7	20	74.07%
class recall	86.54%	95.24%	

Fig. GB7b. Leave-one-out confusion matrix for classifier 1.

ExampleSet (70 examples, 1 special attribute, 21 regular attributes)

Role	Name	Type	Statistics	Range
label	OS_months	nominal	mode = below 24 (49), least = above 24 (21)	below 24 (49), above 24 (21)
regular	Sex	binominal	mode = Male (46), least = Female (24)	Male (46), Female (24)
regular	Hemiparesis	binominal	mode = NO (62), least = YES (8)	NO (62), YES (8)
regular	Behavioral_Changes	binominal	mode = NO (56), least = YES (14)	NO (56), YES (14)
regular	Confusion	binominal	mode = NO (60), least = YES (10)	NO (60), YES (10)
regular	Memory_Disturbance	binominal	mode = NO (57), least = YES (13)	NO (57), YES (13)
regular	Motoric_Symptoms	binominal	mode = NO (53), least = YES (17)	NO (53), YES (17)
regular	Cognitive_Deterioration	binominal	mode = NO (37), least = YES (33)	NO (37), YES (33)
regular	Number_of_locations	binominal	mode = 1.0 (65), least = 2_or_3 (5)	1.0 (65), 2_or_3 (5)
regular	Main_location_of_tumor	polynominal	mode = Temporal_I (28), least = Corp_Call (1)	Parietal_I (10), Temporal_I (28), Frontal_I (24), Occipital_I (7), Corp_Call (1)
regular	Residual_Tumor_Volume	polynominal	mode = RTV=0 (33), least = RTV>2cm³ (11)	RTV<2cm³ (26), RTV=0 (33), RTV>2cm³ (11)
regular	Early/Late_vaccination	binominal	mode = late (36), least = early (34)	late (36), early (34)
regular	WBC(10^9/L)-aph	polynominal	mode = normal (61), least = high (4)	normal (61), low (5), high (4)
regular	Lymphocytes(10^9/L)-P1/V1	polynominal	mode = low (44), least = relapse_before_P1/V1 (10)	low (44), normal (16), relapse_before_P1/V1 (10)
regular	Lymphocytes_difference	polynominal	mode = steady (29), least = rise (3)	steady (29), drop (28), relapse_before_P1/V1 (10), rise (3)
regular	Monocytes(10^9/L)-P1/V1	polynominal	mode = normal (55), least = low (2)	normal (55), relapse_before_P1/V1 (10), low (2), high (3)
regular	Neutrophils(10^9 / L)-aph	polynominal	mode = normal (60), least = high (3)	normal (60), low (7), high (3)
regular	Eosinophils(10^9 / L)-P1/V1	polynominal	mode = normal (56), least = high (4)	normal (56), relapse_before_P1/V1 (10), high (4)
regular	Eosinophils_difference	polynominal	mode = steady (54), least = drop (2)	steady (54), relapse_before_P1/V1 (10), rise (4), drop (2)
regular	Immature platelet fraction (%)-P1/V1	polynominal	mode = normal (48), least = relapse_before_P1/V1 (10)	normal (48), relapse_before_P1/V1 (10), low (12)
regular	Hemoglobin(g/dL)-aph	binominal	mode = low (39), least = normal (31)	low (39), normal (31)
regular	Immature reticulocyte fraction(%)-P1/V1	polynominal	mode = normal (33), least = relapse_before_P1/V1 (10)	low (27), normal (33), relapse_before_P1/V1 (10)

Fig. GB8a. Input features, class and patient utilization (70/82) for classifier 2.

	true below 24	true above 24	class precision
pred. below 24	43	2	95.56%
pred. above 24	6	19	76.00%
class recall	87.76%	90.48%	

Fig. GB8b. Leave-one-out confusion matrix for classifier 2.

ExampleSet (70 examples, 1 special attribute, 20 regular attributes)

Role	Name	Type	Statistics	Range
label	OS_months	nominal	mode = below 24 (49), least = above 24 (21)	below 24 (49), above 24 (21)
regular	Sex	binominal	mode = Male (46), least = Female (24)	Male (46), Female (24)
regular	Hemiparesis	binominal	mode = NO (62), least = YES (8)	NO (62), YES (8)
regular	Confusion	binominal	mode = NO (60), least = YES (10)	NO (60), YES (10)
regular	Memory_Disturbance	binominal	mode = NO (57), least = YES (13)	NO (57), YES (13)
regular	Motoric_Symptoms	binominal	mode = NO (53), least = YES (17)	NO (53), YES (17)
regular	Cognitive_Deterioration	binominal	mode = NO (37), least = YES (33)	NO (37), YES (33)
regular	Number_of_locations	binominal	mode = 1.0 (65), least = 2_or_3 (5)	1.0 (65), 2_or_3 (5)
regular	Main_location_of_tumor	polynominal	mode = Temporal_l (28), least = Corp_Call (1)	Parietal_l (10), Temporal_l (28), Frontal_l (24), Occipital_l (7), Corp_Call (1)
regular	Residual_Tumor_Volume	polynominal	mode = RTV=0 (33), least = RTV>2cm³ (11)	RTV<2cm³ (26), RTV=0 (33), RTV>2cm³ (11)
regular	Early/Late_vaccination	binominal	mode = late (36), least = early (34)	late (36), early (34)
regular	WBC(10^9/L)-aph	polynominal	mode = normal (61), least = high (4)	normal (61), low (5), high (4)
regular	Lymphocytes(10^9/L)-P1/V1	polynominal	mode = low (44), least = relapse_before_P1/V1 (10)	low (44), normal (16), relapse_before_P1/V1 (10)
regular	Lymphocytes_difference	polynominal	mode = steady (29), least = rise (3)	steady (29), drop (28), relapse_before_P1/V1 (10), rise (3)
regular	Monocytes(10^9/L)-P1/V1	polynominal	mode = normal (55), least = low (2)	normal (55), relapse_before_P1/V1 (10), low (2), high (3)
regular	Neutrophils(10^9 / L)-aph	polynominal	mode = normal (60), least = high (3)	normal (60), low (7), high (3)
regular	Eosinophils(10^9 / L)-P1/V1	polynominal	mode = normal (56), least = high (4)	normal (56), relapse_before_P1/V1 (10), high (4)
regular	Eosinophils_difference	polynominal	mode = steady (54), least = drop (2)	steady (54), relapse_before_P1/V1 (10), rise (4), drop (2)
regular	Immature platelet fraction (%)-P1/V1	polynominal	mode = normal (48), least = relapse_before_P1/V1 (10)	normal (48), relapse_before_P1/V1 (10), low (12)
regular	Hemoglobin(g/dL)-aph	binominal	mode = low (39), least = normal (31)	low (39), normal (31)
regular	Immature reticulocyte fraction(%)-P1/V1	polynominal	mode = normal (33), least = relapse_before_P1/V1 (10)	low (27), normal (33), relapse_before_P1/V1 (10)

Fig. GB9a. Input features, class and patient utilization (70/82) for classifier 3.

	true below 24	true above 24	class precision
pred. below 24	44	3	93.62%
pred. above 24	5	18	78.26%
class recall	89.80%	85.71%	

Fig. GB9b. Leave-one-out confusion matrix for classifier 3.

ExampleSet (73 examples, 1 special attribute, 19 regular attributes)

Role	Name	Type	Statistics	Range
label	OS_months	nominal	mode = below 24 (52), least = above 24 (21)	below 24 (52), above 24 (21)
regular	Sex	binominal	mode = Male (48), least = Female (25)	Male (48), Female (25)
regular	Hemiparesis	binominal	mode = NO (65), least = YES (8)	NO (65), YES (8)
regular	Confusion	binominal	mode = NO (61), least = YES (12)	NO (61), YES (12)
regular	Memory_Disturbance	binominal	mode = NO (60), least = YES (13)	NO (60), YES (13)
regular	Motoric_Symptoms	binominal	mode = NO (56), least = YES (17)	NO (56), YES (17)
regular	Cognitive_Deterioration	binominal	mode = NO (38), least = YES (35)	NO (38), YES (35)
regular	Number_of_locations	binominal	mode = 1.0 (68), least = 2_or_3 (5)	1.0 (68), 2_or_3 (5)
regular	Main_location_of_tumor	polynominal	mode = Temporal_l (29), least = Corp_Call (1)	Parietal_l (11), Temporal_l (29), Frontal_l (25), Occipital_l (7), Corp_Call (1)
regular	Early/Late_vaccination	binominal	mode = late (37), least = early (36)	late (37), early (36)
regular	WBC(10^9/L)-aph	polynominal	mode = normal (63), least = low (5)	normal (63), low (5), high (5)
regular	Lymphocytes_difference	polynominal	mode = steady (30), least = rise (3)	steady (30), drop (28), relapse_before_P1/V1 (12), rise (3)
regular	Monocytes(10^9/L)-P1/V1	polynominal	mode = normal (56), least = low (2)	normal (56), relapse_before_P1/V1 (12), low (2), high (3)
regular	Neutrophils(10^9 / L)-aph	polynominal	mode = normal (62), least = high (4)	normal (62), low (7), high (4)
regular	Eosinophils(10^9 / L)-P1/V1	polynominal	mode = normal (57), least = high (4)	normal (57), relapse_before_P1/V1 (12), high (4)
regular	Eosinophils_difference	polynominal	mode = steady (55), least = drop (2)	steady (55), relapse_before_P1/V1 (12), rise (4), drop (2)
regular	Immature platelet fraction (%)-P1/V1	polynominal	mode = normal (49), least = relapse_before_P1/V1 (12)	normal (49), relapse_before_P1/V1 (12), low (12)
regular	Hemoglobin(g/dL)-aph	binominal	mode = low (42), least = normal (31)	low (42), normal (31)
regular	Hemoglobin(g/dL)-P1/V1	polynominal	mode = normal (34), least = relapse_before_P1/V1 (12)	normal (34), low (27), relapse_before_P1/V1 (12)
regular	Immature reticulocyte fraction(%)-P1/V1	polynominal	mode = normal (34), least = relapse_before_P1/V1 (12)	low (27), normal (34), relapse_before_P1/V1 (12)

Fig. GB10a. Input features, class and patient utilization (73/82) for classifier 3.

	true below 24	true above 24	class precision
pred. below 24	46	3	93.88%
pred. above 24	6	18	75.00%
class recall	88.46%	85.71%	

Fig. GB10b. Leave-one-out confusion matrix for classifier 4.

ExampleSet (73 examples, 1 special attribute, 18 regular attributes)

Role	Name	Type	Statistics	Range
label	OS_months	nominal	mode = below 24 (52), least = above 24 (21)	below 24 (52), above 24 (21)
regular	Sex	binominal	mode = Male (48), least = Female (25)	Male (48), Female (25)
regular	Hemiparesis	binominal	mode = NO (65), least = YES (8)	NO (65), YES (8)
regular	Confusion	binominal	mode = NO (61), least = YES (12)	NO (61), YES (12)
regular	Memory_Disturbance	binominal	mode = NO (60), least = YES (13)	NO (60), YES (13)
regular	Motoric_Symptoms	binominal	mode = NO (56), least = YES (17)	NO (56), YES (17)
regular	Cognitive_Deterioration	binominal	mode = NO (38), least = YES (35)	NO (38), YES (35)
regular	Number_of_locations	binominal	mode = 1.0 (68), least = 2_or_3 (5)	1.0 (68), 2_or_3 (5)
regular	Main_location_of_tumor	polynominal	mode = Temporal_l (29), least = Corp_Call (1)	Parietal_l (11), Temporal_l (29), Frontal_l (25), Occipital_l (7), Corp_Call (1)
regular	Early/Late_vaccination	binominal	mode = late (37), least = early (36)	late (37), early (36)
regular	WBC(10^9/L)-aph	polynominal	mode = normal (63), least = low (5)	normal (63), low (5), high (5)
regular	Lymphocytes_difference	polynominal	mode = steady (30), least = rise (3)	steady (30), drop (28), relapse_before_P1/V1 (12), rise (3)
regular	Monocytes(10^9/L)-P1/V1	polynominal	mode = normal (56), least = low (2)	normal (56), relapse_before_P1/V1 (12), low (2), high (3)
regular	Neutrophils(10^9 / L)-aph	polynominal	mode = normal (62), least = high (4)	normal (62), low (7), high (4)
regular	Eosinophils(10^9 / L)-P1/V1	polynominal	mode = normal (57), least = high (4)	normal (57), relapse_before_P1/V1 (12), high (4)
regular	Eosinophils_difference	polynominal	mode = steady (55), least = drop (2)	steady (55), relapse_before_P1/V1 (12), rise (4), drop (2)
regular	Immature platelet fraction (%)-P1/V1	polynominal	mode = normal (49), least = relapse_before_P1/V1 (12)	normal (49), relapse_before_P1/V1 (12), low (12)
regular	Hemoglobin(g/dL)-aph	binominal	mode = low (42), least = normal (31)	low (42), normal (31)
regular	Immature reticulocyte fraction(%)-P1/V1	polynominal	mode = normal (34), least = relapse_before_P1/V1 (12)	low (27), normal (34), relapse_before_P1/V1 (12)

Fig. GB11a. Input features, class and patient utilization (73/82) for classifier 5.

	true below 24	true above 24	class precision
pred. below 24	46	2	95.83%
pred. above 24	6	19	76.00%
class recall	88.46%	90.48%	

Fig. GB11b. Leave-one-out confusion matrix for classifier 5.

ExampleSet (74 examples, 1 special attribute, 14 regular attributes)

Role	Name	Type	Statistics	Range
label	OS_months	nominal	mode = below 24 (53), least = above 24 (21)	below 24 (53), above 24 (21)
regular	Sex	binominal	mode = Male (49), least = Female (25)	Male (49), Female (25)
regular	Hemiparesis	binominal	mode = NO (66), least = YES (8)	NO (66), YES (8)
regular	Behavioral_Changes	binominal	mode = NO (59), least = YES (15)	NO (59), YES (15)
regular	Confusion	binominal	mode = NO (62), least = YES (12)	NO (62), YES (12)
regular	Memory_Disturbance	binominal	mode = NO (61), least = YES (13)	NO (61), YES (13)
regular	Hemianopsia	binominal	mode = NO (65), least = YES (9)	NO (65), YES (9)
regular	Number_of_locations	binominal	mode = 1.0 (69), least = 2_or_3 (5)	1.0 (69), 2_or_3 (5)
regular	Main_location_of_tumor	polynominal	mode = Temporal_l (29), least = Corp_Call (1)	Parietal_l (12), Temporal_l (29), Frontal_l (25), Occipital_l (7), Corp_Call (1)
regular	Early/Late_vaccination	binominal	mode = late (38), least = early (36)	late (38), early (36)
regular	WBC(10^9/L)-aph	polynominal	mode = normal (63), least = low (5)	normal (63), low (5), high (6)
regular	Monocytes(10^9/L)-aph	polynominal	mode = normal (63), least = low (1)	normal (63), low (1), high (10)
regular	Monocytes_difference	polynominal	mode = steady (52), least = rise (2)	steady (52), relapse_before_P1/V1 (12), rise (2), drop (8)
regular	Neutrophils(10^9 / L)-aph	polynominal	mode = normal (62), least = high (5)	normal (62), low (7), high (5)
regular	Immature reticulocyte fraction(%)-P1/V1	polynominal	mode = normal (35), least = relapse_before_P1/V1 (12)	low (27), normal (35), relapse_before_P1/V1 (12)

Fig. GB12a. Input features, class and patient utilization (74/82) for classifier 6.

	true below 24	true above 24	class precision
pred. below 24	50	6	89.29%
pred. above 24	3	15	83.33%
class recall	94.34%	71.43%	

Fig. GB12b. Leave-one-out confusion matrix for classifier 6.

ExampleSet (70 examples, 1 special attribute, 23 regular attributes)

Role	Name	Type	Statistics	Range
label	OS_months	nominal	mode = below 24 (49), least = above 24 (21)	below 24 (49), above 24 (21)
regular	Sex	binominal	mode = Male (46), least = Female (24)	Male (46), Female (24)
regular	Hemiparesis	binominal	mode = NO (62), least = YES (8)	NO (62), YES (8)
regular	Behavioral_Changes	binominal	mode = NO (56), least = YES (14)	NO (56), YES (14)
regular	Confusion	binominal	mode = NO (60), least = YES (10)	NO (60), YES (10)
regular	Memory_Disturbance	binominal	mode = NO (57), least = YES (13)	NO (57), YES (13)
regular	Motoric_Symptoms	binominal	mode = NO (53), least = YES (17)	NO (53), YES (17)
regular	Number_of_locations	binominal	mode = 1.0 (65), least = 2_or_3 (5)	1.0 (65), 2_or_3 (5)
regular	Main_location_of_tumor	polynominal	mode = Temporal_l (28), least = Corp_Call (1)	Parietal_l (10), Temporal_l (28), Frontal_l (24), Occipital_l (7), Corp_Call (1)
regular	Residual_Tumor_Volume	polynominal	mode = RTV=0 (33), least = RTV>2cm³ (11)	RTV<2cm³ (26), RTV=0 (33), RTV>2cm³ (11)
regular	Early/Late_vaccination	binominal	mode = late (36), least = early (34)	late (36), early (34)
regular	WBC(10^9/L)-aph	polynominal	mode = normal (61), least = high (4)	normal (61), low (5), high (4)
regular	Lymphocytes(10^9/L)-P1/V1	polynominal	mode = low (44), least = relapse_before_P1/V1 (10)	low (44), normal (16), relapse_before_P1/V1 (10)
regular	Lymphocytes_difference	polynominal	mode = steady (29), least = rise (3)	steady (29), drop (28), relapse_before_P1/V1 (10), rise (3)
regular	Monocytes(10^9/L)-aph	polynominal	mode = normal (60), least = low (0)	normal (60), low (0), high (10)
regular	Monocytes(10^9/L)-P1/V1	polynominal	mode = normal (55), least = low (2)	normal (55), relapse_before_P1/V1 (10), low (2), high (3)
regular	Neutrophils(10^9 / L)-aph	polynominal	mode = normal (60), least = high (3)	normal (60), low (7), high (3)
regular	Eosinophils(10^9 / L)-P1/V1	polynominal	mode = normal (56), least = high (4)	normal (56), relapse_before_P1/V1 (10), high (4)
regular	Eosinophils_difference	polynominal	mode = steady (54), least = drop (2)	steady (54), relapse_before_P1/V1 (10), rise (4), drop (2)
regular	Immature platelet fraction (%)-P1/V1	polynominal	mode = normal (48), least = relapse_before_P1/V1 (10)	normal (48), relapse_before_P1/V1 (10), low (12)
regular	Hemoglobin(g/dL)-aph	binominal	mode = low (39), least = normal (31)	low (39), normal (31)
regular	Hemoglobin(g/dL)-P1/V1	polynominal	mode = normal (34), least = relapse_before_P1/V1 (10)	normal (34), low (26), relapse_before_P1/V1 (10)
regular	Hematocrite-P1/V1	polynominal	mode = normal (44), least = relapse_before_P1/V1 (10)	normal (44), relapse_before_P1/V1 (10), low (16)
regular	Immature reticulocyte fraction(%)-P1/V1	polynominal	mode = normal (33), least = relapse_before_P1/V1 (10)	low (27), normal (33), relapse_before_P1/V1 (10)

Fig. GB13a. Input features, class and patient utilization (70/82) for classifier 7.

	true below 24	true above 24	class precision
pred. below 24	43	2	95.56%
pred. above 24	6	19	76.00%
class recall	87.76%	90.48%	

Fig. GB13b. Leave-one-out confusion matrix for classifier 7.

ExampleSet (74 examples, 1 special attribute, 17 regular attributes)

Role	Name	Type	Statistics	Range
label	OS_months	nominal	mode = below 24 (53), least = above 24 (21)	below 24 (53), above 24 (21)
regular	Hemiparesis	binominal	mode = NO (66), least = YES (8)	NO (66), YES (8)
regular	Behavioral_Changes	binominal	mode = NO (59), least = YES (15)	NO (59), YES (15)
regular	Confusion	binominal	mode = NO (62), least = YES (12)	NO (62), YES (12)
regular	Memory_Disturbance	binominal	mode = NO (61), least = YES (13)	NO (61), YES (13)
regular	Hemianopsia	binominal	mode = NO (65), least = YES (9)	NO (65), YES (9)
regular	Epilepsy	binominal	mode = NO (53), least = YES (21)	YES (21), NO (53)
regular	Number_of_locations	binominal	mode = 1.0 (69), least = 2_or_3 (5)	1.0 (69), 2_or_3 (5)
regular	Main_location_of_tumor	polynominal	mode = Temporal_l (29), least = Corp_Call (1)	Parietal_l (12), Temporal_l (29), Frontal_l (25), Occipital_l (7), Corp_Call (1)
regular	Early/Late_vaccination	binominal	mode = late (38), least = early (36)	late (38), early (36)
regular	WBC(10^9/L)-aph	polynominal	mode = normal (63), least = low (5)	normal (63), low (5), high (6)
regular	Lymphocytes_difference	polynominal	mode = steady (31), least = rise (3)	steady (31), drop (28), relapse_before_P1/V1 (12), rise (3)
regular	Monocytes(10^9/L)-aph	polynominal	mode = normal (63), least = low (1)	normal (63), low (1), high (10)
regular	Monocytes_difference	polynominal	mode = steady (52), least = rise (2)	steady (52), relapse_before_P1/V1 (12), rise (2), drop (8)
regular	Neutrophils(10^9 / L)-aph	polynominal	mode = normal (62), least = high (5)	normal (62), low (7), high (5)
regular	Neutrophils_difference	polynominal	mode = steady (47), least = rise (4)	steady (47), rise (4), relapse_before_P1/V1 (12), drop (11)
regular	Immature platelet fraction (%)-P1/V1	polynominal	mode = normal (50), least = relapse_before_P1/V1 (12)	normal (50), relapse_before_P1/V1 (12), low (12)
regular	Immature reticulocyte fraction(%)-P1/V1	polynominal	mode = normal (35), least = relapse_before_P1/V1 (12)	low (27), normal (35), relapse_before_P1/V1 (12)

Fig. GB14a. Input features, class and patient utilization (74/82) for classifier 8.

	true below 24	true above 24	class precision
pred. below 24	47	4	92.16%
pred. above 24	6	17	73.91%
class recall	88.68%	80.95%	

Fig. GB14b. Leave-one-out confusion matrix for classifier 8.

ExampleSet (74 examples, 1 special attribute, 17 regular attributes)

Role	Name	Type	Statistics	Range
label	OS_months	nominal	mode = below 24 (53), least = above 24 (21)	below 24 (53), above 24 (21)
regular	Sex	binominal	mode = Male (49), least = Female (25)	Male (49), Female (25)
regular	Behavioral_Changes	binominal	mode = NO (59), least = YES (15)	NO (59), YES (15)
regular	Confusion	binominal	mode = NO (62), least = YES (12)	NO (62), YES (12)
regular	Memory_Disturbance	binominal	mode = NO (61), least = YES (13)	NO (61), YES (13)
regular	Motoric_Symptoms	binominal	mode = NO (57), least = YES (17)	NO (57), YES (17)
regular	Cognitive_Deterioration	binominal	mode = NO (39), least = YES (35)	NO (39), YES (35)
regular	Number_of_locations	binominal	mode = 1.0 (69), least = 2_or_3 (5)	1.0 (69), 2_or_3 (5)
regular	Main_location_of_tumor	polynominal	mode = Temporal_l (29), least = Corp_Call (1)	Parietal_l (12), Temporal_l (29), Frontal_l (25), Occipital_l (7), Corp_Call (1)
regular	Early/Late_vaccination	binominal	mode = late (38), least = early (36)	late (38), early (36)
regular	WBC(10^9/L)-aph	polynominal	mode = normal (63), least = low (5)	normal (63), low (5), high (6)
regular	WBC(10^9/L)-P1/V1	polynominal	mode = normal (48), least = high (1)	normal (48), relapse_before_P1/V1 (12), low (13), high (1)
regular	Lymphocytes_difference	polynominal	mode = steady (31), least = rise (3)	steady (31), drop (28), relapse_before_P1/V1 (12), rise (3)
regular	Monocytes(10^9/L)-P1/V1	polynominal	mode = normal (57), least = low (2)	normal (57), relapse_before_P1/V1 (12), low (2), high (3)
regular	Neutrophils(10^9 / L)-aph	polynominal	mode = normal (62), least = high (5)	normal (62), low (7), high (5)
regular	Eosinophils(10^9 / L)-P1/V1	polynominal	mode = normal (58), least = high (4)	normal (58), relapse_before_P1/V1 (12), high (4)
regular	Immature platelet fraction (%)-P1/V1	polynominal	mode = normal (50), least = relapse_before_P1/V1 (12)	normal (50), relapse_before_P1/V1 (12), low (12)
regular	Immature reticulocyte fraction(%)-P1/V1	polynominal	mode = normal (35), least = relapse_before_P1/V1 (12)	low (27), normal (35), relapse_before_P1/V1 (12)

Fig. GB15a. Input features, class and patient utilization (74/82) for classifier 9.

	true below 24	true above 24	class precision
pred. below 24	46	1	97.87%
pred. above 24	7	20	74.07%
class recall	86.79%	95.24%	

Fig. GB15b. Leave-one-out confusion matrix for classifier 9.

