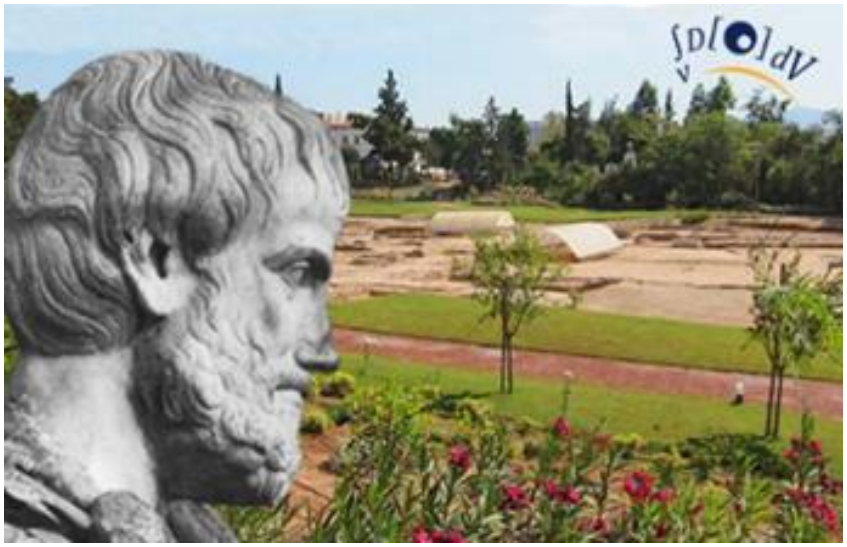


PROCEEDINGS

of the 2014
6th International Advanced Research Workshop
on
In Silico Oncology and Cancer Investigation – The CHIC Project Workshop (IARWISOCI)

An IEEE-EMBS Technically Co-sponsored Conference
Funded by the European Commission
through the CHIC Project
in the Framework of the VPH Initiative

Athens, Greece, 3-4 November 2014

Edited by
G. S. Stamatakos and D. Dionysiou

(Open Access Version)
ISBN: 978-618-80348-1-5

Available at http://6th-iarwisoci.iccs.ntua.gr/

2014 6th International Advanced Research Workshop on *In Silico* Oncology and Cancer Investigation - The CHIC Project Workshop (IARWISOCI)
(open-access version)

Edited by Georgios S.Stamatakos and Dimitra D. Dionysiou

Available at www.6th-iarwisoci.iccs.ntua.gr

ISBN (electronic): 978-618-80348-1-5

Institute of Communication and Computer Systems
National Technical University of Athens
(ICCS-NTUA)
Iroon Polytechniou 9
Zografos
GR-157 80 Greece

Proceedings of the 2014 6th International Advanced Research Workshop on *In Silico* Oncology and Cancer Investigation – The CHIC Project Workshop (IARWISOCI)

Athens, Greece, 3-4 November 2014

Edited by
G. S. Stamatakos and D. Dionysiou

(Open AccessVersion)

ISBN (electronic): 978-618-80348-1-5

Available at www.6th-iarwisoci.iccs.ntua.gr

An IEEE-EMBS Technically Co-sponsored Conference
Funded by the European Commission
through the CHIC Project
in the Framework of the
Virtual Physiological Human (VPH) Initiative

ORGANIZING COMMITTEE

General Chair

G. Stamatakos, PhD, ICCS - National Technical University of Athens (GR)

Members

N. Graf, MD, University of Saarland (DE)

M.Akay, PhD, University of Houston (US)

D.Dionysiou, PhD, ICCS – National Technical University of Athens (GR)

K.Marias,PhD, Foundation for Research and Technology Hellas (GR)

R.Radhakrishnan, PhD, University of Pennsylvania (US)

N.Uzunoglu, PhD, ICCS – National Technical University of Athens (GR)

PAPER CITATION FORMAT
(Open Access Version)

(Co-)Author(s), "Paper Title." In G. Stamatakos and D. Dionysiou (Eds): Proc. 2014 6th Int. Adv. Res. Workshop on *In Silico* Oncology and Cancer Investigation – The CHIC Project Workshop (IARWISOCI), Athens, Greece, Nov.3-4, 2014 (www.6th-iarwisoci.iccs.ntua.gr), pp.xx-xx. (open-access version), ISBN: 978-618-80348-1-5

NOTE

It is noted that the **open-access version** of the proceedings is freely available on the workshop website www.6th-iarwisoci.iccs.ntua.gr

An **IEEE Xplore® version** of the proceedings will also be made available.
IEEE Catalog Number CFP14IAR-ART
IEEE Technically Co-sponsored Conference # 35655

Please clearly mention which proceedings version a paper citation refers to.

CORRESPONDENCE

All correspondence should be addressed to

Georgios S. Stamatakos
Research Professor
In Silico Oncology Group
Institute of Communication and Computer Systems
National Technical University of Athens
Iroon Polytechniou 9, Zografos GR 157 80 , Greece
Tel: (+30) 210 772 2287, Fax: (+30) 210 772 3557
E-Mail: gestam@central.ntua.gr
URL: www.in-silico-oncology.iccs.ntua.gr

TABLE OF CONTENTS
(Open Access Version)

In Silico Medicine: The Paradigm of *In Silico* Oncology

Proceedings of the 6th International Advanced Research Workshop on *In Silico* Oncology and Cancer Investigation – The CHIC Project Workshop*

Editorial

Georgios S. Stamatakos, *Member, IEEE*

I. Introduction

Diseases are *natural phenomena* and consequently are amenable to mathematical and computational description. Clinically driven complex multi-scale disease models are capable of producing realistic spatio-temporal and patient-specific simulations of several clinical interventions. Clinical data-processing procedures and computer technologies play an important role in this context. Following clinical adaptation and validation within the framework of clinico-genomic trials, models are expected to advance the prospect of individualized treatment optimization, this being the long term goal of the emergent scientific, technological and medical discipline of *in silico* medicine. *In silico* oncology has proven to be *inter alia* an excellent didactic, research and clinical paradigm of this new discipline, since cancer is strongly manifested at all scales of biocomplexity.

Treatment optimization is to be achieved through experimentation *in silico* i.e. on the computer. Moreover, provision of improved insight into disease dynamics and optimization of clinical trial design and interpretation constitute short- and mid-term goals of this new domain.

The IEEE-EMBS technically co-sponsored 6th International Advanced Research Workshop on *In Silico* Oncology and Cancer Investigation (6th IARWISOCI) (www.6th-iarwisoci.iccs.ntua.gr), being also the CHIC project workshop (http://chic-vph.eu/), proved an excellent opportunity for contributing to the shaping of the discipline. The presented papers deal with modeling of tumor dynamics and response to treatment from the biochemical to the macroscopic level and from basic science to clinics via information technology and legal and ethical handling. They have been contributed by top international researchers and research groups. This year's workshop was dedicated to Aristotle and special focus was put on the collection, processing, exploitation and legal and ethical aspects of the clinical multiscale data which represent the Aristotelian observable reality for the advancement of *in silico* oncology. A one hour session of the workshop was held on the very site of Aristotle's Peripatetic School. The workshop took place in Athens, Greece on 3-4 November 2014.

II. *In Silico* Oncology

In silico oncology could be formally defined as being "…a complex and multiscale combination of sciences, technologies and clinical medicine intending to simulate malignant tumor growth and tumor and normal tissue response to therapeutic modalities at all biomedically meaningful spatio-temporal scales". Its long term goal is to quantitatively understand cancer and related phenomena and optimize therapeutic interventions by performing *in silico* experiments using clinical, imaging, histopathological, molecular and pharmacogenomic data from individual patients. In order to achieve such an ambitious goal translation of cancer models and *oncosimulators* or more generally *clinical simulators* into the clinical trials arena is a *sine qua non* condition.

III. Organizing Committee

The Organizing Committee of the workshop consisted of the following persons

G. Stamatakos, PhD, ICCS - National Technical University of Athens (GR), General Chair
N. Graf, MD, University of Saarland (DE)
M.Akay, PhD, University of Houston (US)
D.Dionysiou, PhD, ICCS – National Technical University of Athens (GR)
K.Marias,PhD, Foundation for Research and Technology Hellas (GR)
R.Radhakrishnan, PhD, University of Pennsylvania (US)
N.Uzunoglu, PhD, ICCS – National Technical University of Athens (GR)

*The workshop was funded by the European Commission through the transatlantic CHIC project (FP7- Grant Agreement No 600841).

G.S.Stamatakos, the CHIC Project Coordinator, is Research Professor at the Institute of Communication and Computer Systems, National Technical University of Athens and Director of the *In Silico* Oncology and *In Silico* Medicine Group. 9, Iroon Polytechniou, Zofrafos, 157 80, Greece (phone: +30 210 772 2287; fax: +30 210 772 3557; e-mail: gestam@mail.ntua.gr).

Computational Horizons In Cancer (CHIC): Developing Meta- and Hyper-Multiscale Models and Repositories for *In Silico* Oncology – a Brief Technical Outline of the Project*

G.Stamatakos, *Member IEEE*, Dimitra Dionysiou, Fay Misichroni, Norbert Graf, *Member IEEE*, Stefaan van Gool, Rainar Bohle, Feng Dong, Marco Viceconti, Kostas Marias, *Member IEEE*, Vangelis Sakkalis, Nikolaus Forgo, Ravi Radhakrishnan, Helen Byrne, Caterina Guiot, Philippe Buechler, Elias Neri, Anca Bucur, Bernard de Bono, Debora Testi, Manolis Tsiknakis, *Member IEEE*, on behalf of the CHIC consortium

***Abstract*— This paper briefly outlines the aim, the objectives, the architecture and the main building blocks of the ongoing large scale integrating transatlantic research project CHIC (http://chic-vph.eu/).**

I. Introduction

The impressive rate of generation of human biological data during the last decades has dictated the development of numerous statistical, computational and mathematical methods, in order to extract, analyze and exploit the hidden wealth of information. Unquestionably systems biology has been established as a key player in this arena. However, despite its maturation over the last decade a number of obstacles render it difficult for systems biology to be directly exploitable by clinical practice [1]. Recognizing that in most medical conditions crucial biological phenomena are manifested at several spatiotemporal scales, including scales lying far above the subcellular level - which is traditionally addressed by systems biology- researchers have proposed a number of ways to integrate super-cellular levels into systems biology approaches. Such initiatives have taken various forms and names such as systems physiology [2] systems medicine, multiscale modeling [3] and Virtual Physiological Human (VPH).

Cancer in the clinical context dictates the development of integrative hypermodels consisting of simpler and more manageable constituent component models which may already be available. Nevertheless, in order for models generally developed by different modellers or modelling groups to be reusable, there are a number of prerequisites that have to be satisfied. Models should be robust, reproduceable and interoperable. This implies that standardization of model description and operation is a sine qua non necessity if rational, coherent and comprehensive exploitation of the invaluable information hidden within human multiscale biological data is envisaged. Responding to this imperative in the context of both the broad (VPH) initiative and the paradigmal cancer domain, CHIC proposes the development of a suite of tools and services in a secure infrastructure that will support accessibility and reusability of VPH mathematical and computational hypermodels. The proposed objective is primarily centered around the development of a hypermodelling environment which, although will be applicable to the broad VPH space, it will be driven by and

*This work has been supported by the European Commission under the project Computational Horizons In Cancer (CHIC): Developing Meta- and Hyper-Multiscale Models and Repositories for In Silico Oncology (FP7-ICT-2011-9, Grant agreement no: 600841).

G.S.Stamatakos is with the *In Silico* Oncology Group, Institute of Communication and Computer Systems, National Technical University of Athens, Greece (corresponding author, project scientific coordinator: phone:+302107722287, fax:+302107723557, e-mail: gestam@central.ntua.gr).

D Dionysiou is the *In Silico* Oncology Group, Institute of Communication and Computer Systems, National Technical University of Athens, Greece (e-mail:) dimdio@esd.ece.ntua.gr).

F.Misichroni is with the In Silico Oncology Group, Institute of Communication and Computer Systems, National Technical University of Athens, Greece (e-mail: faymisi@mail.ntua.gr).

N.Graf is with the University of Saarland, Pediatric Oncology and Hematology Clinic, Germany (e-mail: Norbert.Graf@uniklinikum-saarland.de).

S. van Gool is with the Catholic University of Leuven, Pediatric Oncology Clinic, Belgium (e-mail: stefaan.vangool@uzleuven.be)

R.Bohle is with the University of Saarland, Dept. of Pathology, Germany (e-mail: Rainer.Bohle@uniklinikum-saarland.de).

F.Dong is with the University of Bedfordshire, UK (e-mail: Feng.Dong@beds.ac.uk).

M.Viceconti is with the University of Sheffield, UK (e-mail: m.viceconti@sheffield.ac.uk).

K.Marias is with the Foundation for Research and Technology, Hellas, Greece (e-mail: kmarias@ics.forth.gr).

V. Sakkalis is with the Foundation for Research and Technology, Hellas, Greece (e-mail: sakkalis@ics.forth.gr).

N.Forgo is with the G.W.Leibnitz University of Hannover, Germany (e-mail: nikolaus.forgo@iri.uni-hannover.de).

R.Radhakrishnan is with the University of Pennsylvania, USA (e-mail: rradhak@seas.upenn.edu).

H.Byrne is with the University of Oxford, UK (e-mail: byrneh@maths.ox.ac.uk).

C.Guiot is with the University of Torino (e-mail: caterina.guiot@unito.it).

P.Buechler is with the University of Bern, Switzerland (e-mail: philippe.buechler@istb.unibe.ch).

E.Neri is with Custodix NV, Belgium (e-mail: elias.neri@custodix.com)

A.Bucur is with Philips Electronics Nederland B.V., The Netherlands (e-mail: anca.bucur@philips.com).

B.de Bono is with University College London, UK (e-mail: b.bono@ucl.ac.uk).

D.Test is with Consorzio Universitario CINECA, Italy (e-mail: d.testi@cineca.it).

M.Tsikankis is with the Technological Educational Institute of Crete, Greece (e-mail: tsiknaki@ics.forth.gr).

originally tested in the cancer domain. In order to ensure clinical relevance and foster clinical acceptance of hypermodelling in the future, the whole endeavour will in practice be driven by the clinical partners of the consortium. Cancer hypermodels to be collaboratively developed by the consortium cancer modellers will provide the framework and the testbed for the development of the CHIC technologies. Clinical adaptation and partial clinical validation [4-5] of hypermodels and hypermodel oncosimulators will be undertaken.

II. Aim

The CHIC proposal aims at developing cutting edge ICT tools, services and secure infrastructure to foster the development of elaborate and reusable integrative models (hypermodels) and larger repositories so as to demonstrate benefits of having both the multiscale data and the correponding models readily available. Although the broader VPH domain is the primary target of the hypermodelling infrastructure to be developed by CHIC, the primary application domain will be cancer and in silico oncology.
In the mid and long term CHIC aims to pave the way for reliable in silico clinical trials, lying at the heart of the vision of in silico medicine, and subsequently for patient individualized treatment optimization based on in silico experimentation [4-5].

III. Objectives

CHIC proposes the development of clinical trial driven tools, services and secure infrastructure that will support the creation of multiscale cancer hyper-models (integrative models). The latter are defined as choreographies of component models, each one describing a biological process at a characteristic spatiotemporal scale, and of relation models/metamodels defining the relations across scales. Integrative models can become component models for other integrative models. The development of a secure hypermodelling infrastructure consisting primarily of a hypermodelling editor and a hypermodelling execution environment is a central generic VPH geared objective of CHIC.

In order to render models developed by different modellers semantically interoperable, an infrastructure for semantic metadata management along with tools and services for ontology-based annotations will be developed. Existing approaches such as the one developed by the EC funded RICORDO project will be exploited and extended. Facilitated operations will range from automated dataset matching to model merging and managing complex simulation workflows. In this way standardization of cancer model and data annotation allowing multiscale hypermodelling will be fostered.

The following entities will also be developed: a hypermodel repository, a hypermodel-driven clinical data repository, a distributed metadata repository and an in silico trial repository for the storage of executed simulation scenarios, an image processing toolkit, a visualization toolkit and cloud and virtualization services.

In order to ensure that the entire project will be clinically driven and clinically oriented, three concrete clinical trials/studies will be adopted and addressed. They concern nephroblastoma treated by combined chemotherapy, glioblastoma treated by immunotherapy in combination with chemotherapy and radiotherapy and non-small cell lung cancer treated by a combination of chemotherapy and radiotherapy.

The multiscale data generated by these trials/studies will be exploited so as to both drive the development of a number of integrative multiscale cancer models (hypermodels) and hypermodel oncosimulators and clinically adapt and partly validate them.

The whole process will be supported by the technological tools, services and infrastructure to be developed and will serve as a paradigm of applicability and usability of the latter. Additional available multiscale data concerning colon and prostate cancer will be exploited in a similar way. The participation of five prominent multiscale cancer modelling groups from both EU and the US covering all spatiotemporal scales (from the molecular up to the organism and from nsecs up to years) and all the fundamental biological processes of cancer as well as some aspects of the treatment response of normal tissues will ensure a comprehensive coverage of the domain of cancer. The latter refers to both the process of annotating component models and hypermodels as well as pertinent multiscale data and the development of exemplary clinically driven and clinically validatable hypermodels.

This is expected to considerably advance the exploitation of both existing models and models to be developed in the future. An integrative platform dictated by the IT architecture of the project will provide access to all hypermodelling tools and services to be developed. Apart from the tools addressing semantic interoperability, a number of data pre-processing tools, services and resources will be developed and/or made available. These will include inter alia image segmentation, three-dimensional reconstruction, several forms of data and model prediction visualization and cloud computing.

The legal and ethical aspects of patients' data handling will be addressed by a workpackage dealing with both the legal and the IT aspects of data anonymization and pseudonymization, patient's consent etc. The same work package will also address the intellectual rights issues arising from the amalgamation of component models potentially developed by different modellers in order to construct integrative models.

The dissemination and exploitation of the CHIC proposal will target all stakeholders, namely clinicians, fundamental science researchers, IT specialists and engineers, industry and patients. Similarly, the project is expected to have a significant impact on all the corresponding domains. More precisely, CHIC aspires to make a breakthrough in multiscale cancer modelling through greatly facilitating multi-modeller cancer hypermodelling and its clinical adaptation and validation. Standardization of model description and model "fusion" will be two of the core means to achieve this goal. The creation of such elaborate and refined hypermodels is expected to sharply accelerate the clinical translation of

multiscale cancer models and oncosimulators following their prospective clinical validation (*in silico* oncology). Addressing intellectual property issues in a multi modeller setting will foster the community spirit in the VPH domain.

IV. Architecture and Main Building Blocks of CHIC

CHIC will develop a variety of tools and repositories that will assist the researcher in searching and retrieving models and data, composing and saving hypermodels, executing models and hypermodels and last but not least validating the outcome of the simulations (Fig.1).

The core reference point for the users will be the CHIC **portal**. All the components of CHIC will reside under the "umbrella" of the **security framework** that will deal with the issues of secure and safe storing, acquisition and sharing of models and data.

Four individual repositories will be implemented in CHIC.

- A **model repository** that will store the multiscale models, the complimentary tools and modules that will be needed in order to construct hypermodels and the hypermodels themselves. In the model repository will also reside the **visualization** and **image processing** tools that will be developed in CHIC.
- A **data repository** that will store the heterogeneous multiscale data coming from clinical environment (clinical trials etc.). Especially for the storage of "sensitive" patient-specific data a special **pseudo-anonymization/anonymization** procedure will be followed in compliance with the **legal and ethical framework.** Due to legal limitations, the CHIC repositories, especially the ones that are dealing with patient data, will be implemented so as to be easily deployable in local or private cloud infrastructures of medical, educational and research institutions.

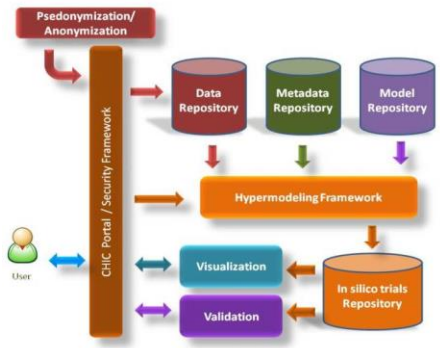


Figure 1 Main technological components of CHIC directly related to the hypermodelling workflow

- A **metadata repository** that will store the machine-readable documentation material that will semantically represent both models and data.
- An ***in silico* trial repository** which will store the input and output of the *in silico* simulations along with the complete profile of each simulation, including the model/hypermodel used in the simulation and its version, the model/hypermodel configuration parameters etc.

The users will upload their models and the complimentary tools in the model repository. In addition the user will use the model annotation framework to add semantic information to his/her models and data. This information will be used later on by the hypermodelling framework in order to construct and execute hypermodels (Fig.2).

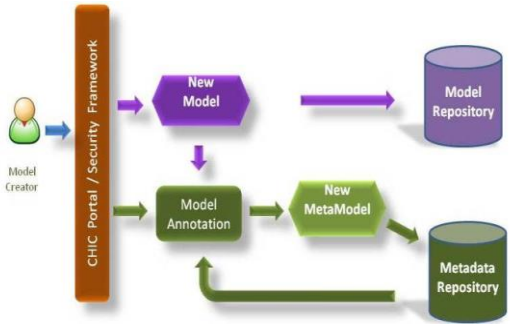


Figure. 2 Model and metamodel creation workflow

The composition and execution of hypermodels will be done by the **Hypermodelling Framework**. This will consist of the Hypermodelling Editor and the Hypermodel Executional Framework. The **Hypermodelling Editor** will communicate with the model and metadata repositories and will guide the user in easily and effectively constructing hypermodels (Fig.3) by

- exposing information about existence and availability of models,
- presenting interconnection possibilities,
- indicating the model/modules that need to be developed in order to fill in the gaps,
- visually constructing the hypermodels, provided that all needed components are available, either as

implemented models/modules or as a " to be implemented" dummy black boxes.

The Hypermodel Executional Framework will communicate with the model, the metadata and the data repository, in order to retrieve the relevant information to be used in the simulation (*in silico* trial). The outcome of the execution will be send to the in silico trial repository for persistent storage. The user will be able to retrieve the results of a simulation from the in silico trials repository (Fig.4).

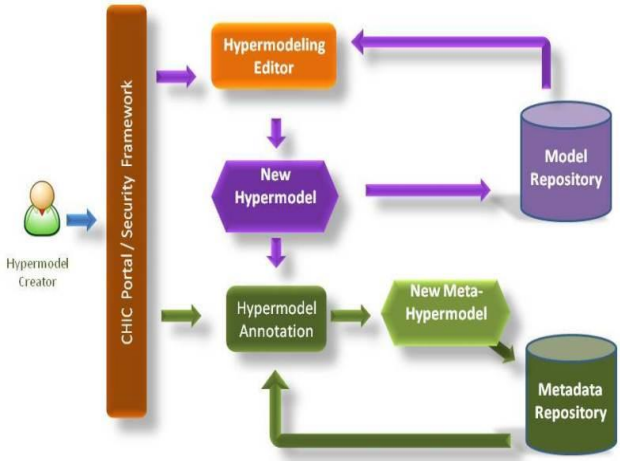


Figure 3. Hypermodel and meta-hypermodel creation workflow

The CHIC image processing tools will be used in the preprocessing of imaging data in order to be prepared for usage in the simulations. The results of the simulations will use the CHIC visualization tools in order to be presented to the user.

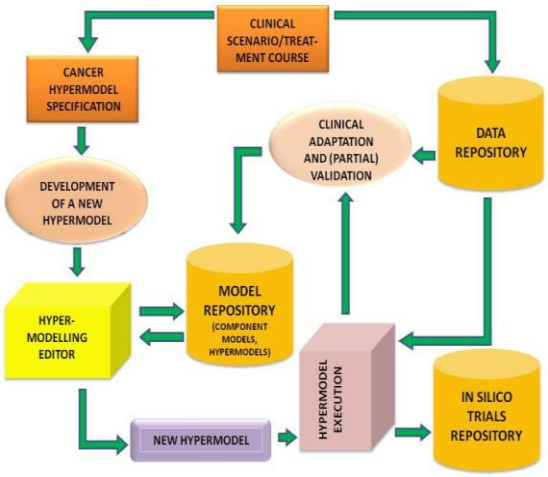


Figure.4 Clinical scenario driven hyper- model development

Fig.5 shows the gross overall CHIC architecture from a clinical study and trial centered perspective.

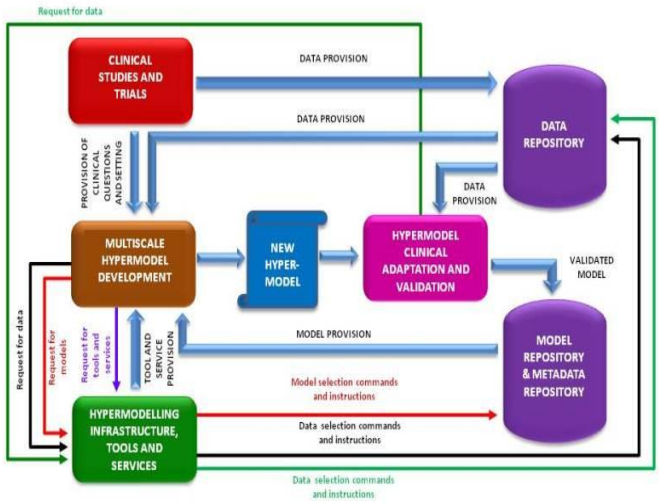


Figure 5. The overall CHIC architecture from a clinical study and trial centered perspective

The Hypermodel Oncosimulator is an extension of the notion and the system of the original Oncosimulator [4-5] so as to make use of cancer and normal tissue hypermodels. The (hypermodel) Oncosimulator is at the same time a concept of multilevel integrative cancer biology, a complex algorithmic construct, a biomedical engineering system and eventually in the future a clinical tool which primarily aims at supporting the clinician in the process of optimizing cancer treatment in the patient individualized context through conducting experiments in silico i.e. on the computer. Additionally it is a platform for simulating, investigating, better understanding and exploring the natural phenomenon of cancer, supporting the design and interpretation of clinicogenomic trials and finally training doctors, researchers and interested patients alike. A synoptic outline of the clinical utilization of a specific version of the Oncosimulator, as envisaged to take place following an eventually successful completion of its clinical adaptation, optimization and validation process is provided in the form of steps (Fig.6).

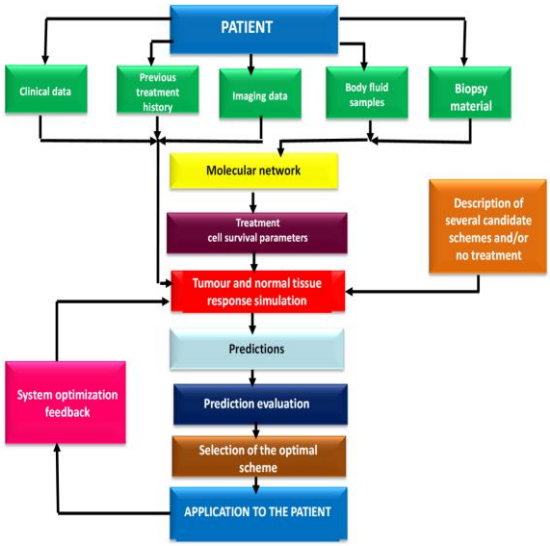


Figure 6. A synoptic diagram of the hypermodel based Oncosimulator

VI. CONCLUSION

In this paper a short description of the main technical architecture and components of the CHIC project has been provided. Initial successful results (http://chic-vph.eu/) have demonstrated that the design of the project is realistic and possesses great potential for the semi-automatic development of cancer hypermodels . Since the technologies developed are quite generic, an extension to domains beyond cancer will be an obvious additional outcome of the project.

ACKNOWLEDGEMENTS

The scientific support of E.Kolokotroni, E.Ouzounoglou, E.Georgiadi all from the In Silico Oncology Group, ICCS, NTUA as well as the contributions of other consortium person-members including C.Hahn, Eurice during the preparation of the CHIC Proposal are duly acknowledged.

REFERENCES

[1] G. Clermont, C. Auffray, Y. Moreau, D. M. Rocke, D. Dalevi, D. Dubhashi, D. R. Marshall, P. Raasch, F. Dehne, P. Provero, J Tegner, B. J. Aronow, M.A. Langston, and M. Benson, "Bridging the gap between systems biology and medicine," *Genome Medicine* vol 1, p. 88, Sep. 2009.

[2] H. Kitano, "Grand challenges in systems physiology," *Frontiers in Physiology*, vol 1, pp. 1-3, May 2010.

[3] http://ecancer.org/tv/pubdate/105

[4] G. Stamatakos, "In Silico Oncology Part I: Clinically Oriented Cancer Multilevel Modeling Based on Discrete Event Simulation," in *Multiscale Cancer* Modeling, T.Deisboeck and G. Stamatakos, Eds. 407-436 2011-01-01 CRC Press, Print ISBN: 978-1-4398-1440-6 eBook ISBN: 978-1-4398-1442-0 DOI: 10.1201/b10407-19 Boca Raton, Florida, USA, 2011

[5] G. Stamatakos, D. Dionysiou, A. Lunzer, R Belleman, E. Kolokotroni, E. Georgiadi, M. Erdt, J. Pukacki, S. Rueping, S. Giatili, A. d`Onofrio, S. Sfakianakis, K. Marias, C. Desmedt, M. Tsiknakis, and N. Graf, "The Technologically Integrated Oncosimulator: Combining Multiscale Cancer Modeling with Information Technology in the In Silico Oncology Context," *IEEE J Biomedical and Health Informatics*, vol. 18, no. 3, pp. 840-854, May 2014. DOI:10.1109/JBHI.2013.2284276

Dendritic Cell Vaccination for Glioblastoma Multiforme: Clinical Experience and Future Directions*

Joost Dejaegher, Lien Solie, Steven De Vleeschouwer and Stefaan W. Van Gool

***Abstract*— Dendritic cell vaccination is an experimental treatment for malignant gliomas, and has been subject of a translational program for more than ten years in our center. *In vitro* research, animal models and clinical trials for relapsed and newly diagnosed patients have been conducted. In this paper, we give an overview of the mechanism and rationale of this treatment for brain cancer. We also briefly discuss recently updated results of our clinical trials. Finally, we mention strategies to select patients for this therapy and additional immunotherapeutic strategies to further enhance the antitumor immune responses.**

I. INTRODUCTION

Glioblastoma multiforme (GBM) is the most common, yet most aggressive primary brain tumor, with an estimated incidence of 2-3 per 100000 per year for adults [1]. Current standard treatment consists of maximal cytoreductive neurosurgery, followed by radiotherapy with concomitant temozolomide chemotherapy and an additional 6 cycles of adjuvant temozolomide. With this trimodal therapy, however, median overall survival is only 15 months and less than 10% of treated patients is alive 5 years after diagnosis [2],[3]. More difficult to measure is the massive impact of the disease and its neurological consequences on the daily life of the patients and their families. GBM is incurable, mainly due to the extensive microscopic spread of infiltrative tumor cells in the apparently normal brain tissue surrounding the heterogeneous tumor mass [4]. Opposite to the dismal epidemiology of the disease are the many research teams focusing on GBM, with clinical trials testing new treatment modalities. As the aforementioned trimodal therapy is considered to be standard, most of these new treatments are investigated in the relapse setting or as an add-on to the standard therapy. The remainder of this article will focus on one particular new treatment in the setting of a newly diagnosed GBM: dendritic cell vaccination.

* This work has been supported by the Olivia Hendrickx Research Fund (www.olivia.be), the Herman Memorial Research Fund (www.hmrf.be) and the James E. Kearney Foundation (www.jekfoundation.org). Part of this project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement No [600841] (CHIC project, http://chic-vph.eu/).

J. Dejaegher is affiliated to the Department of Neurosciences, Laboratory of Experimental neurosurgery and neuroanatomy, KU Leuven and University Hospital Leuven (e-mail: Joost.dejaegher@uzleuven.be).

L. Solie is affiliated to the Department of Microbiology and Immunology, Pediatric Oncology, KU Leuven (e-mail: Lien.solie@med.kuleuven.be)

S. De Vleeschouwer is affiliated to the Department of Neurosciences, Laboratory of Experimental neurosurgery and neuroanatomy, KU Leuven and University Hospital Leuven (e-mail: Steven.devleeschouwer@uzleuven.be)

S. W. Van Gool is affiliated to the Department of Microbiology and Immunology, Pediatric Oncology, KU Leuven and University Hospital Leuven (corresponding author phone: +3216343867 ; fax: +3216343842; e-mail: Stefaan.vangool@uzleuven.be).