ExampleSet (73 examples, 1 special attribute, 21 regular attributes)

Role	Name	Type	Statistics	Range
label	OS_months	nominal	mode = below 24 (52), least = above 24 (21)	below 24 (52), above 24 (21)
regular	Sex	binominal	mode = Male (48), least = Female (25)	Male (48), Female (25)
regular	Hemiparesis	binominal	mode = NO (65), least = YES (8)	NO (65), YES (8)
regular	Confusion	binominal	mode = NO (61), least = YES (12)	NO (61), YES (12)
regular	Memory_Disturbance	binominal	mode = NO (60), least = YES (13)	NO (60), YES (13)
regular	Motoric_Symptoms	binominal	mode = NO (56), least = YES (17)	NO (56), YES (17)
regular	Cognitive_Deterioration	binominal	mode = NO (38), least = YES (35)	NO (38), YES (35)
regular	Number_of_locations	binominal	mode = 1.0 (68), least = 2_or_3 (5)	1.0 (68), 2_or_3 (5)
regular	Main_location_of_tumor	polynominal	mode = Temporal_l (29), least = Corp_Call (1)	Parietal_l (11), Temporal_l (29), Frontal_l (25), Occipital_l (7), Corp_Call (1)
regular	Early/Late_vaccination	binominal	mode = late (37), least = early (36)	late (37), early (36)
regular	WBC(10^9/L)-aph	polynominal	mode = normal (63), least = low (5)	normal (63), low (5), high (5)
regular	Lymphocytes_difference	polynominal	mode = steady (30), least = rise (3)	steady (30), drop (28), relapse_before_P1/V1 (12), rise (3)
regular	Monocytes(10^9/L)-aph	polynominal	mode = normal (63), least = low (0)	normal (63), low (0), high (10)
regular	Monocytes(10^9/L)-P1/V1	polynominal	mode = normal (56), least = low (2)	normal (56), relapse_before_P1/V1 (12), low (2), high (3)
regular	Neutrophils(10^9 / L)-aph	polynominal	mode = normal (62), least = high (4)	normal (62), low (7), high (4)
regular	Neutrophils_difference	polynominal	mode = steady (47), least = rise (4)	steady (47), rise (4), relapse_before_P1/V1 (12), drop (10)
regular	Eosinophils(10^9 / L)-P1/V1	polynominal	mode = normal (57), least = high (4)	normal (57), relapse_before_P1/V1 (12), high (4)
regular	Eosinophils_difference	polynominal	mode = steady (55), least = drop (2)	steady (55), relapse_before_P1/V1 (12), rise (4), drop (2)
regular	Immature platelet fraction (%)-P1/V1	polynominal	mode = normal (49), least = relapse_before_P1/V1 (12)	normal (49), relapse_before_P1/V1 (12), low (12)
regular	Hemoglobin(g/dL)-aph	binominal	mode = low (42), least = normal (31)	low (42), normal (31)
regular	Hematocrite-aph	binominal	mode = low (39), least = normal (34)	normal (34), low (39)
regular	Immature reticulocyte fraction(%)-P1/V1	polynominal	mode = normal (34), least = relapse_before_P1/V1 (12)	low (27), normal (34), relapse_before_P1/V1 (12)

Fig. GB16a. Input features, class and patient utilization (73/82) for classifier 10.

	true below 24	true above 24	class precision
pred. below 24	46	1	97.87%
pred. above 24	6	20	76.92%
class recall	88.46%	95.24%	

Fig. GB16b. Leave-one-out confusion matrix for classifier 10.

Role	Name	Type	Statistics	Range
label	OS_months	nominal	mode = below 24 (53), least = above 24 (21)	below 24 (53), above 24 (21)
regular	Sex	binominal	mode = Male (49), least = Female (25)	Male (49), Female (25)
regular	Behavioral_Changes	binominal	mode = NO (59), least = YES (15)	NO (59), YES (15)
regular	Confusion	binominal	mode = NO (62), least = YES (12)	NO (62), YES (12)
regular	Memory_Disturbance	binominal	mode = NO (61), least = YES (13)	NO (61), YES (13)
regular	Motoric_Symptoms	binominal	mode = NO (57), least = YES (17)	NO (57), YES (17)
regular	Cognitive_Deterioration	binominal	mode = NO (39), least = YES (35)	NO (39), YES (35)
regular	Number_of_locations	binominal	mode = 1.0 (69), least = 2_or_3 (5)	1.0 (69), 2_or_3 (5)
regular	Main_location_of_tumor	polynominal	mode = Temporal_l (29), least = Corp_Call (1)	Parietal_l (12), Temporal_l (29), Frontal_l (25), Occipital_l (7), Corp_Call (1)
regular	Early/Late_vaccination	binominal	mode = late (38), least = early (36)	late (38), early (36)
regular	WBC(10^9/L)-aph	polynominal	mode = normal (63), least = low (5)	normal (63), low (5), high (6)
regular	WBC(10^9/L)-P1/V1	polynominal	mode = normal (48), least = high (1)	normal (48), relapse_before_P1/V1 (12), low (13), high (1)
regular	Lymphocytes_difference	polynominal	mode = steady (31), least = rise (3)	steady (31), drop (28), relapse_before_P1/V1 (12), rise (3)
regular	Monocytes(10^9/L)-P1/V1	polynominal	mode = normal (57), least = low (2)	normal (57), relapse_before_P1/V1 (12), low (2), high (3)
regular	Neutrophils(10^9 / L)-aph	polynominal	mode = normal (62), least = high (5)	normal (62), low (7), high (5)
regular	Eosinophils(10^9 / L)-P1/V1	polynominal	mode = normal (58), least = high (4)	normal (58), relapse_before_P1/V1 (12), high (4)
regular	Immature platelet fraction (%)-P1/V1	polynominal	mode = normal (50), least = relapse_before_P1/V1 (12)	normal (50), relapse_before_P1/V1 (12), low (12)
regular	Immature reticulocyte fraction(%)-P1/V1	polynominal	mode = normal (35), least = relapse_before_P1/V1 (12)	low (27), normal (35), relapse_before_P1/V1 (12)

ExampleSet (74 examples, 1 special attribute, 17 regular attributes)

Fig. GB17a. Input features, class and patient utilization (74/82) for classifier 11.

	true below 24	true above 24	class precision
pred. below 24	46	1	97.87%
pred. above 24	7	20	74.07%
class recall	86.79%	95.24%	

Fig. GB17b. Leave-one-out confusion matrix for classifier 11.

Role	Name	Type	Statistics	Range
label	OS_months	nominal	mode = below 24 (52), least = above 24 (21)	below 24 (52), above 24 (21)
regular	Sex	binominal	mode = Male (48), least = Female (25)	Male (48), Female (25)
regular	Hemiparesis	binominal	mode = NO (65), least = YES (8)	NO (65), YES (8)
regular	Behavioral_Changes	binominal	mode = NO (58), least = YES (15)	NO (58), YES (15)
regular	Confusion	binominal	mode = NO (61), least = YES (12)	NO (61), YES (12)
regular	Memory_Disturbance	binominal	mode = NO (60), least = YES (13)	NO (60), YES (13)
regular	Motoric_Symptoms	binominal	mode = NO (56), least = YES (17)	NO (56), YES (17)
regular	Cognitive_Deterioration	binominal	mode = NO (38), least = YES (35)	NO (38), YES (35)
regular	Number_of_locations	binominal	mode = 1.0 (68), least = 2_or_3 (5)	1.0 (68), 2_or_3 (5)
regular	Main_location_of_tumor	polynominal	mode = Temporal_l (29), least = Corp_Call (1)	Parietal_l (11), Temporal_l (29), Frontal_l (25), Occipital_l (7), Corp_Call (1)
regular	Early/Late_vaccination	binominal	mode = late (37), least = early (36)	late (37), early (36)
regular	WBC(10^9/L)-aph	polynominal	mode = normal (63), least = low (5)	normal (63), low (5), high (5)
regular	Lymphocytes_difference	polynominal	mode = steady (30), least = rise (3)	steady (30), drop (28), relapse_before_P1/V1 (12), rise (3)
regular	Monocytes(10^9/L)-P1/V1	polynominal	mode = normal (56), least = low (2)	normal (56), relapse_before_P1/V1 (12), low (2), high (3)
regular	Neutrophils(10^9 / L)-aph	polynominal	mode = normal (62), least = high (4)	normal (62), low (7), high (4)
regular	Eosinophils(10^9 / L)-P1/V1	polynominal	mode = normal (57), least = high (4)	normal (57), relapse_before_P1/V1 (12), high (4)
regular	Eosinophils_difference	polynominal	mode = steady (55), least = drop (2)	steady (55), relapse_before_P1/V1 (12), rise (4), drop (2)
regular	Immature platelet fraction (%)-P1/V1	polynominal	mode = normal (49), least = relapse_before_P1/V1 (12)	normal (49), relapse_before_P1/V1 (12), low (12)
regular	Hemoglobin(g/dL)-aph	binominal	mode = low (42), least = normal (31)	low (42), normal (31)
regular	Immature reticulocyte fraction(%)-P1/V1	polynominal	mode = normal (34), least = relapse_before_P1/V1 (12)	low (27), normal (34), relapse_before_P1/V1 (12)

ExampleSet (73 examples, 1 special attribute, 19 regular attributes)

Fig. GB18a. Input features, class and patient utilization (73/82) for classifier 12.

	true below 24	true above 24	class precision
pred. below 24	46	2	95.83%
pred. above 24	6	19	76.00%
class recall	88.46%	90.48%	

Fig. GB18b. Leave-one-out confusion matrix for classifier 12.

However, since the sample is small, these classifiers will be further tested in an independent set of 50 patients, to be provided by KU Leuven.

The results up to now agree with the observation that there seems to be a specific group of patients which will benefit from dendritic cell vaccination, namely, the ones that have a high probability of reaching OS>24 months [GB35, GB36, GB37, GB38, GB39]. The methods described above to provide a tool for deciding if a patient is likely to belong in that group, thus providing an answer to research question 2.

In principle, since vaccination schedule is included in the input attributes of the classifiers, the methods described above provide an answer for research question 3. For each specific patient, the two resulting probabilities of reaching OS>24 for early and late vaccination as they are calculated for the classifiers, can be compared. If only one of these probabilities is above 0.5 (threshold of the Bayesian classifier), this will provide an answer. However, since the sample of patients is small and does not contain patients with the same characteristics and different vaccination schedules, it is not clear if the classifiers can provide a confident answer in every occasion (for example, if both vaccination schedules result in OS>24 probability above 0.5).

GB4. References

[GB1] O.Grauer,P.Wesseling G.Adema, *Immunotherapy of diffuse gliomas: Biological Background,Current status and Future Developments* ,Brain Pathology 2009.

[GB2] Kim TG, Kim CH, Park JS, Park SD, Kim CK, Chung DS, Hong YK, *Immunological factors Relating to the Antitumour Effect of Temozolomide Chemoimmunotherapy in a Murine Glioma Model.* Clin Vaccine Immunol. 2010 Jan;17(1):143-53

[GB3] Wainwright DA, Balyasnikova IV, Chang AL, Ahmed AU, Moon KS, Auffinger B, Tobias AL, Han Y, Lesniak MS. , *IDO expression in brain tumours increases the recruitment of regulatory T cells and negatively impacts survival* ,Clin Cancer Res. 2012 Nov 15;18(22):6110-21.

[GB4] Choi BD, Fecci PE, Sampson JH,*Regulatory T cells move in when gliomas say I-DO*, Clin Cancer Res. 2012 Nov 15;18(22):6086-8

[GB5] H.von Boehmer, *Mechanisms of suppression by suppressor T cells,* Nat Immunol. 2005 Apr;6(4):338-44

[GB6] Onishi Y, Fehervari Z, Yamaguchi T, Sakaguchi S, *Foxp3+ natural regulatory T cells preferentially form aggregates on dendritic cells in vitro and actively inhibit their maturation*. Proc Natl Acad Sci U S A. 2008 Jul 22;105(29):10113-8

[GB7] Sakaguchi S, Wing K, Onishi Y, Prieto-Martin P, Yamaguchi T.,*Regulatory T cells: How do they suppress immune responses?* Int Immunol. 2009 Oct;21(10):1105-11.

[GB8] Tadokoro CE, Shakhar G, Shen S, Ding Y, Lino AC, Maraver A, Lafaille JJ, Dustin ML. *Regulatory T cells inhibit stable contacts between CD4+ T cells and dendritic cells in vivo J Exp Med. 2006 Mar 20;203(3):505-11..*

[GB9] Sonabend AM, Rolle CE, Lesniak MS.,*The role of regulatory T cells in malignant Glioma*, Anticancer Res. 2008 Mar-Apr;28(2B):1143-50.

[GB10] Jordan JT, Sun W, Hussain SF, DeAngulo G, Prabhu SS, Heimberger AB. *Preferential migration of regulatory T cells mediated by glioma secreted chemokines can be blocked by chemotherapy*. Cancer Immunol Immunother. 2008 Jan;57(1):123-31.

[GB11 Fecci PE, Mitchell DA, Whitesides JF, Xie W, Friedman AH, Archer GE, Herndon JE 2nd, Bigner DD, Dranoff G, Sampson JH.*Increased regulatrory T cell fraction amidst a diminished CD4 compartment explains cellular immune defects in Patients with malignant glioma. Cancer Res. 2006 Mar 15;66(6):3294-302.*

[GB12] Camara NO, Sebille F, Lechler RI. *Human CD4+CD25+ regulatory cells have marked and sustained effects on CD8+ T cell activation*. Eur J Immunol 2003;33:3473–83

[GB13] Piccirillo CA, Shevach EM. *Cutting edge: control of CD8+ T cell activation by CD4+CD25+ immunoregulatory cells*. J Immunol 2001;167:1137–40.

[GB14] Ogbomo H1, Cinatl J Jr, Mody CH, Forsyth PA. *Immunotherapy in gliomas: Limitations and potential of natural killer cell therapy*. Trends Mol Med. 2011 Aug;17(8):433-41.

[GB15] Albesiano E, Han JE, Lim M. *Mechanisms of local immunoresistance in glioma. Neurosurg Clin N Am. 2010 Jan;21(1):17-29*

[GB16] Crane CA et al, *TGF-beta downregulates the activating receptors NKp30 and NKG2D on NK and CD8+ T cells*. Neuro Oncol 12, 7-13

[GB17] Kane A., Yang I. Interferon-gamma in BrainTumour Immunotherapy Neurosurg Clin N Am. 2010 Jan;21(1):77-86

[GB18] Alberts B. Molecular biology of the cell,chapters 24-25.

[GB19] Yang I, Han SJ, Kaur G, Crane C, Parsa AT. *The role of microglia in Central Nervous System immunity and Immunotherapy*. J Clin Neurosci. 2010 Jan;17(1):6-10

[GB20] da Fonseca AC, Badie B. *Microglia and Macrophages in Malignant Gliomas: Recent Discoveries and Implications for Promising Therapies*. Clin Dev Immunol. 2013;2013:264124.

[GB21] W. Li and M. B. Graeber, *The molecular profile of microglia under the influence of glioma,* Neuro-Oncology 14(8):958–978, 2012

[GB22] Ostrand-Rosenberg S. *Myeloid-derived suppressor cells: more mechanisms for inhibiting anti-tumour immunity*, Cancer Immunol Immunother. 2010 Oct;59(10):1593-600.

[GB23] Kohanbash G, Okada H. *Myeloid-derived Suppressor cells (MDSC's) in Gliomas and Glioma development. Immunol Invest. 2012;41(6-7):658-79.*

[GB24] D'Agostino PM, Gottfried-Blackmore A, Anandasabapathy N, Bulloch K. *Brain dendritic cells: biology and pathology*, Acta Neuropathol. 2012 Nov;124(5):599-614.
[GB25] Banchereau J1, Briere F, Caux C, Davoust J, Lebecque S, Liu YJ, Pulendran B, Palucka K. *Immunobiology of dendritic cells*. Annu Rev Immunol. 2000;18:767-811.

[GB26] Fong L, Engleman EG. *Dendritic Cells in Cancer Immunotherapy*. Annu Rev Immunol. 2000;18:245-73.

[GB27] Mempel TR, Henrickson SE, Von Andrian UH.*T-Cell priming by dendritic cells in lymph nodes occurs in three distinct phases*. Nature. 2004 Jan 8;427(6970):154-9.

[GB28] Wainwright DA, Sengupta S, Han Y, Lesniak MS. *Thymus derived rather than tumour induced regulatory T cells predominate in brain tumours*. Neuro Oncol. 2011 Dec;13(12):1308-[29] Hugues S. *Dynamics of dendritic cell-T cell interactions: a role in T cell outcome*. Semin Immunopathol. 2010 Sep;32(3):227-38.

[GB30] Lämmermann T, Sixt M. *The microanatomy of T-cell responses*. Immunol Rev. 2008 Feb;221:26-43

[GB31] Miller MJ, Safrina O, Parker I, Cahalan MD. *Imaging the single cell dynamics of CD4+ T cell activation by dendritic cells in lymph nodes*. J Exp Med. 2004 Oct 4;200(7):847-56.

[GB32] Ma DY, Clark EA. *The role of CD40 and CD154/CD40L in dendritic cells*. Semin Immunol. 2009 Oct;21(5):265-72.

[GB33] Clarke SR *The critical role of CD40/CD40L in the CD4-dependent generation of CD8+ T cell immunity*. J Leukoc Biol. 2000 May;67(5):607-14.

[GB34] Heath WR, Carbone FR. *Cross-presentation, dendritic cells, tolerance and immunity.* Annu Rev Immunol. 2001;19:47-64.

[GB35] Van Gool SW *Brain Tumour Immunotherapy: What have We Learned so Far?* Front Oncol. 2015 Jun 17;5:98.

[GB36] Dejaegher J, Van Gool S, De Vleeschouwer S *Dendritic cell vaccination for glioblastoma multiforme: review with focus on predictive factors for treatment response* ImmunoTargets and Therapy, 13 March 2014 Volume 2014:3 Pages 55—66

[GB37] Ardon H, Van Gool S, Lopes IS, Maes W, Sciot R, Wilms G, Demaerel P, Bijttebier P, Claes L, Goffin J, Van Calenbergh F, De Vleeschouwer S. *Integration of autologous dendritic cell-based immunotherapy in the primary treatment for patients with newly diagnosed glioblastoma multiforme: a pilot study.* J Neurooncol. 2010 Sep;99(2):261-72.

[GB38] Van Gool S, Maes W, Ardon H, Verschuere T, Van Cauter S, De Vleeschouwer S. *Dendritic cell therapy of high-grade gliomas.* Brain Pathol. 2009 Oct;19(4):694-712

[GB39] Ardon H, Van Gool SW, Verschuere T, Maes W, Fieuws S, Sciot R, Wilms G, Demaerel P, Goffin J, Van Calenbergh F, Menten J, Clement P, Debiec-Rychter M, De Vleeschouwer S. *Integration of autologous dendritic cell-based immunotherapy in the standard of care treatment for patients with newly diagnosed glioblastoma: results of the HGG-2006 phase I/II trial.* Cancer Immunol Immunother. 2012 Nov;61(11):2033-44

[GB40] Garg A D, Vandenberk L, Koks C, Verschuere T, Boon L, Van Gool SW, Agostinis P. *Dendritic cell vaccines based on immunogenic cell death elicit danger signals and T cell–driven rejection of high-grade glioma.* Science Translational Medicine 02 Mar 2016: Vol. 8, Issue 328, pp. 328ra27

[GB41] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, 3rd edition, Morgan Kaufmann, 2011.

[GB42] T. Hastie, R. Tibshirani, J. Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Second Edition (Springer Series in Statistics) 2009.

[GB43] Koller D. and Friedman N. *Probabilistic Graphical Models Principles and Techniques.* The MIT Press, 2009

[GB44] Friedman, N., Geiger, D., Goldszmidt, M. (1997) B*ayesian network classifiers.* Machine Learning, 29, 131–163

[GB45] G. F. Cooper, E. Herskovits. A bayesian method for the induction of probabilistic networks from data. Machine Learning, October 1992, Volume 9, Issue 4, pp 309–347

PR. The Prostate Hypermodel

The development of the Prostate Hypermodel has been led by UNITO. ICCS has fostered the alignment of the hypermodel with the CHIC hypermodelling principles.

In the mainframe of Prostate Cancer (PCa), which is a very diffused pathology addressed by many different therapeutic approaches, we tried to define a scenario in which the main features of the CHIC infrastructure may be foreseen.

After a short presentation of the clinical network in Regione Piemonte that has been built and supported within the CHIC project, attention will be focused on the following points:

1.   the collection and storage of clinical data (EUREKA1 and EUREKA2 studies),

2.   the appropriate ethical and technological tools for effective clinical repositories

3.   the appropriate ICT and computational tools for an efficient model evaluation/validation oriented database

4.   the proposal of predictive hypomodels and their clinical evaluation/validation

5.   the feasibility of clinically oriented hypermodels for PCa

6.   exploitation and dissemination issues

PR1. Introduction

Prostate cancer (PCa) is a slow-proliferating adenocarcinoma with a steadily increasing incidence related to ageing. It is the most common cancer in men and the second most common cause of death from tumours in the male population. In 3 cases on 4 it can be successfully cured, otherwise relapse occurs both locally or inducing distant metastasis.

Like breast cancer for women, hormonal drive is crucial for proliferation and for survival, making PCa highly integrated within the endocrine homeostasis of the whole body.

Together with diffusion and slow proliferation, some other issues make PCa an ideal candidate for modelling and for testing CHIC infrastructures:

1) an easily assessable and cost effective blood marker, the PSA (Prostate Specific Antigen), secreted in the human body exclusively by the prostatic tissue, is currently evaluated in men when PCa is suspected. It is useful for diagnostic purposes and highly reliable for relapse evaluations. Although also gene mutation, protein expression patterns and miRNA regulation are under study and will further disclose prostate cancer physiopathology, prognostic classification and treatment response evaluation, only for PCa, a biomarker is already available on large scales. For prostactomized patients the collection of PSA values may afford a reliable monitoring of tumour growth with no need of specific and repeated follow-up through imaging;

2) the feasibility of needle biopsy and the well consolidated Gleason grading system proposed almost 50 years ago makes it possible a histological analysis of PCa cells i.e. *bioptic Gleason Score (bGS)* which, following radical prostatectomy on the whole prostate specimen, will be

corroborated by the *pathologic Gleason Score (pGS)*.(2005 ISUP Modified Gleason System**)** [PR1]. Both the bGS and the availability of several PSA evaluation add information to the traditional TNM staging system assigned at diagnosis through clinical and radiological exams, i.e. *clinical staging*. For Prostate cancer staging is performed according to American Joint Committee on Cancer (AJCC) 2010, Seventh Edition**.**

3) Prostate cancer is managed by many clinicians in an interdisciplinary way. Practically all the oncology specialists and hospital services have to deal with prostate cancer: Radiotherapists, Urologists and Medical Oncologists for treatment, Pathologists and Radiologists for diagnosis and staging, Epidemiologists and even General Practitioners for the widespread diffusion in the population and for the co-operation during follow-up. Different therapies are also performed, ranging from Surgery, Radiotherapy and drugs (mainly Androgen Deprivation Therapy, ADT). Guidelines for the clinical use of the various treatment modalities have been proposed by several clinical associations, such as European Association of Urology (EAU) [PR2, PR3] for urologists, NCCN (National Comprehensive Cancer Network) [PR4, PR5] for radiotherapists and Piedmont Guidelines [PR6] for the multidisciplinary approach including medical oncologists. In general, surgery (Radical Prostatectomy, RP) is indicated in low and intermediate-risk patients, EBRT (External Beam Radiation Therapy) in all risk patients, joined to ADT in intermediate ad high-risk patients; ADT (eventually with palliative RT) is fostered for metastatic disease, while in castration-resistant patients chemotherapy or new drugs are indicated.

These different therapeutic approaches offers a wider range of possibilities to modelers in order to test several hypo- and hyper-models.

PR2. Predicting (Hypo-)Models for PCa Recurrence After Surgery

In recurrent PCa, after a variable time following the surgery, a so-called "biochemical recurrence" (BCR) is observed, with a progressive rise of the PSA values above or at 0.20 ng/ml after RP (see [PR7]).

Predicting the probability of recurrence of PCa after RP is one of the main goals of studies and researches in this field. Roughly speaking, there are two main ways of thinking: one relates the recurrence probability and its timing to the pre-operative tumour characteristics (e.g. Gleason Score, tumour stage, surgical methods and so on – static models) while the second one investigates the postoperative tumour dynamics mainly based on the PSA growth timing (dynamic models).

Static models are normally validated on huge clinical database, and aim at producing simple and reliable tools for addressing therapeutic decisions. Very popular nomograms have been proposed, starting from the first model of [PR8], the GPSM (Gleason, PSA, Seminal Vesicle and Margin Status) proposed by [PR9], the nomogram of Briganti [PR10] and all their updated versions.

Dynamic models, already proposed by D'Amico [PR11] and independently by Yorke [PR12], and further developed by Dimonte [PR13] focus on the estimation of PSA velocity and doubling time, evaluating the timing of tumour proliferation from serial PSA measurements.

As an alternative to direct evaluation of PSA data dynamics, the use of the Universal Growth Law (UGL) proposed by West et al.[PR14] and formerly applied to tumours by Guiot et al.[PR15], may offer a frame for describing PSA (and tumour) growth predicting its clinical outcome. In particular, two parameters of the equation should be analysed: the growth parameter alpha and the carrying capacity (P in the following equations). According to [PR14, PR15], the carrying capacity should be set depending on the physical problem, i.e. the maximum value achievable by the population taken into account.

Based on the PCa tumour volumes measured along the natural history of PCa [PR12, PR15], the growth parameter alpha estimated by the model is about 0.42 g^0.25 /day. If a few surviving PCa cells start proliferating and producing a detectable quantity of PSA after RP, we assume that the PSA detected in the patient serum should follow the same 'universal law' with a value of alpha which reflects its actual proliferation. Accordingly to West and colleagues, alpha is the ratio between the metabolic energy and the energy needed for duplication of a given cell type. In case of PSA production, it is expected to increase when cells actively produce PSA and reproduce themselves. Moreover, it should decrease when cells produce small amounts of PSA and duplicate slowly.

The parameter alpha can then be added to the standard clinical parameters into a statistical multi-parametric model to estimate timing of recurrence on retrospective data. Moreover, in order to devise a robust and easily manageable tool for clinicians, we investigated how many PSA values are needed to make a good prediction. Predictability of the growth parameter based on the first 3 detectable PSA measurements are investigated to study the feasibility of a rapid estimation procedure.

PR2.1 Mathematical Procedure

According to the model presented by Thompson et al. [PR7], we suppose that the PSA dosage p collected at time t reflect the actual tumour mass at that stage and tumour growth can be described as follow:

$$\frac{dp}{dt} = \alpha\, p^{\frac{3}{4}} \left[1 - \left(\frac{p}{P} \right)^{\frac{1}{4}} \right]$$

where p is the PSA value (in ng/ml), P is its maximum value reached by the population. In our case $P = 100$ ng/ml, t is the timing of the measurement expressed in months after surgery and alpha is the growth parameter for PSA.

Physical data can be re-normalized following simple calculations (see [PR7] for the details), in terms of rescaled tumour fraction $r = (p/P)^{1/4}$ and rescaled time

$$\tau = \frac{\alpha\, t}{4 P^{\frac{1}{4}}} - \ln\left(1 - \left(\frac{p_0}{P} \right)^{\frac{1}{4}} \right)$$

where $p0$ is the initial value of the series. Far from being mathematical tricks, the rescaled units allows us a quick comparison between our experimental data and the parameterless universal curve

$$r = 1 - e^{-\tau}$$

obtained by substituting r and alpha in the previous equations.

Secondly, we can evaluate the alpha parameter value deterministically by averaging the single alpha values corresponding to the measured PSA:

$$\alpha = \frac{1}{n}\sum \alpha_i = \frac{1}{n}\sum 4P^{\frac{1}{4}}\frac{\log(1-\frac{p_0}{P})-\log(1-\frac{p_i}{P})}{t_i}$$

where p_i (= p_i) , t_i (=t_i) are the non null ith PSA values and the times of the measurement respectively and n is the size of the sample. Note that in some cases this method failed and alpha results equal to zero, hence the time to relapse can not be estimated.

In case of adjuvant ADT, which tends to depress the PSA value and the tumour volume during its action, we calculated alpha as the average of α_i (=alpha_i) from the set of PSA values taken after the end of it.

In clinical practice, however, time to relapse should be predicted on the basis of only pre-relapse data! We investigated therefore the reliability of the estimation of alpha using only the a limited number of PSA values collected before relapse, with a maximum of 4. We call this estimation as α_4 (= alpha_4).

The alpha_4 value has the same behaviour of the alpha and the two parameters can be compared as shown in Fig. PR1. The coloured circles represent the values of alpha_4 versus the time to relapse. The intercept (the more marked one) has a p-value p<0.001.
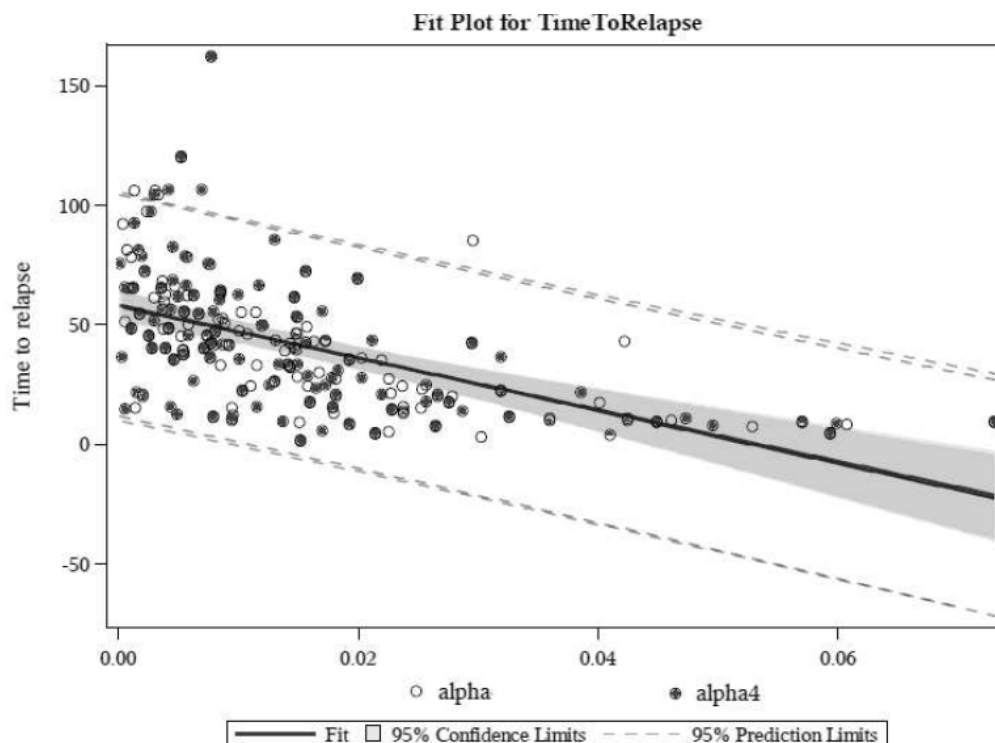
Fig. PR1 - Fit plot of the estimated versus real time to relapse. Black dots are the alpha values estimated with only 4 PSA

PR2.2 Results and Conclusions

As far as patients without any adjuvant therapy are concerned, the statistical analysis shows that the time to relapse is independent from any clinical information apart from the first post-operative PSA value and the alpha parameter. In particular, alpha is predictive of the range of the time of relapse, being the smaller alpha values predicting the longer free from disease time (see [PR27] for details).

Contrary to retrospective studies, waiting for a long PSA series to make a prevision is impractical and ethically questionable in clinical practice: since PSA dosage is usually prescribed every 3 – 6 months, waiting for sixth-eighth PSA values makes predicting early recurrences senseless.

We can conclude that a careful collection of PSA values would be valuable and overcome traditional clinical parameters (such as pGS, pathological staging, etc.) to predict the timing of the tumour recurrence development.

PR3. Predictive (hypo-) Models for PCa Recurrence After Radiotherapy

The study was based on a large database of 2493 patients affected by prostate cancer and treated with EBRT as primary treatment belonging to to the EUREKA-2 retrospective multi-centric database, including 3776 cases of radio-treated prostate cancer cases in North-West Italy between 1997 and 2012, approved by FPO-IRCCS Cancer Center of Candiolo Ethical Committee in July 2013 and amended in November 2014.

In particular, from the whole database were excluded 1283 patients without complete information regarding established pre-treatment factors (PSA, clinical-radiologic stage and bGS) and the number of total and positive biopsy cores.

In all cases, staging evaluation included anamnesis, physical exam with Digital Rectal Examination (DRE), serum PSA and a trans-rectal ultrasound (TRUS) guided needle biopsy of the prostate with GS histologic grading. Radiological examinations (abdominal CT, endo-coil or pelvic MRI and bone scan) were performed according to the patient risk-class, to the physician's opinion and to the available hospital facilities.

All patients were treated with curative 3-Dimensional Conformal EBRT (3D-CRT) or Intensity Modulated Radiation Therapy (IMRT). The fractionation schedules to prostate-GTV (Gross Tumour Volume) varied between traditional fractionation of 1.8-2 Gy per fraction to moderate hypo-fractionation of 2.5-2.7 Gy per fraction; all doses were normalized to Equivalent Dose at 2 Gy per fraction (ED2Gy) using a mean α/β of 2.5 Gy for prostate cancer (according to literature α/β ratio for prostate cancer ranges between 1.5 and 5.7 Gy [PR16-PR18]). Treatment consisted of radiotherapy alone or radiotherapy combined with ADT in 38% and 62% of the cases, respectively.

Median follow-up of the 2493 patients was 50 months. Standard follow-up included PSA and DRE every 3-months for 2 years, every 6-months until the fifth year and annually thereafter.

During the follow-up 453 patients (18%) had a biochemical relapse, 249 (10%) relapsed clinically, 138 (5.5%) had distant metastases, and 233 (9%) died, 72 of these (3% of the total) because of prostate cancer. Time 0 was defined as the last day of EBRT for all patients and PSA failure according to Phoenix consensus definition (i.e. a rise by 2 ng/mL or more above the nadir PSA [PR19]).

Clinical relapse was defined as a recurrence in the prostate bed, regional lymph nodes or distant metastasis shown by radiologic examinations (bone scan, choline-PET-CT, MRI, CT, ultrasound) or by physical examination or by biopsy. Systemic relapse was defined as a distant metastasis, including bone or other visceral organs, shown by radiologic examinations or by physical examination. Prostate cancer specific mortality was defined as death because of prostate cancer, checked by a physician through patients' case history reports, cancer regional registries and, if necessary, phone calls to the patient or to a close relative or General Practitioner of him (if the patient was dead).

Figure 1 – Candiolo nomogram.

Points:

bGS≤6 0 pt, bGS=3+4 35 pt, bGS=4+3 48 pt, bGS=8 76 pt, bGS=9-10 106 pt;

cT1 0 pt, cT2 17 pt, cT3-4 58 pt;

PSA<7 0 pt, PSA7-15 42 pt, PSA>15 96 pt;

%PC 1-20% 0 pt, 21-50% 29 pt, 51-80% 50 pt, 81-100% 81 pt;

age≥70 0 pt, age<70 22 pt;

very-low risk 0-56 pt, low risk 57-116 pt, intermediate risk 117-193 pt, high risk 194-262 pt, very-high risk 263-363 points.
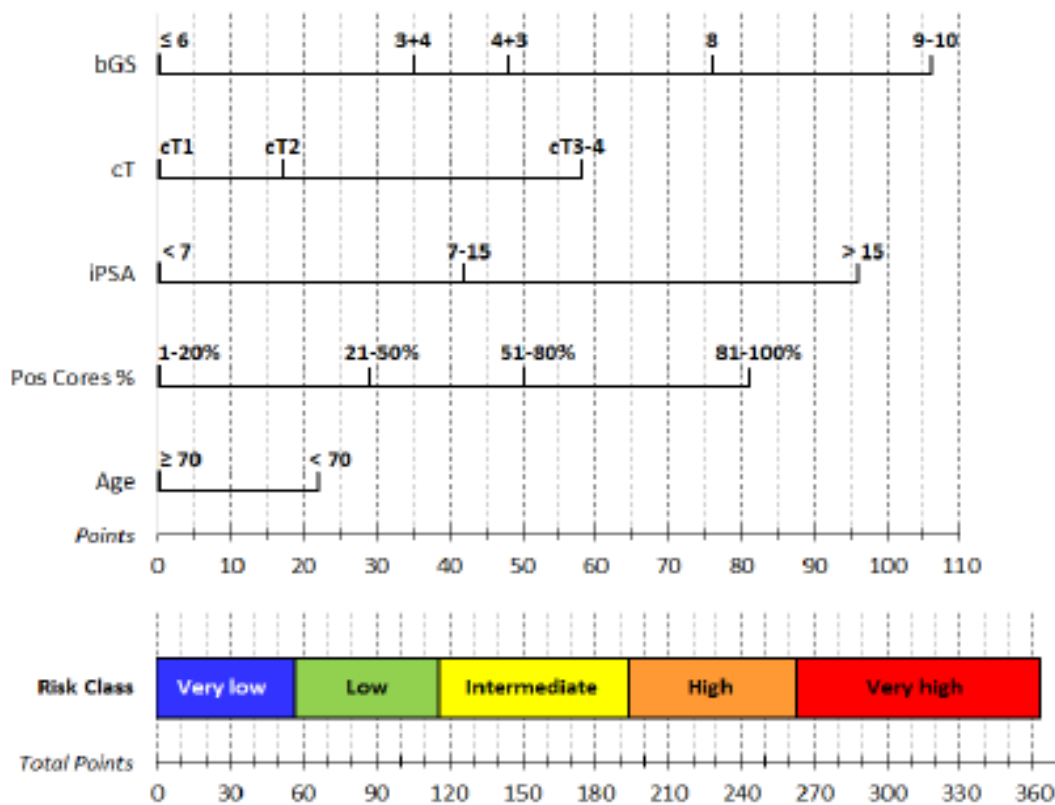
Fig. PR2 – The Candiolo nomogram.

PR4. Nomogram after Surgery

A nomogram on 976 prostatectomized patients is now under construction and validation.

Here we report the draft.
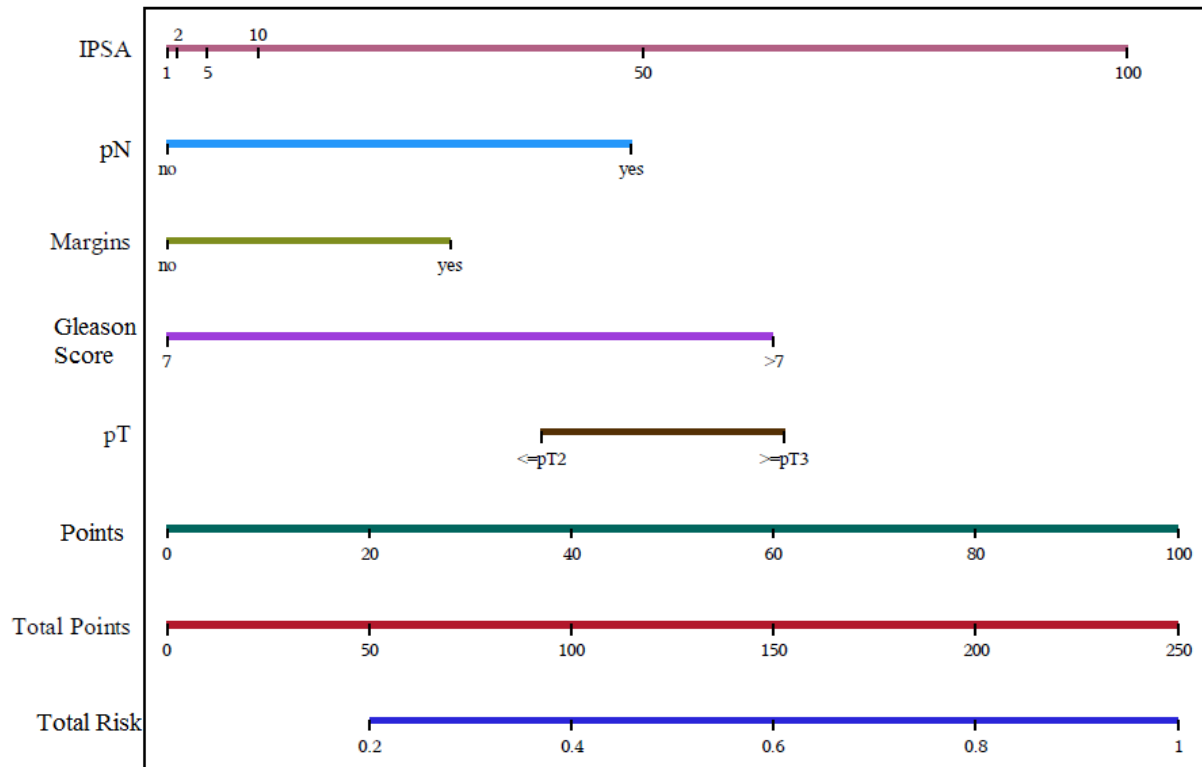
PR5. (Hypo-) Models for Evaluating Resistance Induction by Hormonal Therapies

PR5.1 Introduction

Normally, the prostate cancer cells are hormone-sensitive cells. They produce PSA and they need a hormone to stimulate growth. There is a very small group, however, which is not hormone-sensitive but hormone-resistant [PR23]. This type of cells does not need hormones to duplicate but its growth potential is very low respect to the hormone-sensitive ones.

Nowadays, hormone therapies are very common to contrast the growing of hormone-sensitive tumours, like breast and prostate cancer. In this last case the growth of hormone-resistant cells will finally induce an almost uncontrollable increase after an initial reduction of the tumour volume [PR23]. Any realistic model should therefore take into account the appearance of therapy-induced cell mutations (or phenotypic modifications [PR24]).

The key question in this context is how strong can the interplay of the different cell populations be. Since they are part of the same organism, a ``minimal'' hypothesis states that they share the same overall energetic resources. It is therefore reasonable to assume that the total tumour carrying capacity is limited, and the growth of both cell populations is constrained [PR25, PR26].

We investigated an asymmetrical two-cell-populations model, identifying its equilibria. Their stability or instability expresses the successful cure or the fatal evolution of the tumour. We identify the parameter conditions ensuring the stable configuration, i.e. the situation where the tumour stops growing.

PR5.2 Two Populations Model

We assume that the two populations respond to treatments differently. In particular is proved [PR23] that there is an androgen dependent (AD) cell population, very sensitive to the hormone therapy while an androgen independent (AI) population less sensitive or not sensitive at all to it.

The system is:

$$
\begin{cases}
\dfrac{dN_1}{dt} = c_1(t)N_1 - d_1(t)N_1 \\
\dfrac{dN_2}{dt} = c_2(t)N_2 - d_2(t)N_2 + \epsilon N_2(t)
\end{cases}
$$

where $d_1(t)$ and $d_2(t)$ are the treatment kill rates on the populations N_1 and N_2 respectively and epsilon is a measure of the metabolic rate increment in response to particularly favourable growth conditions. It also expresses an additional growth rate for the second population. In principle, d_1 and d_2 could be functions of time to account for different times of treatments and one- or multi-shot therapies.

The mathematical analysis of the system allows us to predict the behaviour of the two populations, knowing the parameters values. Moreover, we are able to know threshold values for the kill rates in order to eradicate one or both populations. For example, we can say that in the Gompertzian model the AD can be considered eradicated when $d_1 > 10\, r_1$. The best scenario (death of both populations) is not attained so easily, since $d_2 >> epsilon + r_2$.
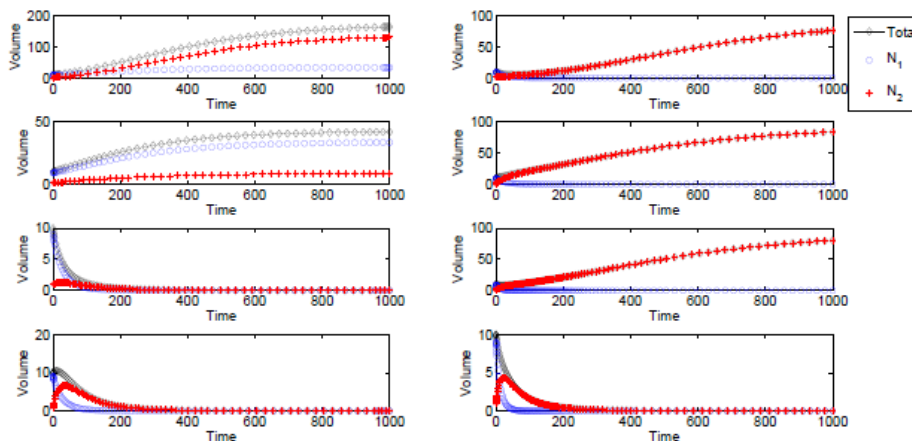
Mutation and treatment in U_1 case. The parameter reference values in all frames, unless otherwise noted, are $r_1 = 0.005, r_2 = 0.003, m = 0.005, d_1 = 0.005, d_2 = 0.003$. Top row left: all parameters are at the low reference values entailing coexistence of the two populations. Top row right: high kill rate of the first population ($d_1 = 0.05$), but the second one survives; Second row left: high kill rate of the second population ($d_2 = 0.03$), the first one prevails; Second row right: high mutation rate ($m = 0.05$), only the second one thrives; Third row left: high kill rates ($d_1 = 0.05, d_2 = 0.03$), both populations are eradicated; Third row right: high kill rate of the first population and high mutation rate ($d_1 = 0.05, m = 0.05$), the first population decreases rapidly; Bottom row left: high kill rate of the second population and high mutation rate ($d_2 = 0.03, m = 0.05$), both populations vanish; Bottom row right: high kill and high mutation rates ($d_1 = 0.05, d_2 = 0.03, m = 0.05$), both populations disappear.

PR5.3 Results and Discussion

The approach of the ``Phenomenological Universalities" allows a satisfactory investigation of the growth of an asymmetrical two-population cancer. Different interactions were studied, corresponding to different clinical scenarios, i.e. the growth of both populations constrained by a fixed total carrying capacity, the response to treatments, the occurrence of spontaneous or induced mutations.

We applied the Gompertzian and West functions to model the growth of the cells populations in a manageable and realistic way. They have been successfully validated on various tumour scenarios, finding analytical solutions whenever possible. Numerical simulations assessed the effectiveness and role of the model parameters in the remaining cases.

Our two-clones model confirms that effective ADT can reduce the AD cell population although the eradication of the AI cells is much more critical. This happen in the presence of spontaneous mutations and, even worse, when mutations are induced or promoted by therapies.

The West growth assumption represents an optimal model for the simulations of the tumour development and response to therapies. This happens because of the biological significance of the growth parameter, related to the cellular metabolic rate and duplication energy of each specific cell population. Moreover, a clinical application is provided.

We want to note that a 'standard' kill rate can be found in literature, but also a 'personalized' kill rate can be used. In particular, UPENN model could provide this value in the cases in which the genetic profile is provided by the hospital.