II. THE CONCEPT OF DENDRITIC CELL VACCINATION

A. *Rationale for dendritic cell-based immunotherapy*

Dendritic cell (DC) vaccination is an immunotherapeutic strategy. Immunotherapy as an anti-cancer treatment has, at least theoretically, great potential. If directed against tumor-specific antigens, it can combine a high degree of specificity for tumor tissue while sparing normal brain tissue. We refer to some excellent reviews for more information about different types of immunotherapy used for brain cancer [5-8]. DC vaccination is an active, specific immunotherapeutic strategy, which means that it activates the patient's own immune system against the tumor. By doing this, it has one more advantage over other types of immunotherapy: the development of immunological memory [9]. DCs are a subset of white blood cells and the most powerful antigen presenting cells [10]. In several cancers, the antigen presenting capacities of these cells have been exploited to activate an immune response of the host against the tumor [11]. To date, only for metastatic prostate cancer FDA approval as standard therapy has been attained [12]. There is good evidence that patients with a GBM are immune suppressed, both locally in the brain as well as systemically [13-15]. Moreover, radiochemotherapy can accentuate the immune dysfunction [16]. Hence, induction of an effective immune response could have a beneficial effect on tumor control and survival.

B. *Loading, administration and function of dendritic cells*

Currently, large amounts of DCs can be generated *ex vivo* from harvested peripheral blood monocytes [17],[18]. These cells are loaded *ex vivo* with tumor antigens, which can be autologous lysates [19-28], autologous or synthetic tumor associated peptides [29-34], or messenger RNA from autologous glioma or cancer stem cells [35],[36]. DCs loaded with this tumor material are injected back into the patient, with or without local immunomodulator, after which they travel to the cervical lymph nodes to induce a T-cell response [37]. DCs are known to prime not only CD4+ T helper cells but also CD8+ cytotoxic T cells [38], because they can present and cross-present antigens in the context of both MHC-Class 2 and Class 1 molecules, respectively [39],[40]. Moreover, they are also able to activate NK cells and NKT cells, which are more and more believed to be an important effector pathway of DC vaccination therapy [27],[41-43].

III. DENDRITIC CELL VACCINATION TRIALS AT LEUVEN UNIVERSITY HOSPITAL

In KU Leuven and University Hospitals Leuven, a translational research program aimed at developing and testing DC vaccines loaded with autologous tumor lysate in patients with relapsed malignant gliomas was started in 2003. These relapse studies were set up as a cohort comparison trial, in which new insights from the previous cohort and from preclinical work and literature were used to optimize the vaccination strategy of new study cohorts. In this way, the former cohorts functioned as an historical cohort for the latter.

This cohort comparison set-up also allowed to give the maximum of patients the experimental treatment. Results up to the fourth cohort have been published [24],[44-46]. Recently, the survival results were updated to July 2014. For adults with relapsed GBM, median Progression Free Survival (PFS) was 2.6 months, and Overall Survival (OS) 9.9 months. To compare, the reported Overall Survival rates in clinical trials with chemotherapy are 5-13 months [47], and 9.3 months for the anti-VEGF antibody bevacizumab [48]. A possible advantage of immunotherapy might be the induction of long term survival in a subset of patients, which is a rare phenomenon in the relapsed setting that has to be investigated further. In our series, we have seen 10% of patients surviving longer than 3 years after relapse. A synergistic effect of DC vaccination with chemotherapeutic agents is suspected in these multi-treated long term survivors, although our clinical trials were not designed to evaluate this. Indeed, immunotherapy could enhance the effect of previously, concomitantly or subsequently administered chemotherapy [51-53].

To further elaborate the possibilities of DC vaccination, we aimed to implement it in the standard of care for patients with newly diagnosed GBM. This adds next to surgery, radiotherapy and chemotherapy a fourth completely different treatment strategy to the primary treatment. DC vaccination was started after concomitant radiochemotherapy. Surgery, necessary to collect tumor antigens and to minimize the immunosuppressive effect of the glioma, was already part of standard treatment. The theoretical concept of immunization at the time of immune reconstitution has been described [49],[50], and vaccination before maintenance temozolomide might enhance its chemotherapeutic effect [51]. The first clinical trial (HGG-2006) included 77 patients with newly diagnosed GBM. This trial proved safety and feasibility of incorporating DC vaccination in standard therapy. This is important, because combination of therapies could theoretically lead to more morbidity or adverse events. Especially in RPA class 3 favorable survival results were seen with a median OS of 39.7 months [25]. As was seen in the relapsed patients, encouraging long term survival was recently documented, with 36% of patients in RPA class 3 alive after 5 year, compared to 28% in the temozolomide landmark trial [3]. This study provided a rationale to move to a randomized trial, which was started in 2010. This HGG-2010 trial compares standard therapy with add-on DC vaccination versus placebo, and has 6 month PFS as primary endpoint. OS, quality of life and immune profiling are secondary endpoints. To study immunotherapy after versus during chemotherapy, patients receiving placebo were treated with real vaccines after completion of adjuvant chemotherapy, *i.e.* at the read-out time point of the primary study endpoint. Inclusion has recently ended and results of the trial are to be expected the next years.

IV. Future directions

Conceptually, improvements in the current vaccination strategy can be done in 2 ways: selecting the patients that will respond to DC vaccination, and adding complementary immunotherapeutic strategies to the current vaccines.

A. Patient selection

It has become clear that not all patients benefit from DC vaccination, but there is a subset of patients that respond to this advanced therapy which is reflected in the appearance of long-term survivors as mentioned earlier. Identifying the subgroup of patients likely to respond, or vice versa, identifying the patients that certainly will not respond to DC vaccination, is of great clinical importance. We have indeed to realize that DC vaccination to date has not been shown to induce severe side effects, but treatment has to be considered invasive anyway: First, an operation has to be performed to start DC vaccination. In the treatment of newly diagnosed GBM surgery is part of the standard therapy, but at relapse the usefulness of a reoperation is unknown from literature. There is certainly a selection bias for patients who get a repeat surgery, and this is usually not performed without adjuvant treatment [54]. Second, patients have to undergo a leukapheresis to collect a large amount of monocytes. At last, during treatment patients should stay off corticosteroids unless in clinical need [55]. To date, there are no trustable parameters to determine which patient will benefit from DC vaccination. To address the problem of proper patient selection, we recently started participation in a European Union's FP7 project 'Computational Horizons In Cancer' (CHIC). As clinical partner of this project, we aim to provide multiscale data collected from the HGG 2010 trial. The data consist of clinical, radiological and biological data. Clinical data are basic patient characteristics as age, gender, etc., but also disease related data as corticosteroid intake, comorbidity, change in other anti-tumor therapy and overall survival. Furthermore, patients in the trial had multiple detailed MRI scans taken at fixed time points or at clinical need. These MRI scans will integrally be provided to the consortium for detailed analysis. Finally, biological data consist of standard and experimental measurements on tumor tissue, DC vaccines and blood samples taken at fixed time points during the trial. These data will be subject to mathematical modeling in the CHIC consortium to build a model to predict patient-specific responses. On one hand, a 'black-box' statistical approach will be used to search for significant interactions. On the other hand, we will try to build a more theoretically based mechanistic model for a better understanding of the underlying mechanisms of treatment success or failure. After building a model, we will test and validate the model on a smaller set of similarly treated patients. The final goal is to be able to answer the YES/NO question whether it is clinically useful for a specific patient to be vaccinated or not, based on a realistic data set at start of treatment.

B. Combination of immunotherapeutic strategies

DC vaccination focuses on effective antigen presentation and activation of mainly T cells. However, this antigen presentation is only one way of enhancing immune responses. Recently, interest has risen for additional strategies to further enhance the patient's own immune response. Much research has been done to overcome the immunosuppressive environment created by the glioma. A main role has been attributed to the function of a subset of CD4+ T cells, named regulatory T cells (Tregs) and characterized by CD25+ and FoxP3 expression [56]. These immunosuppressive T cells have been found in GBM, but are absent in normal brain tissue [57]. Depleting this Tregs could have beneficial effects

additional to DC vaccination, as was proven in preclinical animal research [58]. In humans, depletion of Tregs could be done by low dose metronomic cyclophosphamide [59]. Other immunosuppressive players in the context of GBM with potential therapeutic possibilities include macrophages [60-62], myeloid derived suppressor cells [14] and endothelial and perivascular cells [63]. Another promising target are immune checkpoint inhibitors. These are a family of surface molecules on immune cells through whose activation of inhibition immune activation or inhibition is regulated. Best known is the inhibitory checkpoint CTLA-4 which inhibits T cell proliferation and production of pro-inflammatory cytokines [64]. An inhibitor for this molecule is already FDA approved for treatment in metastatic melanoma [65], but preclinical research suggests also positive effects in malignant gliomas [66]. Other checkpoints under investigation are for example PD-1 [67]. Combination of these molecules with DC vaccination is currently investigated in animal research and hopefully will be able to enhance anti-tumor immune responses combined with DC vaccination.

REFERENCES

[1] T. A. Dolecek, J. M. Propp, N. E. Stroup, and C. Kruchko, "CBTRUS statistical report: primary brain and central nervous system tumors diagnosed in the United States in 2005-2009.," *Neuro. Oncol.*, vol. 14 Suppl 5, pp. v1–49, Nov. 2012.

[2] R. Stupp, W. P. Mason, M. J. van den Bent, M. Weller, B. Fisher, M. J. B. Taphoorn, K. Belanger, A. a Brandes, C. Marosi, U. Bogdahn, J. Curschmann, R. C. Janzer, S. K. Ludwin, T. Gorlia, A. Allgeier, D. Lacombe, J. G. Cairncross, E. Eisenhauer, and R. O. Mirimanoff, "Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma.," *N. Engl. J. Med.*, vol. 352, no. 10, pp. 987–96, Mar. 2005.

[3] R. Stupp, M. E. Hegi, W. P. Mason, M. J. van den Bent, M. J. B. Taphoorn, R. C. Janzer, S. K. Ludwin, A. Allgeier, B. Fisher, K. Belanger, P. Hau, A. a Brandes, J. Gijtenbeek, C. Marosi, C. J. Vecht, K. Mokhtari, P. Wesseling, S. Villa, E. Eisenhauer, T. Gorlia, M. Weller, D. Lacombe, J. G. Cairncross, and R.-O. Mirimanoff, "Effects of radiotherapy with concomitant and adjuvant temozolomide versus radiotherapy alone on survival in glioblastoma in a randomised phase III study: 5-year analysis of the EORTC-NCIC trial.," *Lancet Oncol.*, vol. 10, no. 5, pp. 459–66, May 2009.

[4] A. Claes, A. J. Idema, and P. Wesseling, "Diffuse glioma growth: a guerilla war.," *Acta Neuropathol.*, vol. 114, no. 5, pp. 443–58, Nov. 2007.

[5] J. C. Marsh, J. Goldfarb, T. D. Shafman, and A. Z. Diaz, "Current Status of Immunotherapy and Gene Therapy for High-Grade Gliomas," pp. 43–48, 2013.

[6] X. Xu, F. Stockhammer, and M. Schmitt, "Cellular-based immunotherapies for patients with glioblastoma multiforme.," *Clin. Dev. Immunol.*, vol. 2012, p. 764213, Jan. 2012.

[7] A. B. Heimberger and J. H. Sampson, "Immunotherapy coming of age : What will it take to make it standard of care," *Neuro. Oncol.*, vol. 13, no. 1, pp. 3–13, 2011.

[8] A. A. Thomas, M. S. Ernstoff, and C. E. Fadul, "Immunotherapy for the treatment of glioblastoma," *Cancer J.*, vol. 18, no. 1, pp. 59–68, 2013.

[9] S. De Vleeschouwer, S. W. Van Gool, and F. Van Calenbergh, "Immunotherapy for malignant gliomas: emphasis on strategies of active specific immunotherapy using autologous dendritic cells," *Childs. Nerv. Syst.*, vol. 21, no. 1, pp. 7–18, Jan. 2005.

[10] J. Banchereau, F. Briere, C. Caux, J. Davoust, S. Lebecque, Y. Liu, B. Pulendran, and K. Palucka, "Immunobiology of Dendritic Cells," *Annu. Rev. Immunol.*, vol. 18, pp. 767–811, 2000.

[11] S. Anguille, E. L. Smits, E. Lion, V. F. van Tendeloo, and Z. N. Berneman, "Clinical use of dendritic cells for cancer therapy.," *Lancet Oncol.*, vol. 15, no. 7, pp. e257–67, Jun. 2014.

[12] P. W. Kantoff, C. S. Higano, N. D. Shore, R. E. Berger, E. J. Small, D. F. Penson, C. H. Redfern, A. C. Ferrari, R. Dreicer, R. B. Sims, Y. Xu, M. W. Frohlich, and P. F. Schellhammer, "Sipuleucel-T immunotherapy for castration-resistant prostate cancer," *N. Engl. J. Med.*, vol. 363, no. 5, pp. 411–422, 2012.

[13] a R. Dix, W. H. Brooks, T. L. Roszman, and L. a Morford, "Immune defects observed in patients with primary malignant brain tumors.," *J. Neuroimmunol.*, vol. 100, no. 1–2, pp. 216–32, Dec. 1999.

[14] J. C. Rodrigues, G. C. Gonzalez, L. Zhang, G. Ibrahim, J. J. Kelly, M. P. Gustafson, Y. Lin, A. B. Dietz, P. a Forsyth, V. W. Yong, and I. F. Parney, "Normal human monocytes exposed to glioma cells acquire myeloid-derived suppressor cell-like properties.," *Neuro. Oncol.*, vol. 12, no. 4, pp. 351–65, Apr. 2010.

[15] P. E. Fecci, D. a Mitchell, J. F. Whitesides, W. Xie, A. H. Friedman, G. E. Archer, J. E. Herndon, D. D. Bigner, G. Dranoff, and J. H. Sampson, "Increased regulatory T-cell fraction amidst a diminished CD4 compartment explains cellular immune defects in patients with malignant glioma.," *Cancer Res.*, vol. 66, no. 6, pp. 3294–302, Mar. 2006.

[16] S. A. Grossman, X. Ye, G. Lesser, A. Sloan, H. Carraway, S. Desideri, and S. Piantadosi, "Immunosuppression in patients with high grade gliomas treated with radiation and temozolomide," *Clin. Cancer Res.*, vol. 17, no. 16, pp. 5473–5480, 2012.

[17] B. Thurner, C. Röder, D. Dieckmann, M. Heuer, M. Kruse, A. Glaser, P. Keikavoussi, E. Kämpgen, A. Bender, and G. Schuler, "Generation of large numbers of fully mature and stable dendritic cells from leukapheresis products for clinical application.," *J. Immunol. Methods*, vol. 223, no. 1, pp. 1–15, Feb. 1999.

[18] M. Eyrich, S. C. Schreiber, J. Rachor, J. Krauss, F. Pauwels, J. Hain, M. Wölfl, M. B. Lutz, S. de Vleeschouwer, P. G. Schlegel, and S. W. Van Gool, "Development and validation of a fully GMP-compliant production process of autologous, tumor-lysate-pulsed dendritic cells.," *Cytotherapy*, vol. 16, no. 7, pp. 946–64, Jul. 2014.

[19] R. Yamanaka, T. Abe, N. Yajima, N. Tsuchiya, J. Homma, T. Kobayashi, M. Narita, M. Takahashi, and R. Tanaka, "Vaccination of recurrent glioma patients with tumour lysate-pulsed dendritic cells elicits immune responses: results of a clinical phase I/II trial.," *Br. J. Cancer*, vol. 89, no. 7, pp. 1172–9, Oct. 2003.

[20] R. Yamanaka, J. Homma, and N. Yajima, "Clinical Evaluation of Dendritic Cell Vaccination for Patients with Recurrent Glioma : Results of a Clinical Phase I / II Trial Cancer Therapy : Clinical Clinical Evaluation of Dendritic Cell Vaccination for Patients with Recurrent Glioma : Results of a C," *Clin. Dev. Immunol.*, vol. 11, no. 11, pp. 4160–4167, 2005.

[21] J. S. Yu, G. Liu, H. Ying, W. H. Yong, K. L. Black, and C. J. Wheeler, "Vaccination with Tumor Lysate-Pulsed Dendritic Cells Elicits Antigen-Specific, Cytotoxic T-Cells in Patients with Malignant Glioma," *Cancer Res.*, vol. 64, no. 14, pp. 4973–4979, Jul 2004.

[22] S. Rutkowski, S. De Vleeschouwer, E. Kaempgen, J. E. a Wolff, J. Kühl, P. Demaerel, M. Warmuth-Metz, P. Flamen, F. Van Calenbergh, C. Plets, N. Sörensen, A. Opitz, and S. W. Van Gool, "Surgery and adjuvant dendritic cell-based tumour vaccination for patients with relapsed malignant glioma, a feasibility study," *Br. J. Cancer*, vol. 91, no. 9, pp. 1656–62, Nov. 2004.

[23] C. J. Wheeler, K. L. Black, G. Liu, M. Mazer, X. Zhang, S. Pepkowitz, D. Goldfinger, H. Ng, D. Irvin, and J. S. Yu, "Vaccination elicits correlated immune and clinical responses in glioblastoma multiforme patients," *Cancer Res.*, vol. 68, no. 14, pp. 5955–64, Jul. 2008.

[24] S. De Vleeschouwer, S. Fieuws, S. Rutkowski, F. Van Calenbergh, J. Van Loon, J. Goffin, R. Sciot, G. Wilms, P. Demaerel, M. Warmuth-Metz, N. Soerensen, J. E. a Wolff, S. Wagner, E. Kaempgen, and S. W. Van Gool, "Postoperative adjuvant dendritic cell-based immunotherapy in patients with relapsed glioblastoma multiforme," *Clin. Cancer Res.*, vol. 14, no. 10, pp. 3098–104, May 2008.

[25] H. Ardon, S. W. Van Gool, T. Verschuere, W. Maes, S. Fieuws, R. Sciot, G. Wilms, P. Demaerel, J. Goffin, F. Van Calenbergh, J. Menten, P. Clement, M. Debiec-Rychter, and S. De Vleeschouwer, "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.*, vol. 61, no. 11, pp. 2033–44, Nov. 2012.

[26] C. E. Fadul, J. L. Fisher, T. H. Hampton, E. C. Lallana, Z. Lil, J. Gui, Z. M. Szczepiorkowski, T. D. Tosteson, C. Harker Rhodes, H. A. Wishart, L. D. Lewis, and M. S. Ernstoff, "Immune response in patients with newly diagnosed glioblastoma multiforme treated with intranodal autologous tumor lysate-dendritic cell vaccination after

radiation chemotherapy.,” *J. Immunother.*, vol. 34, no. 4, pp. 382–389, 2011.

[27] S. Pellegatta, M. Eoli, S. Frigerio, C. Antozzi, M. G. Bruzzone, G. Cantini, S. Nava, E. Anghileri, L. Cuppini, V. Cuccarini, E. Ciusani, M. Dossena, B. Pollo, R. Mantegazza, E. a Parati, and G. Finocchiaro, “The natural killer cell response and tumor debulking are associated with prolonged survival in recurrent glioblastoma patients receiving dendritic cells loaded with autologous tumor lysates,” *Oncoimmunology*, vol. 2, no. 3, p. e23401, Mar. 2013.

[28] J. L. Lasky, E. H. Panosyan, A. Plant, T. Davidson, W. H. Yong, R. M. Prins, L. M. Liau, and T. B. Moore, “Autologous tumor lysate-pulsed dendritic cell immunotherapy for pediatric patients with newly diagnosed or recurrent high-grade gliomas,” *Anticancer Res.*, vol. 33, no. 5, pp. 2047–56, May 2013.

[29] J. S. Yu, C. J. Wheeler, P. M. Zeltzer, H. Ying, D. N. Finger, P. K. Lee, W. H. Yong, F. Incardona, R. C. Thompson, M. S. Riedinger, W. Zhang, R. M. Prins, and K. L. Black, “Vaccination of Malignant Glioma Patients with Peptide-pulsed Dendritic Cells Elicits Systemic Cytotoxicity and Intracranial T-cell Infiltration Advances in Brief Vaccination of Malignant Glioma Patients with Peptide-pulsed Dendritic Cells Elicits Systemic,” *Cancer Res.*, vol. 61, no. 3, pp. 842–847, Feb. 2001.

[30] L. M. Liau, R. M. Prins, S. M. Kiertscher, S. K. Odesa, T. J. Kremen, A. J. Giovannone, J. Lin, D. J. Chute, P. S. Mischel, T. F. Cloughesy, and M. D. Roth, “Dendritic Cell Vaccination in Glioblastoma Patients Induces Systemic and Intracranial T-cell Responses Modulated by the Local Central Nervous System Tumor Microenvironment Dendritic Cell Vaccination in Glioblastoma Patients Induces Systemic Nervous System,” *Clin. Cancer Res.*, vol. 11, no. 15, pp. 5515–5525, Aug. 2005.

[31] H. Okada, P. Kalinski, R. Ueda, A. Hoji, G. Kohanbash, T. E. Donegan, A. H. Mintz, J. a Engh, D. L. Bartlett, C. K. Brown, H. Zeh, M. P. Holtzman, T. a Reinhart, T. L. Whiteside, L. H. Butterfield, R. L. Hamilton, D. M. Potter, I. F. Pollack, A. M. Salazar, and F. S. Lieberman, “Induction of CD8+ T-cell responses against novel glioma-associated antigen peptides and clinical activity by vaccinations with {alpha}-type 1 polarized dendritic cells and polyinosinic-polycytidylic acid stabilized by lysine and carboxymethylcellulose in patients with recurrent malignant glioma,” *J. Clin. Oncol.*, vol. 29, no. 3, pp. 330–6, Jan. 2011.

[32] K. Iwami, S. Shimato, M. Ohno, H. Okada, N. Nakahara, Y. Sato, J. Yoshida, S. Suzuki, H. Nishikawa, H. Shiku, A. Natsume, and T. Wakabayashi, “Peptide-pulsed dendritic cell vaccination targeting interleukin-13 receptor α2 chain in recurrent malignant glioma patients with HLA-A*24/A*02 allele,” *Cytotherapy*, vol. 14, no. 6, pp. 733–42, Jul. 2012.

[33] Y. Akiyama, C. Oshita, A. Kume, A. Iizuka, H. Miyata, M. Komiyama, T. Ashizawa, M. Yagoto, Y. Abe, K. Mitsuya, R. Watanabe, T. Sugino, K. Yamaguchi, and Y. Nakasu, “α-type-1 polarized dendritic cell-based vaccination in recurrent high-grade glioma: a phase I clinical trial,” *BMC Cancer*, vol. 12, no. 1, p. 623, Jan. 2012.

[34] S. Phuphanich, C. J. Wheeler, J. D. Rudnick, M. Mazer, H. Wang, M. a Nuño, J. E. Richardson, X. Fan, J. Ji, R. M. Chu, J. G. Bender, E. S. Hawkins, C. G. Patil, K. L. Black, and J. S. Yu, “Phase I trial of a multi-epitope-pulsed dendritic cell vaccine for patients with newly diagnosed glioblastoma,” *Cancer Immunol. Immunother.*, vol. 62, no. 1, pp. 125–35, Jan. 2013.

[35] D. a Caruso, L. M. Orme, A. M. Neale, F. J. Radcliff, G. M. Amor, W. Maixner, P. Downie, T. E. Hassall, M. L. K. Tang, and D. M. Ashley, “Results of a phase 1 study utilizing monocyte-derived dendritic cells pulsed with tumor RNA in children and young adults with brain cancer,” *Neuro. Oncol.*, vol. 6, no. 3, pp. 236–46, Jul. 2004.

[36] E. O. Vik-Mo, M. Nyakas, B. V. Mikkelsen, M. C. Moe, P. Due-Tønnesen, E. M. I. Suso, S. Sæbøe-Larssen, C. Sandberg, J. E. Brinchmann, E. Helseth, A.-M. Rasmussen, K. Lote, S. Aamdal, G. Gaudernack, G. Kvalheim, and I. a Langmoen, “Therapeutic vaccination against autologous cancer stem cells with mRNA-transfected dendritic cells in patients with glioblastoma,” *Cancer Immunol. Immunother.*, vol. 62, no. 9, pp. 1499–509, Sep. 2013.

[37] G. J. Adema, J. M. de Vries, C. J. Punt, and C. G. Figdor, “Migration of dendritic cell based cancer vaccines : in vivo veritas ?” *Curr. Opin. Immunol.*, vol. 17, pp. 170–174, Apr. 2005.

[38] S. R. Clarke, “The critical role of CD40 / CD40L in the CD4-dependent generation of CD8 + T cell immunity,” *J. Leukoc. Biol.*, vol. 67, no. 5, pp. 607–14, May 2000.

[39] K. L. Rock, S. Gamble, and L. Rothstein, “Presentation of Exogenous Antigen with Class I Major Histocompatibility Complex Molecules,” *Science .*, vol. 249, no. 4971, pp. 918–921, Aug. 1990.

[40] K. L. Rock and K. Clark, “Analysis of the Role of MHC Class II Presentation in the Stimulation of Cytotoxic T Lymphocytes by antigens targeted into the exogenous antige-MHC class I presentation pathway,” *J. Immunol.*, vol. 156, pp. 3721–3726, May 1996.

[41] K. M. Dhodapkar, B. Cirignano, F. Chamian, D. Zagzag, D. C. Miller, J. L. Finlay, and R. M. Steinman, “Invariant natural killer T cells are preserved in patients with glioma and exhibit antitumor lytic activity following dendritic cell-mediated expansion,” *Int. J. cancer*, vol. 109, no. 6, pp. 893–9, May 2004.

[42] H. Ogbomo, J. Cinatl, C. H. Mody, and P. a Forsyth, “Immunotherapy in gliomas: limitations and potential of natural killer (NK) cell therapy,” *Trends Mol. Med.*, vol. 17, no. 8, pp. 433–41, Aug. 2011.

[43] C. H. M. J. Van Elssen, T. Oth, W. T. V Germeraad, G. M. J. Bos, and J. Vanderlocht, “Natural Killer Cells : The Secret Weapon in Dendritic Cell Vaccination Strategies Natural Killer Cells : The Secret Weapon in Dendritic Cell,” *Clin. Cancer Res.*, vol. 20, no. 5, pp. 1095–103, Mar. 2014.

[44] S. De Vleeschouwer, F. Van Calenbergh, P. Demaerel, P. Flamen, S. Rutkowski, E. Kaempgen, J. E. Wolff, C. Plets, R. Sciot, and S. W. Van Gool, “Transient local response and persistent tumor control in a child with recurrent malignant glioma: treatment with combination therapy including dendritic cell therapy. Case report,” *J. Neurosurg.*, vol. 100, no. 5 Suppl Pediatrics, pp. 492–7, May 2004.

[45] S. De Vleeschouwer, H. Ardon, F. Van Calenbergh, R. Sciot, G. Wilms, J. van Loon, J. Goffin, and S. Van Gool, “Stratification according to HGG-IMMUNO RPA model predicts outcome in a large group of patients with relapsed malignant glioma treated by adjuvant postoperative dendritic cell vaccination,” *Cancer Immunol. Immunother.*, vol. 61, no. 11, pp. 2105–12, Nov. 2012.

[46] H. Ardon, S. De Vleeschouwer, F. Van Calenbergh, L. Claes, C. M. Kramm, S. Rutkowski, J. E. A. Wolff, and S. W. Van Gool, “Adjuvant Dendritic Cell-Based Tumour Vaccination for Children With Malignant Brain Tumours,” *Pediatr. Blood Cancer*, vol. 54, no. 4, pp. 519–525, Apr. 2010.

[47] M. Weller, T. Cloughesy, J. R. Perry, and W. Wick, “Standards of care for treatment of recurrent glioblastoma - are we there yet?,” *Neuro. Oncol.*, vol. 15, no. 1, pp. 4–27, Jan. 2013.

[48] E. Wong, S. Gautam, C. Malchow, M. Lun, E. Pan, and S. Brem, “Bevacizumab for recurrent glioblastoma multiforme: a meta-analysis,” *J. Natl. Compr. Cancer Netw.*, vol. 9, no. 4, pp. 403–407, Apr. 2012.

[49] W. Asavaroengchai, Y. Kotera, and J. J. Mulé, “Tumor lysate-pulsed dendritic cells can elicit an effective antitumor immune response during early lymphoid recovery,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 99, no. 2, pp. 931–6, Jan. 2002.

[50] J. H. Sampson, K. D. Aldape, G. E. Archer, A. Coan, A. Desjardins, A. H. Friedman, H. S. Friedman, M. R. Gilbert, J. E. Herndon, R. E. Mclendon, D. A. Mitchell, D. A. Reardon, R. Sawaya, R. Schmittling, W. Shi, J. J. Vredenburgh, D. D. Bigner, A. B. Heimberger, D. Neurosurgery, and R. Sc, “Greater chemotherapy-induced lymphopenia enhances tumor-specific immune responses that eliminate EGFRvIII-expressing tumor cells in patients with glioblastoma,” *Neuro. Oncol.*, vol. 13, no. 3, pp. 324–333, Mar. 2011.

[51] C. J. Wheeler, A. Das, G. Liu, J. S. Yu, and K. L. Black, “Clinical Responsiveness of Glioblastoma Multiforme to Chemotherapy after Vaccination Clinical Responsiveness of Glioblastoma Multiforme to Chemotherapy after Vaccination,” *Clin. Cancer Res.*, vol. 10, no. 16, pp. 5316–5326, Aug. 2004.

[52] T.-G. Kim, C.-H. Kim, J.-S. Park, S.-D. Park, C. K. Kim, D.-S. Chung, and Y.-K. Hong, “Immunological factors relating to the antitumor effect of temozolomide chemoimmunotherapy in a murine glioma model,” Clin. vaccine Immunol., vol. 17, no. 1, pp. 143–53, Jan. 2010.

[53] M. H. Andersen, R. B. Sørensen, D. Schrama, I. M. Svane, J. C. Becker, and P. Thor Straten, “Cancer treatment: the combination of vaccination with other therapies.,” Cancer Immunol. Immunother., vol. 57, no. 11, pp. 1735–43, Nov. 2008.

[54] K. L. Chaichana, P. Zadnik, J. D. Weingart, A. Olivi, G. L. Gallia, J. Blakeley, M. Lim, H. Brem, and A. Quiñones-Hinojosa, “Multiple

resections for patients with glioblastoma: prolonging survival," *J. Neurosurg.*, vol. 118, no. 4, pp. 812–20, Apr. 2013.

[55] M. Girndt, U. Sester, K. Harald, F. Hünger, and H. Köhler, "Glucocorticoids inhibit activation-dependent expression of costimulatory molecule B7-1 in human monocytes," *Transplantation*, vol. 66, no. 3, pp. 370–375, Aug. 1998.

[56] A. M. Sonabend, C. E. Rolle, and M. S. Lesniak, "The role of regulatory T cells in malignant glioma," *Anticancer Res.*, vol. 28, no. 2B, pp. 1143–50, Mar.-Apr. 2008.

[57] A. B. Heimberger, M. Abou-Ghazal, C. Reina-Ortiz, D. S. Yang, W. Sun, W. Qiao, N. Hiraoka, and G. N. Fuller, "Incidence and prognostic impact of FoxP3+ regulatory T cells in human gliomas," *Clin. Cancer Res.*, vol. 14, no. 16, pp. 5166–72, Aug. 2008.

[58] W. Maes, T. Verschuere, A. Van Hoylandt, L. Boon, and S. Van Gool, "Depletion of regulatory T cells in a mouse experimental glioma model through anti-CD25 treatment results in the infiltration of non-immunosuppressive myeloid cells in the brain," *Clin. Dev. Immunol.*, vol. 2013, p. 952469, Jan. 2013.

[59] F. Ghiringhelli, C. Menard, P.E. Puig, S. Ladoire, S. Roux, F. Martin, E. Solary, A. Le Cesne, L. Zitvogel and B. Chauffert, "Metronomic cyclophosphamide regimen selectively depletes CD4+CD25+ regulatory T cells and restores T and NK effector functions in end stage cancer patients," *Cancer Immunol Immunother.*, vol. 56, pp. 641–648, May 2007.

[60] A. C. C. da Fonseca and B. Badie, "Microglia and macrophages in malignant gliomas: recent discoveries and implications for promising therapies," Clin. Dev. Immunol., vol. 2013, p. 264124, Jan. 2013.

[61] S. V Kushchayev, Y. S. Kushchayeva, P. C. Wiener, B. Badie, and M. C. Preul, "Monocyte-Derived Cells of the Brain and Malignant Gliomas: The Double Face of Janus," *World Neurosurg.*, no. November, pp. 1–15, Nov. 2012.

[62] S. M. Pyonteck, L. Akkari, A. J. Schuhmacher, R. L. Bowman, L. Sevenich, D. F. Quail, O. C. Olson, M. L. Quick, J. T. Huse, V. Teijeiro, M. Setty, C. S. Leslie, Y. Oei, A. Pedraza, J. Zhang, C. W. Brennan, J. C. Sutton, E. C. Holland, D. Daniel, and J. a Joyce, "CSF-1R inhibition alters macrophage polarization and blocks glioma progression," *Nat. Med.*, vol. 19, no. 10, pp. 1264–72, Oct. 2013.