PR6. Present and Future Work

PR6.1 Model Approach to ADT and Immuno Therapy (IT) Salvage Therapy Treatment Schedules

The hypomodel will predict the tumour response to treatments, according to the collected data. We will take into account also the available knowledge about the adaptations/mutations occurring in the PCa cells undergoing ADT or other CT. Vaccination strategies will be evaluated as well.

A new model is under construction and validation, in order to find differences and similarities between the ADT (most used in Italy) and IT (mostly used in USA).

In this work, the collaboration with UPENN should be very important in order to study the (possible) correspondence between genetic patways and better response to ADT/IT.

PR6.2: Joint Work with BED

Following Radical Prostatectomy a temporary impairment of the urogenital functions is normally suffered. Incontinence problems normally disappears within a few months, but it may continue to occur. This worsen the patient's quality of life, unless specific therapeutic strategy is performed. Moreover, the PSA must be under control for the all life, to monitor every possible case of relapse.
The basic idea of the joint work of UNITO and BED is to integrate two different approaches: a user-friendly visual interface and a more rigorous control by the clinician using quantitative parameters.

As concerns the first approach, we created a set of 3D animations using Blender, then to provide them as 'serious games' to patients. This SW allows planning of personalized ' at home' training sessions, with or without the active surveillance of the clinician.

As concerns the second one, the model described in section PR2 and the nomogram (section PR4) are implemented in conjunction with a MyHealthAvatar project application.

The patient, after surgery, can download (for free) the App, then follow the rehabilitation exercises as videos and insert his own data. The data are checked and seen by the urologist that can run the models and use the results to adjust the therapy or the frequency of the PSA controls.

PR7. References

PR1. Epstein JI, Allsbrook WC Jr, Amin MB, Egevad LL; ISUP Grading Committee. The 2005 International Society of Urological Pathology (ISUP) Consensus Conference on Gleason grading of prostatic carcinoma. Am J Surg Pathol 2005;29(9):1228–1242

PR2. Heidenreich A, Bastian PJ, Bellmunt J, Bolla M, Joniau S, van der Kwast T, et al. EAU Guidelines on prostate cancer. part 1: screening, diagnosis, and local treatment with curative intent-update 2013. Eur Urol. 2014 Jan;65(1):124–37

PR3. Heidenreich A, Bastian PJ, Bellmunt J, Bolla M, Joniau S, van der Kwast T, et al. EAU guidelines on prostate cancer. Part II: Treatment of advanced, relapsing, and castration-resistant prostate cancer. Eur Urol. 2014 Feb;65(2):467–79

PR4. Mohler JL, Kantoff PW, Armstrong AJ, Bahnson RR, Cohen M, D'Amico AV, et al. Prostate cancer, version 2. 2014. J Natl Compr Cancer Netw JNCCN. 2014 May;12(5):686–718

PR5. National Comprehensive Cancer Network, http://www.nccn.org/

PR6. CARCINOMA DELLA PROSTATA, Linee guida clinico organizzative per la Regione Piemonte, Commissione Oncologica Regionale e Centro di Riferimento per l'Epidemiologia e la Prevenzione Oncologica in Piemonte. Luglio 2009

PR7. Thompson,I. M. and Valicenti,R. K. and Albertsen,P. and Davis,B. J. and Goldenberg,S. L. and Hahn,C. and Klein,E. and Michalski,J. and Roach,M. and Sartor,O. and Wolf Jr.,J. S. and Faraday,M. M., *Adjuvant and salvage radiotherapy after prostatectomy: AUA/ASTRO guideline*, Journal of Urology (2013), 190:2, 441-449

PR8. Partin, AW and Kattan, MW and Subong, EN and Walsh, PC and Wojno, KJ and Oesterling, JE and Scardino, PT and Pearson, JD, *Combination of prostate-specific antigen, clinical stage, and Gleason score to predict pathological stage of localized prostate cancer. A multi-institutional update*, JAMA (1997), 18, 1445-51

PR9. Blute,ML. and Jacobsen,SJ. and Kaplan,SA. and Lowe,FC. and O'Leary,MP. and Steers,WD. and Roehrborn,CG., *Conference on Evaluation and Managment of Benign prostatic Hyperplasia, Cambridge, Massachusets, March 31 2001: Introduction*, Urology (2001), 58:6 SUPPL. 1

PR10. Briganti,A. and Chun,FK. and Salonia,A. and Zanni,G. and Scattoni,V. and Valiquette,L. and Rigatti,P. and Montorsi,F. and Karakiewicz,PI., *Validation of a Nomogram Predicting the Probability of Lymph Node Invasion among Patients Undergoing Radical Prostatectomy and an Extended Pelvic Lymphadenectomy*, European urology (2006), 49:6, 1019-1027

PR11. D'Amico,A. and Chen,M. and Roehl,KA. and Catalona,WJ., *Identifying patients at risk for significant versus clinically insignificant postoperative prostate-specific antigen failure*, Journal of Clinical Oncology (2005), 23:22, 4975-4979

PR12. Yorke,E. D. and Fuks,Z. and Norton,L. and Whitmore,W. and Ling,CC., *Preoperative PSA velocity is an independent prognostic factor for relapse after radical prostatectomy*, Journal of Clinical Oncology (2005), 23:25, 6157-62

PR13. Dimonte,G. and Bergstralh,EJ. and Bolander,M. E. and Karnes,RJ. and Tindall,DJ., *Use of tumour dynamics to clarify the observed variability among biochemical recurrence nomograms for prostate cancer*, Prostate (2012), 72:3, 280-290

PR14. West,G.B. and Brown,J.H. and Enquist,B.J., *A general model for ontogenetic growth*, Nature (2001), 413, 628-631

PR15. Guiot,C. and Degiorgis, P.G. and Delsanto,PP. and Gabriele,P. and Deisboeck, TS., *Does Tumour growth follow a 'universal law'?,* Journal of Theoretical Biology (2003), 225, 147-151

PR16. Lin Y-K, Gettle L, Raman JD. Variable prostate-specific antigen management patterns by neurologist providers at a tertiary care medical center. Urology. 2011 Aug;78(2):244–8

PR17. Djavan B, Ravery V, Zlotta A, Dobronski P, Dobrovits M, Fakhari M, et al. Prospective evaluation of prostate cancer detected on biopsies 1, 2, 3 and 4: when should we stop? J Urol. 2001 Nov;166(5):1679–83

PR18. Epstein JI, Allsbrook WC, Amin MB, Egevad LL, ISUP Grading Committee. The 2005 International Society of Urological Pathology (ISUP) Consensus Conference on Gleason Grading of Prostatic Carcinoma. Am J Surg Pathol. 2005 Sep;29(9):1228–42

PR19. Kunz GM, Epstein JI. Should each core with prostate cancer be assigned a separate Gleason Score? Hum Pathol. 2003 Sep;34(9):911–4

PR20. Carter HB, Albertsen PC, Barry MJ, Etzioni R, Freedland SJ, Greene KL, et al. Early detection of prostate cancer: AUA Guideline. J Urol. 2013 Aug;190(2):419–26

PR21. Heidenreich A, Bastian PJ, Bellmunt J, Bolla M, Joniau S, van der Kwast T, et al. EAU guidelines on prostate cancer. part 1: screening, diagnosis, and local treatment with curative intent-update 2013. Eur Urol. 2014 Jan;65(1):124–37

PR22. Portz,T. and Kuang,Y. and Nagy,JD., *A clinical data validated mathematical model of prostate cancer growth under intermittent androgen suppression therapy*, AIP Advances(2012), 2:1

PR23. Chisholm, R. H., T. Lorenzi, A. Lorz, A. K. Larsen, L. N. De Almeida, A. Escargueil, and J. Clairambault. 2015. *Emergence of Drug Tolerance in Cancer Cell Populations: An Evolutionary Outcome of Selection, Nongenetic Instability, and Stress-Induced Adaptation.* Cancer Research 75 (6): 930-939. doi:10.1158/0008-5472.CAN-14-2103

PR24. Hanin,L. and Bunimovich-Mendrazitsky,S., *Reconstruction of the natural history of metastatic cancer and assessment of the effects of surgery: Gompertzian growth of the primary tumour*, Journal of Theoretical Biology (2014), 247, 47-58

PR25. Stura,I. and Gabriele,D. and Guiot,C., *Modeling prostate cancer within CHIC*, Minerva Urologica and Nefrologica, 1:1. 97-98

PR26. Stamatakos,G. S. and Kolokotroni,E. and Dionysiou,D. and Veith,C. and Kim,Y. -. and Franz,A. and Marias,K. and Sabczynski,J. and Bohle,R. and Graf,N. *In silico oncology: Exploiting clinical studies to clinically adapt and validate multiscale oncosimulators*, Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS (2013), p. 5545-5549

PR27. Stura,I. and Gabriele,D. and Guiot,C., *A Simple PSA-Based Computational Approach Predicts the Timing of Cancer Relapse in Prostatectomized Patients*, Cancer Research, to appear (in August)

SE. A Brief Outline of the Process of Semantically Annotating Hypomodels and Hypermodels

The development of the Semantic Annotation of Hypomodels and Hypermodels has been led by UCL. ICCS has provided the specific overarching principles for a coherent and useful annotation of the CHIC hypomodels and hypermodels.

SE1. Metadata Annotation of the Models

The formal machine-processable annotation of hypomodels have been designed and implemented.

The formal, machine processable annotation of hypomodel parameters has been designed.

A mechanism whereby the compositional structure of hypermodels and qualitative relations between hypermodels could be formalised so as to facilitate model selection and model comparison in tools such as CRAF is under development.

SE2. Design and Implementation of Hypomodel Semantic Categorisation

The document D7.3 "Hypermodel Annotation Services" has served as the basis of the related work. D7.3 reviews the requirements relating to the use of semantic technologies within the CHIC infrastructure and CHIC tools. The deliverable also outlines the scope, the content and structure of the formal solution used in order to describe models though semantic annotation. The implementation relies on the use of the W3C Standard Resource Description Framework.

The starting baseline is a manual annotation of selected models in order to validate the implementation of the design and prototype the development of software support for the creation, storage and retrieval of model annotations. A paradigmatic development case has been that of the construction of a nephroblastoma hypermodel from a number of available nephroblastoma hypomodels. The hypomodels have been annotated according to three categorisation perspectives namely perspective I (tissue type), perspective IV (biomechanisms) and perspective V (tumour type). These perspectives are defined in deliverable D6.1 entitled "Cancer Hypomodelling and Hypermodelling Strategies and Initial Component Models." Input from modellers has been used in order to determine the (hypo-)model characteristics.

The formalisation has been carried out in the RDF language and has used the vocabulary (called an 'RDF schema' or sometimes an 'ontology') designed for the CHIC resources called CHICRO: https://github.com/open-physiology/chic/tree/master/ontologies/internal. We relied on a convention defined by the Model Repository in order to refer to model through a globally unique identifier. For example, https://mr.chic-vph.eu/metadata#04e3c5aa-ad45-11e5-bd32-fa163e092aac is the identifier of the ICCS Wilms Oncosimulator hypomodel. Using CHICRO, we have been able to encode the location of such a model within the defined perspectives. While perspectives are defined informally as categorising models using natural language expressions, formal semantic description uses language independent identifiers representing the meaning associated with linguistic descriptions. We have derived these identifiers from existing ontologies and in some cases created new identifiers allowing to fill gaps in the available repertoire of formalised ontologies. For example, in order to point at the category of neoplasm or tumour, we have used an ontological term from the Human Phenotype Ontology with identifier http://purl.obolibrary.org/obo/HP_0002664 . The ascription of the Wilms Tumour model to that value in the Perspective I is formally encoded in the following statement:

<https://mr.chic-vph.eu/metadata#04e3c5aa-ad45-11e5-bd32-fa163e092aac>

<http://www.chic-vph.eu/ontologies/resource#hasPositionIn-1>

<http://purl.obolibrary.org/obo/HP_0002664>

This initial prototyping work has enabled us to refine our knowledge representation framework and to enhance the meta-model for annotation so as to support also the work of applications such as the Model Repository categorisation interface. By storing additional information and constraints of the CHICRO vocabulary, such as which values are allowed for a given perspective, and using services programmatically accessible, the semantic infrastructure has become better integrated and self-contained.

The semantic work carried out from this point onwards consists in extending the initial treatment to other perspectives that could then be handled uniformly by client tools. The manual prototype of the nephroblastoma hypomodels has been replaced with annotations generated through the Graphical User Interface (GUI) where annotation of additional models has taken place.

SE3. Design of Model Parameter Annotations

Deliverable D7.3 outlines the annotation of model parameters. Extending the work to support a similar treatment as explained above is ongoing. The areas of interest include the representation of the logic of model parameters themselves, such as whether a parameter is an input or output parameter for a given model. This characterisation is already supported in CHICRO. An important process, however, is to provide the means for formally recording a number of additional characteristics that are meaningful in the analysis of models and support the composition of hypomodels into hypermodels. The conceptual work done so far points at the description of two aspects specifically: i) units of measurements and ii) semantic interpretation.

Semantic technologies and ontological resources are useful in this context in order to provide *first*: standardised identifiers for units of measurements involved and, *second*: standardised identifiers for the biological interpretation of model parameters. These standardised terms are useful in order to facilitate the semantic integration and to benefit from the formalised meaning associated with the terms. For example, we abstract from different conventional notations when using an identifier for the millilitre unit of volume instead of relying on strings such as 'ml' or 'milliL'. An advantage we would like to derive from the use of a formalised representation of unit of measurement is, however, more significant and consists in recognising equivalent or convertible units as this would greatly facilitate the matching of parameters in the construction of hypermodels. The motivation is very similar for the record of semantic types characterising a parameter. However, the technical difficulty is greater given the open-ended nature of interpretation available. There is no tool allowing the *on the fly* creation of well-defined ontological terms at the time of annotation. Our strategy is therefore somewhat conservative and aims at providing a range of generic reusable terms belonging to well articulated ontologies that may not allow the annotation of a parameter with its most specific and detailed meaning. Despite this apparent limitation, the trade-off is a gain in applicability of the approach. This approach also allows for further development and refinement.

SE4. Exploratory Design of Hypermodel Representation

We have initiated a thread of work regarding how semantic metadata could facilitate the explicit description of qualitative properties of hypermodels and their relations. This strategy would consist in developing a modest vocabulary for characteristics and comparative relations corresponding to heuristics guiding model selection in a context of use. We are not proposing to power a model ranking and selection mechanism with semantics as the work required would be too substantial, However, explicit descriptions of a limited range of information could facilitate enhancements to the user experience.

MR. A Brief Outline of the Model Repository in the Context of Hypomodel Integration

The development of the Model Repository has been led by ICCS.

MR1. The Model Repository Facilitates the Semantic Annotation of the Model Categorization based on 13 Perspectives

The Model and Tool Repository is the CHIC component which permanently hosts the models that have been developed in the context of the CHIC project. It also hosts tools such as linkers and data transformation tools which are necessary for the construction of the hypermodels. For each model/tool, the Model and Tool Repository stores all the related information including:

- descriptive information (information about the models, references, etc.)

- input and output parameters (for proper linking with other models/tools)

- files related to each model (executables, documentation, etc.)

- categorization of the models based on the perspective from which they are viewed in the basic science context.

Currently, most of the information related to models and tools is stored in a relational MySQL database. The use of a relational MySQL database provides many benefits, like the following:
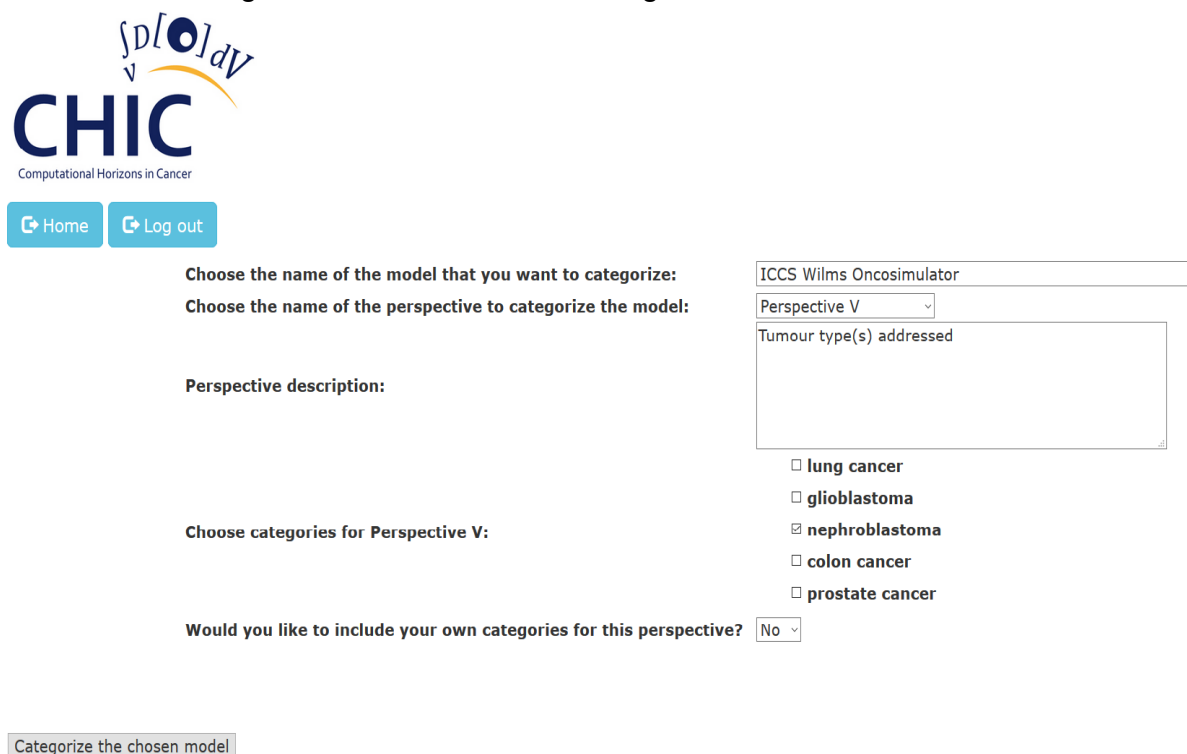
- SQL (Structured Query Language) databases are long-established standard, which is being adopted by ANSI & ISO.

- SQL queries can be used to retrieve large amounts of records from a database quickly and efficiently.

- By using standard SQL it is easier to manage database systems without having to write substantial amount of code.

- SQL is a complete language for a database and it is used to create databases, manage security of a database, update, retrieve and share data with users.

- SQL is used for linking front end computers and back end databases. Thus, it provides a client server architecture.

- SQL supports the latest object based programming.

- SQL is the database language which is used by businesses and enterprises throughout the globe.

Even though there are many reasons for using a relational database, some of the meta-information related to models and tools will be converted to RDF triples so as to be stored in the CHIC

triplestore. RDF triples can be applied equally to all structured, semi-structured and unstructured content. By defining new types and predicates, it is possible to create more expressive vocabularies within RDF in order to describe information related to models. This expressiveness enables RDF to define controlled vocabularies with exact semantics. Furthermore, triplestores have the ability to ingest diverse data, providing flexibility with respect to schema changes and mappings. They also allow for greater freedom, efficient handling of powerful queries and serving unforeseen information needs. Moreover, they employ intelligent data management solutions which combine full text search with graph analytics and logical reasoning to produce deep, rich results. The cost for data integration, management and query definition is much lower than other approaches. It must be also noted that these databases (also known as RDF, OWL, or Graph databases) are now widely used to manage unstructured and structured data in media and publishing, life sciences and financial services.

Consequently, the Model and Tool Repository has been updated in order to be able to automatically store into the CHIC RDF triplestore, information related to the categorization of the models. As stated in the Deliverable 6.1 "Cancer hypomodelling and hypermodelling strategies and initial component models", mathematical and computational cancer models can be categorized depending on the perspective from which they are viewed in the basic science context. The definition of the thirteen perspectives and their indicative values is included in the aforementioned deliverable.

Consequently, the Model and Tool repository and the CHIC semantics infrastructure make use of a common RDF mapping configuration file so as to produce a model (a set of RDF triples) based on the already locally stored relational data. The aforementioned configuration file maps some of the model repository's database tables and columns to CHIC RDF vocabularies and OWL ontologies. This mapping defines the virtual RDF graph that contains some of the information from the Model and Tool repository's MySQL database which is related to the categorization of the models. With this kind of integration between the Model repository and the CHIC triplestore, the user is able to categorize their model by visiting only a single CHIC component. After the submission of the user's data, it is the Model repository's responsibility to store the information related to the categorization of the model both to the repository's relational database and to the CHIC triplestore. The page where the user categorizes their model is shown in Fig. MR1.

Fig. MR1: The web page where the user categorizes their model

As shown in Fig. MR1, the user categorizes the model "ICCS Wilms Oncosimulator" for perspective V, named "Tumour type(s) addressed". Since the "ICCS Wilms Oncosimulator" is an integrated cancer treatment support system modelling the growth of nephroblastoma tumours, the user checks the box "nephroblastoma". After pushing the button "Categorize the chosen model", all the corresponding information of this categorization is going to be stored both in the Model repository and the CHIC semantics infrastructure. The topology of the CHIC infrastructure that handles the semantic annotation of the categorization of the models based on the 13 perspectives that have been defined within CHIC, is shown in Fig. MR2.
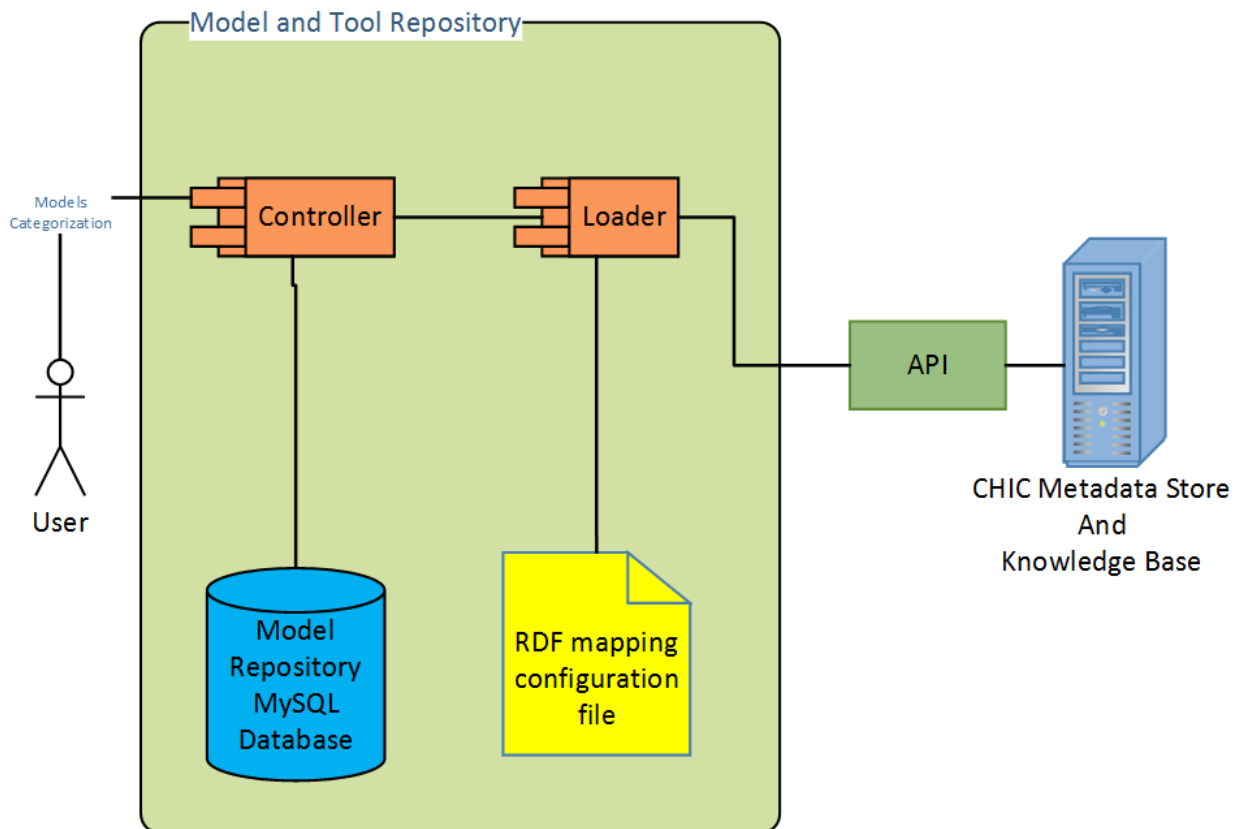
Fig. MR2: Topology of the CHIC infrastructure that handles the semantic annotation of the models

MR2. Modules Used for the Semantic Annotation of Model Categorization

As shown in Fig. MR2 , the following modules are used for the use case of the semantic annotation of the models categorization:

- **Controller:** The controller is the central module of the model repository that consists of many other submodules. It opens the local relational database connection and it handles web requests and presentation details that the user will see. It also calls the Loader module.