[63] N. A. Charles, E. C. Holland, R. Gilbertson, R. Glass, and H. Kettenmann, "The brain tumor microenvironment," *Glia*, vol. 59, no. 8, pp. 1169–80, Aug. 2011.

[64] B. M. E. Krummel and J. P. Allison, "CTLA-4 Engagement Inhibits Il-2 Accumulation and Cell Cycle Porgression upon Activation or Resting T Cells," *J. Exp. Med.*, vol. 183, no. 6, pp. 2533-2540, June 1996.

[65] F. S. Hodi, S. O'Day, M. W. McDermott, R. W. Weber, J. A. Sosman, and W. J. Urba, "Improved survival with ipilimumab in patients with metastatic melanoma," *N. Engl. J. Med.*, vol. 363, no. 8, pp. 711–723, Aug. 2010.

[66] Z. Belcaid, J. A. Phallen, J. Zeng, A. P. See, D. Mathios, C. Gottschalk, S. Nicholas, M. Kellett, J. Ruzevick, C. Jackson, E. Albesiano, N. M. Durham, X. Ye, P. T. Tran, B. Tyler, J. W. Wong, H. Brem, D. M. Pardoll, C. G. Drake, and M. Lim, "Focal Radiation Therapy Combined with 4-1BB Activation and CTLA-4 Blockade Yields Long-Term Survival and a Protective Antigen-Specific Memory Response in a Murine Glioma Model," *PLoS One*, vol. 9, no. 7, pp. 1–9, July 2014.

[67] J. D. Wolchok, A. Hoos, S. O'Day, J. S. Weber, O. Hamid, C. Lebbé, M. Maio, M. Binder, O. Bohnsack, G. Nichol, R. Humphrey, and F. S. Hodi, "Guidelines for the evaluation of immune therapy activity in solid tumors: immune-related response criteria," *Clin.Cancer Res.*, vol. 15, no. 23, pp. 7412–20, Dec. 2009.

Machine Learning Predictions of Cancer Driver Mutations*

E. Joseph Jordan and Ravi Radhakrishnan

***Abstract*—A method to predict the activation status of kinase domain mutations in cancer is presented. This method, which makes use of the machine learning technique support vector machines (SVM), has applications to cancer treatment, as well as numerous other diseases that involve kinase misregulation.**

I. Introduction

Cancer is an evolutionary disease whereby an heterogeneous population of cells acquire a fitness advantage over neighboring cells via such mechanisms as mutations, changes in expression levels, and epigenetic factors, among others. These changes allow cancerous cells to have altered phenotype relative to the parent cells that they derive from, including increased proliferation and invasiveness, as well as less susceptibility to apoptotic signals than non-cancerous cells [1, 2]. The continuing decline in the cost of genome sequencing, as well as the relative ease of interpreting the effects of mutations in many proteins via methods such as activity assays has led to a sustained drive to understand the effects of cancer derived mutations on cancer progression. The challenge of finding mechanistic links between mutations and cancer progression is made even more imperative by the fact that many cancer drugs target mutations that have specific effects, as well as the observation that many clinical trials fail due to patient cohorts that are not suitable for specific therapies [3]. Sequencing efforts as well as the frequent failure of targeted therapies has led to an increasingly well-recognized principle that not all mutations confer selective advantage on cancer cells. These mutations are known as passenger mutations while mutations that confer some advantage are commonly referred to as driver mutations, because they can be seen as driving cancer progression [4].

The growing understanding of the importance of mutations on cancer progression is reflected in the rapid increase in available sequencing data via repositories such as The Catalog of Somatic Mutations in Cancer (COSMIC) and The Cancer Genome Atlas (TCGA). Coterminal with this increase in data has been an increase in efforts to computationally assess the effects of these mutations. These methods generally seek to use machine learning on large datasets of variants that are known to be (or not to be) deleterious. Features that are common among one class of known variants or the other are then used to predict what will be the effect of new variants. One pioneering and still widely cited method to classify mutations called Sorts Intolerant From Tolerant (SIFT) [5]. This tool uses evolutionary sequence conservation to predict whether a single nucleotide polymorphism (SNP), not necessarily cancer related, will be deleterious on protein function. Since the overlap between mutations that have a deleterious effect on protein function and mutations that are cancer drivers may not be complete, this method may not be well suited to predicting the effect of cancer mutations, although many more recent methods have included the results of SIFT as one among many features use to classify mutations as well as used SIFT as a baseline for comparison of classification accuracy and sensitivity.

The most popular technique used to classify mutations that are specific to cancer is a method from machine learning known as support vector machines (SVM) [6]. This essentially geometric method seeks to find combinations of features that are common to mutations of different classes so that mutations of unknown class (i.e. driver or passenger) can be classified. SVM has been used to develop classifiers that predict whether mutations across the whole genome [7] as well as in a specific class of proteins, kinases [8]. These methods report that they make accurate predictions, based on cross-validation, and that they have a high receiver operating characteristic area under the curve (AUC), or the probability of distinguishing between examples of different classes. However, they both suffer from a problem that is seemingly common in the literature whereby they make a priori decisions about which mutations belong in which class, instead of letting the data lead the way. This should make the results of these and any similarly constructed classifiers suspect.

A brief comment on the selection of kinases as a class of protein to receive special attention is warranted. Kinase proteins have an important role in numerous cell signaling processes and have been seen to play an outsized role in human cancers. Indeed, recent work has shown that in COSMIC, kinase domain mutations account for over 20% of non-synonymous coding mutations in cancer, even though they account for less than 3% of protein coding genes in the human genome investigation [9]. Additionally, which kinase domain mutations are driver mutations can be determined by performing kinase activity assays or cell colony transformation assays [10]. These facts, combined with the fact that kinase proteins are the target of numerous drugs both in the clinic as well as in various stages of clinical trial, shows that this is a class of proteins worthy of detailed investigation [9].

* The research leading to these results has received funding from the European Commission grant FP7-ICT-2011-9-600841 (CHIC project) and National Institutes of Health Grant U01-EB016027. Computational resources were provided in part by the National Partnership for Advanced Computational Infrastructure under Grant No. MCB060006 from XSEDE.

E. J. Jordan, is with The University of Pennsylvania Biochemistry and Molecular Biophysics Graduate Group, PA 19104 USA. (e-mail: joea@mail.med.upenn.edu).

R. Radhakrishnan is with The University of Pennsylvania Departments of Bioengineering, Chemical and Biomolecular Engineering, Biochemistry and Biophysics, PA 19104 USA. (corresponding author phone:215-898-0487; fax: 215-573-2071; e-mail: rradhak@seas.upenn.edu).

II. METHODS

A. Kinase Dataset

The dataset was constructed via mining of uniprot and the literature for mutations in a subset of human kinases consisting of 468 kinase domain containing proteins. Left out were kinases that are uncharacterized in terms of kinase activity, several kinases that have large insertions that cause them to have unique structural features, and kinases that are known to be constitutively active. The dataset contains 135 mutations that cause kinase domain activation as well as 413 mutations that do not cause activation.

B. Support Vector Machines (SVM) Construction

The kinase dataset was converted into a feature vector using an in house perl script. The elements of the feature vector include: the wild type and mutant residue identity, the wild type and mutant residue chemistry (i.e. aliphatic, acidic, basic, polar, aromatic), change in Kyte-Doolittle hydropathy from wild type to mutant as well as change of mutant from average Kyte-Doolittle hydropathy at that position, similarly for free energy of solvation, change in van der Waals radius, change in charge, and change in polarity. Also included are the SIFT score, whether the mutation occurs in a functionally important region in the kinase domain such as the αC helix, the nucleotide binding loop, the activation loop, or the catalytic loop. Whether the loop if predicted to be a β sheet, α helix, or unstructured loop, and the predicted solvent accessibility as given by Rost and Sander [11]. Finally, the proportion of a specific mutant occurring at a specific location relative to the total number of mutants in the whole kinase domain of the protein that a mutation occurs in was also included as a measure of oncogenicity. These features were encoded in a form readable by the program SVM^{perf}, which was used for all model creation and validation. The tradeoff between training error and margin (the parameter *c* in SVM^{perf}) was set to 0.25, a 1-slack dual structural learning algorithm was used (parameter *w 3*), a linear classifier without a bias term was used (parameters *t 0* and *b 0* respectively), and a loss function which maximized the area under the ROC curve was used (parameter *l 10*). This set of parameters gave the best outcome on cross-validation.

C. Model Validation

Performance was evaluated via cross-validation. The datasets used for cross-validation were always balanced, as very imbalanced datasets used for training purposes can lead to classifiers which predict the dominant class too often. In order to obtain a balanced dataset from the strongly imbalanced dataset described above, the number of members of minority class, in this case activating mutations, was used to determine the number of members in the majority class for training and testing. This was achieved by taking all members of the minority class and randomly selecting the same number of members from the majority class. Cross-validation was performed by taking 75% of the balanced dataset and using it for model training while using the remaining 25% of the data for testing. This was performed 300 times to account for the fact that data is left out of the balanced training/testing set. For the purposes of evaluating the results, the following abbreviations are used: true positives (TP), true negatives (TN), false positives (FP), false negatives (FN). Here, positive is taken as kinase activating and negative is taken as kinase non-activating As well, the following definitions are useful.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

$$\text{Sensitivity} = \frac{TP}{TP+FN}$$

$$\text{Precision} = \frac{TP}{TP+FP}$$

In words, accuracy is the percentage of predictions which are correct, sensitivity is the percentage of activating mutations predicted correctly relative to the total number of activating mutations, and precision is the percentage of activating mutations predicted correctly relative to the total number of activating mutations predicted. One more measure is worth mentioning here, the AUC, which as noted in the introduction is the probability that given an activating mutations and a non-activating mutation, the classifier will correctly identify which is which. All cross-validation was done with an in house perl script. As a final test of the method, the results were compared to those of a more mechanistically detailed, and computationally expensive, study of several mutations in the kinase domain of the anaplastic lymphoma kinase (ALK) [10].

D. Ranking Feature Vectors

One potentially important result of this study is which features are most useful for classifier performance. In this study, we use the computationally simple method of comparing differences in feature vector means. The mean and standard deviation of each class is used to determine the distance between distributions based on the formula $|m_1-m_2|/(s_1+s_2)$, where m_1 and m_2 are the sample means and s_1 and s_2 are the sample standard deviations for the class, and the elements in the feature vector have been normalized to lie in [-1,1]. Features can then be ranked based on this metric.

A method known as kernel density estimation can be used to determine non-parametric probability distributions for the two classes of each feature. This method is akin to histograms but does not suffer from issues related to selecting bin width or origin. Kernel density estimates were constructed using linear kernels and a width parameter determined by cross-validation via the python package *sklearn*. Visualization was performed with the python package *matplotlib*.

III. RESULTS AND DISCUSSION

A. Classifier Performance

As seen in Table I, the method is fairly accurate and precise, though sensitivity is somewhat lacking. The area under the curve is also quite good. Table I also shows the importance of using a balanced dataset, as there is no metric on which the whole dataset does better than a balanced one. The chief reason that the model (which uses the whole dataset) does poorly on sensitivity and precision is that it too frequently predicts that a mutation is a member of the dominant class. The model was also constructed using an alternative package, a python implementation of *libsvm*, which allow the user to set a parameter to control for imbalanced datasets. Extensive testing of parameters using

this alternate setup did not produce results that were commensurate with those produced by the balanced dataset in SVM^{perf} (results not shown). It remains to be seen if better results could be obtained via the use of a sampling method that synthetically generates examples of the minority class to arrive a balanced dataset such as the synthetic minority oversampling technique (SMOTE) [12]. This remains a future direction.

TABLE I. SVM PERFORMANCE METRICS

	DATASET	
Performance Metric	***Balanced***	***Whole***
Accuracy	78%	78%
Sensitivy	67%	44%
Precision	81%	56%
AUC	91%	82%

B. *Feature Rankings*

A listing of the top six ranked features is given in table II. These six features show the largest differences in sample mean, together accounting for almost half of the cumulative difference of sample mean across all features.

TABLE II. TOP SIX RANKED SVM FEATURES

Feature Name	$\frac{\lvert m_1 - m_2 \rvert}{s_1 + s_2}$
Mutant A	0.144
Wild Type K	0.128
Kyte-Doolittle Hydropathy	0.089
Polarity Difference	0.084
Mutant Acidic	0.078
Wild Type Basic	0.075

The largest difference in sample means is found to have alanine as the mutant residue. This is depicted in Figure 1A. The likely reason that the ranked list is topped by having the mutant residue be an alanine is that many of the mutations found in uniprot are the product of alanine scanning mutagenesis. This should make us wary of making assumptions on the character of cancer driver mutations without a closer look at the data and how it is generated. Given the cost of making kinase domain mutants and performing activity assays or cell transformation assays, the reliance on datasets with some systematic bias such as uniprot will continue for the foreseeable future. The second, fifth, and sixth ranked features are all related and tell an interesting story. These features are all related in that they involve a charged residue. This is significant as kinase domains in the inactive state tend to be stabilized by salt bridges that are broken in the active state, which in some cases has its own stabilizing salt bridges. Mutations that result in charge loss or reversal could easily disrupt the delicate balance that holds the kinase domain in the inactive conformation, or alternatively could add stabilizing interactions to the active state. Either of these could result in constitutive kinase activation and aberrant signaling. Interestingly, this result is also in line with a recent study of kinase domain mutations in the COSMIC database. This study showed that many of the most frequently observed driver mutations in kinase domains are the result of loss or reversal of charge and concomitant inactive state destabilization [9]. The distribution of mutations by class for which the wild type residue is lysine is shown in figure 1B. The third and fourth ranked features, shown in figure 1C and 1D, are also related and may offer their own unique insight into which kinase domain mutations are likely to be drivers. These features both are related to polarity. Large changes in the polarity of even one residue can have a destabilizing effect on an entire protein by causing reorganization of a hydrophobic core. This reorganization could easily be responsible for a shift in population from inactive to active in a kinase domain. Figure 1C and 1D show that activating mutations tend to increase hydropathy scores and decrease polarity. Both of these can result in decreased potential for hydrogen bonding or salt bridge formation. This is yet further evidence that kinase domain activating mutations tend to destabilize the interactions that maintain the inactive conformation.

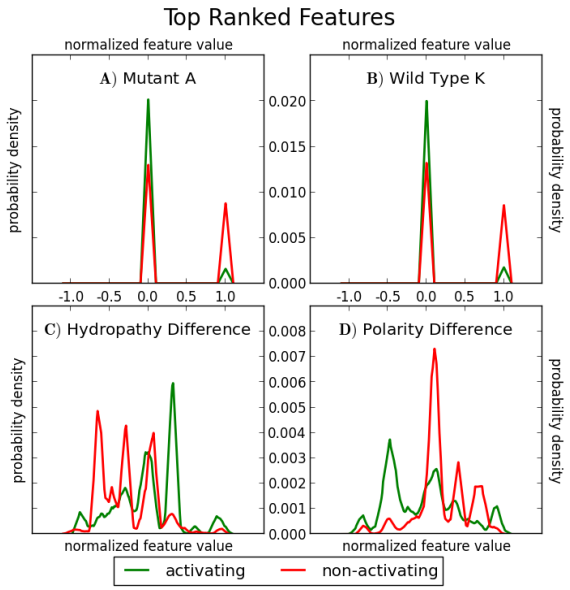


Figure 1. Kernel density estimates for the top ranked SVM features as determined by difference in mean feature value for each class.

C. *Comparison to Mechanistic Methods of Mutation Classification*

One recent study of kinase domain mutations involved performing molecular dynamics (MD) simulations on a series of 22 clinically observed mutants of ALK. This study used various criteria to assess whether mutants were destabilizing the inactive state of ALK, such as changes in hydrogen bonding, solvent accessible surface area, and proximity to conserved kinase structural motifs such as activation loop or αC helix [10]. A comparison of the results of the SVM classifier outlined above and the MD classifier of Bressler et al. [10] is given in Table III.

TABLE III. COMPARISON TO MD BASED CLASSIFIER

	Method	
Contingency	***SVM***	***MD***
TP	10	11
TN	6	6
FP	3	3
FN	3	2

As is made clear by Table III, these methods both do quite well on predicting kinase domain activating mutations, with the MD method doing slightly better. What is striking about this table though is the computational costs associated with the two methods. The MD method takes thousands of hours of processing time on a scientific compute cluster to generate the MD trajectory, and some additional time to run analysis scripts. This must be done for each mutation. In contrast, the SVM method takes a few minutes of compute time on a commodity desktop to generate the feature vector, and then a few seconds to train the model and test the mutations. Although the MD method offers mechanistic insight that cannot be gained from the SVM method, the speedup of several orders of magnitude seems to point to SVM as the more efficient model, even if it is slightly less accurate. Moreover, as outlined below, a combination of the two approaches can yield an optimal (accurate as well as fast) strategy for profiling the effects of mutations.

IV. CONCLUSION

Here we have outlined a method to predict whether clinically observed cancer mutations are driver mutations. The method, based on support vector machines was shown to be fairly reliable at predicting the effect of kinase domain mutations, especially when a balanced dataset is used. It is also faster than recent efforts to use MD to predict the effect of kinase domain mutations, and does not make any *a priori* decisions about which mutations are driver mutations, as many recent machine learning efforts have. Interestingly, we find that ability to affect salt bridge formation can be an important factor in determining whether a given mutation is likely to be a driver. Similarly, mutations that change residue polarity, possibly leading to altered packing, are also important.

While the method outlined here is robust, it could still be improved upon. One class of features that are notably absent from this classifier is ones relating to the three dimensional structure and dynamics of a protein. Inclusion of this sort of information could result in even more reliable classifiers. One particular avenue of future research that may prove useful is the addition of structural information in the form of protein structure graphs. This could also be a way to include insight from MD simulations into the method, as features of protein structure graphs that vary between simulations of activating and non-activating mutations could be easily included in an SVM as long as the appropriate structural information is available (in the form of crystal or computationally generated structures of kinases). This goal is something that we are actively pursuing.

ACKNOWLEDGMENT

We thank G. S. Stamatakos, N. Graf, and members of the Radhakrishnan Laboratory for insightful discussions.

REFERENCES

[1] T. Tian, S. Olson, J. M. Whitacre, and A. Harding, "The origins of cancer robustness and evolvability," *Integr Biol (Camb)*, vol. 3, pp. 17-30, Jan. 2011.

[2] D. Hanahan and R. A. Weinberg, "Hallmarks of cancer: the next generation," *Cell*, vol. 144, pp. 646-74, Mar. 2011.

[3] N. Normanno, A. M. Rachiglio, C. Roma, F. Fenizia, C. Esposito, R. Pasquale, M. L. La Porta, A. Iannaccone, F. Micheli, M. Santangelo, F. Bergantino, S. Costantini, A. De Luca, "Molecular diagnostics and personalized medicine in oncology: challenges and opportunities," *J Cell Biochem*, vol. 114, pp. 514-24, Mar. 2013.

[4] B. Vogelstein, N. Papadopoulos, V. E. Velculescu, S. Zhou, L. A. Diaz, Jr., and K. W. Kinzler, "Cancer genome landscapes," *Science*, vol. 339, pp. 1546-58, Mar. 2013.

[5] P. C. Ng and S. Henikoff, "SIFT: Predicting amino acid changes that affect protein function," *Nucleic Acids Res*, vol. 31, pp. 3812-4, July 2003.

[6] V. N. Vapnik, *Narure of Statistical Learning Theory*: Springer, New York, NY, 1995.

[7] E. Capriotti and R. B. Altman, "A new disease-specific machine learning approach for the prediction of cancer-causing missense variants," *Genomics*, vol. 98, pp. 310-7, Oct. 2011.

[8] J. M. Izarzugaza, A. del Pozo, M. Vazquez, and A. Valencia, "Prioritization of pathogenic mutations in the protein kinase superfamily," *BMC Genomics*, vol. 13, suppl 4, p. S3, June 2012.

[9] E. J. Jordan and R. Radhakrishanan, "In silico profiling of activating mutations in cancer," *Integrative Biol.*, submitted for publication.

[10] S. Bressler, D. Weiser, P. J. Huwe, J. H. Park, K. Krytska, H. Ryles, M. Laudenslanger, E. F. Rappaport, A. C. Wood, P. W. McGrady, M. D. Hogarty, W. B. London, R. Radhakrishnan, M. A. Lemmon, Y. P. Mossé, "Integrative functional assessment of ALK mutations for therapeutic stratification in neuroblastoma," *Cancer Cell*, vol. 26, pp. 682-694, Nov. 2014.

[11] B. Rost and C. Sander, "Improved prediction of protein secondary structure by use of sequence profiles and neural networks," *Proc Natl Acad Sci U S A*, vol. 90, pp. 7558-62, Aug. 1993.

[12] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," *J. Artificial Intelligence Research*, vol. 16, pp. 321-357, June 2002.

Simulating Tumour Vasculature at Multiple Scales*

J. A. Grogan, P. K. Maini, J. Pitt-Francis and H. M. Byrne

***Abstract*— The vasculature plays an important role in tumour development and treatment, acting as a conduit for nutrients, waste products and therapeutics. Simulating transport and network structure evolution in malformed tumour networks is a challenging multi-scale problem. Current approaches for modelling the vasculature at distinct size scales are described here, followed by a discussion of current efforts in developing integrated multi-scale modelling approaches for simulating the growth and treatment of vascular tumours.**

I. Introduction

The vessel network transports nutrients such as glucose and oxygen to tissues and provides a mechanism for waste product removal. In tumours the vasculature can become dysfunctional, losing its hierarchical structure and increasing in permeability. The resulting poor perfusion of tumour tissue can lead to hypoxia, which is associated with malignancy [1].

Current therapeutic strategies rely on tumour tissue perfusion, either directly for chemotherapeutic drugs to reach diseased regions or indirectly due to the dependence of radiotherapy effectiveness on tissue oxygenation. Normalization strategies, which improve the transport efficiency of the tumour vessel network prior to subsequent treatments, are showing promise in the clinic [2].

One of the challenges when adopting individual or combined therapies is predicting suitable dosings and timings for individual patients. Particularly in the case of combined therapies there may be periods of time during which synergistic effects can be exploited [3-4]. Imaging forms an important role in identifying such time periods, however the spatial resolution of functional imaging is relatively coarse, meaning information about the state of the vasculature can be limited.

Simulation of transport and network evolution in the vasculature in response to tumour cells or the application of normalization therapies allows more detailed predictions of suitable dosing strategies for patients. This is particularly the case when simulations are closely linked with functional imaging data, such as with PET and fMRI [3]. One of the challenges in developing simulations which can utilise the information given by functional imaging is that the length scales are significantly larger than those associated with individual cells or small capillaries. Since it is at this scale that the physical mechanisms used in the simulations are characterized, multi-scale approaches are required to link imaging and cell scales. This work is focused on the development of such approaches.

We first briefly overview existing macro-scale (imaging or tissue scale) and micro-scale (cell and capillary scale) modelling strategies. Efforts in developing multi-scale modelling approaches are then discussed, in particular those being developed as part of the Computational Horizons in Cancer (CHIC) Project.

II. MICRO-SCALE VASCULATURE MODELLING

The diameters of vessels in capillary beds are on the order of 10 μm, similar to the size of most surrounding cells [5]. It is within this micro-vasculature that species exchange primarily takes place in tissue. At this size scale individual cells and vessels can be modelled as discrete entities.

Discrete modelling approaches for cells at this scale typically involve representing them using cellular automata [4] or lattice free approaches, with the latter being able to account for mechanical interactions [6]. Since cells are modelled individually there is scope for including sub-cellular details such as cell-cycle and signalling pathway models [4] and for characterising cell behaviours based on mutation status. This allows detailed predictions of how specific mutations, or therapies targeted at cells at specific stages in their cycle, affect tumour development [7].

For the vessels, a typical approach at this scale is to model them as 1-D pipes with blood flow within the system approximated as Hagen-Poiseuille flow [8]. Within this framework flow rates, red blood cell concentrations (haematocrit) and wall shear stresses can be predicted in individual capillaries. Since individual vessels are being modelled it is possible to predict the influence of network topology on tumour development.

The vessels and surrounding cells form an interdependent system, which is relatively straightforward to model at the micro-scale. In a general sense, cells require nutrients from the vasculature. If they are deficient their cycle can be altered and they release angiogenic factors that stimulate new vessel formation. Through modelling the transport of nutrients and growth factors at the micro-scale it is possible to couple the vessel and cell layers [4].

The vessel network adapts according to the metabolic needs of the surrounding tissue, perfusion or in response to therapeutics [9]. This can be done through the change of vessel diameters, the regression of existing vessels or the development of new vessels. New vessels can form through angiogenesis and vasculogenesis. At the micro-scale these processes can be modelled explicitly [4]. Angiogenesis can be stimulated by growth factors secreted by individual hypoxic cells. Individual sprouts can be tracked as endothelial tip cells respond to mechanical and chemical gradients in their environment. Micro-scale modelling of this type gives useful

*This work has received funding from the European Union's Seventh Programme for research, technological development and demonstration under grant agreement No [600841] – CHIC project.

J.A.G., H.M.B. and P.K.M. are with the Wolfson Centre for Mathermatical Biology (WCMB), Mathematical Institute, University of Oxford, UK (corresponding author phone: +44-1865-615162; fax: +44-1865-273583; e-mail: james.grogan@maths.ox.ac.uk).

H.M.B and J.P.F. are with the Computational Biology Group, Dept. of Computer Science, University of Oxford, UK.

predictions of tumour behaviour in a confined volume [10], however it is infeasible to apply the same approach when predicting the behaviour of an entire tumour or when comparing predictions with clinical imaging, both of which operate on size scales on the order of 10 mm or more.

III. Macro-scale vasculature modelling

At the size scale of a whole tumour or at the resolution provided by current functional imaging technology it is not practical to model individual capillaries and cells. Instead a number of approaches exist in which the vasculature is described in a continuum sense [11-12]. In [11] the vasculature is treated as a diffusible species within a continuum reaction-diffusion framework. The spreading of the vasculature is captured through a diffusion term and the proliferation is captured through a source term dependent on the local concentration of angiogenic factor. Interaction with cells is mediated by allowing the rate at which cells become hypoxic to depend on local vessel concentration. A limitation of this approach is that the mechanics of the spreading tumour and phases are neglected, however for applications in modelling diffuse tumours such as glioblastoma this may not be significant. Due to the macro-scale nature of this approach it is natural to combine it with functional imaging data [13].

In [12] multi-phase continuum mechanics approaches are used to model the vasculature and tumour. In this approach the constituents of the tumour are divided into distinct spatial regions with uniform properties (phases) representing the vasculature, healthy cells, cancer cells and extra-cellular matrix. A volume of space may contain multiple phases, which are described through their relative volume fractions and which can interact mechanically and chemically. The interdependence of vascular and cellular phases is treated in a similar manner to that in [11], however in addition to species transport these models also allow the mechanical environment of the tumour to be included by applying momentum balances to the individual phases. This is important given the relatively high interstitial pressures known to exist in typical tumour environments. This approach also allows simulation of tumour growth in confined environments.

A challenge in the development of tissue scale models is investigating the scenarios and size scales at which a continuum level description of a discrete vessel network is appropriate [14]. Given that current continuum descriptions do not account for hierarchy and heterogeneity in vessel networks their predictive capability remains to be established. A further challenge is the development of suitable rules at the tissue scale to account for complex structural changes in the network at the micro-scale, such as branching and anastamosis [15].

IV. Developing integrated multi-scale models

The CHIC project is focused on the development of integrated multi-scale cancer models (hypermodels) which can ultimately be used as a clinical tool. In this framework models developed at different scales (from molecular to compartment models) and by different research groups are linked to provide insight into clinical problems. An example of the planned framework is shown in Fig. 1, with the angiogenesis and vasculature component being developed in the present work highlighted.

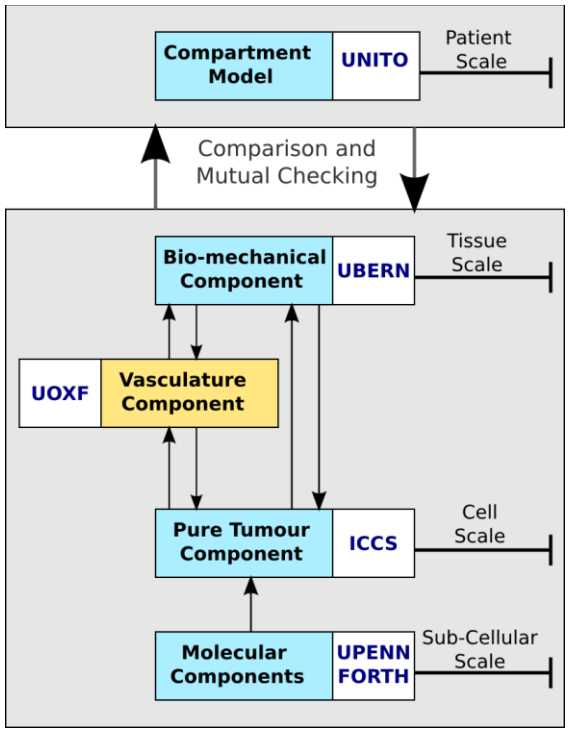


Figure 1. Schematic of the planned modelling framework for the CHIC project with the angiogenesis/vascular component highlighted. The project partners included here are the University of Oxford (UOXF), the Institute of Communications and Computer Systems (ICCS) Greece, the University of Bern (UBERN), the University of Turin (UNITO), the Foundation for Research and Technology, Helas (FORTH), and the University of Pennsylvania (UPENN).

The development of an integrated model of this type brings new challenges. In addition to the need to link tissue and cell scale vasculature models, each vasculature model must also link with models representing different aspects of the tumour environment. In particular, as shown in Fig. 1, the vasculature modelling component communicates with a macro-scale bio-mechanical component, currently being developed at UBERN and a micro-scale tumour growth component being developed at ICCS. Addressing the challenge of linking across size scales and model domains will require the development of new theoretical and computational strategies. Two approaches currently being considered are now described.

In the first approach, shown in Fig. 2 the tumour and its environment are divided into three domains. Ω_B is the domain of tissue surrounding the diseased region and is modelled as an isotropic elastic material using the model of UBERN. Ω_o is the domain of diseased tissue and is modelled using the previously described multiphase-fluid model of Hubbard and Byrne [12], implemented by UOXF. To realise this implementation careful consideration of the interface $\partial\Omega_{OB}$ is required. The macro-scale multi-phase model and micro-scale discrete model are linked by discretising the macro-scale model into subdomains Ω_I. Within each subdomain the macro-scale model provides an average vascular oxygenation and growth factor uptake rate while the micro-scale model provides summed cellular oxygen uptake and growth factor release rates for a unified time step.

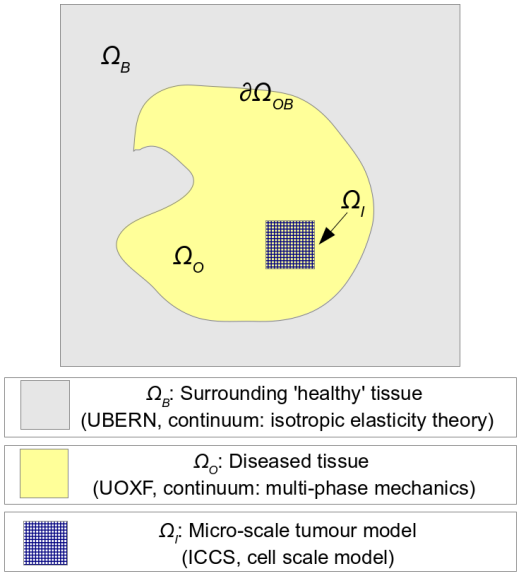


Figure 2. Schematic showing how to interface the macro-scale vasculature model with the bio-mechanical and cell scale models.

In the second approach, shown in Fig. 3, the tumour and its environment are again divided into three domains. In this case the macro-scale mechanics of both the tumour and the environs are modelled using the macro-scale bio-mechanical model of UBERN. The macro-scale model calculates the pressure field in the tumour based on the local density of cells, given by the micro-scale cell model of ICCS. This approach has previously been described in [16]. Within this framework a micro-scale model of the vasculature developed at UOXF, based on [4], is used to provide spatial oxygen and growth factor sink and source information to the ICCS cell model, as well as evolving constraints on cell movement. In turn the cell model supplies growth factor distributions to the vessel model, while the macro-scale model provides an average interstitial pressure. This pressure can be used to predict vessel collapse.