- **Loader:** The Loader is in charge of converting MySQL data into RDF property values that will be provided to the CHIC semantics infrastructure web services. It also loads the RDF

mapping configuration file and calls the application programming interfaces of the CHIC metadata store.

- **RDF mapping configuration file:** This file includes the necessary information for mapping MySQL table and columns of the CHIC model repository to RDF properties, vocabularies and OWL ontologies of the CHIC metadata store.

- **API:** This module consists of all the web annotation services that are exposed from the CHIC metadata store and are being used, among others, for the semantic annotation of the models' categorization.

Table MR1presents the result of the semantic annotation of the categorization of the model named "ICCS Wilms Oncosimulator" for perspective V, in the form of subject-predicate-object expressions. The subject denotes the resource, and the predicate denotes traits or aspects of the resource and expresses a relationship between the subject and the object. The RDF statements that are included in Table MR1 represent the following knowledge base:

- The CHIC resource with the URI https://mr.chic-vph.eu/metadata#04e3c5aa-ad45-11e5-bd32-fa163e092aac, represents a CHIC hypomodel.

- The aforementoned CHIC hypomodel has the name "ICCS Wilms Oncosimulator"

- The aforementioned CHIC hypomodel has the unique identifier "04e3c5aa-ad45-11e5-bd32-fa163e092aac"

- The aforementioned CHIC hypomodel addresses the tumour type named "Nephroblastoma". As stated in the fifth row and third column of Table MR1, the "Nephroblastoma" term has the URI "http://purl.obolibrary.org/obo/HP_0002667" which has been derived from the human phenotype ontology.

TABLE MR1: The RDF statements that represent the semantic annotation of the categorization of the model named "ICCS Wilms Oncosimulator" for perspective V

Subject	Predicate	Object
<https://mr.chic-vph.eu/metadata#04e3c5aa-ad45-11e5-bd32-fa163e092aac>	<http://www.chic-vph.eu/ontologies/resource#hasCHICuuid>	"04e3c5aa-ad45-11e5-bd32-fa163e092aac"
<https://mr.chic-vph.eu/metadata#04e3c5aa-ad45-11e5-bd32-fa163e092aac>	<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>	<http://www.chic-vph.eu/ontologies/resource#Model-ChicHypomodel>
<https://mr.chic-vph.eu/metadata#04e3c5aa-ad45-11e5-bd32-fa163e092aac>	<http://www.chic-vph.eu/ontologies/resource#hasName>	"ICCS Wilms Oncosimulator"
<https://mr.chic-vph.eu/metadata#04e3c5aa-ad45-11e5-bd32-fa163e092aac>	<http://www.chic-vph.eu/ontologies/resource#hasPositionIn-5>	<http://purl.obolibrary.org/obo/HP_0002667>

The semantic knowledge base that has been described in Table MR1, has been produced from the Model Repository and has been stored in the CHIC triplestore in order for the other CHIC client components, like the editor or the CRAF (Clinical Research Application Framework), to be able to recognize the model named "ICCS: Wilms Oncosimulator" as a model that simulates the growth of nephroblastoma tumour, and more specifically, according to the human phenotype ontology, the growth of the neoplasm of the kidney that primarily affects children.

The same procedure can be applied for the semantic annotation of the categorization of any new model, through the model repository. Nonetheless, since new CHIC RDF vocabularies and OWL ontologies may be incorporated in the future in order to represent new perspective values, the upgrade and the maintenance of the Model Repository and the CHIC semantics infrastructure is essential in order to ensure the correct mapping between the Model Repository's relational database and the CHIC RDF schema.

CR. The Clinical Research Application Framework (CRAF) in the Context of Utilizing Integrated Hypermodels

The development of the Clinical Research Application Framework (CRAF) has been led by FORTH. USAAR and KUL have provided the clinical user requirements and clinical feedback. ICCS has provided the basic science requirements, coordinated the provision of simulation software and provided feedback.

The standardized cancer hypermodels are available to the clinicians through the CRAF (Clinical Research Application Framework) application. In the following paragraphs we describe in detail the workflow which has to be performed by a clinician in order to configure and execute a standardized cancer hypermodel using CRAF, and finally how to view the resulting report of the execution.

CR1. Description and Versions

The CRAF application comes in two versions, as a native desktop application (desktop CRAF) and as a web application (web CRAF). The reason for having two different versions has to do with the fact that the native application comes as an integrated and single package, complete with other native applications of the project which provide functionalities impossible to provide (yet) in the context of a web application. A use case in order to clarify the necessity of the desktop CRAF is that it is the execution base for the CCGVis application which provides real-time 3D visualization of the simulation results, along with the functionality to define 2D snapshots from the actual 3D model in order to include them in the reports and also the generation of a video which clearly shows the evolution of the simulation outcome in time.

The desktop version of the CRAF application, is created using open source components and it is based on the technologies of Java SE 8 [CR1] and JavaFX [CR2]. The adoption of Java renders CRAF independent of the operating system of the execution system and it has been tested both on Microsoft Windows (versions 8+ are supported) and on Apple OSX computers (version El Capitan).

The web CRAF application is a "Single Page Application" (SPA) [CR3] web application based on state of the art HTML5 technologies. The front-end of the application is based on the AngularJS [CR4] web framework, which is a full client-side framework that has the advantage that all the rendering process happens to the client and not the server. AngularJS's templating system is based on bidirectional UI data binding. Data-binding is an automatic way of updating the view whenever the model changes, as well as updating the model whenever the view changes. The HTML template is compiled in the browser. The compilation step creates pure HTML, which the browser re-renders into the live view. The step is repeated for subsequent page views. In traditional server-side HTML programming, concepts such as controller and model interact within a server process to produce new HTML views. In the AngularJS framework, the controller and model states are maintained within the client browser. Therefore, new pages are capable of being generated without any interaction with a server.

CRAF has a fully responsive layout which adapts to the screen size of any device, and it is suitable for usage from a personal computer, a tablet device or a mobile phone. CRAF is designed using the Material Design [CR5] principles and controls. The layout is flat and it presents only the necessary information to the end user.

CR2. Web CRAF Clinical Workflow

CR2.1 Login and the Options of the Main Screen

In order to access the CRAF the clinician has to provide the proper credentials at the login form (Fig. CR1) and click the login button.
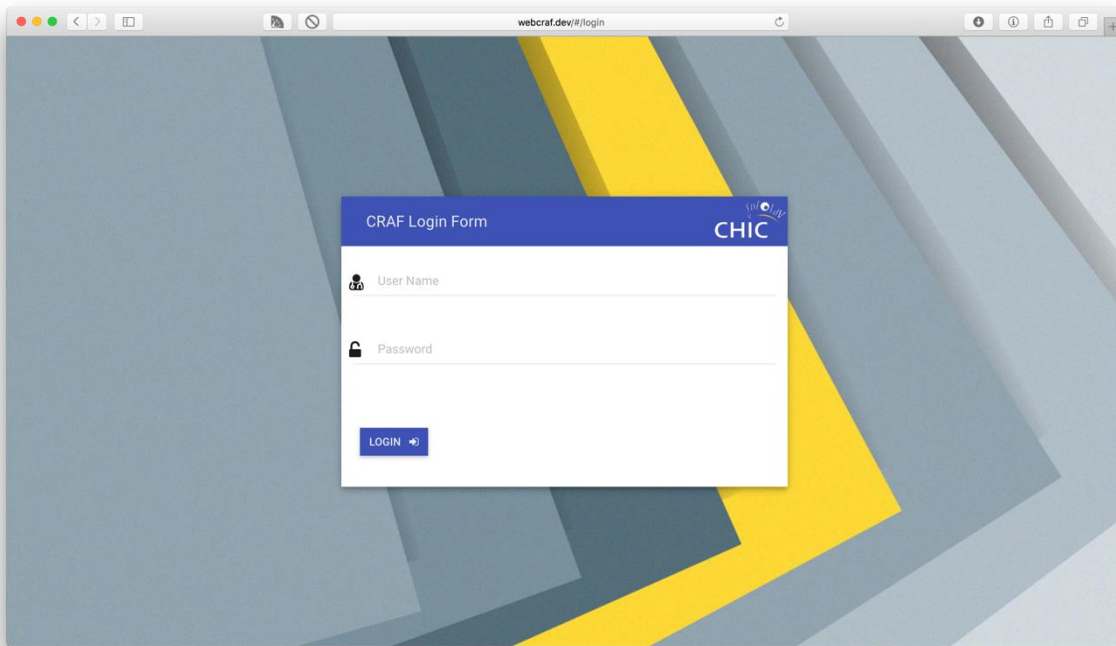

Fig. CR1. CRAF Login Form

if the credentials provided by the user are correct then CRAF successfully verifies the user and provides him permission to use the full functionalities of the CRAF application and access the data which are available specifically to him (data are filtered based on the user's permissions).

Once the user is logged in, CRAF displays the main screen (home), which provides two main options. The clinician is asked to choose whether to define a "Domain specific" execution or a "Patient specific" execution (Fig. CR2). Both selections will result to the configuration and the complete definition of a hypermodel to be executed for a selected patient using the patient's data and the clinician's requests (e.g. duration of the simulation) resulting to a personalized outcome for the selected patient. The main difference among these two options, is that they are actually filters which initiate the same wizard like interface for configuring a hypermodel but using different order for the wizard's steps.
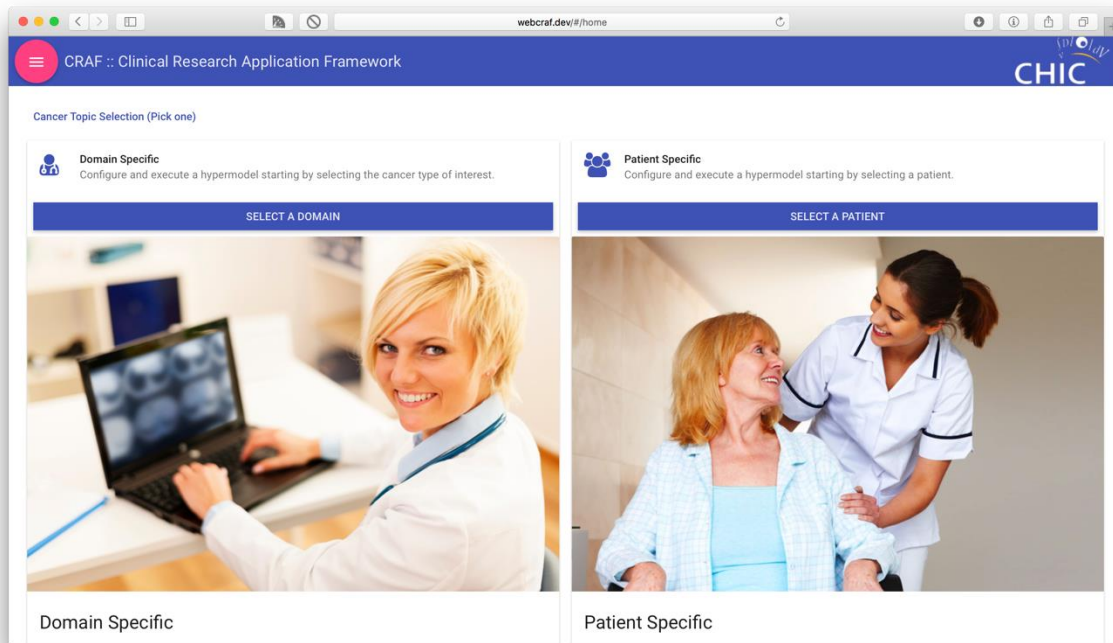
Fig. CR2. Main CRAF window

CR2.2 Configuring a Hypermodel Starting with the Patient Selection
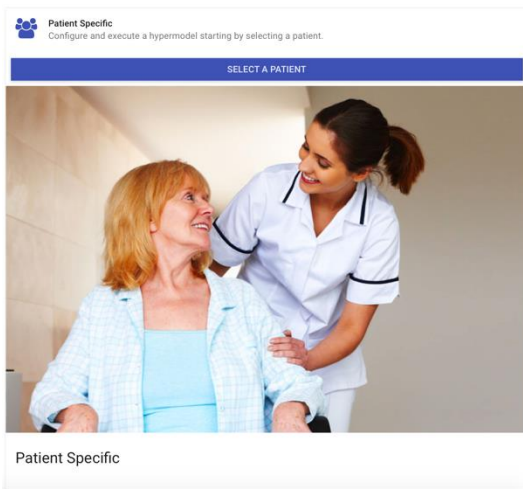
Fig. CR3 Selecting the patient specific option

Once the clinician selects the Patient Specific option (Fig.CR3) from the main screen, then a wizard like procedure begins which has as a first step the selection of a patient from the patient list (Fig. CR4).

As it is shown in Fig. CR4 CRAF displays the list of patients as a list of radio buttons. The patient list is interactive, and as the user picks another patient from the list, the patient's card at the right side is updated, in order to display the patient's corresponding data. When the desired patient is selected, the user proceeds to the next step (Fig. CR5), where he/she has to select the question of interest for the patient of interest. The questions displayed in this step are relative to the cancer type of the selected patient.
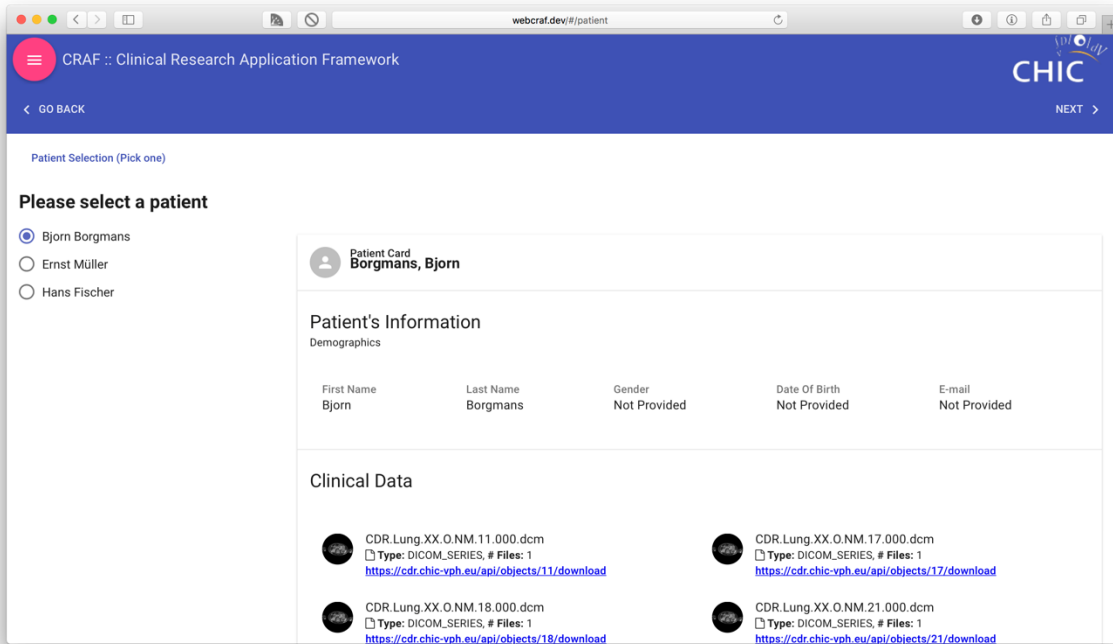
Fig. CR4 Patient selection
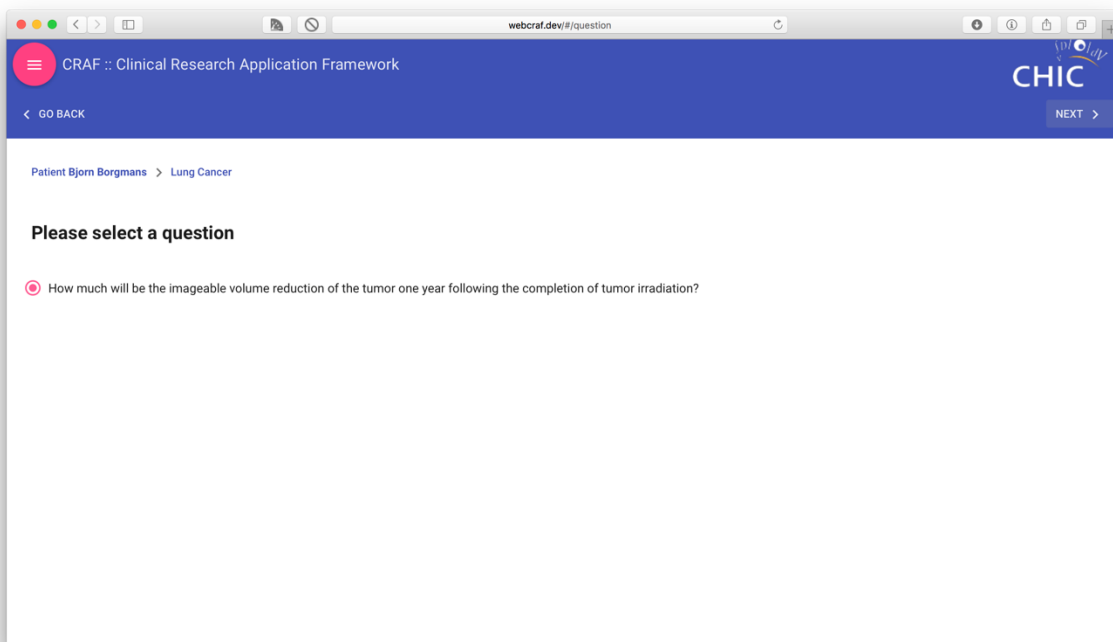
Fig. CR5 Selection of question for the chosen patient

When the patient and the question have been selected, then at the next step (Fig. CR6) CRAF displays all the available hypermodels that can be selected. At this case there is only one hypermodel available, and it is already selected. At this point there is also available the choice to create a new hypermodel at the "Hypermodel editor". By proceeding with the selected hypermodel, at the next step the clinician overviews the input/output parameters for the hypermodel (Fig. CR7), and if

necessary can make any adjustments directly from the CRAF application. When everything is set, the user can proceed to the execution of the model by clicking the "Execute" button. A window appears which notifies the user about the configuration status of the hypermodel (complete or not) and about the status of its execution (successful initialization of the execution of the hypermodel or failure to start). After the user closes the dialog window, CRAF returns to the main screen, ready to configure and execute a new hypermodel.
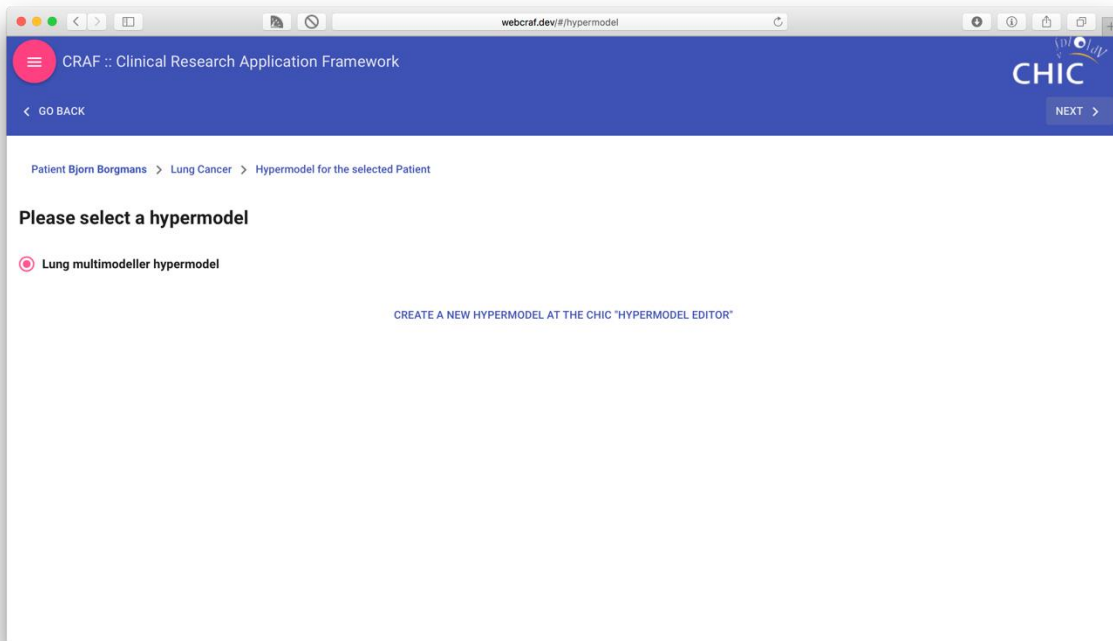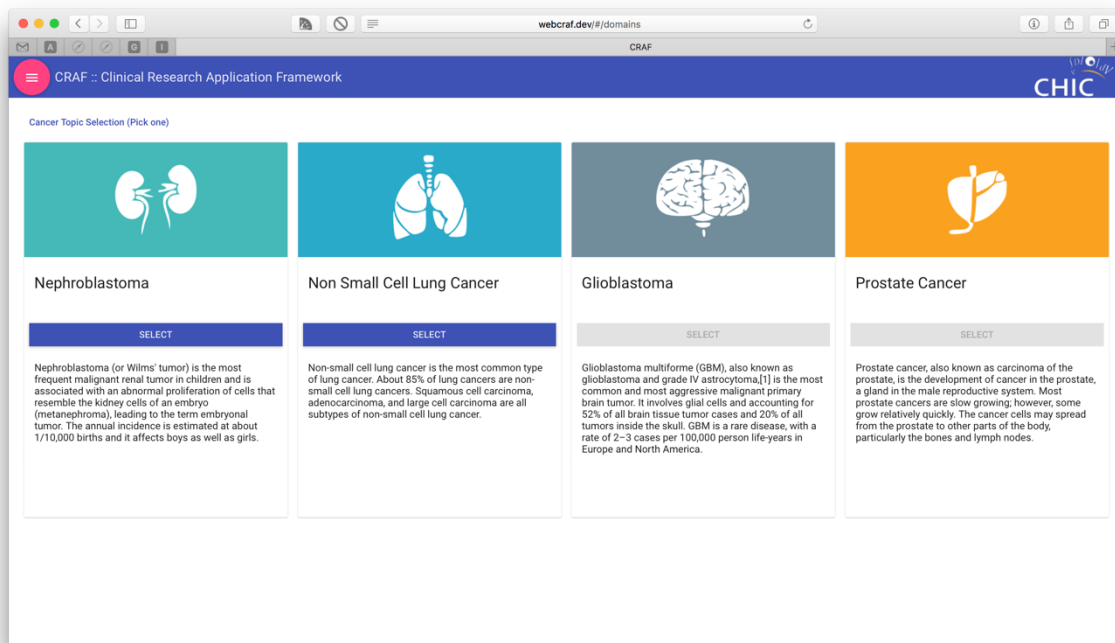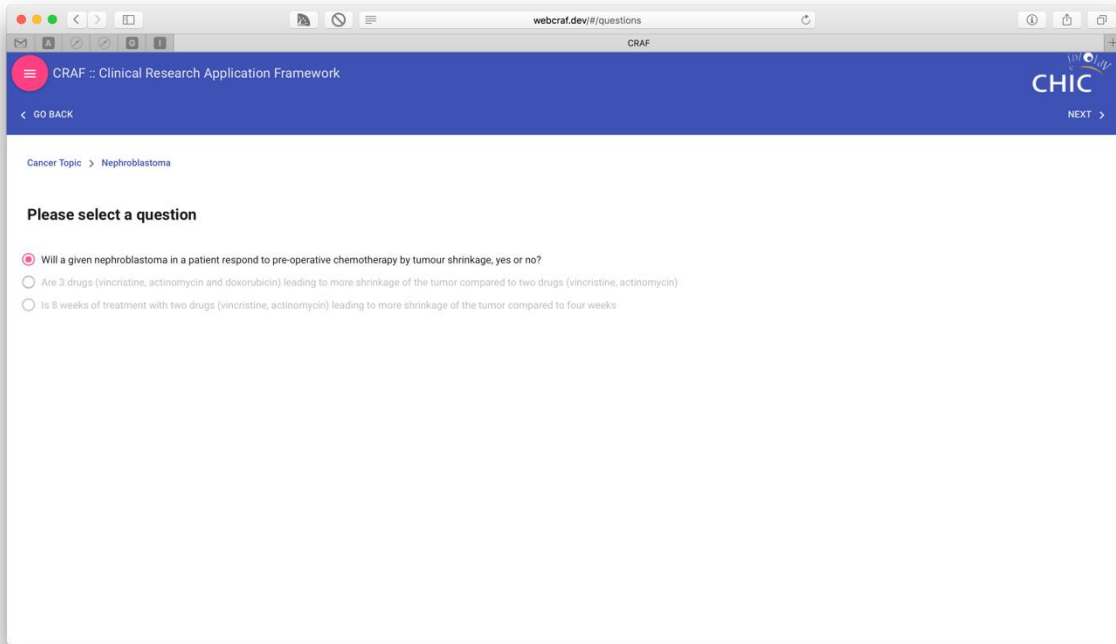