Another important aspect of integration when developing models of the type proposed in the CHIC project is the incorporation of experimental data. The vasculature component of the model is being developed in tandem with real-time micro-scale imaging studies of vessel network development in diseased mice. This allows validation of individual components of the model which has in the past not been possible, for example blood flow rate and haematocrit distribution predictions in tumour vessel networks. Combining modelling and experimental imaging in this manner also facilitates the assessment of the overall predictive capabilities of the model with reference to a relatively controlled tumour environment. For example, available imaging data allows for side-by-side comparisons of model predictions of tumour growth and network development with experimental images right from the time of injection of the initial tumour cells through to the development of a malformed tumour vasculature.

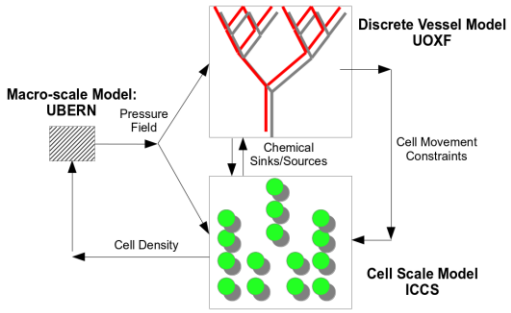


Figure 3. Schematic showing how to interface the macro-scale bio-mechanical model with the discrete vessel and cell models.

When integrating the experimental data in the modelling framework care must be taken to ensure that the effects of per-processing operations, performed on the images in order to make them suitable for modelling, are quantified relative to model predictions. In this sense it is useful to explicitly integrate image analysis and processing directly within the modelling framework. This is currently being explored for the micro-scale modelling approach, as shown in Fig. 4. Subsequent work will focus on the incorporation of macro-scale imaging data such as CT and functional imaging.

V. CONCLUSIONS

The difference in length scales between clinical imaging data and that of individual cells and capillaries makes predicting suitable therapeutic strategies challenging. Computer simulations can aid in resolving this difference, however challenges remain in linking discrete and continuum representations of vessel networks and the surrounding tissue. These challenges are being investigated as part of the modelling component of the CHIC project with the aim of facilitating the integration of different models at different length scales and over different domains.

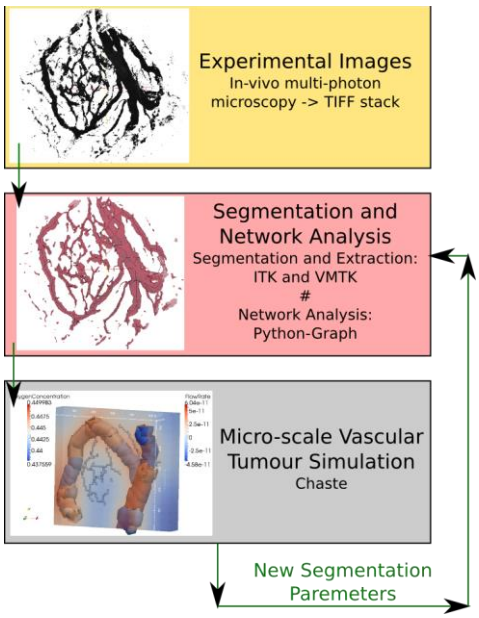


Figure 4. Integrating image analysis and processing with modelling, a sample workflow.

REFERENCES

[1] M. Hockel and P. Vaupel, "Tumor hypoxia: Definitions and current clinical, biologic, and molecular aspects", *J. Natl. Cancer Inst*., vol. 93, no. 4, pp. 266-276, Feb. 2001.

[2] T. T. Batchelor, E. R. Gerstner, K. E. Emblem, D. G. Duda, J. Kalpathy-Cramer, M. Snuderl, M. Ancukiewicz, P. Polaskova, M. C. Pinho, D. Jennings, S. R. Plotkin, A. S. Chi, A. F. Eichler, J. Dietrich, F. H. Hochberg, C. Lu-Emerson, A. J. Iafrate, S. P. Ivy, B. R. Rosen, J. S. Loeffler, P. Y. Wen, A. G. Sorensen, and R. K. Jain, "Improved tumor oxygenation and survival in glioblastoma patients who show increased blood perfusion after cediranib and chemoradiation", *Proc. Natl. Acad. Sci*., vol. 110, no. 47, pp. 19059-19064, Nov. 2013.

[3] A. L. Baldock, R. C. Rockne, A. D. Neal, A. Hawkins-Daarud, D .M. Corwin, C. A. Bridge, L. A. Guyman, A. D. Trister, M. M. Mrugala, J. K. Rockhill, and K. R. Swanson, "From patient-specific mathematical neuro-oncology to precision medicine", Front. Oncol., vol. 3, doi:10.3389/fonc.2013.00062, 2013.

[4] M. R. Owen, I. J. Stamper, M. Muthana, G. W. Richardson, J. Dobson, C. E. Lewis, and H. M. Byrne, "Mathematical modelling predicts synergistic antitumor effects of combining a macrophage-based, hypoxia-targeted gene therapy with chemotherapy", *Cancer Res*., vol. 71, no. 8, pp. 2826-37, Apr. 2011.

[5] R. K. Jain, "Determinants of tumor blood flow: A review", *Cancer Res*., vol. 48, no. 10, pp. 2641-58, May 1988.

[6] P. Macklin, M. E. Edgerton, A. M. Thompson, and V. Cristini, "Patient-calibrated agent-based modelling of ductal carcinoma in situ (DCIS): From microscopic measurements to macroscopic predictions of clinical progression", *J. Theor. Biol*., vol. 301, pp. 122-40, May 2012.

[7] G. G. Powathil, D. J. A Adamson, and M. A. J. Chaplain, "Towards predicting the response of a solid tumour to chemotherapy and radiotherapy treatments: Clinical insights from a computational model", *PloS Comput. Biol*., vol. 9, no. 7, doi: 10.1371/journal.pcbi.1003120, 2013.

[8] A. R. Pries, T. W. Secomb, T. Gessner, M. B. Sperandio, J. F. Gross, and P. Gaehtgens, "Resistance to blood flow in microvessels in vivo," *Circ. Res*., vol. 75, no.5, pp. 904-915, Nov. 1994.

[9] T. W. Secomb, J. P. Alberding, R. Hsu, M. W. Dewhirst ,and A. R. Pries, "Angiogenesis: An adaptive dynamic biological patterning problem," *PloS Comput. Biol*, vol. 9, no. 3, doi: 10.1371/journal.pcbi.1002983, Mar. 2013.

[10] H. Perfahl, H. M. Byrne, T. Chen, V. Estrella, T. Alarcon, A. Lapin, R. Gatenby, R. Gillies, M. C. Lloyd, P. K. Maini, M. Reuss, and M. R. Owen, "Multiscale modelling of vascular tumour growth in 3D: The roles of domain size and boundary conditions," *PloS ONE*, vol. 6, doi: 10.1371/journal.pone.0014790, Apr. 2011.

[11] K. R. Swanson, R. C. Rockne, J. Claridge, M. A. Chaplain, E. C. Alvord Jr., and A. R. A. Anderson, "Quantifying the role of angiogenesis in malignant progression of gliomas: In silico modeling integrates imaging and histology", *Cancer Res*., vol. 71, no. 24, pp. 7366-75, Dec. 2011.

[12] M. E. Hubbard and H. M. Byrne, "Multiphase modelling of vascular tumour growth in two spatial dimensions," *J. Theor. Biol*., vol. 316, pp. 70-89, Jan. 2013.

[13] S. Gu, G. Chakraborty, K. Champley, A. M. Alessio, J. Claridge, R. Rockne, M. Muzi, K. A. Krohn, A. M. Spence, E. C. Alvord, A. R. Anderson, P. E. Kinahan, and K. R. Swanson, "Applying a patient-specific bio-mathematical model of glioma growth to develop virtual [18F]-FMISO-PET images", *Math. Med. Biol.*, vol. 29, no. 1, pp. 31-48, Mar. 2012.

[14] R. J. Shipley and S. J. Chapman, "Multiscale modelling of fluid and drug transport in vascular tumours," *Bulletin Math. Biol.*, vol. 72, no. 6, pp. 1464-91, Aug. 2010.

[15] F. Spill, P. Guerrero, T. Alarcon, P. K. Maini, and H. M. Byrne, "Mesoscopic and continuum modelling of angiogenesis," *J. Math. Biol.*, doi:10.1007/s00285-014-0771-1, Mar. 2014.

[16] S. Bauer, C. May, D. Dionysiou, G. Stamatakos, P. Büchler, and M. Reyes, "Multiscale modeling for image analysis of brain tumor studies," *IEEE Trans. Biomed. Eng.*, vol. 59, no.1, pp. 25-29, Jan. 2012.

Modeling Glioblastoma Growth and Inhomogeneous Tumor Invasion with Explicitly Numerically Treated Neumann Boundary Conditions*

Stavroula G. Giatili and Georgios S. Stamatakos, *Member, IEEE*

***Abstract*—A couple of multiscale spatiotemporal simulation models of glioblastoma multiforme (GBM) growth and invasion into the surrounding normal brain tissue is presented. Both models are based on a continuous and subsequently finite mathematical approach centered around the non-linear partial differential equation of diffusion-reaction referring to glioma tumour cells. A novel explicit, strict and thorough numerical treatment of the three dimensional adiabatic Neumann boundary conditions imposed by the skull is also included in both models. The first model assumes a homogeneous representation of normal brain tissue whereas the second one, assuming an inhomogeneous representation of normal brain tissue, distinguishes between white matter, grey matter and cerebrospinal fluid. The predictions of the tumour doubling time by both models are compared for specific data sets. Clinical observational data regarding the range of the GBM doubling time values are utilized in order to ensure the realism of both models and their predictions. We assume that the inhomogeneous normal brain tissue representation is a virtual rendering of reality more credible than its homogeneous counterpart. The simulation results for the cases considered show that using the homogeneous normal brain based model may lead to an error of up to 10% for the first 25 simulated days in relation to the predictions of the inhomogeneous model. However, the error drops to less than 7% afterwards. This observation suggests that even by using a homogeneous brain based model and a realistic weighted average value of its diffusion coefficient, a rough but still informative estimate of the expected tumour doubling time can be achieved. Additional *in silico* experimentation aiming at statistically testing and eventually further supporting the validity of this hypothesis is in progress. It is noted that the values of the diffusion coefficients and the cell birth and death rates of the model are amenable to refinement and personalization by exploiting the histological and molecular profile of the patient. Work on this aspect is in progress.**

*This work has been supported in part by the European Commission under the projects p-Medicine: Personalized Medicine (FP7-ICT-2009.5.3-270089), CHIC: Computational Horizons in Cancer: Developing Meta- and Hyper-Multiscale Models and Repositories for *In Silico* Oncology (FP7-ICT-2011-9-600841), DR THERAPAT: The Digital Radiation Therapy Patient (FP7-ICT-2011-9-600852) and MyHealthAvatar: A Demonstration of 4D Digital Avatar Infrastructure for Access of Complete Patient Information (FP7-ICT-2011-9-600929).

Stavroula G. Giatili is with the *In Silico* Oncology and *In Silico* Medicine Group, Institute of Communication and Computer Systems, School of Electrical and Computer Engineering, National Technical University of Athens. (e-mail: giatili@mail.ntua.gr).

Georgios S. Stamatakos is with the *In Silico* Oncology and *In Silico* Medicine Group, Institute of Communication and Computer Systems, School of Electrical and Computer Engineering, National Technical University of Athens, 9, Iroon Polytechniou, Zografos, GR-157 80, Greece (corresponding author email: gestam@central.ntua.gr;phone: +302107722287; fax: +302107723557).

I. Introduction

Glioblastoma multiforme (GBM) is the most common and most aggressive type of primary brain tumour in humans with a notoriously poor prognosis. This is partly due to the highly invasive character of this glioma-type neoplasm. Tendrils of tumour extend onto the normal surrounding brain parenchyma rendering these tumours incurable by local therapy. Multiscale mathematical and computational modeling of glioblastoma growth and invasion using inter alia the tomographic data of the patient can considerably support an in depth quantitative understanding of the disease. Cancer modelling may also be exploited in order to support treatment planning - such as radiotherapeutic treatment planning - optimization in the patient individualized context. From a mathematical standpoint GBM growth and brain infiltration can be viewed as a boundary value problem strongly dependent on the values assigned to the physical boundaries of the definition domain. In this paper a continuous mathematics based spatiotemporal approach to model GBM growth and infiltration (invasion) into the surrounding normal brain tissue is briefly outlined. Special attention is paid to a novel explicit consideration of the adiabatic Neumann boundary conditions imposed by the highly complex structure of the skull. The non linear partial differential equation of diffusion-reaction referring to the generation and movement of tumour cells is numerically solved using the Crank-Nicolson finite difference – time domain technique. The paper builds on previous work published by the authors [1]. Two scenarios of modelling the human brain are considered: a homogeneous and an inhomogeneous brain representation. The predictions of the doubling time by both models are compared with each other. Suggestions regarding the implications of the corresponding limited error along with more general comments are provided.

II. Mathematical Treatment

GBM tumour growth and brain infiltration can be expressed by the following mathematical statement in textual form [2]: “rate of change of tumour cell population= diffusion (motility) of tumour cells + net proliferation of tumour cells - loss of tumour cells due to treatment.” If Ω is the brain domain, the previous statement can be symbolically formulated through the following differential equation:

$$\left\{\begin{array}{c}\frac{\partial c(\vec{x},t)}{\partial t}=\nabla\cdot\left(D(\vec{x})\,\nabla c(\vec{x},t)\right)+\rho\,c(\vec{x},t)-G(t)c(\vec{x},t)\ \text{ in }\Omega\\ c(\vec{x},0)=f(\vec{x}),\qquad\text{initial condition}\\ \hat{n}\cdot D(\vec{x})\,\nabla c(\vec{x},t)=0\ \ on\ \ \partial\Omega,\quad\text{Neumann boundary condition}\end{array}\right\} \tag{1}$$

The variable c denotes the tumour cell concentration at any spatial point defined by the position vector $\vec{x}$ and time t. The parameter D denotes the diffusion coefficient and represents the active motility of tumour cells. In the inhomogenous case three values of the parameter D are considered: D_g if $(\vec{x})$ belongs to grey matter, D_w if $(\vec{x})$ belongs to white matter and D_{CSF} if $(\vec{x})$ belongs to the cerebrospinal fluid (CSF). In the homogenous scenario D does not depend on position $\vec{x}$ provided that $\vec{x}$ is located within the intracranial space. The term ρ represents the net rate of tumour growth including tumour cell proliferation, loss and death, n̂ is the unit vector normal to the boundary $\partial\Omega$ of the domain and $f(\vec{x})$ is a known function that defines the initial spatial distribution of malignant cells. The term G(t) accounts for the temporal profile of treatment. In order to numerically apply the Neumann boundary condition at each node of the discretizing mesh (grid) that belongs to the boundary, a number of "fictitious nodes" equal to the number of the adjacent nodes that belong to the skull is considered [1]. An indicative case of numerically applying the boundary condition at the boundary point (x_i, y_j, z_k) in the negative x direction is the following:

$$-\left.\frac{\partial c}{\partial x}\right|_{(x_i,y_j,z_k)}=0\Rightarrow c_{i+1,j,k}=c_{F_{i-1,j,k}} \tag{2}$$

where $F_{i,j,k}$ denotes a fictitious node.

The number of the different cases considered in this work regarding nodes that possess boundary node(s) as their neighbour(s) is 26. Due to the high complexity of the biological system, the diffusion equation has to be solved numerically. By implementing the Crank - Nicolson scheme, 26 equations are produced. For the homogeneous scenario an indicative equation at the boundary grid point (x_i, y_j, z_k) where skull tissue is found only in the positive x direction is the following:

$$[1+6\lambda-\frac{\Delta t}{2}(\rho-G)]c_{i,j,k}^{t+1}-\lambda(2c_{i-1,j,k}^{t+1}+c_{i,j+1,k}^{t+1}+c_{i,j-1,k}^{t+1}+c_{i,j,k+1}^{t+1}+c_{i,j,k-1}^{t+1})=$$

$$[1-6\lambda+\tfrac{\Delta t}{2}(\rho-G)]c_{i,j,k}^{t}+\lambda(2c_{i-1,j,k}^{t}+c_{i,j+1,k}^{t}+c_{i,j-1,k}^{t}+c_{i,j,k+1}^{t}+c_{i,j,k-1}^{t}) \tag{3}$$

For the inhomogeneous scenario the corresponding equation is the following:

$$\left[1+6\lambda_{i,j,k}-\frac{\Delta t}{2}(\rho-G)\right]c_{i,j,k}^{t+1}-2\lambda_{i,j,k}c_{i-1,j,k}^{t+1}-$$

$$\left(\lambda_{i,j,k}+\frac{\lambda_{i,j+1,k}-\lambda_{i,j-1,k}}{4}\right)c_{i,j+1,k}^{t+1}-\left(\lambda_{i,j,k}-\frac{\lambda_{i,j+1,k}-\lambda_{i,j-1,k}}{4}\right)c_{i,j-1,k}^{t+1}-$$

$$\left(\lambda_{i,j,k}+\frac{\lambda_{i,j,k+1}-\lambda_{i,j,k-1}}{4}\right)c_{i,j,k+1}^{t+1}-\left(\lambda_{i,j,k}-\frac{\lambda_{i,j,k+1}-\lambda_{i,j,k-1}}{4}\right)c_{i,j,k-1}^{t+1}=$$

$$\left[1-6\lambda_{i,j,k}+\frac{\Delta t}{2}(\rho-G)\right]c_{i,j,k}^{t}+2\lambda_{i,j,k}c_{i-1,j,k}^{t}+$$

$$\left(\lambda_{i,j,k}+\frac{\lambda_{i,j+1,k}-\lambda_{i,j-1,k}}{4}\right)c_{i,j+1,k}^{t}+\left(\lambda_{i,j,k}-\frac{\lambda_{i,j+1,k}-\lambda_{i,j-1,k}}{4}\right)c_{i,j-1,k}^{t}+$$

$$\left(\lambda_{i,j,k}+\tfrac{\lambda_{i,j,k+1}-\lambda_{i,j,k-1}}{4}\right)c_{i,j,k+1}^{t}+\left(\lambda_{i,j,k}-\tfrac{\lambda_{i,j,k+1}-\lambda_{i,j,k-1}}{4}\right)c_{i,j,k-1}^{t} \tag{4}$$

where $c_{i,j,k}^{t}$ is the finite difference approximation of c at the grid point $(x_i\ y_j, z_k)$ at time t, Δt is the time step size for the time discretization, h is the space step size at each axis of the gridding scheme for the space discretization,

$\lambda = D\Delta t\,/\,[2(h)^2]$ and $\lambda_{i,j,k} = D_{i,j,k}\Delta t\,/\,[2(h)^2]$.

The resulting system of equations may be written equivalently in the form

$$\overleftrightarrow{A}\,\vec{x}=\vec{b} \tag{5}$$

where $\vec{x}$ denotes a vector that contains an approximation of the solution c at the mesh nodes at time $t = t_n$. In the homogeneous case, the matrix $\overleftrightarrow{A}$ is a sparse, *symmetric* matrix and the algorithm selected for the solution is the non - stationary iterative Conjugate Gradient method (CG). In the inhomogeneous case, matrix $\overleftrightarrow{A}$ a sparse *non-symmetric* and *non positive definite* matrix. The algorithm selected in order to handle this more general linear system, is the biconjugate gradient method (BiCG) [3].

III. Clinical Validation Aspects

For an initial gross clinical validation of the models a real normal human head has been considered. MRI based head imaging data freely available on the internet (www.slicer.org) has been used. The structures of white matter, grey matter, CSF and skull have been segmented. A fictitious growing virtual spherical glioblastoma tumour of radius equal to 0.7 cm has been virtually placed inside the intracranial cavity. The concentration of tumour cells within the initial tumour has been arbitrarily assumed uniform and equal to 10^6 cells/mm^3. Diffusion phenomena have been ignored before the time point corresponding to the start of the simulation. The following parameter values have been used:

h=0.1cm,

Δt=0.5d,

and net tumor growth rate

$\rho=0.012d^{-1}$.

For the inhomogeneous scenario the value of the space dependent diffusion coefficient $D_{i,j,k}$, has been calculated as

the average value of the growing diffusion coefficient and the migrating diffusion coefficient [4]. The value of D for the homogeneous case has been estimated as the weighted average value of the diffusion coefficient for white matter, grey matter and CSF. The virtual tumour grows for 180 days after the initialization time point. A two and a three dimensional snapshot of the fictitious GBM tumour on the first and the 180th simulated day for the inhomogeneous and homogeneous scenario respectively is depicted in Fig. 1. The spatial pattern of the simulated tumour - in particular in the inhomogeneous brain scenario - is in good agreement with actual clinical observations reported in literature [5]. According to [6] doubling times for gliomas have been estimated to span from 1 week to 12 months. Predictions of the doubling time for both models and for the previously mentioned data have been produced and compared. During the simulated periods the doubling times for both scenarios fall within the lower part of the range of doubling times for gliomas. This is in agreement with the fact that GBM is a high grade glioma. For example a doubling time of two months is observed on the 39th day of the simulated period. The temporal dependence of doubling time is in agreement with the typical Gompertzian tumour growth curve.

We assume that the inhomogeneous normal brain tissue representation is a more credible virtual rendering of reality than its homogeneous counterpart. The simulation results for the cases considered (Fig. 2) show that using the homogeneous normal brain based model may lead to an error of up to 10% for the first 25 simulated days in relation to the predictions of the inhomogeneous model. However, the error drops to less than 7% afterwards. This observation suggests

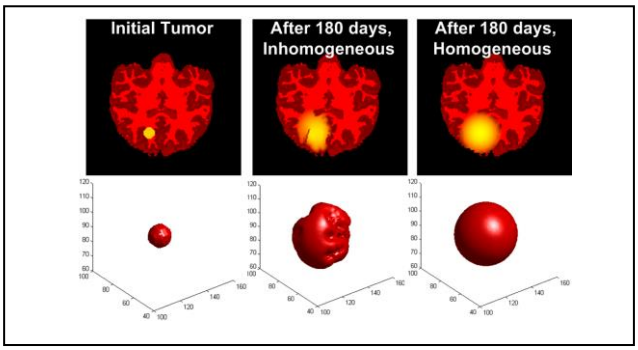


Figure 1. Schematic representation of the growth of a virtual glioblastoma tumour in vivo for the inhomogeneous and homogeneous case after 180 days. Upper panels: The colour intensity level depends logarithmically on the tumour cell concentration. Lower panels: The tumour cell concentration threshold below which a voxel is not painted red (indicating that the voxel belongs to the bulk tumour component) is 1 cell/mm³.

that even by using a homogeneous brain based model and an adequately weighted average value of its diffusion coefficient, a rough but still informative estimate of the expected tumour doubling time can be achieved.

It is noted that the values of the diffusion coefficients and the cell birth and death rates of the model are amenable to refinement and personalization by exploiting the histological and molecular profile of the patient. This type of work undertaken by our research group is in progress.

IV. Conclusion

The phenomenon of tumour invasion into the surrounding tissues, constituting a hallmark of cancer, has been exemplified by the paradigm of glioblastoma multiform. Two four dimensional spatiotemporal clinically driven and oriented models have been developed and compared. Both models (i.e the homogeneous brain based one and its inhomogeneous counterpart) are based on a finite difference – time domain numerical solution of the diffusion-reaction equation in conjunction with a novel explicit numerical treatment of the adiabatic Neumann boundary conditions imposed by the skull. Clinical observational data regarding the range of the GBM doubling time values have been utilized in order to ensure the realism of both models and their predictions. Additionally, the behaviour of both models is in good agreement with pertinent GBM imaging data available in literature. The numerical results presented in this paper suggest that even by using by a homogeneous brain based model and a realistic weighted average value of its diffusion coefficient, a rough but still informative estimate of the expected tumour doubling time can be achieved. Additional in silico experimentation aiming at statistically testing and eventually supporting the validity of this hypothesis is in progress.

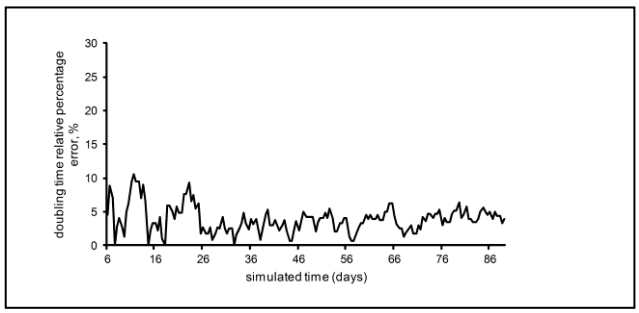


Figure 2. Relative percentage error of the doubling time calculated by comparing the values generated by the model for the homogeneous with the ones obtained for the inhomogeneous case.

The models presented are to serve as two of several modeling components of an Oncosimulator [7] dedicated to neuro-oncology. In order for such a system to be translated into clinical practice [8-9], a strict clinical adaptation and validation procedure has to be successfully completed. Part of this procedure is currently taking place within the framework of the large scale transatlantic (EU-US) EU funded integrating research project entitled "CHIC:Developing Meta- and Hyper-Multiscale Models and Repositories for In Silico Oncology" [http://chic-vph.eu/ , Project Identifier: 600841]. Prof. S. Gool, MD, Catholic University of Leuven, has undertaken the task of providing pertinent multiscale data for the clinical adaptation and validation of the CHIC Neuro-Oncosimulator.

Acknowledgment

The authors would like to thank Prof. Dr med N. Graf, Director of the Pediatric Oncology and Hematology Clinic,

University Hospital of Saarland, Germany for fruitful discussions.

REFERENCES

[1] S. G. Giatili and G. S. Stamatakos, "A detailed numerical treatment of the boundary conditions imposed by the skull on a diffusion–reaction model of glioma tumour growth. Clinical validation aspects," *Applied Mathematics and Computation*, vol. 218, no. 17, pp. 8779-8799, May 2012.

[2] K. R. Swanson, E. C. Alvord, and J. D. Murray, "Virtual brain tumours (gliomas) enhance the reality of medical imaging and highlight inadequacies of current therapy," *Br. J. Cancer*, vol. 86, no. 1, pp. 14 – 18, Jan. 2002.

[3] B. Bradie, *A friendly Introduction to Numerical Analysis*. Pearson International Ed. Unites States of America, 2006.

[4] S. E. Eikenberry, T. Sankar, M. C.Preul, E. J. Kostelich, C.J. Thalhauser, and Y. Kuang, "Virtual glioblastoma: growth, migration and treatment in a three-dimensional mathematical model," *Cell Prolif*. Vol. 42, no. 4, pp. 511-528, Aug. 2009.

[5] A. D. Waldman, A. Jackson, S. J. Price, C. A. Clark, T. C. Booth, D. P. Auer, P. S. Tofts, D. J. Collins, M. O. Leach, and J. H. Rees, "Quantitative imaging biomarkers in neuro-oncology," *Nat. Rev. Clin. Oncol*., vol. 6 , no. 8, pp. 445-454, Aug. 2009.

[6] E. C. Alvord Jr and C. M. Shaw, "Neoplasm affecting the nervous system in the elderly," *in The Pathology of the Aging Human Nervous System*, S. Duckett, Ed. Philadelphia: Lea & Febiger, 1991, pp. 210–286.

[7] G. Stamatakos, D. Dionysiou, A. Lunzer, R. Belleman, E. Kolokotroni, E. Georgiadi, M. Erdt, J. Pukacki, S. Rueping, S. Giatili, A. d`Onofrio, S. Sfakianakis, K. Marias, C. Desmedt, M. Tsiknakis, and N Graf, l, "The Technologically Integrated Oncosimulator: Combining Multiscale Cancer Modeling With Information Technology in the In Silico Oncology Context," *IEEE J Biomedical and Health Informatics*, vol. 18, no. 3, pp. 840-854, May 2014.

[8] G. Stamatakos, N. Graf, and R. Radhakrishnan, "Multiscale Cancer Modeling and In Silico Oncology: Emerging Computational Frontiers in Basic and Translational Cancer Research," *Journal of Bioengineering & Biomedical Science*, vol. 3:e114. May 2013.

[9] G. 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*, vol. 6, no. 3, e17594, Mar. 2011

The Importance of Grid Size and Boundary Conditions in Discrete Tumor Growth Modeling*

Georgios Tzedakis, Giorgos Grekas, Eleftheria Tzamali, Kostas Marias, *Member, IEEE*, and Vangelis Sakkalis

***Abstract*— Modeling tumour growth has proven a very challenging problem, mainly due to the fact that cancer is a very complex process that spans multiple scales both in time and space. The desire to describe interactions in multiple scales has given rise to modeling approaches that use both continuous and discrete variables, called hybrid. The biochemical processes occurring in tumour environment are usually described by continuous variables. Cancer cells tend to be described as discrete agents interacting with their local neighborhood, which is comprised of their extracellular environment and nearby cancer cells. These interactions shape the microenvironment, which in turn acts as a selective force on clonal emergence and evolution. In this work, we study the effects of grid size and boundary conditions of the continuous processes on the discrete populations. We perform various tests on a simplified hybrid model with the aim of achieving faster execution runtimes. We conclude that we can reduce the grid size while maintaining the same dynamics of a larger domain by manipulating the boundary conditions.**

I. Introduction

Modeling, analysis and simulation are expected to play crucial roles in explaining complex biological systems and help turn biology in a predictive science. One such complex system is the uncontrolled growth of cancer cells over multiple temporal and spatial scales. Solid tumours are believed to arise from few mutated cells that proliferate without supervision. At the first stages, the tumour is supplied with nutrient from the nearby preexisting healthy tissue. As tumors grow in size, they gradually pass from the simple avascular phase to much more complex interactions with their environment, recruiting and reforming the vascular network.

Recent efforts have been made to model tumour growth and invasion using an interdisciplinary approach combining mathematical models of cancer, *in silico* and *in vitro* experiments and clinical data [1-3]. The aim of these efforts has been to shed light on the root causes of solid tumour invasion and metastasis, to aid in the understanding of experimental and clinical observations and to help design both new experiments and treatment strategies. Traditionally mathematical models belong into one of two broad categories based on how the tumor tissue is represented: discrete cell-based models and continuum models.

Discrete models describe tumour cells as individual entities, discrete elements that evolve through local rules in discrete space and time [4-6]. Although useful for studying phenomena such as carcinogenesis, natural selection, genetic instability and interactions of individual cells with each other and the microenvironment, discrete models are computationally demanding and thus mainly applied to small systems. On the other hand, continuum models approximate tumour cells as continuous fields by means of partial differential equations [7-9]. This approach tends to draw principles from continuum mechanics to describe cancer and its environment. Although the continuum and discrete modeling approaches have each provided important insight into cancer-related processes occurring at particular spatial and temporal scales, the complexity of cancer and the interactions between the cell- and tissue-level scales may be elucidated further by means of a multiscale (hybrid) approach that uses both continuum and discrete representations of tumour cells and components of the peritumoral microenvironment [10, 11]. Both cellular and microenvironmental factors act both as tumour morphology regulators and as determinants of invasion potential by controlling the mechanisms of cancer cell proliferation and migration [12] and hybrid models provide a way to integrate this multiscale nature of cancer [13].

The aim of this work is to improve the speed of a simple hybrid model by decreasing the computational domain. We introduce a simple hybrid model where tumour growth is described using discrete cells while the nutrient supplying the tumour cells is perceived as a continuous variable that can be modeled by a partial differential equation (PDE) of reaction-diffusion type. We assume that the aforementioned vital nutrient that allows tumour cells to survive and promotes proliferation is Oxygen. In this hybrid model, but also in many models that follow similar approaches [14], the need to describe phenomena in different scales imposes a problem. Specifically, within the hybrid framework, the spatial domain is discretized to very small lattice sites that correspond to the size of individual cells. However, such discretization for the continuous variables is unnecessarily dense and causes the continuous model to act as a computational bottleneck for the whole system.

In order to amend that problem, Dirichlet boundary conditions are used in order to mimic the flux behavior of larger spatial domain size. It is supposed that the density values of oxygen outside of the boundaries are close to the maximum density value of the tissue, as it is occurred in healthy tissues. The aforementioned assumption, in diffusion equation, results in a continuous influx from the boundary to the internal domain, when internal oxygen densities in the neighborhood of boundary are lower than the maximum value.

* This work was supported by the Computational Horizons In Cancer – CHIC (FP7-ICT-2011.5.2-600841) project.

G.Tzedakis is with the Institute of Computer Science, FORTH, Vassilika Vouton GR-70013 Heraklion, Crete, Greece (corresponding author phone: 0030-2810-392441; fax: 0030-2810-391428; email: gtzedaki@ics.forth.gr).

G. Grekas, E. Tzamali, K. Marias and V. Sakkalis are with the Institute of Computer Science, FORTH, Vassilika Vouton GR-70013 Heraklion, Crete, Greece ({ggrekas; tzamali; kmarias; sakkalis}@ics.forth.gr).

Comparing Dirichlet, with boundary values equal to the maximum value density, and Neumann, with no flux, boundary conditions negligible differences on oxygen flow are observed, when high oxygen consumption from tumor cells occur far from the boundary. Thus, for the large domain Neumann boundary conditions with zero flux are normally used.

II. Model Description

To evaluate the effect of boundary conditions and grid size on tumor evolution, we assume a simple hybrid model. As already mentioned, the model consists of two components, a continuous and a discrete one. The discrete component is used to describe the evolution of the cancer cells. The continuous component, on the other hand, is used to describe the environment of the tumour. This particular model is inspired by the work of Anderson [14].

A. Continuous Part

For simplicity, from all the elements that can comprise the complex tumour micro-environment, we only consider here the effects of the Oxygen concentration. We also assume that the tumour is well vascularized, thus oxygen is produced from every lattice site. However, the rates of oxygen production and consumption affect the oxygenation levels within the tumour and increased metabolic demands of tumour cells can prove vasculature inadequate leading to increased tumour hypoxia.

$$\frac{\partial o}{\partial t} = \nabla \cdot (D_o \nabla o) - bo - \gamma_{i,j} o + \beta \qquad (1)$$

Equation (1) describes the spatiotemporal evolution of the oxygen concentration. The value of the concentration $o(\boldsymbol{x}, t)$ represents the oxygen concentration at point $\boldsymbol{x}$ and time t. The parameters $D_o, b, \beta, \gamma_{i,j}$ indicate the oxygen diffusion coefficient, the natural oxygen decay, the oxygen production rate by healthy tissue and the oxygen consumption rate by tumour cells, respectively. The value of $\gamma_{i,j}$ depends on the type and the state of the cell at the grid point i, j.

Equation (1) is solved on a $[0, L] \times [0, L]$ square grid. The parameter values for the continuous and the discrete parts are rendered dimensionless in such a way that a square of the grid represents $25\,\mu m$, which we assume is the approximate diameter of a single cell. That condition is important for the interaction of the continuous and the discrete parts of the model.

B. Discrete Part

The discrete part of the model involves tracking each cancer cell separately. Every cell lives, proliferates and dies independently, while it interacts with its microenvironment locally. The cell life in brief can be seen at the flow chart at Fig. 1. For simplicity, cell movement is not taken into account.

At each time step the cell 'checks' if the total oxygen level is below a certain threshold o_{death} and if that condition holds true then the cell cannot continue living and dies. As soon as a cell dies, the space it was occupying is treated as vacant and can be filled by another cancer cell.

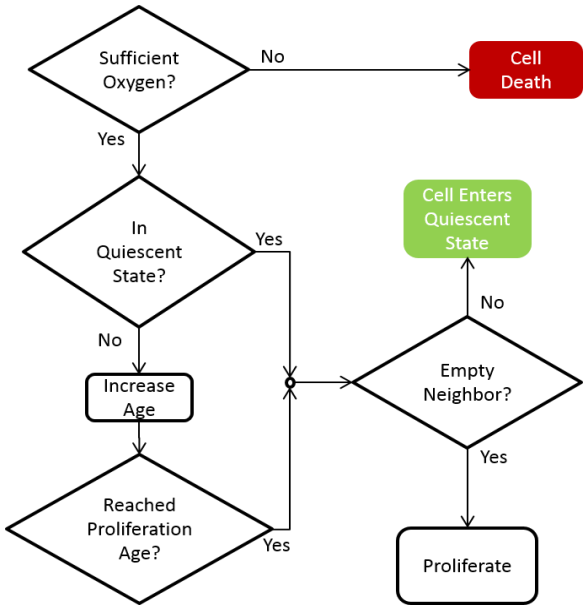


Figure 1. Cell life flow chart

Tumor cells with sufficient oxygen prepare for proliferation, increasing their age at each time step. We also assume that the increment of the cellular age depends linearly on local oxygen availability such that limited oxygen results in a decreased cellular proliferation rate. Specifically, if the cell age is updated every t seconds the cell ages are incremented by $o(i,j) \cdot t$, where $o(i,j)$ is the value of the normalized oxygen concentration at the cell.

When a cell has reached its proliferation age, the cell resets its age, an empty neighboring site is chosen at random, and an exact copy of the cell is placed at that site. In accordance to [14], if no adjacent empty space can be found to place the daughter cell, then the cell enters the quiescent state and waits until an empty site is found in its neighbor. Thus, cells in quiescent state are ready to proliferate and when a neighboring site becomes empty, they immediately proliferate. Because quiescent cells no longer prepare to proliferate it is assumed that they consume half the nutrients than the non-quiescent cancer cells.

III. Implementation

As there is no analytical solution to the particular equation, we approximated the solution numerically. In particular, for the solutions of the equation alternating direction implicit (ADI) finite difference method was used, resulting in the solution of tridiagonal linear systems. Using ADI the PDE can be solved in $O(L^2)$, where L denotes the edge of our grid [15]. To ensure numerical accuracy and stability the time step used was one fifth of the spatial step [16].

In order to avoid artefacts that would occur if all cells started proliferating at the same time, initial cell ages are randomly assigned to them. Additionally, to ensure no bias towards a certain cell, at each time step the life cycle is executed for all cells but the succession is chosen randomly, which dramatically reduces the possibility that a cell has the

same priority for consequent turns.

The parameters used are in accordance to [14] and can be found in Table I. If no units are mentioned then the values provided are the non-dimensionalized ones. Non-dimensionalization was done using $\tau = 8$ hours for the temporal dimension and 0.5, 1 and 2 cm for the length scale L'', L' and L, respectively.

TABLE I. PARAMETER VALUES

Symbol	Description	Values
L'', L', L	Grid size	200,400,800
h	Spatial step	25 μm
τ	Iteration time step	8 hours
D_o	Oxygen diffusion parameter	10^{-5} cm^2 s^{-1}
β	Oxygen production rate	0.25
γ	Cancer cell Oxygen uptake	107.4
α	Oxygen decay	0.0125
M	Cancer cell proliferation age	8 hours

IV. RESULTS

In all experiments we initialize the tumour at the center of the grid as a square of 15×15 cells. The tumour is allowed to grow for 400 hours where its size increases from approximately 0.4 mm to less than 0.5 cm in diameter. We first run the model for a grid size $L = 800$ (2 cm), which is large relatively to the tumor size, using no flux boundary conditions and we use these simulations for the ground truth. We then perform simulations for reduced grid sizes $L' = 400$ (1 cm) and $L'' = 200$ (0.5 cm) using Dirichlet boundary conditions, fixing the boundary values for the Oxygen at its maximum value of 1. Finally, we also run a simulation for the grid L' using no flux boundary conditions. The simulations were run for 50 iterations or 400 modeling hours and the final tumor sizes are illustrated in Fig. 2.

All the figures contain the central 200×200 part of the domain. In all the cases, we notice that the center of the tumour is comprised of trapped quiescent cells, while on the rim it is comprised of proliferating cells. We see that the simulated tumours using the Dirichlet boundary condition (Fig. 2- 1st row) are larger than the ones using Neumann conditions (Fig. 2- 2nd row).

In Fig. 3 we can see the tumour growth curves for all the experiments. Since no necrosis has been detected, the observed differences in growth are mainly because of the dependence of the proliferation time of the cells on the local oxygen levels. As mentioned before, the large 800×800 grid with no flux boundary conditions is taken as ground truth for the other simulations (depicted with the black dotted line in Fig. 3). As we can see in Fig.3 (shown with cyan dotted line), when reducing the grid size to half without switching the boundary conditions, the growth curve starts to substantially deviate from the ground truth after 200 fictitious hours. However, when we change the boundary conditions to Dirichlet, the gap between the curves is considerably reduced (solid green line). Further reduction of the grid size even with Dirichlet boundary conditions (solid blue line) causes the tumour to grow faster that the assumed ground truth.

The large deviation observed in the 200×200 grid size as compared to the ground truth can be attributed to the fact that the core assumption made -that outside of the grid the oxygen concentration is maximum- is no longer a valid assumption.

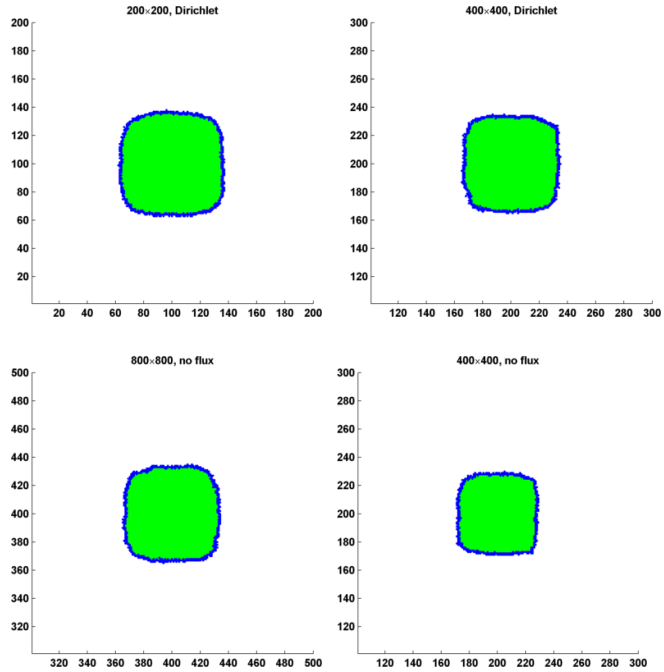


Figure 2. Resulting tumour morphology for all the experiments. The cells shown are the in the 200×200 central part. Proliferating and quiescent cells are marked with blue and green respectively. In the top row we can see the experiments using Dirchlet boundary conditions and in the bottom the results of no flux Neumman condition. The grid sizes are 200, 400, 800 and 400 starting from top left and continuing clockwise.

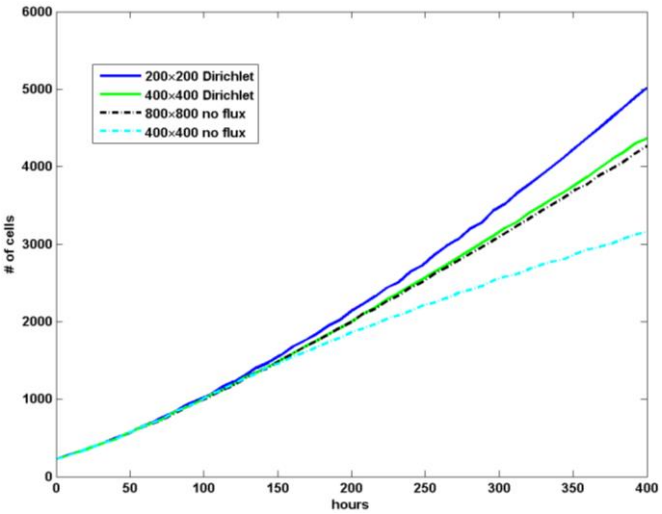


Figure 3. Tumour growth curves for all experiments.

In reality, it is neither valid for the 400×400 grid, which can be confirmed by the less than maximum Oxygen values of the ground truth case that are more than 200 lattice points away from the center. However, these values are close to the maximum concentration. Considering the smallest grid, on the other hand, the ignored values are not close enough to the maximum because the grid is not large enough to hold a big portion of the oxygen concentration curve. This can be observed when running experiments for the smallest grid with

no flux boundary conditions (not shown here). Thus, we can conclude that this grid is not big enough to make the assumption that the boundary retains the maximum oxygen value, valid. Surprisingly though, the smallest grid achieves better results that the average grid with Dirichlet boundary conditions.

In Table II we present the final cell populations and the absolute relative error (ARE) when compared to our ground truth case. The execution time and the relative improvement (RI) of the execution times are also included in Table II. We can see that the increase in simulation speed as we change from 800 to 400 lattice is 86% while introducing 2.5% relative error to the cell population number. It is noteworthy that if we had kept the no flux boundary conditions the speed up percentage would be similar but the error would be 25%. If we do the same calculations with the 800 and 200 edged grids we get 96% decrease in speed and 17.6% relative error to the population count with Dirichlet boundary conditions. However we can observe from the Fig. 3 that the error tends to increase as the simulation is running.

TABLE II. FINCAL CELL POPULATIONS AND EXECUTION TIMES

Grid Size & B.C.	**200×200, Dirichlet**	**400×400, Dirichlet**	**800×800, no flux**	**400×400, no flux**
ARE	17.6%	2.5%	0%	25%
(total cells)	(5014)	(4366)	(4262)	(3156)
RI	96%	86%	0 %	86%
(execution time)	(363 sec)	(1285 sec)	(9572 sec)	(1264 sec)

V. CONCLUSION

In this work we sought to improve the speed of a simple hybrid model by decreasing the computational domain. In general, the reduction of the domain sped up the computations but the cancer growth was impeded from the nutrient shortage. To compensate for the missing domain we changed the boundary conditions to provide a nutrient influx, essentially replacing the missing sources for the cancer growth. That significantly reduced the resulting error when compared to the original larger lattice.

Future work could include the dynamic adaptation of the Dirichlet boundary condition in such a way that the growth curves do not drift apart as time goes by. One possible way to do that might be by measuring the average Oxygen levels and somehow compensate accordingly. Another intriguing problem to tackle would be investigating how the boundary conditions would change when the domain is supplied with nutrient non-uniformly and even time-variant for example when angiogenesis is introduced.

Real tumours are highly more heterogeneous and complex and considerably less well-defined and symmetric to allow any direct comparison at this point, however, improvements in the processing time allow hybrid models to better cope with the complex and multiscale nature of tumour growth. Undoubtedly, the next step includes the introduction of a more sophisticated model that describes the complex interaction between cells and their microenvironment, which is comprised of the extracellular matrix and vasculature as well as the incorporation of cellular motility, which is affected by cell-matrix interactions and plays a critical role in tumour invasion.

VI. REFERENCES

[1] Y. Kam, K. A. Rejniak, and A. R. Anderson, "Cellular modeling of cancer invasion: integration of in silico and in vitro approaches," *J Cell Physiol*, vol. 227, pp. 431-8, Feb 2012.

[2] T. E. Yankeelov, N. Atuegwu, D. Hormuth, J. A. Weis, S. L. Barnes, M. I. Miga, *et al.*, "Clinically relevant modeling of tumor growth and treatment response," *Sci Transl Med*, vol. 5, p. 187ps9, May 29 2013.

[3] V. Sakkalis, S. Sfakianakis, E. Tzamali, K. Marias, G. Stamatakos, F. Misichroni, *et al.*, "Web-based workflow planning platform supporting the design and execution of complex multiscale cancer models," *IEEE J Biomed Health Inform*, vol. 18, pp. 824-31, May 2014.

[4] J. v. Neumann and A. W. Burks, "Theory of self-reproducing automata," 1966.

[5] Y. Jiao and S. Torquato, "Emergent Behaviors from a Cellular Automaton Model for Invasive Tumor Growth in Heterogeneous Microenvironments," *PLoS Comput Biol*, vol. 7, p. e1002314, 2011.

[6] T. Alarcon, H. M. Byrne, and P. K. Maini, "A cellular automaton model for tumour growth in inhomogeneous environment," *J Theor Biol*, vol. 225, pp. 257-74, Nov 21 2003.

[7] E. Tzamali, G. Grekas, K. Marias, and V. Sakkalis, "Exploring the competition between proliferative and invasive cancer phenotypes in a continuous spatial model," *PLoS One*, vol. 9, p. e103191, 2014.

[8] S. M. Wise, J. S. Lowengrub, H. B. Frieboes, and V. Cristini, "Three-dimensional multispecies nonlinear tumor growth—I: Model and numerical method," *Journal of Theoretical Biology*, vol. 253, pp. 524-543, 8/7/ 2008.

[9] K. R. Swanson, R. C. Rockne, J. Claridge, M. A. Chaplain, E. C. Alvord, Jr., and A. R. Anderson, "Quantifying the role of angiogenesis in malignant progression of gliomas: in silico modeling integrates imaging and histology," *Cancer Res*, vol. 71, pp. 7366-75, Dec 15 2011.

[10] V. Cristini and J. Lowengrub, *Multiscale modeling of cancer : an integrated experimental and mathematical modeling approach*. Cambridge ; New York: Cambridge University Press, 2010.

[11] K. Marias, D. Dionysiou, V. Sakkalis, N. Graf, R. M. Bohle, P. V. Coveney, *et al.*, "Clinically driven design of multi-scale cancer models: the ContraCancrum project paradigm," *Interface Focus*, vol. 1, pp. 450-61, Jun 6 2011.

[12] S. Sanga, H. B. Frieboes, X. M. Zheng, R. Gatenby, E. L. Bearer, and V. Cristini, "Predictive oncology: A review of multidisciplinary, multiscale in silico modeling linking phenotype, morphology and growth," *Neuroimage*, vol. 37, pp. S120-S134, 2007.

[13] K. A. Rejniak and A. R. A. Anderson, "Hybrid models of tumor growth," *Wiley Interdisciplinary Reviews-Systems Biology and Medicine*, vol. 3, pp. 115-125, Jan-Feb 2011.

[14] A. R. A. Anderson, "A hybrid mathematical model of solid tumour invasion: the importance of cell adhesion," *Mathematical Medicine and Biology-a Journal of the Ima*, vol. 22, pp. 163-186, Jun 2005.

[15] J. J. Douglas and D. W. Peaceman, "Numerical solution of two-dimensional heat-flow problems," *AIChE J.*, vol. 1, pp. 505-512, 1955.

[16] A. Roniotis, G. C. Manikis, V. Sakkalis, M. E. Zervakis, I. Karatzanis, and K. Marias, "High-grade glioma diffusive modeling using statistical tissue information and diffusion tensors extracted from atlases," *IEEE Trans Inf Technol Biomed*, vol. 16, pp. 255-63, Mar 2012.

A Two Population Model of Cancer Growth with Fixed Capacity*

Ilaria Stura, Domenico Gabriele, and Caterina Guiot

***Abstract*— Cancer is not a homogenous tissue, but a very complex mix of different cell populations; moreover, a delicate equilibrium exists between these components of a tumour mass. In this work we address prostate cancer although the methods presented can be generalized to most tumour types. The aim of our work is to model the behaviour of the different cell populations within the tumour and simulate changes which occur during natural evolution and treatments.**

I. Introduction

Prostate cancer is the most prevalent tumour affecting the male population and in approximately 30% of cases it develops chronically for a long period of time (i.e. years), requiring surgical and/or medical interventions (e.g. Androgen-Deprivation Therapies with anti-androgens drugs or LHRH analogues).

Usually, the tumour reacts to the hormonal therapy for a highly variable time span, from some months to more than ten years, finally becoming hormone-resistant. This behavior is related to the co-existence in prostate cancer of cells with a different sensibility to hormones: in particular, simplifying the cellular spectrum, we may suppose a balance between two groups of hormone-sensitive versus hormone-resistant cells. Under the pressure of an anti-androgen drug, the second group becomes gradually dominant and causes the clinical relapse and the emergence of a disseminated castration-resistant prostate cancer.

Two main models were developed to depict the evolution of the ratio between the sensitive and resistant cells: the 'adaptation' model and the 'clonal selection' model ([1],[2],[3]). The adaptation model suggests that prostate cancer is initially composed by homogeneous cells, in terms of their androgen requirement, and castration resistance emerges through genetic or epigenetic switches of androgen-dependent cells to androgen-independent cells. On the contrary, the 'clonal selection' model states that primary prostate cancer cells are heterogeneous regarding their androgen requirement, and so a minority of them is a clone of pre-existing castration-resistant cells. In an androgen-deprived environment, these castration-resistant cells are selected for their survival and proliferative advantage.

*This work has been supported by the European Commission under the project Computational Horizons In Cancer (CHIC): Developing Meta- and Hyper-Multiscale Models and Repositories for In Silico Oncology (FP7-ICT-2011-9, Grant agreement no: 600841).

I. Stura is with the Department of Neuroscience, University of Turin, Italy (corresponding author phone: 0116638198; e-mail: ilaria.stura@unito.it).

D. Gabriele is with the Department of Neuroscience, University of Turin, Italy (e-mail: domenico.gabriele@unito.it).

C. Guiot is with the Department of Neuroscience, University of Turin, Italy (e-mail: caterina.guiot@unito.it).

Our model describes such scenarios and simulates interactions between clones taking into account that the general amount of nutrients and the anatomical space available are limited (fixed carrying capacity) and the different effects that treatments have on various cell populations. In section II there is a brief introduction on the Phenomenological Universalities approach (or PUN) to simulate a cancer cell population, in section III we will explain how to model the tumor as two populations system, and in section III-B our results on how the behaviour of a population could change in relation to the other one are reported.

II. PUN

The Phenomenological Universalities (PUN) approach (see [4], [5] for details) describes tumor growth with the equations:

$$\frac{dN}{dt} = c(t)N(t)$$
$$\frac{dc}{dt} = \sum \beta_i c^i \qquad (1)$$

N is the cancer cells population numerosity, *c(t)* is the growth rate function and *n* is the degree of its Taylor expansion. This approach generalizes the most used equations in population growth, in fact: for $n = 0$, *c(t)* constant, *N* grows as an exponential law; for $n = 1$, as a gompertzian law and for $n = 2$ as West/logistic growth law. We will refer to a particular class of PUN using the notation '*Un*' with *n*=0,1,2,.., e.g. *U1* instead of gompertzian model.

We focus our attention on the case of $n = 1$ (gompertzian growth law): this function has an exponential initial growth that progressively tends to slow and finally it reaches its carrying capacity. This behavior reflects very realistically the trend of a tumour: in an early stage, cells grow faster because they have a lot of nutrient and space, then the cancer core becomes necrotic and hypoxic and/or the tumour reaches some physical barrier such as tissue or bones and it cannot swell again. The dynamic system, in U1 case, is:

$$\frac{dN}{dt} = c(t)N(t)$$
$$\frac{dc}{dt} = \beta c$$

or

$$\frac{dN}{dt} = c_0 e^{\beta t} N(t) \qquad (2)$$

The solution of (2) is:

$$N(t) = N_0 e^{\frac{c_0}{\beta}(e^{\beta t}-1)} \quad (3)$$

β is inversely proportional to the carrying capacity that tumour can reach and c_0 is the growth rate. Note that in this case the carrying capacity depends on the initial condition N_0. We will normalize this equation as:

$$V(t) = \frac{N(t)}{N_0} = e^{\frac{c_0}{\beta}(e^{\beta t}-1)}$$

To stress the importance of the carrying capacity, equation (3) can be rewritten as:

$$N(t) = N_\infty e^{ze^{-rt}} \quad (4)$$

where N_∞ is the carrying capacity and r the growth rate. We can easily transform (3) in (4) and vice versa with $r = \beta$, $N_1 = N_0 e^{c_0/\beta}$ and $z = c_0/\beta$.

In the next section we will show a two populations model based on PUN.

III. Two Populations Model

Two different populations of cells, due to genetic (e.g. two clones of the same tumour) or epigenetic (e.g. necrotic center of the mass) relation, are considered and described using the PUN approach.

If the two populations are gompertzian and independent of each other, the system becomes:

$$\frac{N_1(t)}{dt} = c_1(t)N_1(t)$$
$$\frac{N_2(t)}{dt} = c_2(t)N_2(t) \quad (5)$$

and the solutions are independent, so each population grows as in (1). The only constraint will be that the total amount of the carrying capacities is a constant: $N_{\infty 1} + N_{\infty 2} = N_\infty$, $N_{\infty 1}$ and $N_{\infty 2}$ being the carrying capacities of N_1 and N_2 respectively in the system (5).

A. The two populations model with mutation

A more complex system, where population 1 could mutate in 2 but 2 couldn't mutate in 1, as occurs, for example, if 2 is a more resistant and/or more aggressive clone than 1, is described by:

$$\frac{N_1(t)}{dt} = c_1(t)N_1(t) - mN_1(t)$$
$$\frac{N_2(t)}{dt} = c_2(t)N_2(t) + mN_1(t) \quad (6)$$

where m is the mutation rate.

1) Analytical solutions: The first equation has no dependencies from the second, so we can find easily the solution $N_1 = \exp\{\int^t c_1(s) - m\, ds\}$.

The second equation depends both on N_1 and N_2 and is a linear non homogeneous equation:

$$N_2 = e^{\int c_2(s)ds}\left(\int e^{\int c_2(l)dl} m e^{\int c_i(l) - mdl} ds + C\right)$$

Note that in some cases N_2 does not have a closed form, but it could be approximated with numerical analysis tool, e.g. MATLAB (MATLAB and Statistics Toolbox 7.0, The MathWorks, Inc., Natick, Massachusetts, United States.) or Octave.

2) Equilibrium points: Forcing equations (6) to zero and solving the system, we could easily find that the equilibrium points are $P_1=(0,0)$ and $P_2=(n_1, -mn_1/c_2(t))$, with $N_1(t^*)=n_1$ and $c(t^*)=m$; by studying the Jacobian matrix and the eigenvalues $\lambda_{1,2}$, we know that points are stable or unstable depending on $\Delta=(c_2(t)-c_1(t)+m)^2$ and $\lambda_{1,2}$:

1. if $\Delta=0$ then $\lambda_{1,2}=c_2(t) \rightarrow \lambda_{1,2}>0 \rightarrow$ unstable point
2. if $\Delta>0$ then $\lambda_1 \neq \lambda_2$ and $\lambda_{1,2}$ are real
3. if $\lambda_{1,2}>0 \rightarrow$ unstable point
4. if $\lambda_{1,2}<0 \rightarrow$ stable point
5. if $\lambda_1<0$, $\lambda_2>0$ or $\lambda_1>0$, $\lambda_2<0 \rightarrow$ saddle point (unstable)

3) U1 case: In case of U1 (gompertzian growth law) we have $c_1(t)=c_1\exp(\beta_1 t)$, $c_2(t)=c_2\exp(\beta_2 t)$ with c_1, c_2, β_1 and β_2 constants. The analytical solutions are:

$$N_1 = e^{\int c_1(s) - mds} = N_0 e^{\frac{c_1}{\beta_1}e^{\beta_1 t} - mt - \frac{c_1}{\beta_1}} = N_{\infty,1} e^{-r_1 e^{\frac{t}{N_{\infty,1}}} - mt + z_1}$$
$$N_2 = e^{c_2 e^{\beta_2 t} - c_2}\left(m\int e^{\frac{-c_2}{\beta_2}e^{\beta_2 s} + \frac{c_2}{\beta_2}} N_1(s)ds + C\right)$$

The equilibrium points are $P_1=(0,0)$, $P_2=(N_1(t^*), -mN_1(t^*)/c_2\exp(\beta_2 t^*))$and they all are unstable.

Studying the limiting behaviour for $t \rightarrow \infty$, we can see that N_1 tends to zero when $m>0$; for N_2 it is too complicate to find a close solution and its relative limit, but plotting the solution using MATLAB we saw that the limit of N_2 depends to $c_1c_2/(\beta_1\beta_2)$.

Re-writing the system using eq. (4) we have:

$$\frac{dN_1(t)}{dt} = -r_1 N_1(t) log\left(\frac{N_1(t)}{N_{\infty,1}}\right) - mN_1(t)$$
$$\frac{dN_2(t)}{dt} = -r_2 N_2(t) log\left(\frac{N_2(t)}{N_{\infty,2}}\right) + mN_1(t)$$

In this case $N_1=N_{\infty 1}\exp(-m/r_1\exp(\log(y_0/N_{\infty,1})\exp(m/r_1 - r_1 t/(1-m))))$ which tends, for $t \rightarrow \infty$, to the new carrying capacity $N_{\infty,1^*}$ that depends on the old one $N_{\infty,1}$ and on the mutation rate: $N_{\infty,1^*}=N_{\infty,1}\exp(-m/r_1)$.

The carrying capacity of the mutated population N_2 is $N_{\infty,2^*}=mk\exp(-m/r_1)/(r_2 \log(N_{2^*}/N_{\infty,2})$, where $N_{2^*}=N_2(t^*)$ and t^* is the time in which N_1 reaches its own carrying capacity $N_{\infty,1^*}$. We can see in the phase portrait that the system tends to the equilibrium point (N_{1^*},N_{2^*}); other variations of the parameters do not lead to a phase change.

B. Two populations model following treatments

When cancer is detected in a patient, in the majority of cases the tumour will be treated to reduce its volume. There is a large variety of treatments: surgery, chemotherapy,

radiotherapy and hormone therapy are the most common. We suppose that the two populations will respond in different manners to a treatment, in particular we assume the cell population 1 to be very sensitive and population 2 to be less sensitive or not sensitive at all (e.g. clone resistant to hormone therapy or a metastasis that will not be eradicated during surgery). The system (5) becomes:

$$\frac{dN_1}{dt} = c_1(t)N_1 - d_1N_1$$
$$\frac{dN_2}{dt} = c_2(t)N_2 - d_2N_2$$

where d_1 and d_2 are the kill rates of the treatment on the populations 1 and 2 respectively; $d_1 >> d_2$ because 1 is more sensitive and d_2 could even vanish if the treatment has no effect of 2.

1) Analytical solutions: The equations are independent to each other and have the same solutions of the first equation of (6):

$$N_1 = e^{\int c_1(s) - d_1 ds}$$
$$N_2 = e^{\int c_2(s) - d_2 ds}$$

2) Equilibrium points: The equilibrium points are P_1=(0,0), P_2=(0,N_2*), P_3=(N_1*,0) and P_4=(N_1*,N_2*) with N_1*=N_1(t_1*) where t_1* is the time in which $c_1(t_1^*)=d_1$ and N_2*=N_2(t_2*) where t_2* is the time in which $c_2(t_2^*)=d_2$.

Studying the Jacobian matrix we can see that all the equilibrium points are unstable, in fact:

1. in P_1=(0,0) eigenvalues depend to $c_1(t)$, $c_2(t)$, d_1, d_2
2. in P_2=(0,N_2*) one eigenvalue is 0 and the other is $1/(d_1 - c_1(t))$
3. in P_3=(N_1*,0) one eigenvalue is 0 and the other is $1/(d_2 - c_2(t))$
4. in P_4=(N_1*,N_2*) both eigenvalues are zero

3) U1 case: In case of *U1* (gompertzian growth law) we have $c_1(t)=c_1\exp(\beta_1 t)$, $c_2(t)=c_2\exp(\beta_2 t)$ with c_1, c_2, β_1 and β_2 constants.

The analytical solutions are:

$$N_i = N_{0,i} e^{\frac{c_i}{\beta_i} e^{\beta_i t} - d_i t - \frac{c_i}{\beta_i}} = N_{\infty,i} e^{-r_i e^{\frac{t}{N_{\infty,i}} - d_i t + z_i}}$$

Concerning the equilibrium points we have: $c_2\exp(\beta_2 t^*)=d_2 \leftrightarrow t^*=\log(d_2/c_2)/\beta_2$; equilibrium points are all unstable.

Studying the limit, we can see that N_1 tends to zero when $d_1>0$ and N_2 tends to zero when $d_2>0$.

Re-writing the system using eq. (4) we have:

$$\frac{dN_1(t)}{dt} = -r_1N_1(t)log\left(\frac{N_1}{N_{\infty,1}}\right) - d_1N_1(t)$$
$$\frac{dN_2(t)}{dt} = -r_2N_2(t)log\left(\frac{N_2}{N_{\infty,2}}\right) - d_2N_2(t)$$

In this case both populations admit the solution:

$$N_i = N_{\infty,i} e^{\frac{-d_i}{r_i}} e^{log\left(\frac{N_{0,i}}{N_{\infty,i}}\right) e^{\frac{-d_i}{r_i} - \frac{r_i t}{1-d_i}}}$$

and tend to $N_{\infty,i}^* = N_{\infty,i}\exp(-d_i/r_i)$.

IV. General Description of the Two Populations Model with Fixed Carrying Capacity

In a real situation, of course, both mutations and treatments, sometimes responsible for mutation themselves, are present (see Fig. 1). Moreover, we expect that the total carrying capacity N_∞ will be fixed by physical and/or energetic constraints such as total nutrient present in the body or space available between two membranes. So, if a population is reduced or eliminated, we expect that the other population will increment its carrying capacity.