Fig. CR6 Selection of the available hypermodels for the chosen patient and the chosen question
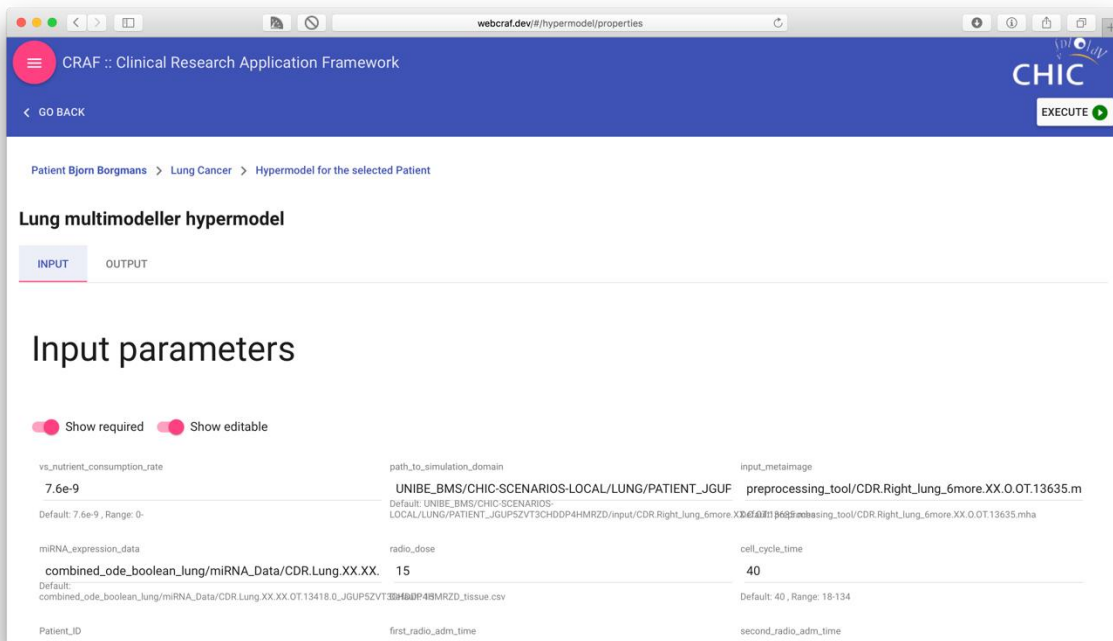
Fig. CR7 Overview of the input/output parameters before the execution of the chosen hypermodel for the selected patient and the chosen question

CR2.3 Configuring a Hypermodel Starting with the Cancer Domain
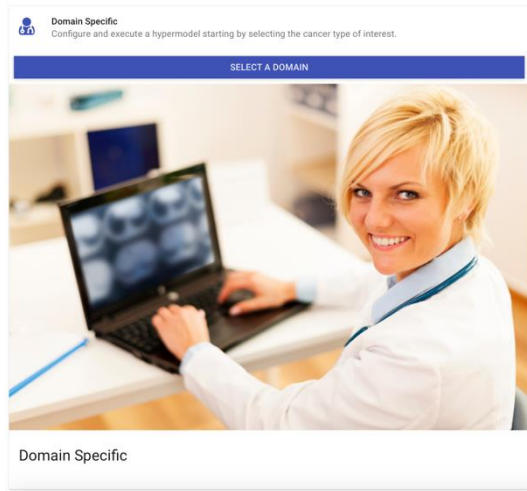
Fig. CR8.

If the clinician selects the Domain Specific option (Fig. CR8) from the main screen, then the wizard starts as before but this time has as the first step the selection of the cancer domain (Fig. CR9). As the user selects the cancer domain of interest, the corresponding questions appear (Fig. CR10) and by picking one of them, CRAF shows the list of patients which fulfil the selected criteria (Fig. CR11) (patients that do not fit, can be hidden using the corresponding switch, in order not to distract the clinician with the excess of information). One step before the wizard completes, the user has to select among the available hypermodels (Fig. CR12). At the last step, once the hypermodel had been selected, the user has an overview of the input/output arguments of the hypermodel (Fig. CR13), and if necessary can make corrections/changes before starting the execution of this hypermodel. As above, when everything is set, the user can proceed to the execution of the model by clicking the "Execute" button. A window appears which notifies the user about the configuration status of the hypermodel (complete or not) and about the status of its execution (successful initialization of the execution of the hypermodel or failure to start). After the user close the dialog window, CRAF returns to the main screen, ready to configure and execute a new hypermodel.

Fig. CR9 Available cancer domains

Fig. CR10. Selection of question for the chosen cancer domain
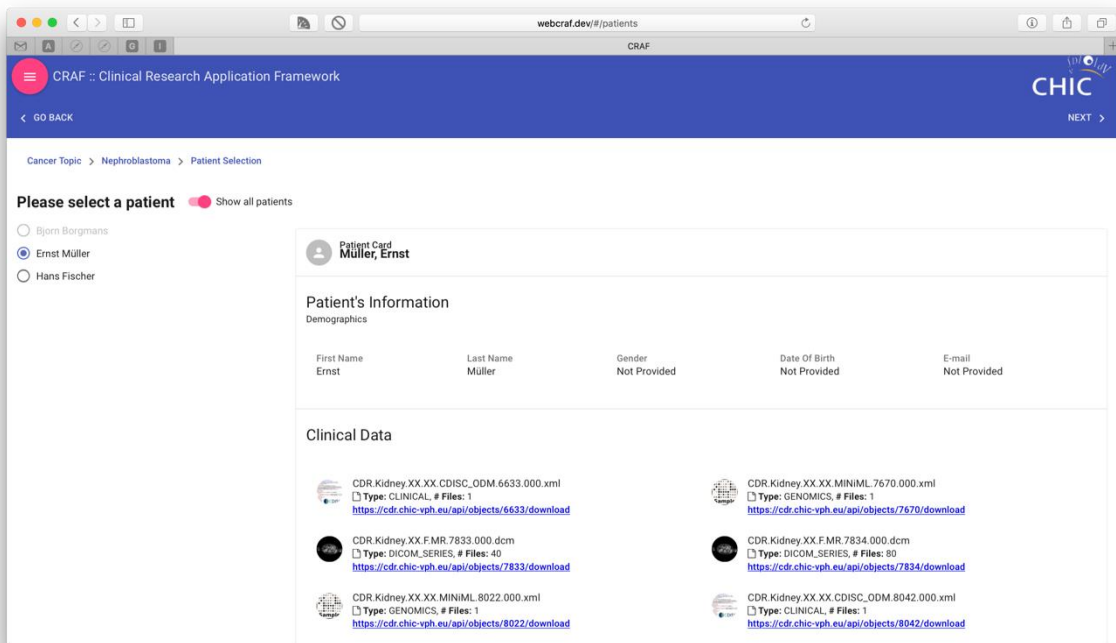
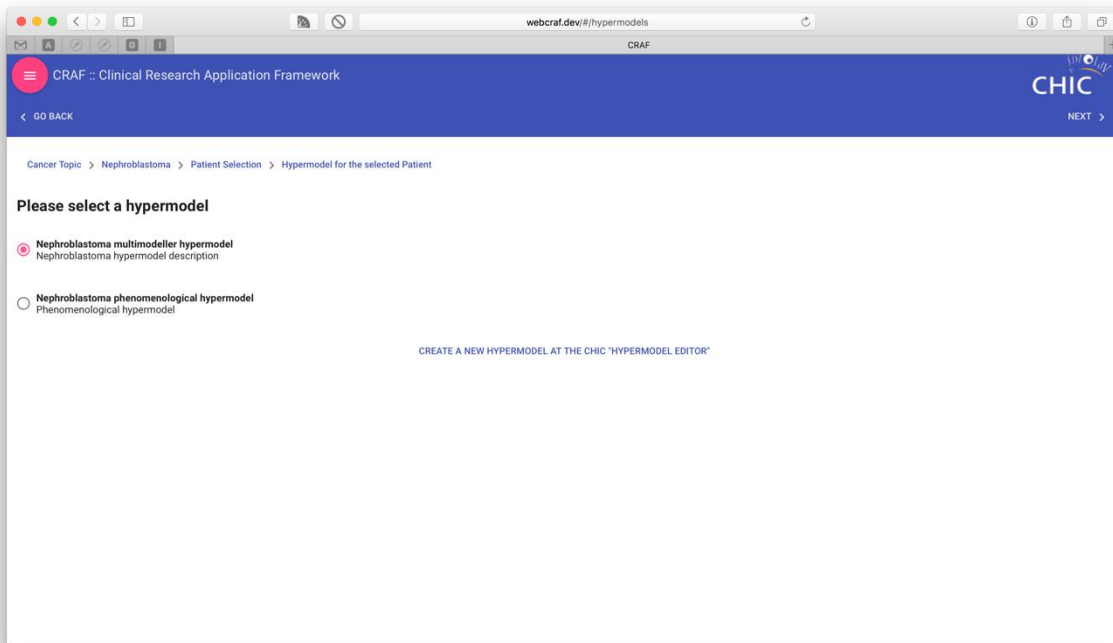Fig. CR11 Selection of the patient for the chosen question and cancer domain

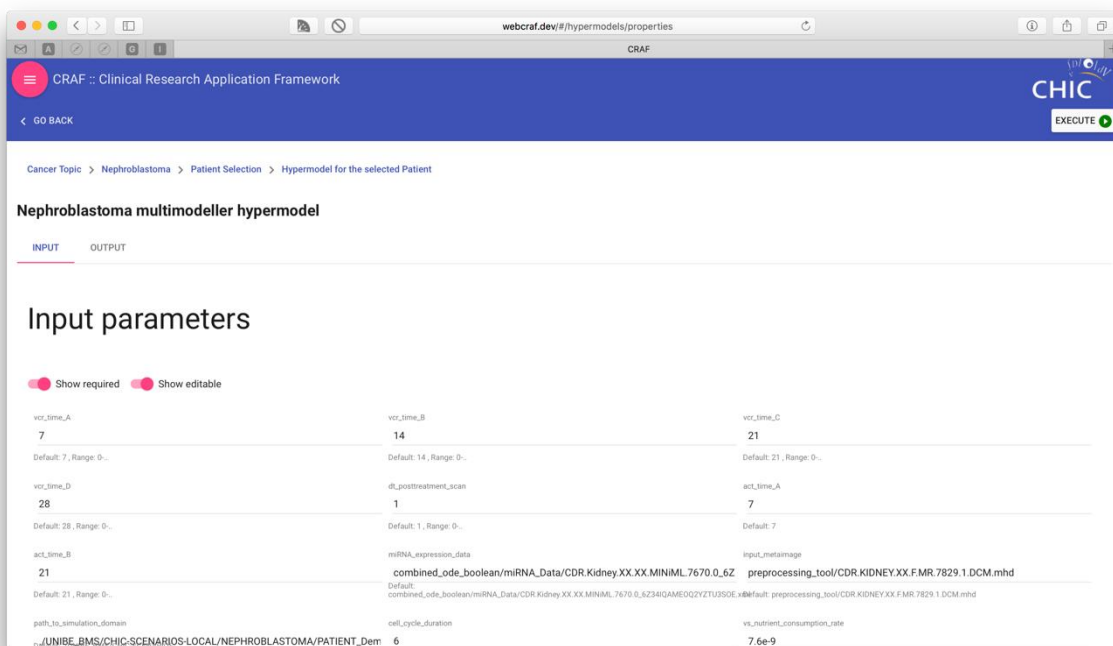Fig. CR12. Selection of the hypermodel for the chosen patient, question and cancer domain

Fig. CR13. Overview of the input/output parameters before the execution of the chosen hypermodel for the selected patient and the chosen question

CR2.4 List of Execution Results

In order for the user to view the list of the executions of the hypermodels, he/she has to access the side menu and select the History option (Fig. CR14). The page of the list of the previous executions of the logged in user appears (Fig. CR15), where the user can create/view a report for the selected

execution or all of its outputs. Upon clicking the Outputs button, a dialog window appears which displays 3 tabs:

1. **Details**

   In the Details tab (Fig. CR16), the user can see detailed information regarding the selected row (selected execution) form the history list. There is information regarding the full name of the patient (all the names in this document are pseudo anonymized and are not the real names of the patients), when the execution started and finished and the time needed in order to complete, and the corresponding cancer domain.

2. **Reports**

   In the Reports tab (Fig. CR17), there is the full list of the reports that have been generated for the selected execution. Each report is in a pdf format and can be directly opened through the browser or downloaded locally to the clinician's PC.

3. **Outputs**

   In the Outputs tab (Fig. CR18), there is the list of the output files of the hypermodel. The user can view directly on the browser all of the simple type parameters (such as numerics, strings etc.) or download any output files that cannot be displayed in the browser (such as dat files, other binary files etc.). There is also the option to download in a single zip file all the outputs, by clicking the "Download" button.
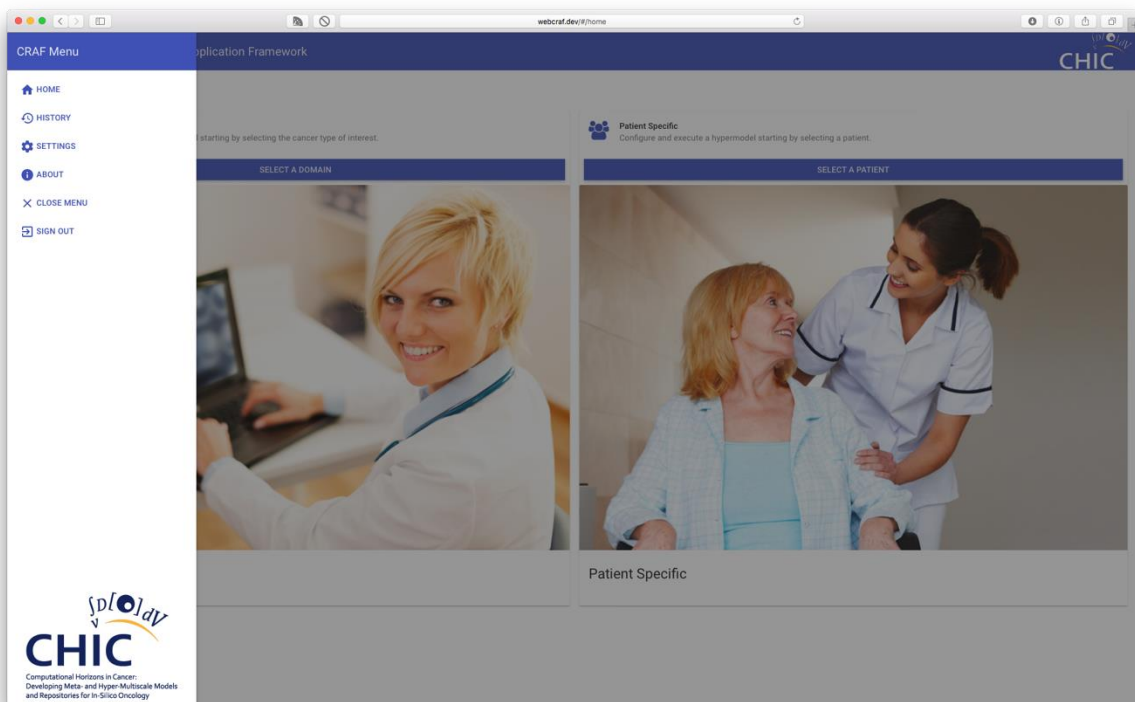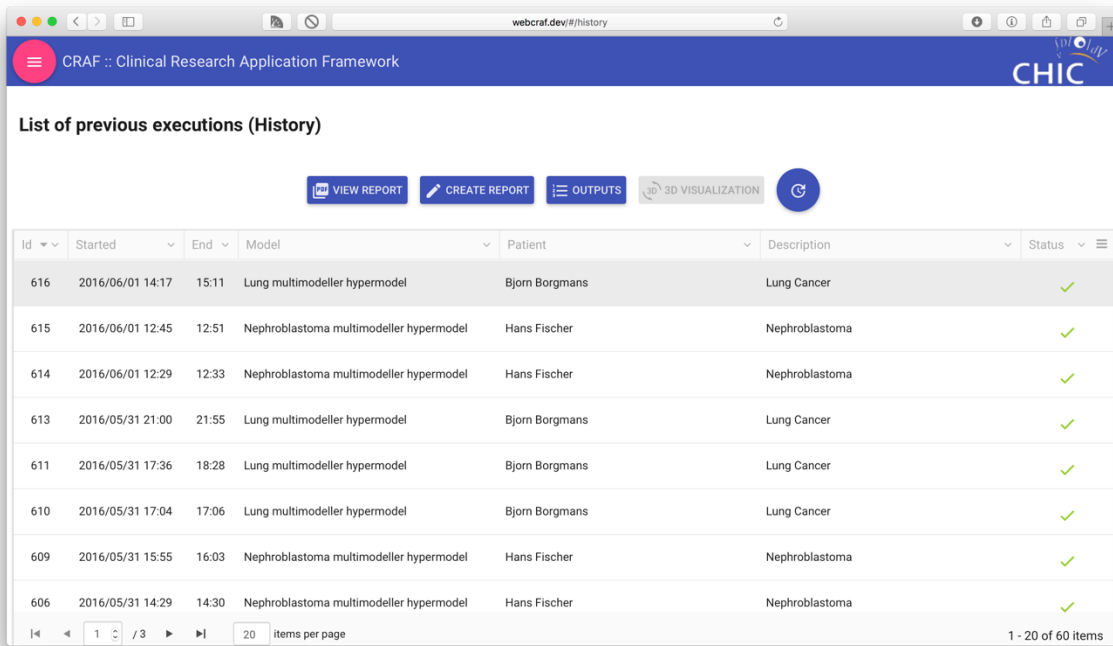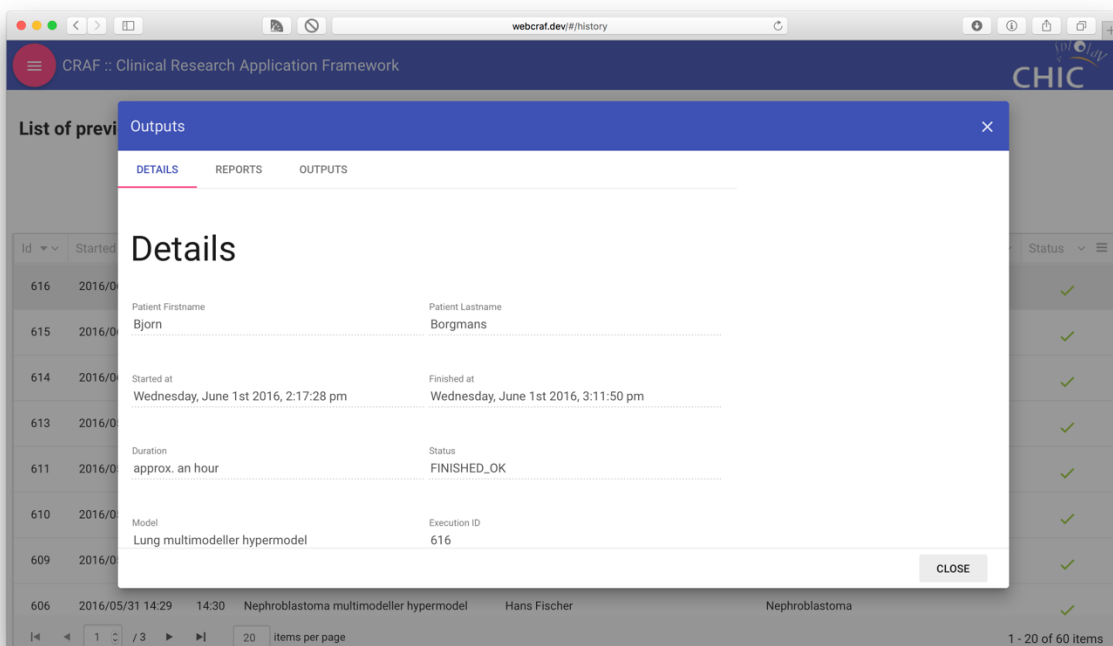

Fig. CR14. Main menu

Fig, CR15. List of previous execution (History)

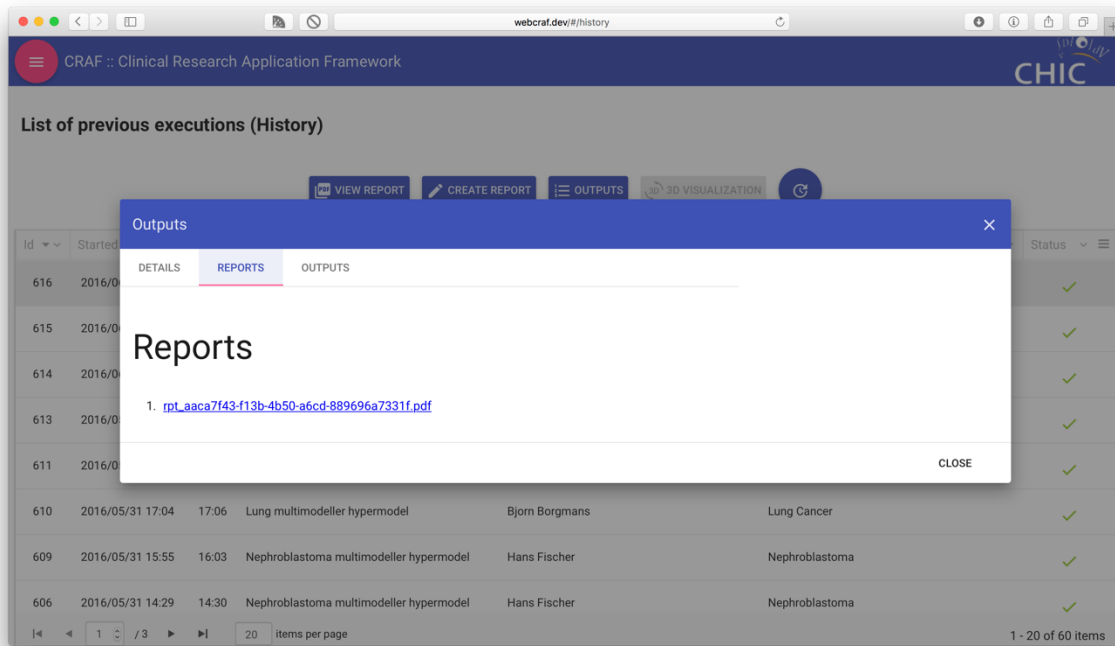Fig. CR16. Detailed overview of a chosen execution from the history list

Fig. CR17. The list of the available reports for the chosen execution from the history list
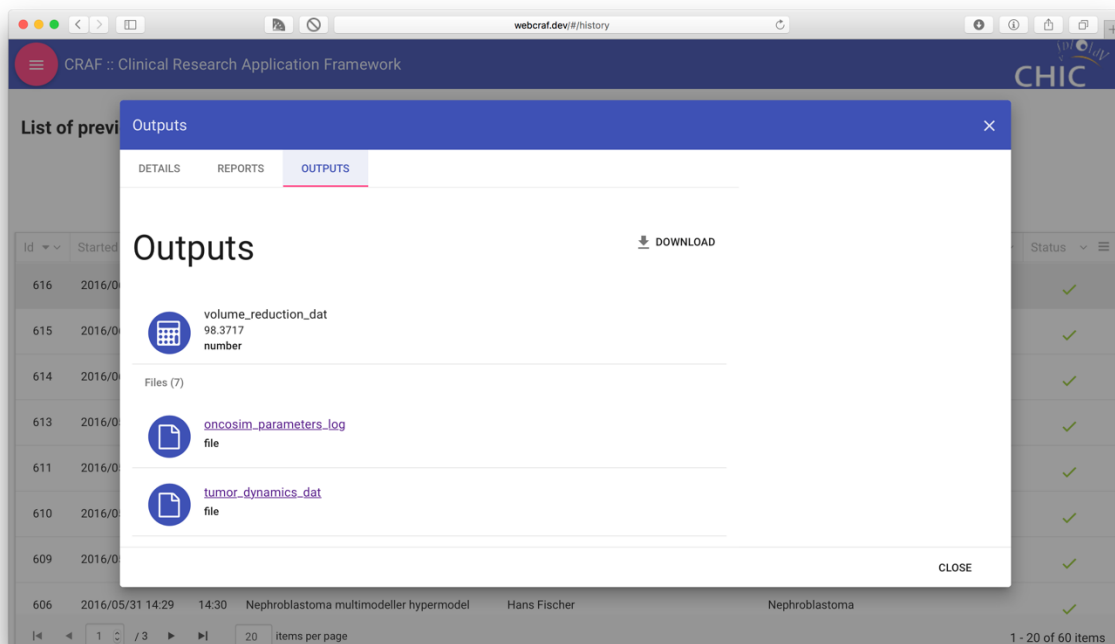
Fig. CR18. Available outputs of the chosen execution from the history list. All the outputs can be downloaded as a zipped file.