The final system will be:

$$\frac{dN_1(t)}{dt} = -r_1N_1(t)log\left(\frac{N_1(t)}{N_{\infty,1}}\right) - mN_1(t) - d_1N_1(t)$$
$$\frac{dN_2(t)}{dt} = -r_2N_2(t)log\left(\frac{N_2(t)}{N_{\infty,2}}\right) + mN_1(t) - d_2N_2(t) + cN_2(t)$$

Where m is the mutation term, $d_{1,2}$ the kill rates and c is the term that must be introduced to satisfy the condition $N_{\infty,1}^* + N_{\infty,2}^* = N_\infty$. In Table I we summarize how the carrying capacities change during mutation, treatment and treatment with mutation respectively; we can also calculate how c will vary to allow N_∞ to be constant, i.e. the growth spurt of N_2 in the event of death or drastic reduction of N_1. This approach could be useful in case of treatments that

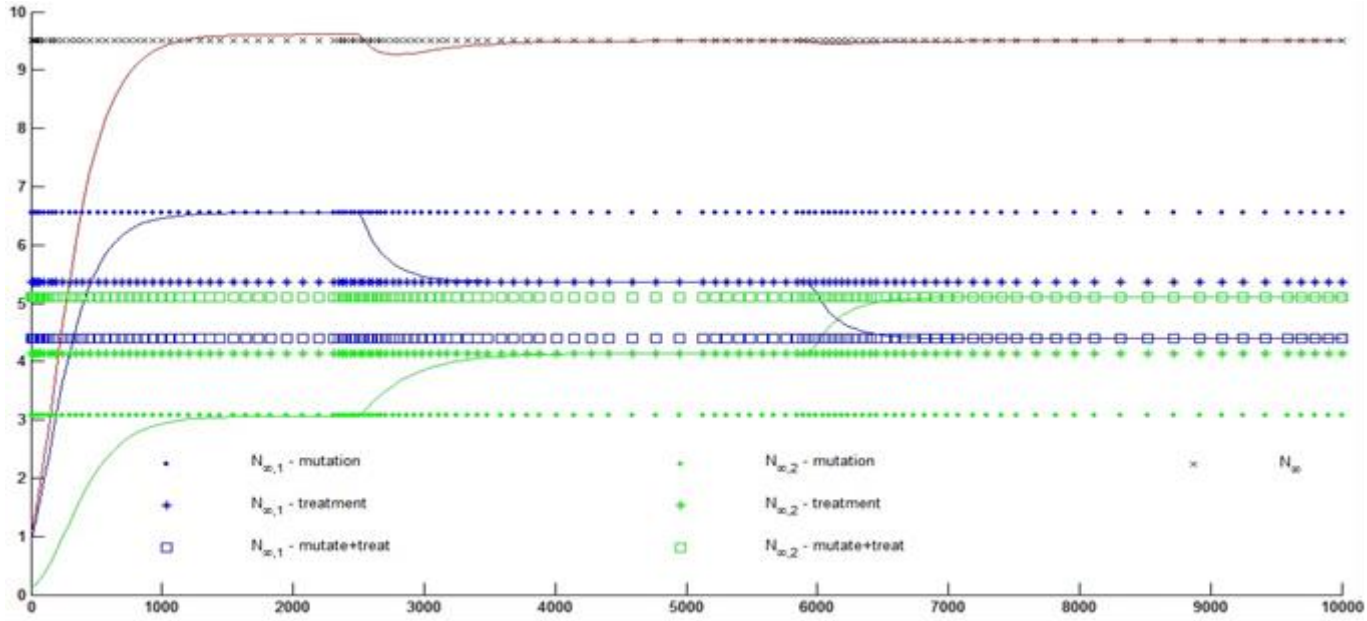


Figure 1: In the first period N_1 mutate in N_2 without treatment; in the second both populations are killed by a treatment but there isn't mutation; in the third period both mutation and treatment are involved. The *c* term has been set to allow the condition of fixed carrying capacity.

Table I. Values of final carrying capacities and of c in mutation, treatment and combined cases.

Parameters	Mutation case only	Treatment case only	Mutation with treatment case
$N^*_{\infty,1}$	$N_{\infty,1}e^{\frac{-m}{r_1}}$	$N_{\infty,1}e^{\frac{-d_1}{r_1}}$	$N_{\infty,1}e^{\frac{-(m+d_1)}{r_1}}$
$N^*_{\infty,2}$	$N_{\infty,2}e^{\frac{c+m\frac{N_1^*}{N_2^*}}{r_2}}$	$N_{\infty,2}e^{-\frac{d_2-c}{r_2}}$	$N_{\infty,2}e^{-\frac{d_2-c-m\frac{N_1^*}{N_2^*}}{r_2}}$
c	$r_2\log(\frac{N_\infty-N_{\infty,1}e^{-\frac{m}{r_1}}}{N_{\infty,2}})-m\frac{N_1^*}{N_2^*}$	$r_2\log(\frac{N_\infty-N_{\infty,1}e^{-\frac{d_1}{r_1}}}{N_{\infty,2}})-d_1$	$r_2\log(\frac{N_\infty-N_{\infty,1}e^{-\frac{m+d_1}{r_1}}}{N_{\infty,2}})-m\frac{N_1^*}{N_2^*}-d_1$

affect only population 1: we can simulate both the rapid increase of N_2 and the rapid decrease of N_1 (e.g. surgical removal of primary tumor and rapid increase in metastasis). Further simulations will seek to maximize the d_2 value in order to eradicate the second population too.

V. Conclusion

The PUN approach provides both interesting analytical results and an easy way to model the tumor growth also in more realistic, non homogeneous cancers.

A two populations model could explain the behaviour of the cancer during treatment and allows us to simulate different scenarios in order to choose the best therapy for the patient. Further investigation is required to estimate the parameters values using real data and to understand more into details the role of the critical times t*, in order to find, whenever possible, the 'best time for the therapy'. Moreover, in real situations *d* and *m* are time dependent and the functions *d(t)* and *m(t)* will be considered instead of constant parameters.

References

[1] M. Ahmed and L.C. Li, "Adaptation and clonal selection models of castration-resistant prostate cancer: Current perspective," *Int. J Urol.* vol.20, no. 4, pp. 362-371, Apr. 2013.

[2] R. R. Fiñones, J. Yeargin, M. Lee, A. P. Kaur, C. Cheng, P. Sun, C. Wu, C. Nguyen, J. Wang-Rodriguez, A. N. Meyer, S. M. Baird, D. J. Donoghue, and M. Haas, "Early Human Prostate Adenocarcinomas Harbor Androgen-Independent Cancer Cells," *PLOS ONE*, vol. 8, no. 9, e74438, Sep. 2013.

[3] L. Hanin and S. Bunimovich-Mendrazitsky, "Reconstruction of the natural history of metastatic cancer and assessment of the effects of surgery: Gompertzian growth of the primary tumor," *Math. Biosci.*, vol. 247, pp. 47-58, Jan. 2014.

[4] C. Guiot, P. G. Degiorgis, P. P. Delsanto, P. Gabriele, and T. S. Deisboeck, "Does Tumor growth follow a 'universal law?", *J. theor Biol*, vol.225, pp. 147-151, 2003.

[5] P. Castorina, P. P. Delsanto, and C. Guiot, "A Classification Scheme for Phenomenological Universalities in Growth Problems in Physics and Other Sciences," *Phys. Rev. Lett*., vol. 96, 188701, May 2006.

Simulation of Cervical Cancer Response to Radiotherapy*

Christos A. Kyroudis, Dimitra D. Dionysiou, Eleni A. Kolokotroni, Jesper F. Kallehague, Kari Tanderup and Georgios S. Stamatakos, *Member, IEEE*

***Abstract*— The goal of this article is to present basic scientific principles and core algorithms of the simulation module of the CERvical cancer ONCOsimulator (CERONCO) developed within the context of the DrTherapat project (FP7-ICT-600852). CERONCO simulates the response of cervical tumours to radiotherapy treatment (external beam radiotherapy followed by brachytherapy) with concomitant weekly cisplatin, in the patient-individualized context. Results from a preliminary clinical adaptation study based on the MR imaging data of a clinical case are presented as well.**

I. Introduction

The CERvical cancer ONCOsimulator (CERONCO) developed within the context of the DrTherapat project (FP7-ICT-600852) is a predominantly discrete (discrete time - discrete entities - discrete events), clinically oriented, multiscale model of cervical cancer response to radiotherapy treatment (external beam radiotherapy followed by brachytherapy) with concomitant weekly cisplatin in the patient-individualized context. The model stems from previous work of the In Silico Oncology Group (ISOG), Institute of Communication and Computer Systems (ICCS), National Technical University of Athens (NTUA) (e.g. [1],[2]). The version of the simulation model presented in this paper deals with the radiotherapy component of the treatment. Future versions of CERONCO will include a module for cisplatin chemotherapy simulation.

The clinical orientation of the model has been a fundamental guiding principle throughout its development. Available medical data can be exploited in order to strengthen patient-individualized modelling.

A ‘‘top-down’’ simulation approach is formulated; the method starts from the macroscopic imaging data (a high biocomplexity level) and can proceed towards lower biocomplexity levels, exploiting all available clinical and literature information. The communication among the biocomplexity levels is based on the use of information pertaining to lower biocomplexity levels, whenever available, to perturb values of specific biomechanism parameters of higher biocomplexity scales.

* The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement n°270089 (Dr Therapat Project).

C. A. Kyroudis, D.D. Dionysiou, E.A. Kolokotroni and G.S.Stamatakos (corresponding author) are with the *In Silico* Oncology Group, Institute of Communication and Computer Systems, National Technical University of Athens, Greece. (chriskyr@central.ntua.gr; dimdio@esd.ece.ntua.gr; ekolok@mail.ntua.gr; gestam@central.ntua.gr; corresponding author phone: 00302107722287; fax: 0030210 772 3557; e-mail: dimdio@esd.ece.ntua.gr).

J.F. Kallehauge is with the Department of Medical Physics, Aarhus University Hospital, Denmark. (e-mail: jespkall@rm.dk; karitand@rm.dk).

K.Tanderup is with Department of Clinical Medicine, Aarhus University, Denmark and Department of Oncology, Aarhus University Hospital, Denmark (email: karitand@rm.dk)

II. Core Algorithms

A. Equivalence Classes and Cytokinetic model

The tumour region (Gross Tumour Volume, GTV) as defined after adequate processing of the imaging data is represented by a three-dimensional discretization mesh, i.e. a grid of groups of cells. The elementary volume of the mesh is called geometrical cell (GC). Each GC of the tumour accommodates initially a number of biological cells (NBC), which is defined based on typical solid tumour cell densities (e.g. 10^9 cells/cm^3) [3], unless more specific spatial information for a particular tumour will be available (e.g. through DW-MRI).

Each GC is essentially a cluster of heterogeneous cells. A finite number of cell states (categories) and a set of biological and geometrical cell evolution and interaction rules are defined in order to simulate the behaviour of all tumour cells residing in the mesh, i.e. the transitions between these states, as well as cell movement throughout the tumour volume. The output of CERONCO at any given instant is the distribution of the tumour cells of the various categories throughout the tumour region. Various quantities of interest (e.g. tumour volume, growth fraction, hypoxic fraction, total number of living tumour cells, total number of dead cells, etc.) can be calculated and presented as output at various time points throughout a simulation.

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 proliferating cell cycle phases (G1: Gap 1 phase, S: DNA Synthesis phase, G2: Gap 2 phase, M: Mitosis) and the dormant (G0) phase constitute subclasses in which stem or LIMP cells may reside.

For any given instant the biological cells belonging to the same cell category and cell cycle phase within a given GC are assumed synchronized. However, biological cells belonging to different GCs or to different categories and cell cycle phases within the same GC are not assumed synchronized.

“Fig.1” depicts a cytokinetic model, which resulted from previously developed cytokinetic models by ICCS after adequate adaptation, and dictates the transitions between cell categories with a time step of one hour. The cytokinetic model incorporates several biological phenomena that take place at the cellular level: Cycling of proliferating cells

through the successive cell cycle phases G1, S, G2, M; Symmetric and asymmetric modes of stem cell division; Terminal differentiation of committed progenitor cells after a number of mitotic divisions; Transition of proliferating cells to the dormant phase, G0, due to inadequate supply of oxygen and nutrients; Reentering of dormant G0 cells into the active cell cycle due to local restoration of oxygen and nutrient supplies; Spontaneous apoptosis; Necrosis of inadequately nourished tumour cells; Irradiation induced cell death through necrosis; Chemotherapy treatment induced cell death through apoptosis.

A concise description of the corresponding model parameters is given in Table I. Most importantly, the model parameters are not arbitrarily defined, but are related to the specific biological mechanisms listed above, thereby enabling the independent handling and study of each of the above phenomena.

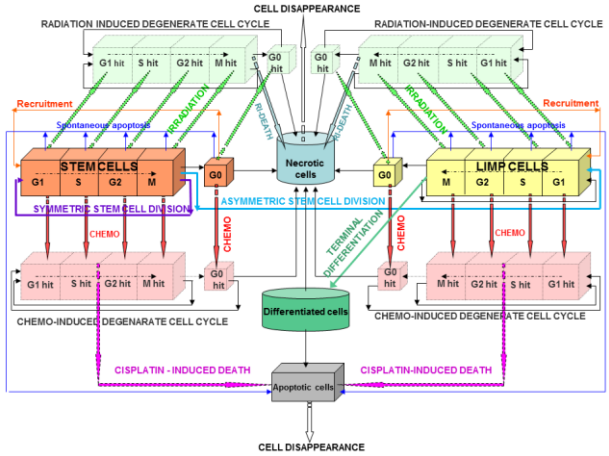


Figure 1. CERONCO's cytokinetic model for cervical tumour growth and/or response to radiotherapy and chemotherapy treatment. Within CERONCO the cytokinetic model functions in intervals of 1h. Arrows represent possible transitions between cell states, which are determined by corresponding model parameters (see Table I). LIMP: LImited Mitotic Potential cells. G1: Gap 1 phase. S: DNA synthesis phase. G2: Gap 2 phase. M: Mitosis. G0: dormant phase. Hit: cells lethally hit by irradiation/drug. RI-death: Radiation Induced death.

TABLE I. TUMOUR DYNAMICS PARAMETERS

Parameter symbol (units)	Description	Value resulting from model adaptation (Range reported in literature)	Indicative References
T_C (h)	Cell cycle duration	70 (16-70)	[4] [5] [6]
T_{G0} (h)	G0 (dormant) phase duration, i.e. time interval before a dormant cell enters necrosis	364 (96-240)	[5]
T_N (h)	Time needed for necrosis to be completed and its lysis products to be removed from the tumour	256 (52-2215)	[7]
T_A (h)	Time needed for apoptosis to be completed and its products to be removed from the tumour	6 (0-25)	[8]
R_A (h^{-1})	Apoptosis rate of living stem and LIMP tumour cells (fraction of cells dying through apoptosis per h)	0.001	
R_{NDiff} (h^{-1})	Necrosis rate of differentiated tumour cells	0.0001	
R_{ADiff} (h^{-1})	Apoptosis rate of differentiated tumour cells	0.00085	
P_{G0toG1} (h^{-1})	Fraction of dormant (stem and LIMP) cells that re-enter cell cycle per time unit	0.01	
P_{sleep}	Fraction of cells entering the G0 phase following mitosis	0.365	
P_{sym}	Fraction of stem cells that perform symmetric division	0.76	
N_{LIMP}	Number of mitoses performed by LIMP cells before becoming differentiated	16 (up to 18)	[9]
α_P (Gy^{-1}), β_P (Gy^{-2})	Radiosensitivity LQ model parameters for the cell cycle phases G1,G2, M (values for S phase and G_0 cells are derived as described in section IIIB)	0.3 (0.03-0.7), 0.03(0.001-0.05)	[10]
OER	Oxygen Enhancement Ratio	2 (1.4-2.5)	[11]
$T_{1/2}$ (h)	Sublethal damage repair half-time	1.5 (0.26-5.7)	[11]

B. *External Beam Radiotherapy Treatment (EBRT) Modelling*

Cell killing by EBRT treatment can be described by the Linear Quadratic or LQ Model, which is widely used in the pertinent literature [3],[10] and constitutes the modelling basis for EBRT therapy in CERONCO:

$$SF(d) = \exp[-(\alpha d+\beta d^2)] \quad (1)$$

where SF(d) is the surviving fraction after a single (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 and correspond to the amount of lethal and sub-lethal cell damage, respectively.

The degenerate cell cycle depicted in "Fig.1" is used in order to simulate the experimental observation that lethally hit

cells usually complete a small number of mitoses (e.g. two mitotic divisions) before ultimate death and removal from the tumour through a lysis process [3].

There is evidence that cell radiosensitivity varies throughout the cell cycle [3]. The S phase is regarded as the most resistant. CERONCO currently assigns different values of radiosensitivity parameters for the following cases: proliferating cells in the G1, G2, or M phase (α_P, β_P); hypoxic cells in G0 ($\alpha_{G0}=\alpha_{prolif}/OER$, $\beta_{G0}=\beta_{G0}/OER^2$) [11]; proliferating cells in the S phase ($\alpha_s=0.6\alpha_{prolif}+0.4\alpha_{G0}$, $\beta_S = 0.6\beta_{prolif}+0.4\beta_{G0}$).

C. Brachytherapy Treatment (BT) modelling

The basis of tumour response to BT treatment modelling is the modified Linear Quadratic model with correction for incomplete repair of Pulsed Dose Brachytherapy [10]. Considering a fraction of pulsed brachytherapy consisting of N pulses of dose d and an inter-pulse interval of the order of one hour, sub-lethal damage may not completely be repaired and the cell survival fraction is given by:

$$SF_N(d)=\exp[-(\alpha Nd+\beta NG_N d^2)] \quad (2)$$

where G_N, the Lea and Catcheside factor, is calculated from the temporal characteristics of the dose distribution. Equation (2) assumes monoexponential repair kinetics of the beta component of radiation damage.

D. Tumor spatiotemporal initialization and evolution

The basis for the initialization of the mesh is the provided patient-specific imaging data. The modelling technique permits the consideration of spatially-varying tumour and treatment characteristics. For example, it supports the division of 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 on the other hand, no spatial information for any tumour characteristics is available through the imaging data, the analysis proceeds considering homogeneous tumours.

The cells residing within each GC of the mesh at the start of a simulation are distributed into the five classes and subclasses mentioned above. The initial distribution of the proliferating cells throughout the cell cycle phases (G1, S, G2, M) is assumed to be proportional to the corresponding cell cycle phase durations.

At each time step the discretizing mesh is scanned and the basic rules that govern the spatiotemporal evolution of the tumour are applied for each GC of the mesh. Practically, each complete scan can be viewed as consisting of two mesh scans, as summarized below.

The first mesh scan aims at updating the state of each GC, by applying the rules of the cytokinetic model of "Fig.1". The second mesh scan deals with the geometrical aspects of the simulation problem, governing the movement of cells throughout the tumour region. It serves to simulate tumour expansion or shrinkage, based on the principle that, 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.

III. A PRELIMINARY ADAPTATION STUDY

This section presents a preliminary adaptation study based on a clinical data set provided in the context of the DrTherapat project (preprocessed T2W-MRI data, EBRT and BT treatment data, other clinical data) corresponding to a squamous cell carcinoma cervical tumour, treated with EBRT (45Gy in 25 fractions, five fractions per week, no irradiation during weekends) with concomitant weekly cisplatin (5 cycles, starting week 1), followed by two PDR-BT fractions (20 pulses each, 1h inter-pulse interval, pulse duration provided in spatial dose distribution raw files). The available MRI data consisted of the GTV tumour at the time of diagnostic MRI (13 days before start of EBRT) and the GTV tumour accompanied by the spatial distribution of the dose at the start of the 1st and the 2nd BT fractions.

It should be noted that the parametric investigation performed up to now is only of an indicative nature, since more patient data sets will be needed before any systematic simulation results interpretation and adaptation effort can begin. In addition, the cisplatin simulation module will be included in the next version of CERONCO, and therefore in all subsequent simulations the effect of cisplatin is assumed to be included into the cytotoxic action of radiotherapy.

A thorough literature review preceded the simulation study so as to define -in conjunction with accumulated basic science and clinical experience- reference values and plausible value ranges of the various model parameters (see Table I) for the case of cervical cancer treatment addressed within the project. Based on the preceding extensive literature review and on the available clinical data, the primary goal was to derive, by appropriate selection of the model parameters, a virtual tumour in agreement with the imaging data in terms of treatment-induced volume reduction while at the same time keeping all the parameter values and resulting tumour characteristics within the constraints of the studied literature. In Table I the parameter values resulting from the adaptation of the model to the provided imaging data in terms of tumour volume are given, as well as the parameter value ranges considered biologically plausible based on the literature review.

Table II. presents the comparison of the simulated tumour volume reduction percentage with the clinical one, at the two available time points (start of first BT fraction, start of second BT fraction). The simulation results are particularly satisfying at the first time point (start of first BT fraction), when the deviation of the virtual and clinical volume reduction percentage is only 0.6%. At the subsequent time-point (start of 2nd BT fraction) the deviation is 13.6%. The virtual tumour has the following characteristics at the pre-therapy time point, which are also in accordance with the relevant literature: volume doubling time = 45d [12],[13], growth fraction (percentage of proliferating cells) = 15% [14], stem cell fraction = 6% [15], hypoxic fraction = 30% [16].

"Fig.2" represents the evolution over time of several tumour populations of interest. The approximately exponential therapy-induced tumour regression is evident, in accordance with relevant literature [17]. Overall, although additional data are needed in order to reach any safe conclusions, biologically and clinically relevant tumour behaviour is observed.

TABLE II. PRELIMINARY ADAPTATION STUDY. TUMOUR VOLUME REDUCTION PERCENTAGE

	Start of 1st BT fraction	Start of 2nd BT fraction
MRI Imaging data	76.2%	93.7%
Simulation	75.6%	80.1%

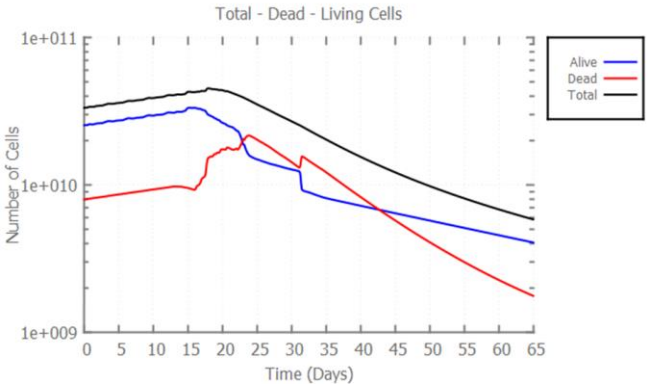


Figure 2. Evolution over time of the total, living and dead tumour cell populations. Pre-therapy MRI scan: day 0, EBRT start: day 13, 1st BT fraction start: day 56, 2nd BT fraction start: day 62.

IV. CONCLUSION

CERONCO simulates the response of cervical tumours to radiotherapy treatment (external beam radiotherapy followed by brachytherapy), in the patient-individualized context. Results from a preliminary adaptation study for a cervical cancer test case have been presented, which generally indicated biologically and clinically relevant tumour behaviour. The acquisition of more patient data sets with imaging data at several time points throughout treatment, during the next phase of the project, will render systematic adaptation studies feasible and will permit a thorough investigation of the interplay of the involved biological mechanisms, which often cannot be grasped intuitively.

ACKNOWLEDGMENT

The authors thank Dr. Steffen Renisch and Dr. Sven Prevrhal from Philips Technologie GmbH Innovative Technologies, for providing coordination related support in the context of the DrTherapat project.

REFERENCES

[1] G. Stamatakos, E. Georgiadi, N. Graf, E. Kolokotroni, and 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*, vol. 6, no.3, e17594, March 2011.

[2] D. Dionysiou, G. Stamatakos, N. Uzunoglu, and K. Nikita, "A computer simulation of in vivo tumour growth and response to radiotherapy: new algorithms and parametric results," *Comp Biol Med*, vol. 36, pp. 448-464, 2006.

[3] G. Steel. Basic Clinical Radiobiology. London: Arnold, 2002. pp. 9-10

[4] R.A. Britten, H.M. Warenius, A.V. Carraway, D. Murray, Differential modulation of radiosensitivity following induction of cis-platinum resistance in radiation-sensitive and radiation-resistant human tumor cells. Rad Oncol Inv 2:25-31, 1994.

[5] R.E. Durand and E. Sham, "The lifetime of human tumor cells," *Int. J. Radiat. Oncol. Biol. Phys*., vol. 42, no. 4, pp. 711-715, Nov. 1998.

[6] R.W. Tsang, A.W. Fyles, M. Milosevic, A. Syed, M. Pintilie, W. Levin, L.A. Manchul, Interrelationship of proliferation and hypoxia in carcinoma of the cervix. Int J radiat Oncol Biol Phys 46(10, 95-99, 2000.

[7] Z. Huang, NA. Mayr, WTC Yuh. Prediciting outcomes in cervical cancer: a kinetic model of tumor regression during radiation therapy. Cancer Res 2010; 70:463-470.

[8] B. Ribba, T. Colin, S. Schnell (2006) A multiscale mathematical model of cancer, and its use in analyzing irradiation therapies. Theor Biol Med Model 3:7. doi:10.1186/1742-4682-3-7.

[9] S. Bernard, J. Belair, MC. Mackey, Oscillations in cyclical neutropenia: new evidence based on mathematical modeling. J theor Biol 223, 283-298, 2003.

[10] R. Pötter, C. Haie-Meder, E. Van Limbergen, I. Barillot, M. De Brabandere, J. Dimopoulos, I. Dumas, B. Erickson, S. Lang, A. Nulens, P. Petrow, J. Rownd, C. Kirisits, Recommendations from gynaecological (GYN) GEC ESTRO working group (II): Concepts and terms in 3D image-based treatment planning in cervix cancer brachytherapy -3D dose volume parameters and aspects of 3D image-based anatomy, radiation physics, radiobiology. Rad Oncol 2006; 78:67-77.

[11] D. J. Carlson, R. D. Stewart, V. A. Semenenko (2006a). Effects of oxygen on intrinsic radiation sensitivity: A test of the relationship between aerobic and hypoxic linear-quadratic (LQ) model parameters. Med Phys 33(9): 3105-3115.

[12] R. M. Wyatt, A. H. Beddoe, R. G. Dale, The effect of delays in radiotherapy treatment on tumour control. Phys Med Biol 48: 139-155, 2003.

[13] G. M. Zharinov , V. A. Gushchin, The rate of tumor growth and cell loss in cervical cancer. Vopr Onkol 1989; 35(1):21-5 [article in Russian].

[14] E. L. Levine, A. Renehan, R. Gossiel, S. E. Davidson, S. A. Roberts, C. Chadwick, D. P. Wilks, C. S. Potten, J. H. Hendry, R. D. Hunter, CML West. Apoptosis, intrinsic radiosensitivity and prediction of radiotherapy response in cervical carcinoma. Rad Oncol 1995; 37: 1-9.

[15] S-L Zhang, Y-S Wang, T Zhou, X-W Yu, Z-T Wei, Y-L Li. Isolation and characterization of cancer stem cells from cervical cancer HeLa cells. Cytotechnology 64:477-484, 2012

[16] A. W. Fyles, M. Milisevic, Wong R. MC Kavanagh, Pintilie M, Sun A, Chapman W, Levin W, Manchul L, Keane TJ, Hill RP. Oxygenation predicts radiation response and survival in patients with cervix cancer. Rad Oncol 48: 149-156, 1998.

[17] A. Huang, N. A. Mayr, M. Gao, S.S. Lo, J. Wang, G. Jia, WTC Yuh. Onset time of tumor repopulation for cervical cancer: first evidence from clinical data. Int J Radiat Oncol Biol Phys 84(2), 478-484, 2012

A Model of Tumor Growth Coupling a Cellular Biomodel with Biomechanical Simulations*

Farhad Rikhtegar, Eleni Kolokotroni, Georgios Stamatakos and Philippe Büchler

***Abstract*— The aim of this paper is to present the development of a multi-scale and multiphysics approach to tumor growth. An existing biomodel used for clinical tumor growth and response to treatment has been coupled with a biomechanical model. The macroscopic mechanical model is used to provide directions of least pressure in the tissue, which drives the geometrical evolution of the tumor predicted at the cellular level. The combined model has been applied to the case of brain and lung tumors. Results indicated that the coupled approach provides additional morphological information on the realistic tumor shape when the tumor is located in regions of tissue inhomogeneity. The approach might be used in oncosimulators for tumor types where the morphometry information plays a major role in the treatment and surgical planning.**

I. Introduction

Modeling the evolution of the tumors inside the brain and the lung is expected to significantly support the optimization of treatment planning and delivery. At the same time it can provide an improved understanding of the underlying multi-scale mechanisms of tumor development and dynamics. The major modeling approaches include reaction-diffusion models and discrete entity – discrete event cellular level based simulations focusing on cell cycling, necrosis, apoptosis, etc. Although these models are able to provide a detailed description of the cellular evolution of the tumor, they neglect the mechanical component and, for the cellular models, assume a conformal expansion or shrinkage of the tumor. These approaches may provide a plausible first approximation regarding tumor morphology, but may also be accompanied by an error in the detailed tumor shape prediction. For these reasons, the present work investigates the inclusion of biomechanical information to better predict the spatial distribution of the tumor cells. A macroscopic model of the tumor biomechanics will be presented. This macroscopic model has been coupled with a detailed biomodel able to describe the cellular evolution of the tumor and its reaction to treatment.

*The research leading to these results has received funding from the European Union's Seventh Programme for research, technological development and demonstration under grant agreement No 600841 (CHIC Project).

F. Rikhtegar was with the Institute for Surgical Technology & Biomechanics, University of Bern, Stauffacherstrasse 78,3014 Bern, Switzerland (e-mail: f.rikhtegar@gmail.com).

E. Kolokotroni is with the Institute of Communication and Computer Systems, National Technical University of Athens, 9, Iroon Polytechniou 157 80, Greece (e-mail: ekolok@mail.ntua.gr)

G. S. Stamatakos is with the Institute of Communication and Computer Systems, National Technical University of Athens, 9, Iroon Polytechniou 157 80, Greece (e-mail: gestam@mail.ntua.gr)

P. Büchler is with the Computational Bioengineering Group, Institute for Surgical Technology & Biomechanics, University of Bern, Stauffacherstrasse 78,3014 Bern, Switzerland (corresponding author phone: +41 31 631 5959; e-mail: philippe.buechler@istb.unibe.ch).

II. Materials and Methods

A. Biomechanical model

The biomechanical model of the tumor and healthy tissues relies on the theory of continuum mechanics. When an elastic body is deformed by an external force, the work done by the force is stored in the form of strain energy within the body. When the material undergoes a small strain, the Kirchoff-St Venant hyper-elastic model can be considered. The strain-energy function for this model, W, is presented by:

$$W = \frac{\lambda}{2}(\mathrm{Tr}\ \mathbf{E})^2 + \mu \mathrm{Tr}(\mathbf{E}^2) \tag{1}$$

where **E** is the Green-Lagrange strain tensor, λ and μ are the Lamé constants which represent the mechanical properties of the material. These constants can be directly related to the Young's modulus E and Poisson ratio ν:

$$\lambda = \frac{\nu E}{(1+\nu)(1-2\nu)},\ \mu = \frac{E}{2(1+\nu)} \tag{2}$$

For an isotropic elastic body, the stress vector **S** is calculated as a derivative of energy density function with respect to Lagrange-Green strain tensor:

$$\mathbf{S} = \lambda \mathrm{Tr}\mathbf{E}\ \mathbf{I} + 2\mu\mathbf{E}. \tag{3}$$

To model the growth, the volume change is considered to be a uniform-and-isentropic strain added in the normal direction of elastic formulation for each element. The growth could be interpreted as an internal pressure of the tissue. The strain formulation is updated to:

$$\mathbf{E} = \mathbf{E}_{\mathrm{mech}} - E_{\mathrm{growth}}\mathbf{I}. \tag{4}$$

E_{growth} represents the mechanical strain generated in the tissue by the change of volume of the tumor. The stress and strain calculations in the tissues have been done using the finite element method. The open source software FEBio developed by the University of Utah was used. A modified material model was implemented as a user subroutine to calculate the stress distribution in the tissue resulting from the change in volume of the cancerous tissues.

An automatic smooth mesh generator was developed to generate a finite element model from segmented medical images. An image-based voxel-mesh algorithm was used to generate the computational domain. The approach has been chosen because it can be fully automated and is robust. However, in order to avoid the jagged edges on boundaries, which can result from this method [1, 2], a smoothing step was added to the procedure. The distorted elements are split

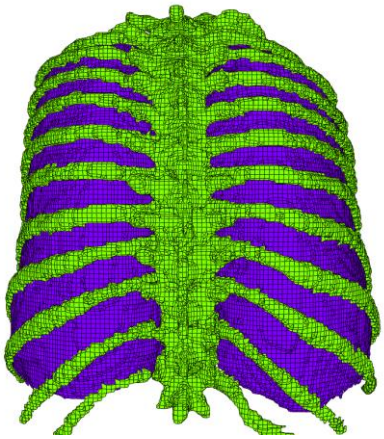


Figure 1. Finite element mesh of the lung. The mesh of about 250,000 elements was automatically generated from the segmented medical images.

into prisms to significantly improve the quality of the near-boundary mesh at a comparably – low computational cost. The outer surface of the mesh is extracted and smoothed according to the geometric signal processing approach of Taubin [3]. This algorithm is based on a Fourier-like decomposition of the geometry, calculated from the Laplacian operator. The degree of smoothing (defined here as smoothing frequency, k) is mainly limited by the fact that inverted elements might appear with extensive smoothing. To improve the quality of the mesh and avoid distorted elements, hexahedral elements with a large angle between faces are divided into prism elements. This strategy aims at preventing the large angle between faces of hexahedrons that produce inaccuracies in the formulation of the element. The number of elements in the mesh can be controlled by an appropriate resampling of the input images (Fig. 1). This approach was successfully applied on brain and lung cases, including tumor..