CR2.5 Desktop CRAF Clinical Workflow

The desktop version of the CRAF application, offers most of the functionalities of the web version described above, but it is also a lot more enhanced as it is able to communicate with other native applications such as CCGVis. The applications are seamlessly integrated in a single package. As with

the web version of the CRAF application, the user has to login first in order to be able to access the data associated with the user and the functionalities of the tool(s) (Fig.. CR19). The main screen of the desktop CRAF (Fig. CR20) resembles in functionality and in layout to the main screen of the web CRAF (Fig. CR9). The similarities among the two flavours of the CRAF application (desktop app and web app) extend also to the screens that display the list of the previous executions (history list) (Fig. CR21 and Fig. CR15). However, in the desktop CRAF application, once the clinician has selected a completed execution from the list of the executions, the available options (Fig. CR22) include a "3D visualization" button which opens the CCGVis application, automatically initialized with the parameters and data from the selected execution.
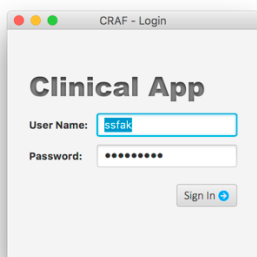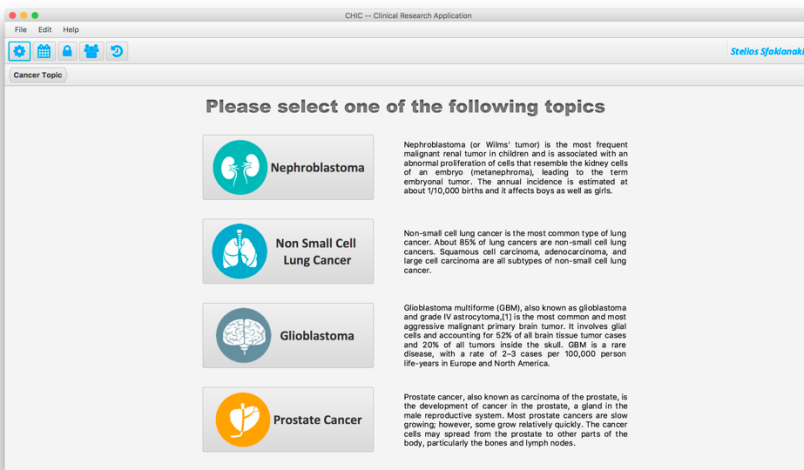
Fig. CR19. Desktop CRAF login

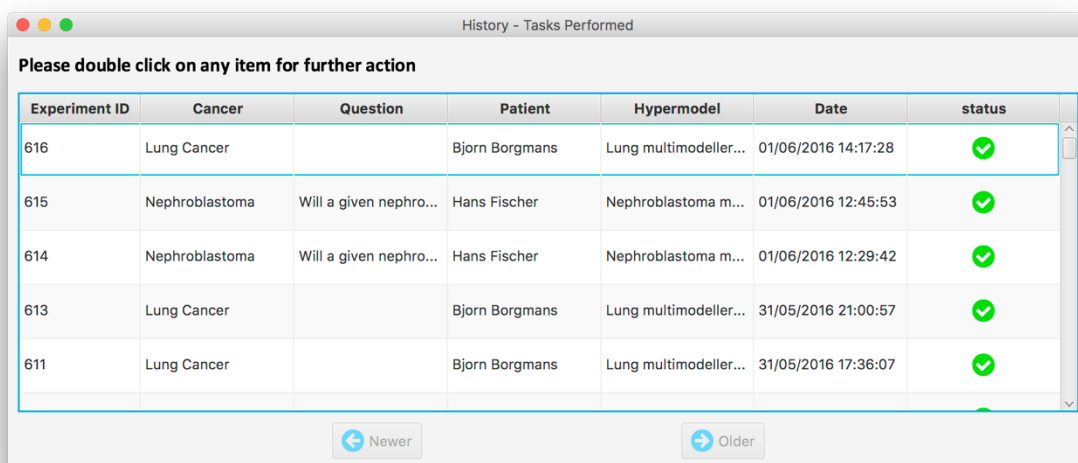Fiig. CR20. Desktop CRAF, selection of cancer domain

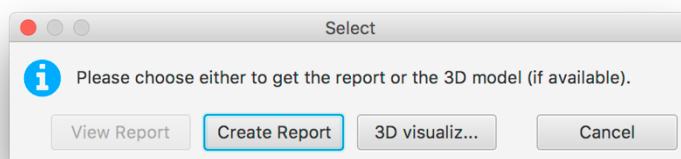Fig. CR21. Desktop CRAF: List of executions

Fig. CR22. Desktop CRAF available options for a completed execution

CR2.6 CCGVis Integration and 3D Visualization

The CCGVis application provides visualization of the imaging data of the hypermodel (both input and output) both in 2D as well as in 3D. It allows visualization in sagittal, transverse or coronal projections, through a user friendly interface. CCGVis is a C++ application based on the open source VTK system [CR6] for creation and manipulation of the 3D objects and on the Qt cross-platform application framework [CR7]. The CCGVis app communicates with the CRAF app through in a two-way data exchange among the two applications. The CRAF application provides to the CCGVis application all the necessary data and parameters (such as imaging data, output files of the executed hypermodels, etc.) and the CCGVis returns to the CRAF application all the screenshots of the output images which the clinician has defined, the graphs and the figure captions to be included in the dynamic reports and all the other available metadata (Fig. CR23)).
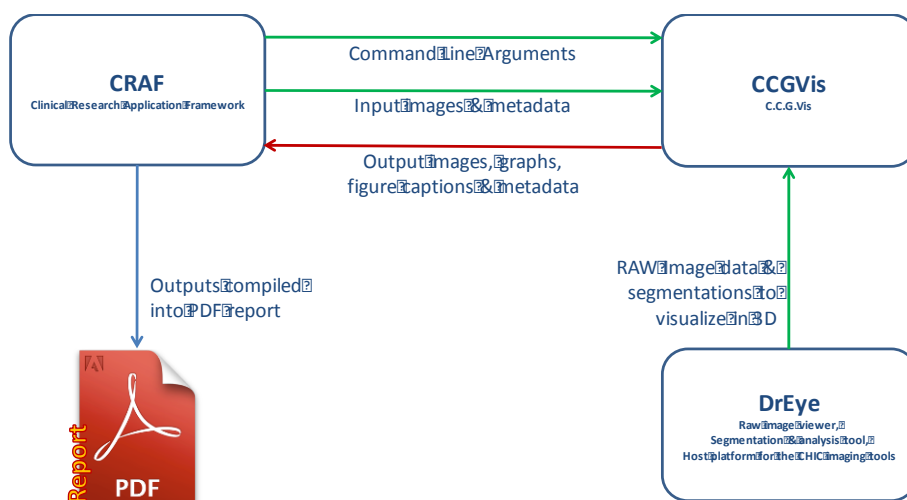
Fig. CR23. Schematic of interaction among CRAF and CCGVis (and also CCGVis and DrEye)

The CCGvis functionalities, apart of the 2D / 3D visualization also include the automatic generation of plots based on the outputs of the hypermodels, a side to side comparison view among the segmented tumour volume and the simulated tumour volume which provides an immediate qualitative overview of the tumour's progression, a save button which sends the current image/snapshot or plot directly to the output in order to be included in the report. The application is continuously evolving (in a rapid pace) to include new features.
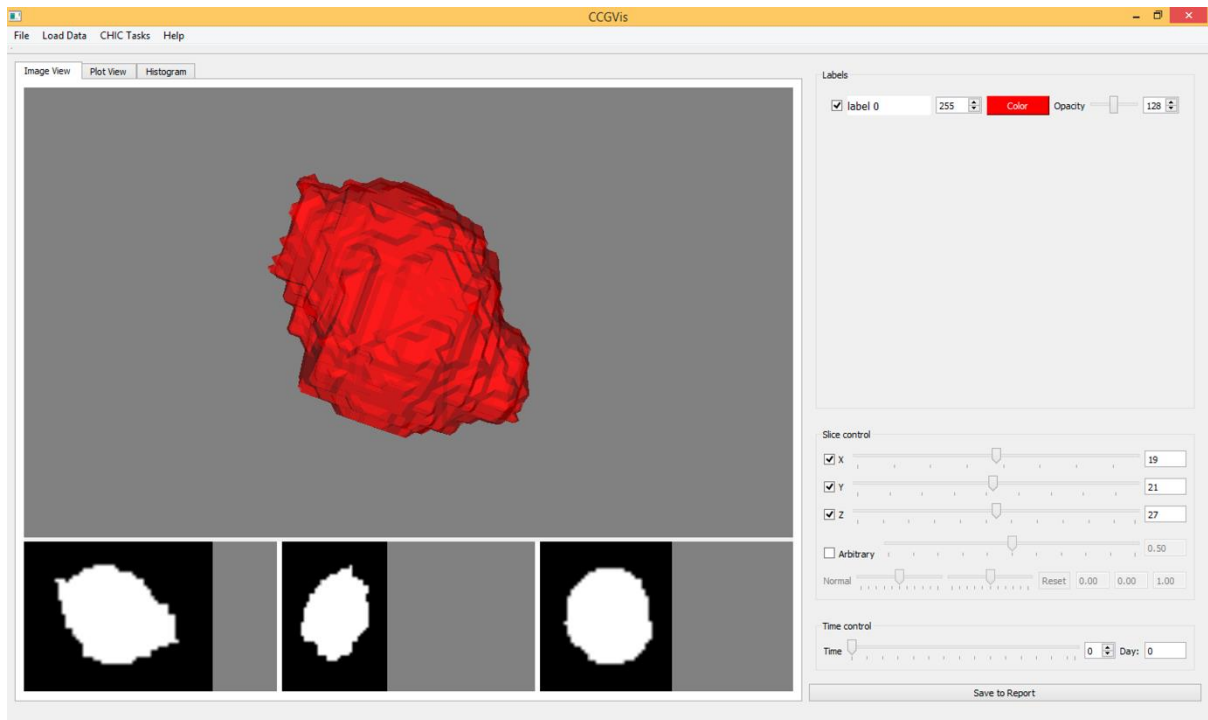
Fig. CR24. 3D visualization of predicted tumour
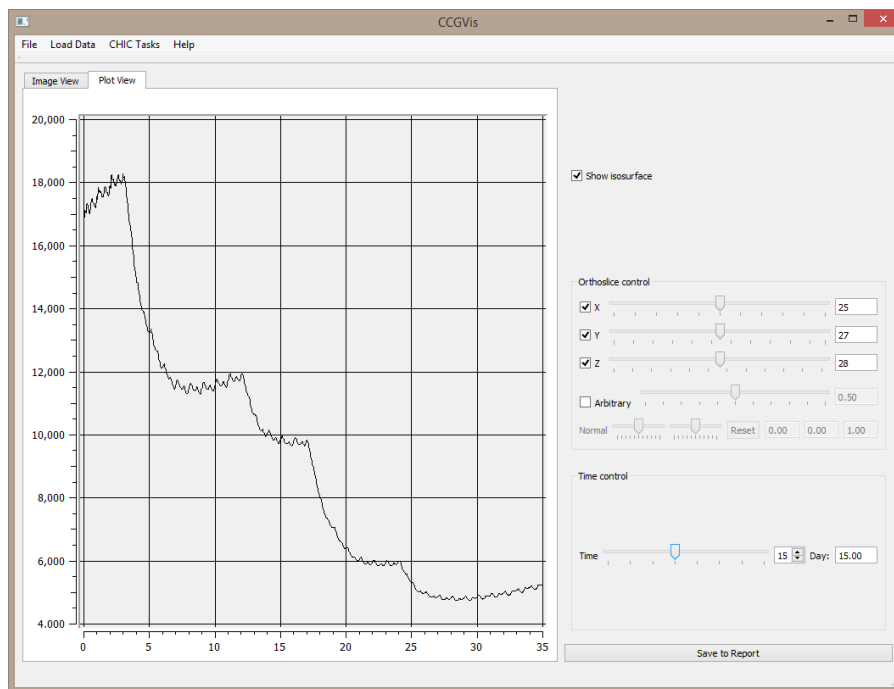
Fig. CR25. Graph of tumour volume vs time
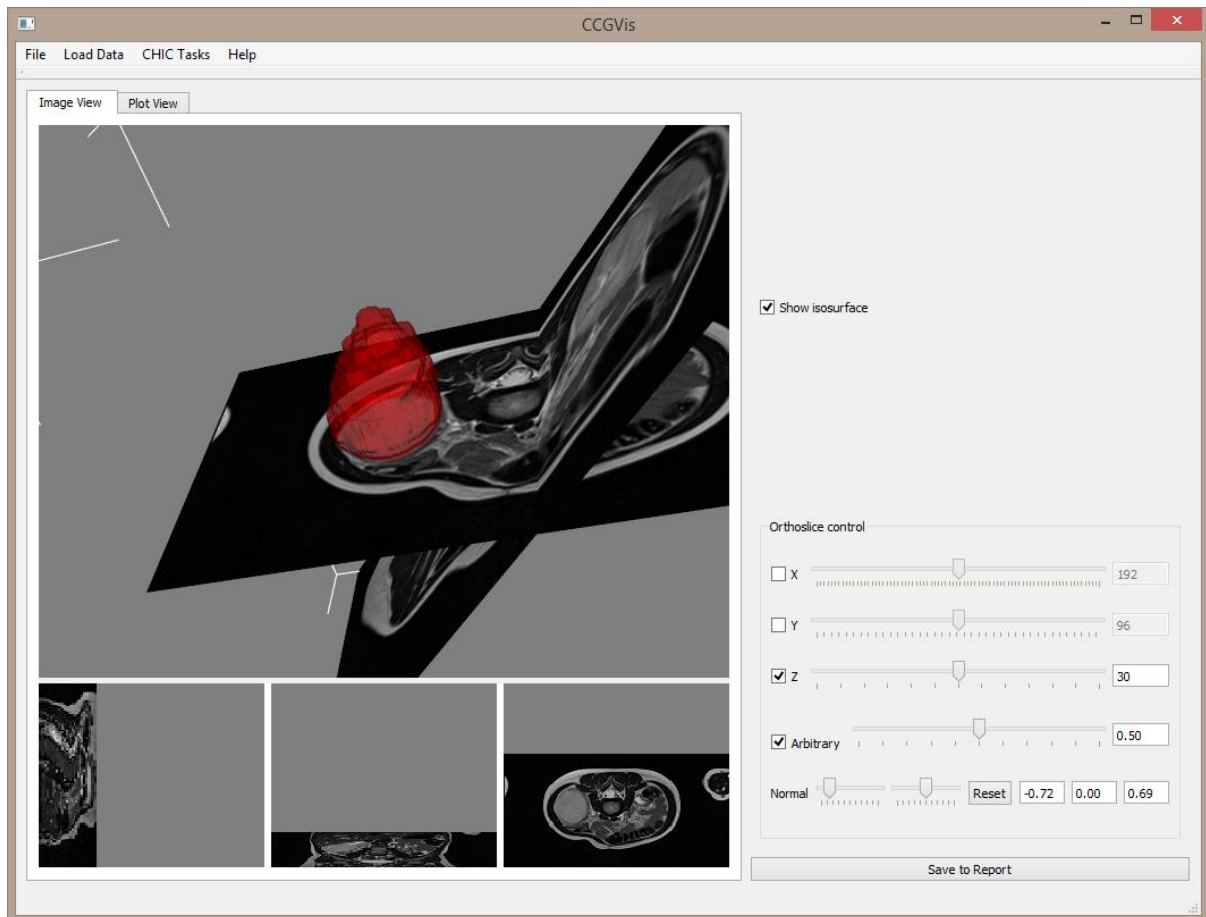
Fig. CR26. 3D slice view of nephroblastoma, showing segmented tumour volume

CR2.7 Custom Reports

Based on the selected patient, question, hypermodel and all the other configuration options every execution has different outcome and results. The CRAF application automatically generates pdf reports based on the outputs of the hypermodel's execution.

The images below are a representative example of the report for the output of the execution of a lung cancer hypermodel.

In silico study of cancer response to treatment

Patient Name: Borgmans, Bjorn
Date of birth: [Not available]
Started at: 2016-06-01 11:17   Finished at: 2016-06-01 12:11

Clinical question:
How much will be the imageable volume reduction of the tumor one year following the completion of tumor irradiation?
Simulated Tumour Volume reduction percentage: **98,37 %**

Simulation predictions

Treatment onset (in respect to simulation time 0) (day): **104**
Reference time point (in respect to simulation time 0) (day): **480**

Time evolution of tumor volume

The CHIC project   **page: 1 / 5**
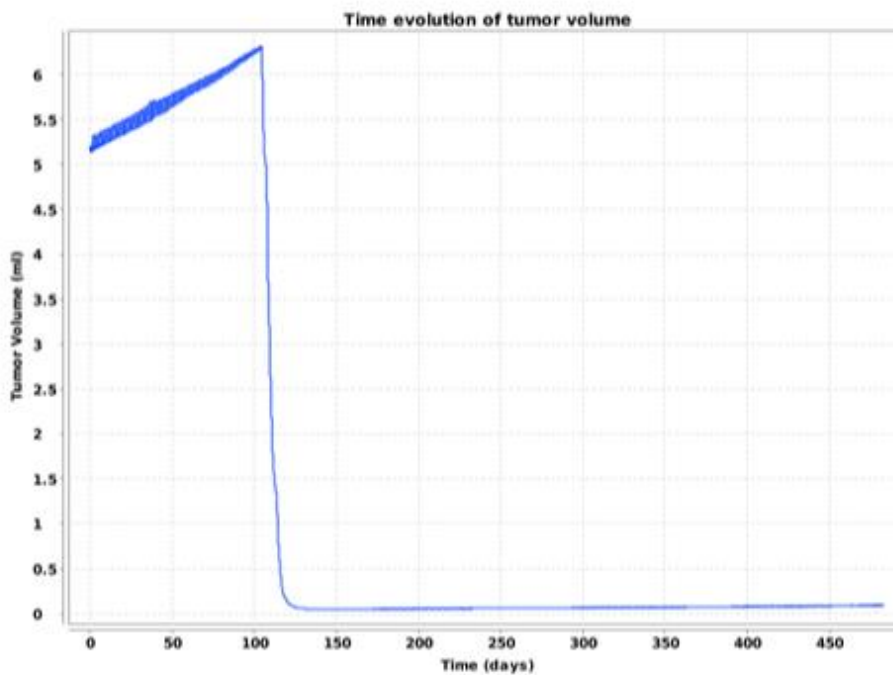
Hypermodel Description

The CHIC project

page: 2 / 5

Input values of hypermodel parameters

Default value 0.396 mm^2/hr from A HYBRID MODEL FOR TUMOR SPHEROID GROWTH IN VITRO I: THEORETICAL DEVELOPMENT AND EARLY RESULTS, Kim et al. 2007.: 0.396

Rate of glucose consumption per cell per hour.: 7.6e-9

Glucose concentration in non tumour regions. Output is normalized by this value.: 0.9

Tumour cell concentration used for pressure computation.: 1000000

Fraction of stem cells that divide symmetrically in necrotic regions: 0.322

Vtk-readable segmentation file with organ labels.   Labels for Lung scenario 1: right lung 2: left lung 3: bone 4: soft tissue outside 5: other internal organs 6: bronchi 255: tumour : UNIBE_BMS/ CHIC-SCENARIOS-LOCAL/ LUNG/ PATIENT_JGUP5ZVT3CHDDP4HMRZD/ input/ CDR.Right_lung_6more.XX.O.OT.13635.mha

Number of limited mitotic potential (LIMP) cell stages before differentiation occurs = number of LIMP cell mitoses before differentiation occur: 22

Defines the threshold for the cropping (float between 0 and 1): 0.5

Tumor cell density in number of biological cells per 1mm^3.: 1000000

The spacing to be used for the isotropic interpolation.: 1

: 1

A valid metaimage which contains the segmentations with the tumor value to be 255: preprocessing_tool/ CDR.Right_lung_6more.XX.O.OT.13635.mha

: 1

Name of output file which will contain the interpollated and cropped result: output

Cell kill rate of cisplatin: 0

Csv or minml file which contains the miRNA expression data.: combined_ode_boolean_lung/ miRNA_Data/ CDR.Lung.XX.XX.OT.13418.0_JGUP5ZVT3CHDDP4HMRZD_tissue.csv

Cell kill rate of vinorelbine: 0
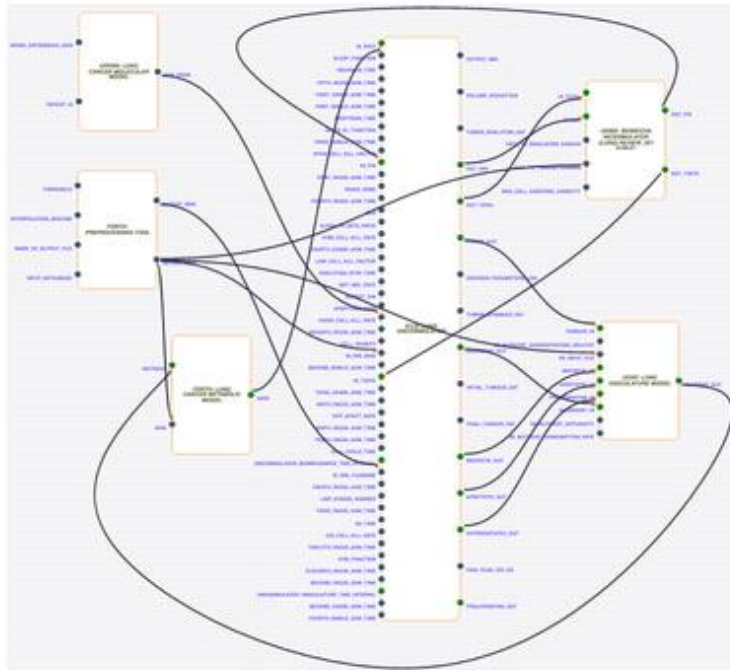
alpha/beta ratio, where alpha: radiosensitivity parameter of the Linear Quadratic model in Gy^-1 beta: radiosensitivity parameter of the Linear Quadratic model in Gy^-2: 10