B. Cellular level simulator

The cellular level simulation for solid tumor free growth and response to therapy has been previously described [4-8]. In summary, the modeling approach is discrete in space and time and based on the concept of cellular automata. The model has been developed to support and incorporate individualized clinical data such as imaging data including the definition of the tumor contour and internal metabolic tumor regions, histopathologic and the genetic data.

The algorithmic approach is outlined as follows: The anatomic region of interest is discretized by a virtual cubic mesh of which the elementary cube is termed geometrical cell and corresponds to a volume of 1 mm^3. The geometrical cells belonging to the tumor are initiated with a cluster of 10^6 heterogeneous cancerous biological cells (typical solid tumor cell densities 10^9 cells/cm^3 [9]). A hypermatrix corresponding to the anatomic region of interest is subsequently defined, which describes the local biological, physical and chemical dynamics of the region based on the available patient-specific medical data. Specific details regarding the mathematical treatment of the imageable versus the non-imageable part of the tumor are available in [4].

C. Multi-scale coupling

The macro-scale biomechanical model of tissue described above was combined with the cellular-level simulator. The microscopic cell-level model provides the local concentration of tumor cells to the macroscopic biomechanical solver. In addition, the macroscopic model simulates the mechanical stresses developed within the organ, while a cancerous tumor is growing inside it. The distribution of the mechanical stresses is then exported back into cellular simulation to precisely predict the direction of the tumor cells proliferation.

The code integration is done to establish a self-consistent computing tool to couple the cellular-level model with the macroscopic biomechanical solver. The cellular-level model requires the direction, along which the cell proliferation and tumor growth will happen. This information is provided by the mechanical model calculating the direction against which the cells sense a minimum pressure [10, 11]. On the other hand, the biomechanical solver needs the information for the growth of the tumor and the cell number inside the element to calculate the stress distribution as a result of tumor growth. The biological microscopic model provides the cell number information for the biomechanical solver [10]. Since the computational grids differ for the calculations performed at macroscopic and microsopic scales, spatial interpolation is required to transfer the results from one scale to the other. Efficient interpolation techniques are utilized in order to avoid significant accumulation of errors as a result of interpolation.

The FEBio software is used to model the mechanical interaction of the tumor and the lung tissue. A linear relationship was defined between the reference cell concentration, which has been defined to be 10^6 cells/mm^3, and volumetric growth of the tumor:

$$E_{\text{growth}}(c) = \sqrt[3]{\frac{c}{c_{\text{ref}}}} - 1 \tag{5}$$

where c is the concentration of tumor cell, c_{ref} corresponds to the reference concentration and V_0 represents the volume at the reference concentration c_0 (here set to 3 10^6 cells mm^{-3}) .

The mechanical pressure in the tissue calculated using the finite element simulation was used to guide the evolution of the tumor at the cellular level. Each element of the cellular model is given the value of the pressure calculated as the trace of the stress tensor:

$$p = \frac{1}{3}\text{Tr}\ \mathbf{S} \tag{6}$$

A map containing the directions of less pressure for each element can then be build such as the negative gradient:

$$\vec{d} = -\frac{\nabla p}{\|\nabla p\|} \tag{7}$$

The simulation of the cellular model only takes a couple of minutes while each biomechanical calculation take about 20 minutes (for a problem of about 600,000 unknowns). Therefore, most of the calculation time is taken be the mechanical simulation, which should only be executed when a significant change in the concentration has taken place. For this reason, the biomechanical simulation was only executed

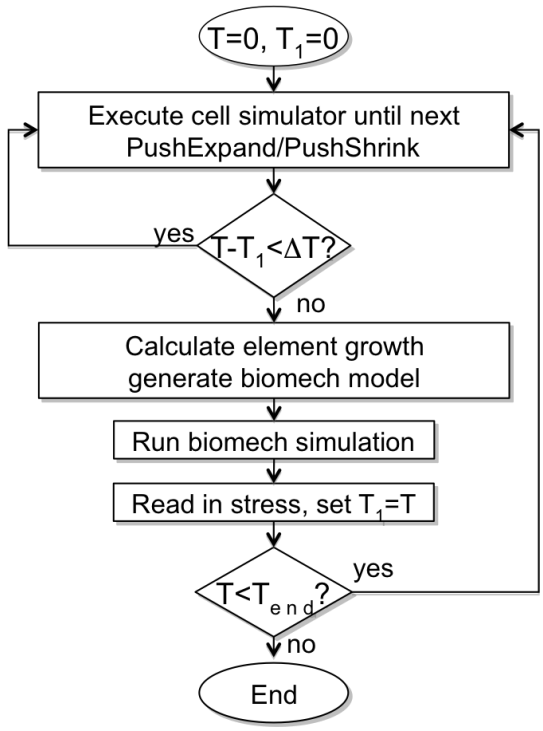


Figure 2. Flowchart of the diffusion/mass effect coupling simulation used to simulate tumor growth. Since the cellular effects occur at a different time scale than the mechanical reaction, the biomechanical calculations occur only after a larger number of cycles of the cell simulator biomodel.

after several iterations of the cellular simulation. Between successive evaluations of the mechanical state of the tissue, a constant map of pressure gradient is used by the cellular simulator to determine the direction of expansion/shrinking of the tumor (Fig. 2). The frequency of the update is difficult to determine precisely, but it seems than one update of the biomechanical status every week is sufficient. This limitation is not critical for the overall accuracy, since the time scale of the process occurring at different scale differ significantly.

III. Results

The coupled model has been applied to lung and brain cancer cases. The mechanical stresses and displacements due to the presence of realistic shaped tumor in the tissue were calculated (Fig. 3).

If we compare the results of mechanical interaction of tumor growth inside the lung with the brain tissue, more symmetric distribution of variables are observed in the lung. This is mainly because of uniformity in the lung tissue and the fact that the boundaries are located quite far from where the tumor grows. On the other hand, inside the brain the stress and displacement magnitudes are very asymmetric due to the different properties for gray and white matter as well as due to the ventricles filled with cerebrospinal fluid (CSF). As shown in [11], the tumor growth is sensitive to the boundaries of the organs and to spatial change of the mecha nical properties. In these cases, a non-spherical shape must be considered during the simulations.

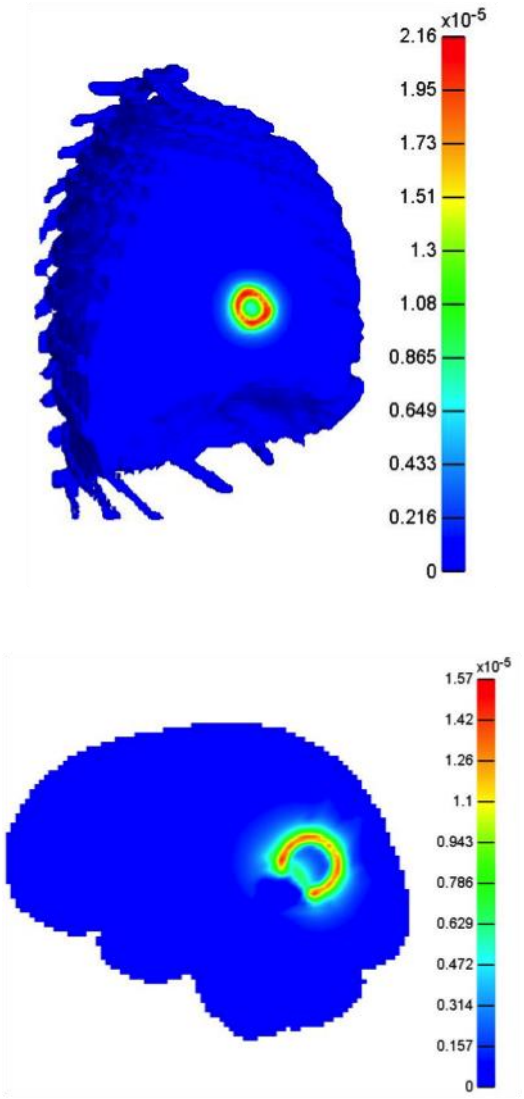


Figure 3. Stress distribution due to the growth of the tumor inside the lung (top) and brain (bottom). The stress distribution is more asymmetric in the brain due to non-uniformity of the domain and proximity of the boundaries

IV. Discussion

A framework to combine biomechanical simulations with cellular modeling has been proposed. This approach enables to include macroscopic information concerning the tumor position within the cellular model, which otherwise would produce perfectly symmetrical tumor evolution. Results showed that the location of the tumor within the organ affects its degree of symmetry during evolution.

This initial approach indicated that the coupling is feasible and that the total simulation time remains within acceptable limits (a few hours). However, the mechanical formulation contains several limitations. First, the tissue is modeled as isotropic, while it is clear that the fiber orientation in the brain will affect the stress response. In addition, the mechanical simulations only consider the mass effect, neglecting the infiltration of the tumor cells within the neighboring tissues. In order to solve this problem, the mechanical simulation should be coupled with a model of the diffusion of the cells in

the tissue following the reaction/diffusion equation for tumor growth:

$$\frac{dc}{dt} = \nabla \cdot (D\nabla c) + S(c,t) - T(c,t) \tag{8}$$

where D is the diffusion tensor indicating the preferential direction of diffusion in the tissue, and S(c,t) and T(c,t) are the source and sink terms for tumor cell proliferation and therapy-related cell death, respectively. This additional simulation step could be performed after updating the cell concentration and before evaluating the mass effect using the biomechanical model.

Finally, the Lagrangian formulation used in our finite element approach reaches its limitations under excessive growth of geometrical cells. Therefore, further work is required in the context of large deformations. This can be solve either using re-meshing techniques or by implementing the mechanical simulation using an Eulerian formulations of the problem, where the nodes of the mesh remain fixed and material has to be advected according to calculated displacements.

REFERENCES

[1] S. K. Boyd and R. Müller, "Smooth surface meshing for automated finite element model generation from 3D image data," *Journal of Biomechanics*, vol 39, no. 7, pp. 1287-1295, 2006.

[2] G. Taubin, "A signal processing approach to fair surface design," in *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*, pp. 351-358, 1995.

[3] G. Stamatakos,, "In silico oncology: PART I Clinically oriented cancer multilevel modeling based on discrete event simulation," in *Multiscale Cancer Modeling* T. Deisboeck and G. Stamatakos, Eds.,.Boca Raton: Chapman & Hill/CRC, 2011

[4] D. Dionysiou, G. Stamatakos, N. Uzunoglu, K. Nikita, and A. Marioli,, "A four dimensional in vivo model of tumour response to radiotherapy: parametric validation considering radiosensitivity, genetic profile and fractionation," *J. Theor. Biol.*, vol 230, pp. 1-20, Sep. 2004

[5] G. Stamatakos, V. Antipas, and N. Uzunoglu, "A spatiotemporal, patient individualized simulation model of solid tumor response to chemotherapy in vivo: the paradigm of glioblastoma multiforme treated by temozolomide," *IEEE Trans. Biomed. Eng.*, pp. 1467-1477, Aug. 2006

[6] G. Stamatakos, D. Dionysiou, E. Zacharaki, N. Mouravliansky, K.Nikita, and N. Uzunoglu, "*In silico* radiation oncology: combining novel simulation algorithms with current visualization techniques," *IEEE Proc. Special Issue on Bioinformatics: Advances and Challenges*. vol. 90, pp. 1764-1777, Nov. 2002.

[7] G. Stamatakos, E. Kolokotroni, D. Dionysiou, E. Georgiadi, and C. Desmedt, "An advanced discrete state discrete event multiscale simulation model of the response of a solid tumor to chemotherapy: mimicking a clinical study," *J. Theor. Biol.*, vol 266, pp. 124-139, Sep. 2010.

[8] S. Sell, "Stem cell origin of cancer and differentiation therapy," *Crit. Rev. Oncol. Hematol.*, vol 51, pp.1-28, July 2004.

[9] S. Bauer, C. May, D. Dionysiou, G. Stamatakos, P. Büchler, and M. Reyes, "Multiscale Modeling for Image Analysis of Brain Tumor Studies," *IEEE Transactions on Biomedical Engineering*, vol 59, no 1, pp. 25-29, Jan. 2012.

[10] S. Bauer, H. Lu, C. May, L. Nolte, P. Büchler, and M. Reyes, "Integrated segmentation of brain tumor images for radiotherapy and neurosurgery," *International Journal of Imaging Systems and Technology*, vol. 23, no. 1, pp. 59-63, Feb. 2013.

[11] C. May, E. Kolokotroni, G. Stamatakos, and P. Büchler, "Coupling biomechanics to a cellular level model: An approach to patient-specific image driven multi-scale and multi-physics tumor simulation," *Progress in Biophysics & Molecular Biology*, vol 107, no 1, pp. 193-199, Oct. 2011.

A collaborative central reviewing platform for cancer detection in digital microscopy images*

I. Karatzanis, A. Iliopoulos, M. Tsiknakis, *Member, IEEE*, V. Sakkalis, and K. Marias, *Member, IEEE*

***Abstract*— Telepathology, the practice of pathology at a long distance, has advanced continuously since 1986. Today, almost 3 decades later, virtual slide telepathology has become a promising tool for providing re-review of surgical pathology cases as part of a quality assurance program but also for educational purposes. In this paper we present the Central Review for Pathology images platform (CRP), developed by the Computational Medicine Laboratory at FORTH-ICS. The CRP is a secure cloud platform, which tries to address current issues that hamper the wider use of virtual pathology. The system offers an easy upgradable multi-format support for virtual slide files from different slide scanner vendors, enhanced collaboration capabilities and scheduling tools, a sophisticated mechanism for defining custom templates for reporting forms which adapts to all user needs and a virtual microscope viewer for the digital slides.**

I. Introduction

Traditionally, laboratories have been exchanging microscope slides and the pathologists had to travel in order to perform a central review. The advent of the digital scanners for microscope slides and the ability to view remotely the digital microscopy images provide new opportunities for central reviewing of pathology data.

Data quality is a central concern in clinical trials because poor data quality can lead to biased estimates of important clinical parameters and compromise the validity of whole studies. In particular, diagnostic accuracy and inter-observer and intra-observer variability can be powerful confounders that weaken the outcome of the analyses [1], [2], [3].

In the following paragraphs we present a collaborative platform for central reviewing of digital pathology images allowing multiple reviewers, which addresses several common problems and offers a novel solution to collaborative telepathology.

* Research supported from the INTEGRATE project funded by the European Commission under the 7th Framework Programme.

I. Karatzanis is with the Computational Medicine Laboratory (CML) of the Institute of Computer Science (ICS) in the Foundation for Research & Technology - Hellas (FORTH), Vassilika Vouton, P.O Box 1385, GR-71110 Heraklion, Crete, Greece (corresponding author phone: +30-2811-391618; fax: +30-2810-391428; e-mail: karatzan@ics.forth.gr).

A. Iliopoulos (ailiop@ics.forth.gr), M. Tsiknakis (tsiknaki@ics.forth.gr), V. Sakkalis (sakkalis@ics.forth.gr) and K. Marias (kmarias@ics.forth.gr) are with the Computational Medicine Laboratory (CML) of the Institute of Computer Science (ICS) in the Foundation for Research & Technology - Hellas (FORTH).

I. Karatzanis and A. Iliopoulos contributed equally to this work.

II. Common Status in Pathology

A. Histopathology Slides

When a patient has a biopsy or a surgery, the surgeon often removes diseased tissue (a "tissue block") in order to be examined by a pathologist. The pathologist will slice the tissue block into very thin layers that are placed on a glass slide and examined under a microscope.

B. Storage

Prepared microscope slides are usually stored in a cool and dark location away from heat and bright light, such as a closed cabinet in a temperature-controlled room. Stained slides naturally fade over time. Keeping them in a cool, dark location helps slow down the process. The slides should be carefully positioned in order to avoid shifting out of position of the cover glass or of the specimen.

C. Sharing

Typically the reviewing of the glass slides is a procedure in which either a pathologist goes to the site where the slides are stored or the slides are sent to the pathologist. This of course is a time consuming and costly procedure and poses the risk of damaging the samples during transport.

III. The Era of Digital Pathology

With the advent of Whole-Slide Imaging, the field of digital pathology has advanced to one of the most promising fields of diagnostic medicine in order to achieve even better, faster and cheaper diagnosis, prognosis and prediction of cancer.

A. The Central Reviewing Platform (CRP)

The Central Review for Pathology (CRP) platform provides all the necessary tools & functionality to assist and speed up the review of digital pathology images by multiple reviewers. The platform promotes the collaboration among pathologists either in real time or by logging and sending notifications, scheduling their tasks and providing communication in many other ways.

B. Digital slides

Using digital slide scanners a digital image file of an entire glass slide is created (whole slide image), and it is stored in a server. The digital slide files are high resolution images which are resistant to being damaged or broken over time (stain fading etc.). A minor downside is that they are relatively large files [4], often exceeding 1 gigabyte in size,

but as the technology advances the price per gigabyte is dropping relatively fast, and their storage cost will soon be trivial. On the other hand there is a gain in laboratory space because thousands of digital slides can be stored in a hard disk. CRP provides an uploading tool, which transfers slide files from compatible scanners (currently two widely used file formats are supported while new ones are relatively easy to implement) and acts as a centralized system for storing and viewing them. The native support of the CRP platform for multiple formats of virtual slides, provides an important advantage as it frees laboratories from being bound to a specific slide scanner manufacturer and its proprietary format.

C. *Remote access & viewing*

The digital slide files stored in the CRP platform can be navigated remotely, over the Internet (or in an intranet), using an html 5 compatible browser. Security is a critical point in medical systems, and for that reason CRP uses authentication for all the users of the platform, through secure connection. Additionally, the viewer is based on a tiles mechanism which displays only the necessary image information fully anonymized, and therefore although one can navigate through the digital slide image like operating a virtual microscope, the information remains secure in the server, and it is not downloaded on the client side. Only authorized users are permitted to download a histopathology image.

Fig.1 depicts the tiles mechanism of the interactive virtual microscope viewer of the CRP platform, showing the converted digital slide file, composed of tiles of PNG images at different resolutions that make up an image pyramid. The tile size is typically 256x256. This procedure reduces the time required for initial load by downloading only the region being viewed and/or only at the resolution it is displayed at. Subsequent regions are downloaded as the user pans to (or zooms into them); animations are used to hide any jerkiness in the transition.

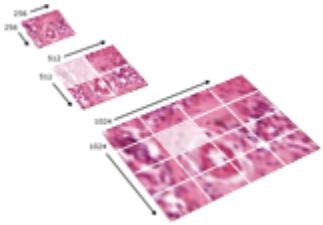


Figure 1. A converted slide image

While navigating the user can insert various types of markers on the pathology image, mark regions of interest (RoIs) using different types of shapes (basic and free form) and add some information (annotations) in the form of free text on those RoIs (Fig. 2). All these elements can be saved in separate layers on the image and can be re-used if needed.

D. *Remote Collaboration*

The platform offers a powerful collaborative environment that promotes communication between histopathologists. The system supports messaging and email, notifications & alerts, real-time discussion among the reviewers who are simultaneously on the same slide image, as well as annotating and highlighting RoIs. If the CRP is configured to be securely available from the internet, then the users may connect from anywhere, and even from remote laboratories (even in other countries) as long as they are registered in the system.

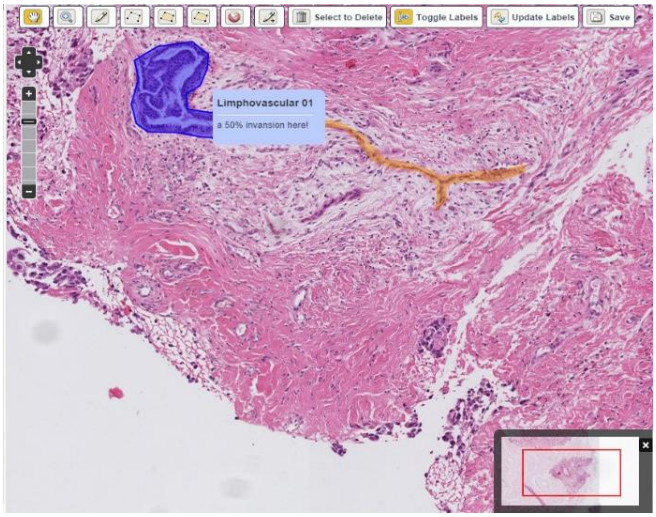


Figure 2. Screenshot of the viewer with an active annotation layer

There is no need for traveling or for packing and for sending slides to other laboratories, and in comparison with the traditional ways, everything is instantaneous, secure and more productive.

IV. FUNCTIONALITY & HIGHLIGHTS OF CRP

A. *Used technologies*

CRP is developed using Java Servlets, Java Server Faces 2, Java Hibernate ORM and other java based technologies. It is based on Liferay portal [5][1], Primefaces JSF framework [6], jQuery [7], jQuery UI [8], Open Layers [9], OpenSlide [10] and GDAL [11]. The result is a robust platform which provides an interactive html 5 user interface, with minimum hardware requirements from the end users side.

B. *Stakeholders*

The target group of the CRP tool is broad and includes any type of specialists who use digital pathology images. Collaboration capabilities offered by the platform also enable knowledge share and the implementation of diverse business scenarios in flexible way. Some representative cases include histopathologists, microscope laboratories, specialists working in clinical trials or in an academic environment for educational purposes [12].

C. *Roles*

The system is built for providing all the necessary tools for the definition and process of a review protocol. The review process addresses a specific workflow of tasks and incorporates three generic user roles.

1) *Uploader*

An uploader is the user who initially uploads the raw pathology images from a remote location (e.g. a laboratory)

to the CRP platform. This is the primary action required for an image to be available for use in a review protocol.

2) Moderator

A moderator is the user who administers the registration and conduction of a review protocol. His is also responsible for initial system configuration which includes the definition of the custom templates for the review forms.

3) Reviewer

A reviewer is the primary user of the platform, who acquires and completes the review tasks (review a pathology image).

D. Workflow

The following text describes the workflow for the different roles of the platform, in a concise and straightforward manner. For all of the following we assume that the users have been authenticated and have successfully logged in the platform.

A laboratory uploads digital slide files to the CRP platform. Once the slides are uploaded, a tile generation service automatically converts the virtual slide in an anonymized series of PNG images appropriate for displaying though a browser. The newly uploaded images are available to the moderators of the platform who can create a new review protocol using a wizard, in order to assign reviewers to the images. Once the moderator verifies the new review protocol the platform automatically sends emails and notifications in the reviewers, with a report stating their pending tasks. The reviewers log in to the platform, have a quick overview of their tasks from a sortable and searchable task list, they pick one and then they proceed with the review of the image which is associated with the task. The pathologist navigates through the digital slide like using a virtual microscope, marks regions of interest, and fills all the fields of the appropriate report (Fig. 3).

In case of a conflict (which is noted by the moderator) the reviewers can communicate by exchanging messages (even real time) until they resolve the conflict. If there is no conflict among the reviewers the image is archived for future access.

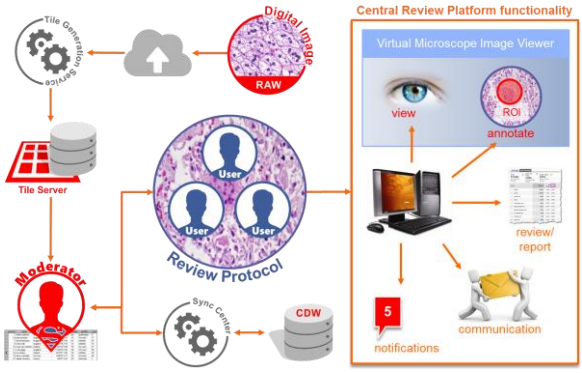


Figure 3. Schematics of the basic workflow of CRP.

E. Highlights

1) Upload mechanism

The CRP platform comes with an interface for uploading digital pathology images to the imaging server, as an independent (parallel) service. The communication protocol used for file transfer is the SSH File Transfer Protocol (SFTP). Uploading mechanism is extendable for including additional communication protocols such as FTP over SSL (FTPS).

2) Custom templates for review forms

The CRP platform enables Moderators to define any type of review protocol. Additionally, all existing types of review protocols can be extended according to any future usage requirements. Moderators can add, update, delete and associate any item (Image Types, Collections & Variables) (Fig. 4 & 5). These elements and their associations define the review forms.

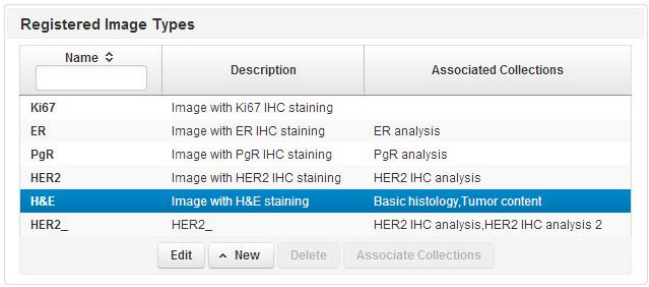


Figure 4. An "H&E" image type can be consisted of collections "Basic Histology" and "Tumor Content"

Image Type: is a specific type of a review protocol and it corresponds to a specific type of digital pathology image. An image type (Fig. 4), consists of one or more collections of variables.

Collection (of Variables): is a set of variables (Fig. 5). Its aim it to group together variables (form fields) into reusable sets of information. A collection of variables can be used by multiple image types.

Variable (or Form Field): is a feature of interest in a review protocol. These variables are the form fields which the reviewer has to score in his/her review task.

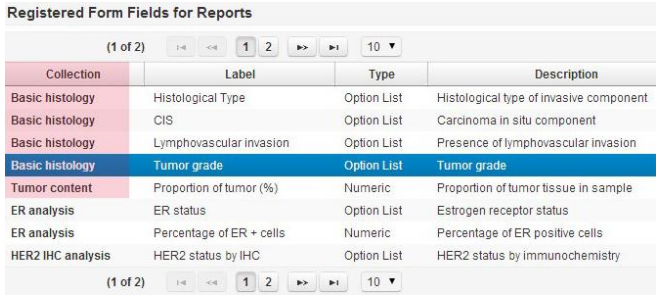


Figure 5. "Basic Histology" collection is consisted of 4 features of interest (form fields) in a review form for a pathology image of Type H&E

3) Creation and management of reviews

CRP tool has a built-in, composite mechanism for the registration of new review protocols.

The platform has also built-in mechanisms for managing a review protocol in its lifecycle by implementing a workflow that defines the steps and the options that are available to the stakeholders.

One of the central modules of the CRP platform assists the moderators to administer registered protocols, allowing them to close or mark tasks with different states, mark specific task features (form fields) for which the reviewers should re-consider their scoring, notify the stakeholders about required actions and close protocols that are completed.

4) Review mechanism

When a reviewer selects a task (an image) to review for the first time, an associated report form is generated. The user interface of the report form consists of a form containing the variables which need to be scored, the image viewer and the annotator.

The form is dynamically generated, based on the features defined during the review protocol type definition process. The reviewers can also access their tasks at any time, e.g. in case of a conflict, for revising an already submitted case.

Image viewer helps the reviewers to perform more accurate scoring, by providing an in detail view of pathology images as well as tools for annotating and sharing their observations (create annotations, define RoIs).

V. Evaluation & Metrics

The evaluation and validation procedure described below focusses not only on technical aspects, but mostly on fitness-of-purpose for the end user. The tool should fit in the envisioned workflow, and provide sufficient performance and functionality to the end user.

A. Evaluation Methodology & Framework

The quantitative evaluation was carried out according to ISO/IEC 25023 [13] of ISO 25000 series International Standard [14].

End-user evaluation of the platform was conducted through a number of selected scenarios covering the anticipated usage of it. The steps which compose each of the scenarios, correspond to criteria that were used to objectively rate the degree of success of CRP's modules.

B. Quantitative Evaluation

The quantitative evaluation session involved 5 pathologists from different countries/institutions.

1) Setup (Evaluation Scenarios)

Prior to the core evaluation session, users: a) were asked to fill in a Consent Form and b) they had a short demo of the CRP tool.

The quantitative evaluation was different for each of the two major roles involved in the platform. Each participant evaluated the platform by executing two evaluation scenarios respectively (as a reviewer and as a moderator).

Each of the two evaluation sessions had the same structure which was composed by the following parts: a) An introductory session with operating instructions and familiarization time, b) Users where asked to perform representative tasks following a specific workflow, c) After performing each task and completing the overall scenario of use, they were asked to fill in the evaluation questionnaire, d) Finally after finishing the evaluation, a discussion took place about users' overall impression, their remarks about platform's weaknesses and their proposals for improvements.

During the test, the user screen and voice was recorded for getting both quantitative and qualitative measures.

2) Representative Tasks

The guided test was aimed to check if the user could easily navigate through the platform and perform all the necessary steps. These steps exploited the core functionality of the platform per user role as also its core business flow and objective. The aim was to simulate the process of:

1. The definition of a review protocol by the moderator.

2. The conduction of a review protocol where a group of pathologists review and annotate the digital pathology images and a moderator administers the whole procedure.

For the two major roles in this procedure, the scenarios were set and the instructions were given per role (table I).

TABLE I. Evaluation Scenario Representative Steps

Reviewer	***Moderator***
TASK 1: Find and view the review task for a specific patient **TASK 2:** Navigate on the digital pathology image & annotate RoIs, related with the scoring **TASK 3:** Based on the observations on the image, proceed to its scoring **TASK 4:** Check that the scoring information is saved	**TASK 1:** Check if there are any new patients (patient images) for participating in review protocols. **TASK 2:** Create a new Review Protocol for a clinical trial. **TASK 3:** Check and manage the answers of the specialists for a specific review protocol. Find any conflicting scoring for a particular patient and inform users that they have to review their scoring

C. Results

The participants had no difficulty in understanding on how to use the platform for accomplishing their tasks. Specific points were the users had some difficulties were noted and will be taken into consideration for future improvements.

1) Usability

The System Usability Scale (SUS) [15][16][17] was used, as a generic tool for measuring the usability. The average of the SUS scores are 77.5 for the tool when used by a Reviewer and 78.5 for the tool when used by a Moderator.

2) Efficiency

Was measured by the tasks completion rates and times. The completion rate and the mean completion time per scenario and task are listed in table II.

Efficiency was also rated in a scale of 5 per task. Moderator's task 3 (review manager module) was the one where users had the most difficulties.

TABLE II. EVALUATION SESSION METRICS PER SCENARIO & TASK

		Task 1	*Task 2*	*Task 3*	*Task 4*
Reviewer	**Completion Rate (%)**	100	100	100	100
	Mean Time (min)	1' 37''	3'	1' 21''	30''
	Efficiency rating (mean)	4.8	4.7	4.6	4.5
Moderator	**Completion Rate (%)**	100	100	100	
	Mean Time (min)	56''	4' 39'	14'	
	Efficiency rating (mean)	5	4.7	3.5	

3) Usefulness

Usefulness was evaluated qualitative by the answers of the post-test questionnaires in a scale of 5. Tool usefulness when used from a "Reviewer" and a "Moderator" was rated respectively (mean values) 4.3 and 4.0.

4) Learnability

Learnability was evaluated by the answers of the post-test questionnaires in a scale of 5. Tool learnability when used from a "Reviewer" and a "Moderator" was rated respectively (mean values) 4.6 and 4.1. Useful qualitative conclusions were also drawn by examining the screen and audio captures.

5) User Satisfaction & Aesthetics

Evaluators were also asked to rate the "look and feel" of the tool in a scale of 5. The mean values of their scores was 4.8 for using the tool as "Reviewer" and 4.5 as a "Moderator".

VI. CONCLUSION & FUTURE PLANS

This paper presented the CRP tool developed to enable remote viewing, scoring and collaboration for digital pathology Based on the evaluation of the users and the feedback gathered by associate pathologists, users from workshops and many experts, CRP has demonstrated a solid foundation as a centralized platform to manage reviews of virtual slide images among multiple reviewers (locally or remotely), with strong collaborative characteristics. The main envisaged application is in multi-centric clinical trials where remote collaboration is mandatory in order to ensure validity of results through multiple reviewers and also reduce the risk of systematic biases from specific sites.

Future work will extend CRP to support a variety of virtual slide file formats from different manufacturers which will enable the platform to also act as a pathology image warehouse and global image viewer.