Enhancement of therapeutic or detrimental effect of ionizing radiation due to the presence of oxygen: 3

The CHIC project   **page: 3 / 5**

Dose D (Gy) of radiation to a population of cells: 15

Cell cycle duration of stem and LIMP cells (G0 phase not included): 40

Dormant (G0) phase duration of stem and LIMP cells : 168

Time before necrosis products are eliminated: 23

Time before apoptosis products are eliminated in necrotic regions: 4

Spontaneous apoptosis rate corresponding to transition to apoptosis from any of the G1, S, G2, M, G0 phases of the stem and LIMP cells : 0.0001

Spontaneous apoptosis rate corresponding to transition to apoptosis from the differentiated cell state: 0.017

Rate to enter necrosis for differentiated cells : 0.025

Fraction of dormant cancer stem cells re-entering the G1 phase after a time interval equal to the G0 duration in necrotic regions: 0.1

Fraction of cells that will enter G0 following mitosis in necrotic regions: 0.263

Anonymized ids of patients selected for the trial. In the model demo run precomputed results are available for only patient JGUP5ZVT3CHDDP4HMRZD: JGUP5ZVT3CHDDP4HMRZD

Time point after initialization (in days) when the 1st combination chemotherapy takes place: -1

Time point after initialization (in days) when the 2nd combination chemotherapy takes place (= -1 if total cycles less than 2): -1

Time point after initialization (in days) when the 3rd combination chemotherapy takes place (= -1 if total cycles less than 3): -1

Time point after initialization (in days) when the 4th combination chemotherapy takes place (= -1 if total cycles less than 4): -1

Time point after initialization (in days) when the 1st single chemotherapy takes place: -1

Time point after initialization (in days) when the 2nd single chemotherapy takes place (= -1 if total cycles less than 2): -1

Time point after initialization (in days) when the 3rd single chemotherapy takes place (= -1 if total cycles less than 3): -1

Time point after initialization (in days) when the 4th single chemotherapy takes place (= -1 if total cycles less than 4): -1

Time point after initialization (in days) when the 1st irradiation takes place: 104

The CHIC project   **page: 4 / 5**

Time point after initialization (in days) when the 2nd irradiation takes place (= the last administration time point if total irradiations less than 2): 107

Time point after initialization (in days) when the 3rd irradiation takes place (= the last administration time point if total irradiations less than 3): 110

Time point after initialization (in days) when the 4th irradiation takes place (= the last administration time point if total irradiations less than 4): 113

Time point after initialization (in days) when the 5th irradiation takes place(= the last administration time point if total irradiations less than 5): -1

Time point after initialization (in days) when the 6th irradiation takes place (= the last administration time point if total irradiations less than 6): -1

Time point after initialization (in days) when the 7th irradiation takes place (= the last administration time point if total irradiations less than 7): -1

Time point after initialization (in days) when the 8th irradiation takes place (= the last administration time point if total irradiations less than 8): -1

Time point after initialization (in days) when the 9th irradiation takes place(= the last administration time point if total irradiations less than 9): -1

Time point after initialization (in days) when the 10th irradiation takes place(= the last administration time point if total irradiations less than 10): -1

Time point after initialization (in days) when the 11th irradiation takes place(= the last administration time point if total irradiations less than 11): -1

Time point after initialization (in days) when the 12th irradiation takes place(= the last administration time point if total irradiations less than 12): -1

Execution stop time after initialization: 480

Output directory: ./ ICCS_LungOncosimulator/ output

The CHIC project   **page: 5 / 5**

CR3. References

[CR1] https://java.com/en/about/

[CR2] http://docs.oracle.com/javase/8/javase-clienttechnologies.htm

[CR3] SPA: Single Page Application. The application fits on a single web page with the goal of providing a more fluid user experience similar to a desktop application. In a SPA, either all necessary code is retrieved with a single page load, or the appropriate resources are dynamically loaded and added to the page as necessary, usually in response to user actions. The page does not reload at any point in the process, nor does control transfer to another page. Interaction with the single page application involves dynamic communication with the web server behind the scenes.

[CR4] AngularJS: Open source web app framework by Google https://www.angularjs.org

[CR5] The visual language of the Material Design, https://material.google.com

[CR6] Visualization ToolKit (VTK), http://www.vtk.org/

[CR7] Qt application framework, https://www.qt.io

HE. The Hypermodelling Editor (HME) as a Supportive Technological Platform for the Integration of Hypermodels

The development of the Hypermodelling Editor has been led by FORTH. ICCS has provided the major basic science requirements for the linking of hypomodels for the creation of hypermodels.

The Hypermodelling Editor is a graphical (web based) tool for the basic scientists and the modellers that greatly helps in the the *design* of new hyper models in a visual and graphical way. The editor presents a "box and arrows" representation of the hypermodel, where each hypomodel is shown as a rectangle and each data exchange link between two hypomodels is depicted as line connecting them. Fig. HE1 shows the editor and an exemplary hyper model:

Figure HE1. The main drawing "canvas" of the CHIC Hypermodelling Editor.

The design of a new hyper model involves the selection of the needed hypo models and the specification of the connections between them. The selection of the hypomodels to include in a new hypermodel can be greatly facilitated if the hypomodels have been properly annotated so that the users are able to filter them. In the current stage of implementation all of the CHIC hypomodels have been annotated according the Perspectives I, IV, and V as defined in Deliverable 6.1:

- Perspective I ("TUMOUR-AFFECTED NORMAL TISSUE MODELLING") refers to whether the model tries to simulate the tumour, normal tissue, oedima, or treatment affected tissue

- Perspective IV ("BIOMECHANISM(S) ADDRESSED") refers to the cell processes modelled, such as the apoptosis, metabolism, necrosis, etc.

- Perspective V ("TUMOUR TYPE(S) ADDRESSED") refers to the specific cancer (tumour) type addressed by the model, such as the lung cancer, glioblastoma, prostate cancer, etc.

Fig. HE2 depicts the user interface in the Hypermodelling Editor that supports the selection of the models to be shown based on these criteria. Fig. HE3 specifically shows the selection of the Nephroblastoma cancer type and the results (5 models) that satisfy this selection criterion. When the user locates the subset of the models that can possibly use for constructing the hyper model, s/he can press the "add" (plus) button to include them into the drawing area.
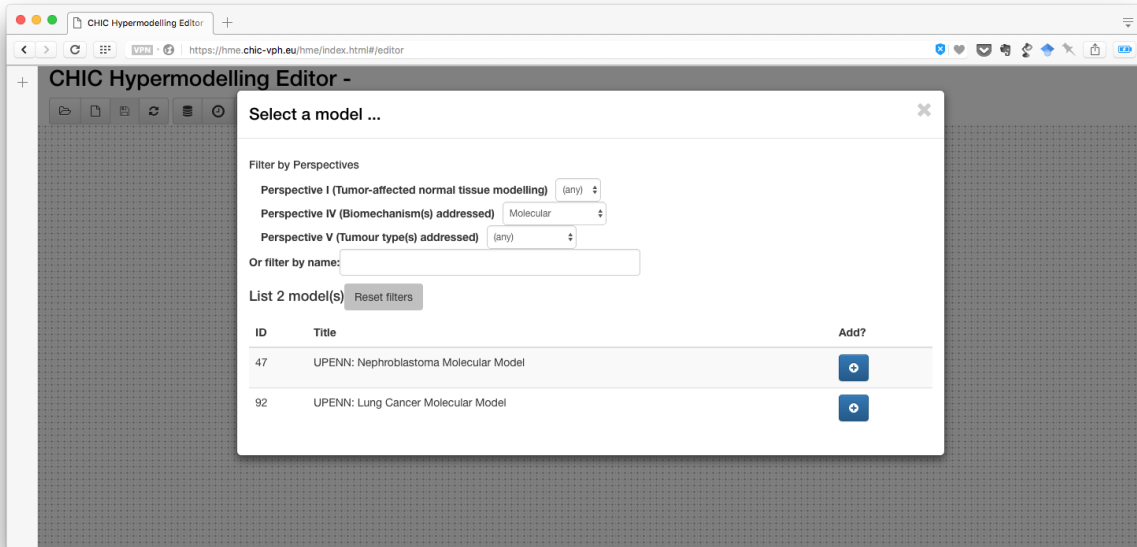

Fig. HE2. Filtering and selection of hypomodels using their semantic, perspective-based metadata annotations
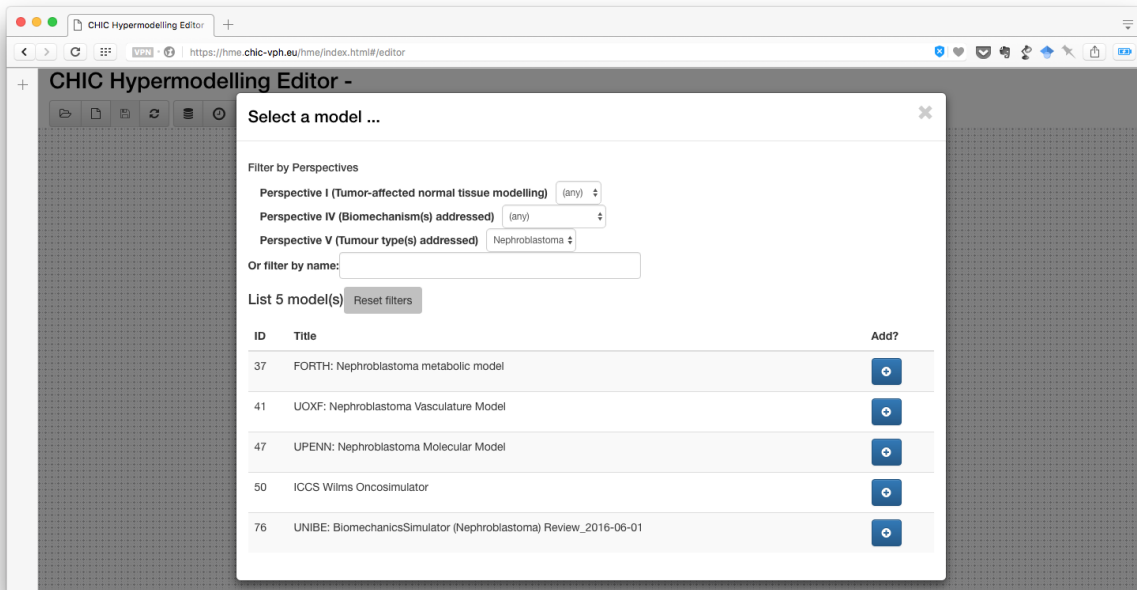

Fig. HE3. Filtering hypomodels based on the specific tumour type addressed

When a new hypo model is selected and put into the main drawing area of the Editor, then user can "double-click" it and a window with more information about the model appears (Fig. HE4). The additional information includes information about the inputs and the outputs of the model, such as the data types of the parameters (e.g. number, string, etc), the units used, and the default values used if an input is not set explicitly to a value, or connected with the output of another model. Furthermore, the perspective-related metadata annotations are shown as "tags" under the description of the model.
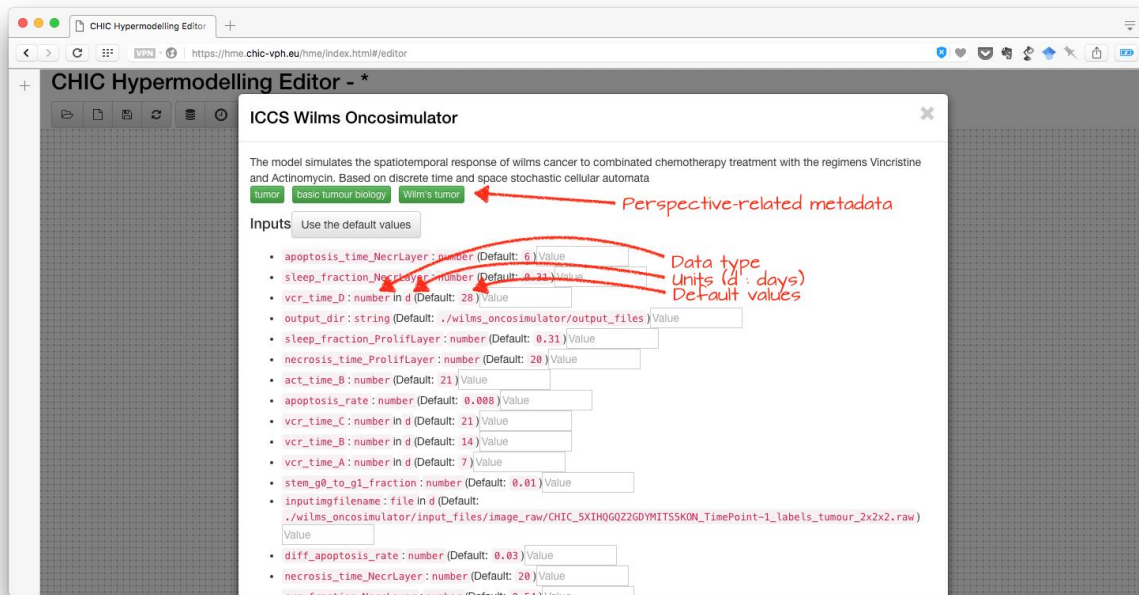
Fig. HE4. Metadata for a specific hypomodel and its inputs and outputs

The next step is to connect the models selected according to the user's envisaged hypermodel and research question. To this end, there's ongoing effort to take advantage of the metadata annotations at the parameters level in order to ease the selection of the model to connect to, and the validation of these connections.

EH. Execution of Hypermodels

The development of the process for the Execution of Hypermodels has been led by CINECA and USFD.

The most common hypermodel execution related scenario in CHIC is that a user, following appropriate authentication, has uploaded patient-specific data that has been pre-processed to a central repository so as to render it suitable inputs for predictive models. From the CRAF web interface or the CRAF desktop application the user selects an existing hypermodel (Nephroblastoma, Lung cancer, Glioblastoma, Prostate Cancer), or from the hypermodelling editor links available hypomodels into a new hypermodel, and then requests its execution from the hypermodelling framework (VPH-HF) with a certain input set. When the execution is finished, the hypermodelling framework notifies CRAF and the In Silico Trial Repository of the status of the execution and uploads the output files centrally in the In Silico Trial Repository. During the run the user can monitor the status of the execution and retrieve the hypermodel output set once it has finished to be further analysed using pertinent visualization tools (Fig. EH1) .
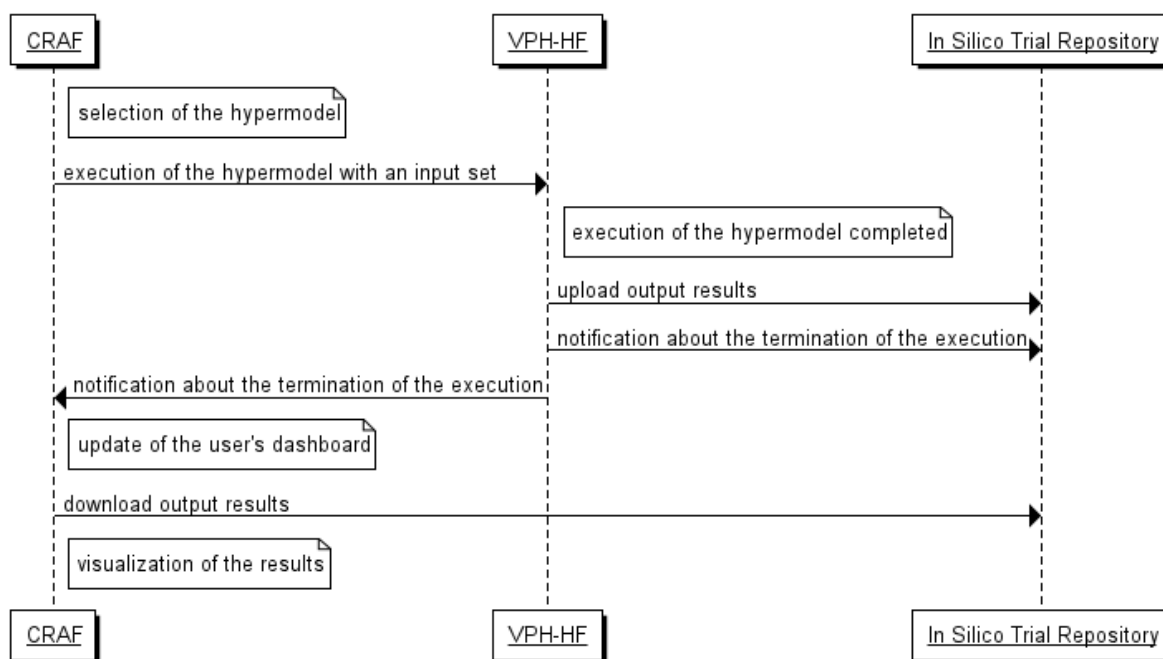

Fig. EH1. Hypermodel Execution Workflow

DI. Discussion

The four novel paradigmatic CHIC hypermodels as well as their component models (hypomodels) have been outlined in their initial standardized forms. More precisely, the following components and processes have been outlined: i) the nephroblastoma oncosimulator hypomodel and the nephroblastoma integrated hypermodel, ii) the non small cell lung cancer oncosimulator hypomodel and the non small cell lung cancer integrated hypermodel, iii) the biomechanical hypomodel for nephroblastoma and non small cell lung cancer, iv) the angiogenesis hypomodel for nephroblastoma and non small cell lung cancer, v) the metabolic hypomodel for nephroblastoma and non small cell lung cancer, vi) the molecular hypomodel for nephroblastoma and non small cell lung cancer, vii) the gross phenomenological hypermodel for nephroblastoma and non small cell lung cancer, viii) the glioblastoma hypermodel, ix) the prostate hypermodel, x) a brief outline of the process of semantically describing hypomodels and hypermodels, xi) a brief outline of the model repository in the context of hypomodel integration, xii) the clinical research application framework (CRAF) in the context of utilizing integrated hypermodels, xiii) the hypermodelling editor as a supportive technological platform for the integration of hypermodels and xiv) the process of hypermodel execution.

The entire process of the development of the initial standardized versions of the CHIC hypemodels as well as their handling and execution appears to be a successful one. Extensions of the technological tools and services including *inter alia* the handling of missing or incomplete data, supporting the prospective hypermodeller to create a new hypermodel using the hypermodelling editor will be mainly presented in the corresponding subsequent technological deliverables.

CO. Conclusions

In this document the initial standardized versions of the four paradigmatic novel cancer hypermodels developed by the CHIC project concerning nephroblastoma, non small cell lung cancer (NSCLC), glioblastoma multiforme and prostate cancer have been outlined. The component hypomodels, their static and/or dynamic coordination and their links have also been outlined. Mechanistic multiscale mathematical and computational modelling and several statistics based methods of machine learning have been recruited in order to develop both the hypomodels and the hypermodels. Clinically relevant questions dictate the precise mathematical and computational strategy to be adopted for each scenario. Pertinent technological procedures and components such as the procedure of semantically annotating hypomodels and hypermodels, the model repository, the Clinical Research Application Framework (CRAF), the hypermodelling editor, and the procedure for the execution of hypermodels have been briefly outlined in conjunction with the hypermodel integration process. Initial prediction results produced by the hypermodels, some of which demonstrated during the recent project reviews, have also been presented. Remarks on several facets of the initial standardized versions of the hypermodels have further enlightened the process of developing cancer hypermodels. The content of the deliverable suggests that the process of the development of the initial standardized versions of the CHIC hypermodels as well as their handling and execution has been a successful one. Clinical adaptation and partial validation of hypermodels are to be presented in deliverable D6.4.

Appendix 1 – Abbreviations and acronyms

AD	Androgen Dependent
ADC	Adenocarcinoma
ADSCC	Adenosquamous Cell Carcinoma
ADT	Androgen Deprivation Therapy
AJCC	American Joint Committee on Cancer
Akt	Protein kinase B (PKB)
ALK	Anaplastic Lymphoma Kinase
AMD	Advanced Microdevices
ANSI	American National Standards Institute
API	Application Program Interface
ATP	Adenosine Triphosphate
AUC	Area Under Curve
BED	University of Bedfordshire
bGS	biopsy Gleason Score
BMS	Bio-Mechanical Simulator
BS	Biomechanics Simulator
CGAL	Computational Geometry Algorithms Library
CHIC	Computational Horizons in Cancer
CINECA	Consorzio Interuniversitario del Nord Est Italiano Per il Calcolo Automatico (Interuniversity Consortium for High Performance Systems)
CKP	Cell Kill Probability
CKR	Cell Kill Rate
CNS	Central Nerous System
COSMIC	Catalog of Somatic Mutations in Cancer
CRAF	Clinical Research Application Framework (CRAF)
CS	Cell Simulator
CSF	Cerebrospinal Fluid
CSS	Cancer Stem Cell
CSV	Comma Separated Values
CT	Computed Tomography
DC	Dendritic Cell
DGM	Diffusion Coefficient of Grey Matter
DICOM	Digital Imaging and Communications in Medicine
DIFF	Terminally Differentiated Cell
DRE	Digital Rectal Examination
DWM	Diffusion Coefficient of White Matter
EAU	European Association of Urology
EBRT	External Beam Radiation Therapy
ED	Equivalent Dose
EGF	Epidermal Growth Factor
EGFR	Epidermal Growth Factor Receptor
ERBB2	erb-b2 Receptor Tyrosine Kinase 2
ERK	Extracellular Signal-Regulated Kinases
FEM	Finite Element Method
FORTH	Foundation for Research and Technology Hellas
GBM	Glioblastoma Multiforme

GF	Growth Fraction
GC	Geormetrical Cell
GPSM	Gleason, PSA, Seminal Vesicle and Margin Status
GS	Gleason Score
GUI	Graphical User Interface
HE	Hypermodelling Editor
HER3	Human Epidermal Growth Factor Receptor 3
HTML	Hypertext Markup Language
ICCS or ICCS-NTUA	Institute of Communication and Computer Systems – National Technical University of Athens
IMRT	Intensity Modulated Radiation Therapy
ISO	International Organization for Standardization
KUL	Catholic University of Leuven
LADC	Lung Adenocarcinoma
LCC	Large Cell Carcinoma
LIMP	LImited Mitotic Potential
LQ	Linear Quadratic
LSCC	Lung Squamous Cell Carcinoma
MAPK	Mitogen-Activated Protein Kinase
MD	Molecular Dyamics
MRI	Magnetic Resonance Imaging
MUSCLE	Multiscale Coupling Library and Environment
MUT	Mutant
NBC	Number of Biological Cells
NCCN	National Comprehensive Cancer Network
NCI	National Cancer Institute
NGCT	Neighbour Geometrical Cells belonging to the Tumour
NIH	National Institutes of Health
NK	Natural Killer
NSCLC	Non Small Cell Lung Cancer
NSG	NOD-$scid$ $IL2r\gamma^{null}$ Mouse Model of Human Skin
OER	Oxygen Enhancement Ratio
OFAT	One Factor at A Time
OS	Oncosimulator
OWL	Web Ontology Language
pAKT	phospho-AKT
PCa	Prostate Cancer
PDE	Partial Differential Equation
PSA	Prostate Specific Antigen
PUN	Phenomenological Universalities (Approach)
RDF	Resource Description Framework
RP	Radical Prostatectomy
RT	Radiotherapy
RTK	Receptor Tyrosine Kinase
SASA	Solvent Accessible Surface Area
SBML	Systems Biology Markup Language
SCC	Squamous Cell Carcinoma

SCID	Severe Combined ImmunoDeficient
SCLC	Small Cell Lung Cancer
SQL	Structured Query Language
STAT	Signal Transducer and Activator of Transcription or Signal Transduction And transcription
SVM	Support Vector Machines
TCGA	The Cancer Genome Atlas
TKI	Tyrosine Kinase Inhibitors
TRUS	Trans-Rectal Ultrasound
UBERN	University of Bern
UCL	University College London
ULC	Undifferentiated Large Cell Carcinoma
UML	Unified Modeling Language
UNITO	University of Turin
UOXF	University of Oxford
UPENN	University of Pennsylvania
USAAR	University of Saarland
USFD	University of Sheffield
VEGF	Vascular Endothelial Growth Factor
VPH	Virtual Physiological Human
VTK	Visualization ToolKit
WT	Wild Type
WT	Wilms Tumour = Nephroblastoma
XML	EXtensible Markup Language