ACKNOWLEDGMENT

The authors acknowledge support for this work from the INTEGRATE project [18] funded by the European Commission under the 7th Framework Programme (http://www.fp7-integrate.eu/). The authors would like to thank all the collaborators of the project and specially Alexandre Irrthum from the Breast International Group (BIG).

REFERENCES

[1] R. S. Weinstein, M.R. Descour, C. Liang, et al., "Telepathology overview: from concept to implementation", Human Pathology, 2001, vol. 32, pp. 1283-99

[2] A. R. Graham, A.K. Bhattacharyya, K. M. Scott, F. Lian, L. L. Grasso, L. C. Richter, J. T. Henderson, J. B. Carpenter, A. M. Lopez, G. P. Barker, R. S. Weinstein, R. S., "Virtual slide telepathology for an academic teaching hospital surgical pathology quality assurance program", Human Pathology, 2009, vol. 40, Issue 8, pp. 1129–1136

[3] R. S. Weinstein, A. M. Graham, L. C. Richter, G. P. Barker, E. A. Krupinski, A. M. Lopez, Y. Yagi, J. R. Gilbertson, A. K. Bhattacharyya et al., "Overview of telepathology, virtual microscopy and whole slide imagining: Prospects for the future", Human Pathology, 2009, vol. 40, Issue 8, 1057–1069

[4] Katharina Glatz-Krieger, Dieter Glatz, Michael J Mihatsch, "Virtual slides: high-quality demand, physical limitations, and affordability", Human Pathology, Oct. 2003, vol. 34, Issue 10, pp. 968–974

[5] Liferay (www.liferay.com)

[6] Primefaces (www.primefaces.org)

[7] jQuery (jquery.com)

[8] jQuery UI (jqueryui.com)

[9] OpenLayers (openlayers.org)

[10] OpenSlide (openslide.org)

[11] GDAL - Geospatial Data Abstraction Library (http://www.gdal.org/)

[12] F. R. Dee, "Virtual microscopy in pathology education.", Human Pathology, Aug. 2009, vol. 40, pp. 1112-21

[13] 25020, ISO/IEC. Software engineering, Software product Quality Requirements and Evaluation (SQuaRE), Measurement reference model and guide.

[14] 2500n, ISO/IEC. Software engineering, Software product Quality Requirements and Evaluation (SQuaRE), Guide to SQuaRE

[15] A. Bangor, P. T. Kortum, JA Miller, "An empirical evaluation of the System Usability Scale (SUS).", International Journal of Human-Computer Interaction, 2008, vol. 24, Issue 6, pp. 574-594

[16] J. Brooke, "SUS: a 'quick and dirty' usability scale.", in P. W. Jordan, P. Thomas, B. A. Weerdmeester, A. L. McClelland (Eds.), "Usability Evaluation in Industry", London: Taylor and Francis, 1996

[17] J. Sauro and J. R. Lewis, "Quantifying the user experience: Practical statistics for user research.", Morgan Kaufmann, Waltham MA, USA, 2012

[18] The INTEGRATE project funded by the European Commission under the 7th Framework Programme (http://www.fp7-integrate.eu/)

A Modular Semantic Infrastructure Layout for the Management of Hypermodel-Pertinent Metadata in the Context of *In Silico* Oncology*

Nikolaos A. Christodoulou and Georgios S. Stamatakos, *Member, IEEE*

***Abstract*— Over the previous years, semantic metadata have largely contributed to the management, exchange and querying of health-related data, including mathematical and computational disease simulation model descriptions, implementations and output results. In this paper, we present a proposal for an abstract semantic metadata infrastructure layout, indicating its modularity, and thus its capability to operate with different combinations of software tools. Its potential contribution for the purposes of the CHIC project is also reported.**

I. Introduction

State of the art in cancer modeling involves the development of clinically oriented and driven models based on data from pertinent medical tests, and the combination of such models into hypermodels in order to provide multiscale simulations of the phenomenon [1]. Thus, the abundance of heterogeneous data and the subsequent effort of modelers for detailed simulation of an ever-growing number of mechanisms leads to implementations containing several processes, which increases the requirements for both computational power to execute the generated models and more sophisticated storage facilities and their respective data management methods. Furthermore, the nature of the data and the development of models as software raise legal and ethical issues, e.g. regarding the intellectual property rights for each model description or implementation, and the protection of the anonymity of patients from which the data originate. These issues often restrict data access and reduce exchanging capabilities.

The use of Semantic Web technologies addresses the aforementioned limitations, since metadata produced to describe existing resources can be as descriptive as necessary to satisfy them. Thus, information exchange is facilitated because legal and ethical frameworks that are in effect can be respected and the amount of exchanged information becomes comparatively less. In addition, use of widely known biomedical ontologies and compliance with established annotation standards like MIRIAM [2] to produce the metadata provide the required integration that addresses the problem of inherent resource heterogeneity. Thus, a need emerges for the creation of appropriate infrastructures similar to those holding the actual data, which will store and manage the produced metadata, making them freely available through querying and producing new knowledge via inferencing.

In this paper, we present a proposal for the architecture of a general purpose semantic metadata infrastructure. This proposal comes in a purely abstract form consisting of independent modules. We will demonstrate the connections between these modules, describe the functionality of each one and show how it can support and contribute to the purposes of the CHIC project (http://www.chic-vph.eu/), which includes a pertinent ongoing task.

II. Components Overview

As shown in Fig. 1, the overall system is comprised by seven main modules. The Initial Access Point, the Extract-Transform-Load (ETL) unit, two repositories (RDF data and knowledge base) and three front end applications which provide users with access to the repositories and their data (annotation management, querying, knowledge base management).

A. User Roles

Individuals that will use the system will be given one of two set of access rights. The *Common User* will be used by people who want to only retrieve stored metadata by querying the RDF repository such as citizens and clinicians and consequently are allowed to access only the querying application. The *Special User* role is reserved for personnel such as modelers and IT experts that update and maintain the information stored in the system. They have access to all of the front end components and can thus modify the contents of both repositories (Fig. 1).

B. Entry Points

The system includes 2 “gateways” to communicate with the outside world. The ETL unit takes on the task of transforming the data contained in the Model Repository in Resource Description Framework (RDF) statements, using mappings between the schema of the former and ontology terms. Depending on the implementation, the ETL unit may be either a separate application, or integrated into the repository and implement the Software as a Service (SaaS) model, by being called using a REST API. Due to the volume of data (thousands or millions of records to be converted) and the potentially high required time to conduct the process, the use of this unit from any other person except for the administrators of the system and the model repository is discouraged. Alternatively, it can be run periodically to keep

*This work has been supported in part by the European Commission under the project Computational Horizons In Cancer (CHIC): Developing Meta- and Hyper-Multiscale Models and Repositories for *In Silico* Oncology (FP7-ICT-2011-9, Grant agreement no: 600841)

N.A. Christodoulou (nikchris@central.ntua.gr). and Georgios S. Stamatakos (corresponding author) (phone: +30 210772 2287; fax: +30 2107733557; e-mail: gestam@central.ntua.gr) are with the *In Silico* Oncology and *In Silico* Medicine Group, Institute of Communication and Computer Systems, National Technical University of Athens, 9 Iroon Polytechniou, GR 157 80, Zografos, Greece).

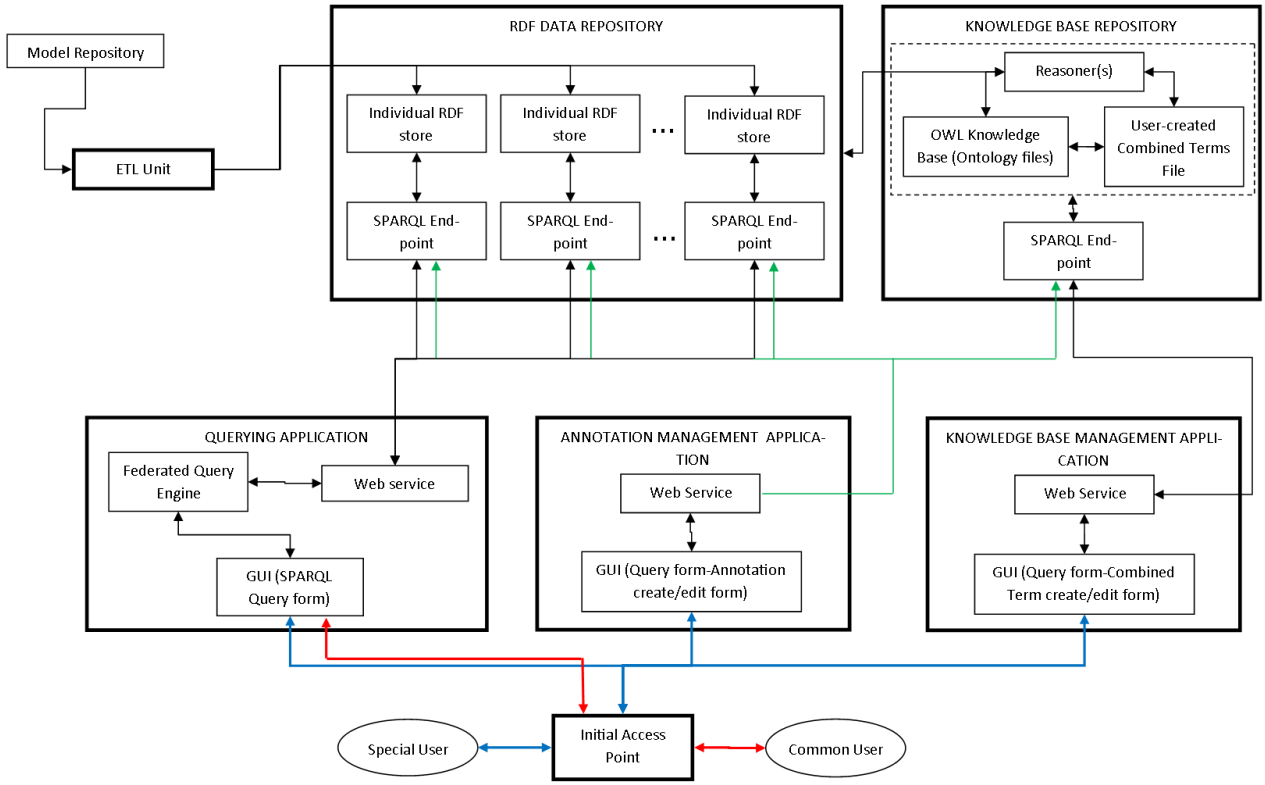


Figure 1. Schematic diagram of the proposed architecture layout.

the system updated. The Initial Access Point is the first module encountered by anyone who wants to log into the system. It identifies the users that request access and enforces the user roles by granting access to the proper modules.

C. User Communication Components

The front end of the system consists of three applications, which are responsible for user interaction with the system's repositories. Access to the contents of the former is provided, through user query submissions. Furthermore, only special users are allowed to insert, edit, or delete data. These applications connect remotely to the repositories through web services which communicate with the formers' available SPARQL endpoints.

The annotation management application allows a special user to create mappings between model resources and terms from ontologies contained in the knowledge base repository. After entering the application, the user is provided with the ability to complete the "annotation create / edit" form with the ID's of a resource and an ontology term in order to create the desired mapping. Alternatively, they can submit queries to the appropriate form for the retrieval of any of the two elements. For this reason the application is required to be connected with both repositories. Finally, a query can be submitted for the retrieval of an existing annotation so that the user can edit it. It should be noted at this point that this application is connected to only one of the RDF repository's SPARQL endpoints at any given session and can use data and annotations published to that specific endpoint. Since the RDF repository is intended to hold data from different sources (hospitals, health organizations, research institutes, etc.) separately, each user must state their employer upon registering to the system. As a result, any user actions will affect only their employer's respective endpoint. Should a person have multiple employers, they must still choose one of them when they sign in, in order for their work to affect only one endpoint.

The knowledge base management application has similar basic components to those of the previous application. However, this application communicates only with the knowledge base repository in order to combine terms from the different ontologies stored there, using logical operations, in order to represent complex biological concepts. These terms can later be used by the annotation management application to effectively annotate these concepts. A user can either directly use the create/edit form to input ontology term ID's, the desired operations and create a combined term, search for said ID's if needed via the query form, or retrieve a combined term in order to modify or delete it. The new terms are stored in a separate RDF file, which resides at the knowledge base repository and also contains links to the stored ontologies. This prevents multi-recording a combined term ID in each of the ontology files, the terms of which it is made from.

The query application is the most crucial part of the front end, as it is responsible for the communication between the users and the RDF repository. It is accessible by all individuals regardless of their role. It includes a query form, in which the user submits the query to the repository in SPARQL language. In this module, the system's federated query engine is included. This tool receives the input query from the query form and breaks it into individual sub-queries

in order to submit them to the possibly more than one different SPARQL endpoints which publish the repository's data. This provides transparent access to the contents of the latter, since the non-use of SERVICE and BINDINGS clauses in the original query's body means that no prior knowledge of the data origin (which information is published by which endpoint) or how they will be retrieved, is required by the end users.

D. Storage Components

This is the backbone of the entire system, which is practically divided into two parts due to the partially different kinds of data stored (changeable totals RDF statements as opposed to OWL ontologies), and the different access regulations that apply to each part which result from the number and type of users that request to log in.

The RDF repository stores the bulk of the metadata in the form of RDF statements. These metadata are derived from the operation of the ETL unit and the annotation management application, and can reach millions of statements in number. It also provides the ability to create SPARQL endpoints, through which user-submitted remote queries are handled and their responses are returned. This repository is expected to be used by a number of different institutions which handle data of the scientific field. The basic requirement is to make these data available to the public. At the same time, however, each institution seeks to independently maintain control of their own information and bear responsibility of keeping them up to date. A proposed solution is a virtual "partition" of the repository and assigning each part (otherwise called an individual RDF store for convenience) in a separate SPARQL endpoint. That way, each stakeholder can seamlessly perform any desired changes. This solution dictates the use of the federated query engine, so that the end user is given the impression that there is only one repository.

The knowledge base repository is smaller in size than the RDF and is accessible only by special users. It contains the ontologies, the terms of which are used for the annotation of resources, the file that contains the combined terms created by the knowledge base management application and one or more semantic reasoners, which are used to produce new statements based on the existence of others, which are regarded as axioms. In addition to any known reasoners (Fact++ [3], Pellet [4], etc.) which are suitable for OWL ontologies, additional rule files may be stored, which extend the former, and are based on the specific characteristics of the system. The reasoners can be also available to the RDF repository to allow the application of their rules directly on the RDF data or analyze the SPARQL queries that posed to them, so the answer contains all the necessary additional statements, or by applying incremental reasoning [5]. Finally, the existence of a SPARQL endpoint serves as the module's communicator, with both front end applications with which it cooperates.

III. USE WITHIN THE CHIC PROJECT

The aim of the project "Computational Horizons In Cancer (CHIC): Developing Meta- and Hyper-Multiscale Models and Repositories for *In Silico* Oncology", is to address the complexity of cancer and to describe the phenomena that are being caused by it at the various biological levels of the human body (molecular, cellular, tissue, organ, etc.). As the representation of the entire disease with a single model is not possible, hypermodels are used; hypermodels are models made of elementary component models, or otherwise called, hypomodels. This leads to the creation of hyper-terms which must be annotated, while the resources that hypomodels can be annotated with, are expected to be combined in the same way as the latter, thus resulting in the formation of hyper-resources.

For that reason, the project includes a task whose objective is to create an infrastructure capable of coping with the management of all this metadata. The heart of this infrastructure was agreed to be the solution developed in the context of the RICORDO project [6]. The goal is to extend the provided infrastructure in order to meet the needs of CHIC. This proposal seeks to achieve this goal through the decomposition of the given RICORDO infrastructure. This is done in order to enrich it with any additional required software components or to upgrade some of the existing ones. In this case, the example of the hypermodelling creation process, the research for which is ongoing, is followed.

A pertinent investigation is currently being carried out giving special attention to the RDF repository and the federated query machine. In order to fully exploit the project's private cloud, distributed open source solutions are considered for the repository, such as HDRS [7], 4store (http://4store.org/) and Virtuoso server's [8] clustered edition. For the federated query engine, free, open source software packages such as SPARQL-DQP [9], ANAPSID [10], FedX (currently at version 3.0) [11] and ADERIS [12] are considered. Each of them implements a different approach to achieve SPARQL endpoint federations, and the capabilities for contribution in the overall result are being explored. Consequently, any changes might affect other components, such as the knowledge base, currently being implemented by a combination of OWLlink server (http://owllink-owlapi.sourceforge.net/) with the Pellet reasoner.

IV. CONCLUSION

In this paper, we presented a proposed abstract layout for a metadata production and management infrastructure. Using the RICORDO results as a starting point, we explained how its further modularization by means of decomposition may help towards achieving the goals of the CHIC project. Along with the aforementioned tools, a more general research is under way, which will last for the entirety of the project, in order to explore the possible combinations of software components for all layout parts.

REFERENCES

[1] G. Stamatakos, *Member, IEEE,* D. Dionysiou, A. Lunzer, R. Belleman, E. Kolokotroni, E. Georgiadi, M. Erdt, J. Pukacki, S. Rüeping, S. Giatili, A. d' Onofrio, S. Sfakianakis, K. Marias, *Member, IEEE,* C. Desmedt, M. Tsiknakis, *Member, IEEE,* and N. Graf, *Member, IEEE,* "The Technologically Integrated Oncosimulator:

Combining Multiscale Cancer Modeling With Information Technology in the In SilicoOncology Context," *IEEE Journal of Biomedical and Health Informatics*, vol. 18, no. 3, pp. 840–854, May 2014.

[2] N. Le Novère, A. Finney, M. Hucka, U. S Bhalla, F. Campagne, J. Collado-Vides, E. J Crampin, M. Halstead, E. Klipp, P. Mendes, P. Nielsen, H. Sauro, B. Shapiro, J. L. Snoep, H. D. Spence, and B. L. Wanner, "Minimum information requested in the annotation of biochemical models (MIRIAM)," *Nature Biotechnology*, vol. 23, no.12, pp. 1509-1515, Dec. 2005.

[3] D. Tsarkov and I. Horrocks , "FaCT++ Description Logic Reasoner: System Description," in *Automated Reasoning*, vol. 4130, U. Furbach, N. Shankar, Ed. Berlin Heidelberg: Springer, 2006, pp. 292–297.

[4] E. Sirin, B. Parsia, B. Cuenca Grau, A. Kalyanpur and Y. Katz, "Pellet: A practical OWL-DL reasoner," *Web Semantics: Science, Services and Agents on the World Wide Web*, vol. 5, no.2, Jun 2007, pp. 51–53.

[5] B. Parsia, C. Halaschek-Wiener, and E. Sirin, "Towards incremental reasoning through updates in OWL-DL," presented at the 2006 15th Int. World Wide Web Conf, Edinburgh, Scotland.

[6] S. M. Wimalaratne, P. Grenon, R. Hoehndorf, G. V. Gkoutos and B. de Bono, "An infrastructure for ontology-based information systems in biomedicine: RICORDO case study," *Bioinformatics*, vol. 28, no. 3, Nov 2011, pp. 448–450.

[7] C. Böhm, D. Hefenbrock, and F. Naumann, "Scalable peer-to-peer-based RDF management," in *Proc. 8th Int, Conf. on Semantic Systems*, New York, 2012, pp. 165–168.

[8] O. Erling and I. Mikhailov, "RDF Support in the Virtuoso DBMS," in *Networked Knowledge – Network Media*, vol. 221, T. Pellegrini, S. Auer, K. Tochtermann and S. Schaffert, Ed. Berlin Heidelberg: Springer, 2009, pp. 7–24.

[9] C. Buil-Aranda, M.Arenas and O.Corcho, "Semantics and Optimization of the SPARQL 1.1 Federation Extension,", in *The Semantic Web: Research and Applications*, vol. 6644, G. Antoniou, M.Grobelnik, E. Simperl, B. Parsia, D. Plexousakis, P. De Leenheer and J.Pan, Ed. Berlin Heidelberg: Springer, 2011, pp. 1–15.

[10] M. Acosta, M. Vidal, T. Lampo, J. Castillo and E. Ruckhaus, "ANAPSID: An Adaptive Query Processing Engine for SPARQL Endpoints," in *The Semantic Web – ISWC 2011*, vol. 7031, L. Aroyo, C. Welty, H. Alani, J. Taylor, A. Bernstein, L. Kagal, N. Noy and E. Blomqvist, Ed. Berlin Heidelberg: Springer, 2011, pp. 18–34.

[11] A. Schwarte, P. Haase, K. Hose, R. Schenkel and M. Schmidt, "FedX: Optimization Techniques for Federated Query Processing on Linked Data," in *The Semantic Web – ISWC 2011*, vol. 7031, L. Aroyo, C. Welty, H. Alani, J. Taylor, A. Bernstein, L. Kagal, N. Noy and E. Blomqvist, Ed. Berlin Heidelberg: Springer, 2011, pp. 601–616.

[12] S. J. Lynden, I. Kojima, A. Matono, and Y. Tanimura, "ADERIS: An adaptive query processor for joining federated SPARQL endpoints," in *On the Move to Meaningful Internet Systems: OTM 2011*, vol. 7045, R. Meersman, T. Dillon, P. Herrero, A. Kumar, M. Reichert, L. Qing, B. C. Ooi, E. Damiani, D. C. Schmidt, J. White, M. Hauswirth, P. Hitzler, M. Mohania, Ed. Berlin Heidelberg: Springer, 2011, pp. 808–817.

Development of the p-medicine Oncosimulator as a Parallel Treatment Support System*

Marek Blazewicz, Eleni Ch. Georgiadi, Juliusz Pukacki, and Georgios S. Stamatakos, *Member, IEEE*

***Abstract*— The purpose of this research was the parallelization of the Wilms' *Oncosimulator*, an integrated cancer treatment support system modeling the growth of nephroblastoma tumors and their *in vivo* response to chemotherapeutic modalities. In this concept, the Oncosimulator has been optimized in order to perform efficient computations on the newest heterogeneous parallel architectures: the CPU and GPU based computing architectures. The simulator has been implemented using a novel solution for distributed computing on heterogeneous architectures – the *Cactus* computational toolkit with *CaKernel* as the module for the computations performed on computing accelerators. In this publication the challenges faced during the process of porting the Oncosimulator onto the aforementioned architectures (within CaKernel framework) are addressed, and the performance benefits of such approaches are analyzed. The successful parallelization of the Oncosimulator advances its computational efficiency and enhances its reusability as well as its eventual translation into clinical practice. The research was performed in the context of the p-medicine project*.**

I. Introduction

Wilms' tumor is the most common malignant tumor of the kidney in children [1]. The Wilms' Oncosimulator is an integrated software system modeling the growth of nephroblastoma tumors and their *in vivo* response to chemotherapeutic modalities within the clinical trials environment, aiming to support clinical decision making in individual patients [2]. The modeling core algorithms of the Oncosimulator, which is a predominantly discrete clinically-oriented multiscale cancer model, have been developed by the *In Silico* Oncology and *In Silico* Medicine Group (ISO&ISMG), Institute of Communication and Computer Systems (ICCS), National Technical University of Athens (NTUA) [3], [4], [5].

The Oncosimulator constitutes a "top-down" simulation approach, starting from the macroscopic imaging data (a high biocomplexity level) and proceeding towards the lower biocomplexity levels. After adequate processing of the imaging data, the tumor volume is spatially initialized by superimposing a three-dimensional discretization mesh over the anatomical region of interest. The elementary volume of the mesh is called *Geometrical Cell* (*GC*). Each GC belonging to the tumor accommodates initially a number of biological cells, which is defined based on typical solid tumor cell densities. Predominantly discrete modeling considers several discrete states in which cells may be found and possible transitions between them, governed by "decision calculators", such as cytokinetic diagrams and agent-based techniques. In this way, several cancer-related biological phenomena that have been reported in literature to take place at the cellular level are incorporated in the model such as proliferation, quiescence, differentiation and death (normal and chemotherapy-induced). At each time step the discretizing mesh is scanned and the basic rules that govern the spatiotemporal evolution of the tumor are applied for each GC of the mesh.

Practically, each complete scan can be viewed as consisting of two mesh scans: at the first scan the temporal evolution of the tumor cell populations is simulated while at the second scan the spatial evolution (shrinkage-expansion) of the tumor mass is modeled. As a great number of registers are used to describe the state of each GC occupied by the tumor, the computational and memory resources are proportional to the number of GCs that define the tumor area. This characteristic of the Oncosimulator makes it a memory-bound application. Although the computational algorithms performed by the Oncosimulator are not restrictively complex, the large number of registers used, increases the computational demands in terms of accessing memory.

Most of the computations performed over a single GC are independent from those performed on its neighbors. This makes the p-medicine Oncosimulator ideal to parallelize. Moreover, this computational pattern along with the data layout and data dependencies makes the simulation a subset of stencil computations. Many scientists have been researching and testing different approaches on how to efficiently compute stencils on a variety of different architectures [6], [7], [8]. The main challenge of stencil computations is the efficient utilization of memory bandwidth. It can be achieved by fulfilling several computational schemes. The most important one is maximizing cache reusage. This is a difficult task, and dependent on the computing architecture. For instance, different CPUs have different caches (L1, L2, L3) with different sizes and length of fetched lines, and therefore require different blocking techniques and loop traversing. On the other hand GPUs and other dedicated accelerators have specific cache memory types that need to be accessed in an

*The research leading to these results has received funding from the European Union's Seventh Programme for research, technological development and demonstration through the p-medicine project (grant agreement No [PI: 270089]).

M. Blazewicz is with the Applications Department, Poznan Supercomputing & Networking Center, Poznan, Poland and the Poznan University of Technology, Poznan, Poland (corresponding author phone: +48 61 858 2517; fax: +48 61 852-59-54; email: marqs@man.poznan.pl).

E. Ch. Georgiadi and G. S. Stamatakos are with the *In Silico* Oncology and *In Silico* Medicine Group, Institute of Communication and Computer Systems, School of Electrical and Computer Engineering, National Technical University of Athens, Athens, Greece (egeorg@central.ntua.gr; gestam@mail.ntua.gr).

J. Pukacki is with the Applications Department, Poznan Supercomputing & Networking Center, Poznan, Poland (pukacki@man.poznan.pl).

explicit manner. In order to compute efficiently on all architectures, the programmer has to write different version of the code for every different architecture. This requires additional programming effort and detailed optimizations for each of the architectures separately.

A solution to the problem might be the use of a certain abstraction allowing for automatic optimizations specific to different architectures. The only requirement usually is concerning following some rules about the methods of data access. The framework which is responsible for handling of such an abstraction usually requires some additional meta-information about the computations that are actually being performed. For example, the size of the boundaries or some hints about memory usage.

One of the frameworks giving mechanisms for such an abstraction is Cactus [9] with a plugin handling heterogeneous architecture – CaKernel [10], [11], [12]. Because of the good performance of this framework and its extensibility, it has been utilized in the p-medicine project. Moreover, Cactus with CaKernel plugin facilitates the programming and the execution of applications in a distributed environment by providing a set of mechanisms for automatic, inter-node synchronization and exchange of neighboring data.

II. Methods: Parallelization

During the performance measurements the two main routines (mesh scans) of the Oncosimulator which are performed per simulation time step and consume 99% of the overall computational time were taken into consideration:

- 1st mesh scan – updating the state of each GC by applying the evolution rules (as defined by a cytokinetic diagram [3], [4]);
- 2nd mesh scan – updating the shape of the tumor mass by applying the rules that govern the movement of cells throughout the tumor region [3], [4];

Although the first routine consists of mathematical operations independent of its neighbors in the computational domain, the second mesh scan performs transitions considering the whole tumor at once. It was very difficult to preserve the logic of that routine and in the same time parallelize it efficiently. That is why we have focused on minimizing the sequential operations to the number required to preserve the consistent and deterministic evolution of the tumor. Motivated by the difficulties of the parallelization of the second mesh scan routines, we have performed multiple performance benchmarks. In particular, we have performed several tests on simulations:

- Tests on the parallelization of the first mesh scan: the tumor was evolved without being reshaped;
- Tests on the parallelization of the first mesh scan with standard sequential second mesh scan: In these tests the second mesh scan is performed independently of the underlying execution architecture, sequentially on CPU. In case the computations were performed in parallel on GPU, the whole simulation data had to be copied back to the CPU (to perform the second mesh scan) and forth to the GPU (to continue the simulation of the first mesh scan);
- Tests on the parallelization of the first mesh scan with second mesh scan performed in a hybrid manner (CPU+ GPU): In these tests the second mesh scan is performed in a hybrid manner (CPU + GPU). The number of performed sequential operations on CPU has been minimized; from GPU to CPU only some data was copied, required to preserve deterministic and consistent evolution. The whole reshaping process finally was performed in a parallel manner on GPU out of transformation values precomputed on the CPU.

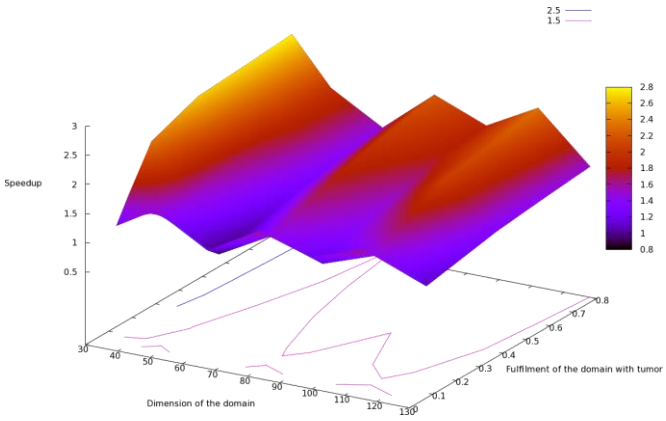


Figure 1. The speedup of the multithreaded computations over the legacy sequential code (first mesh scan tests).

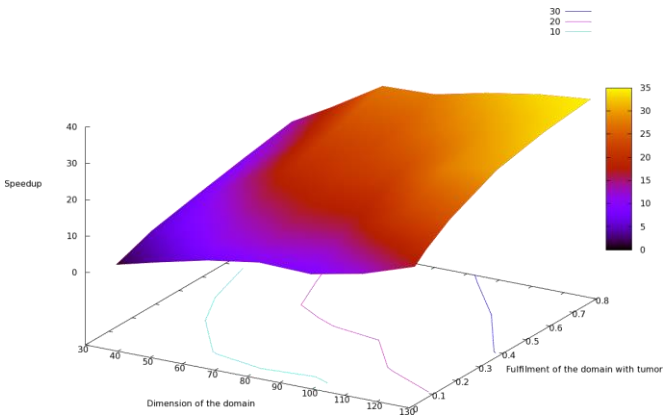


Figure 2. The speedup of the GPU computations over the legacy sequential code (first mesh scan tests).

III. Results: Tests & Developments

The CPU tests were performed on Intel 4-core x86 architecture (Intel Core 2 Quad CPU Q9550, 2.83GHz; 8GB of DDR3 RAM, dual channel, 1333MHz). Multithreaded computations were performed with the use of OpenMP. The GPU tests were perfomed on NVIDIA GTX680 with the use of CUDA programming language. For simplicity, in this work, GPU and CPU stand for the aforementioned architectures.

Some architectures (like GPUs) require sufficiently large domain in order to fully utilize their computational potential (all computational units should be fully occupied). For this reason the tests were performed with the consideration of different sizes of the cubic domain. One important property of the GPU architecture is also the fact that for maximal utilization of the available resources, it is required that most

of the threads perform homogeneous operations. Any variations from this situation require threads to divergent the execution branches, which results in lower performance of the overall computational process. Moreover, lack of tumor in parts of the computational domain imposes that different blocks have varying amount of work to perform. This introduces the need of additional, advanced load balancing, which is not a trivial task. That is why the computational performance has also been measured in respect to the ratio of the amount of the tumor within the computational domain.

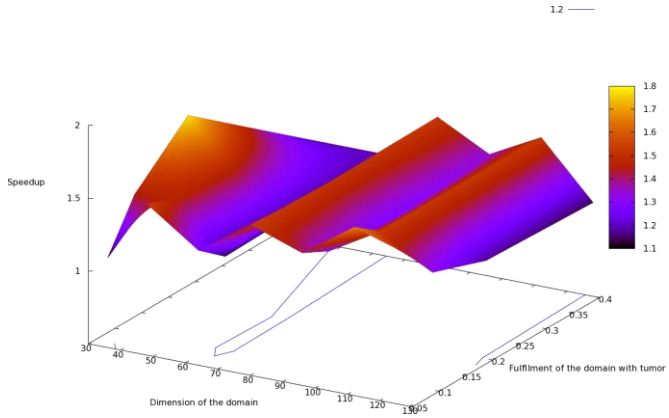


Figure 3. The speedup of the multi-threaded computations over the legacy sequential code (with reshaping).

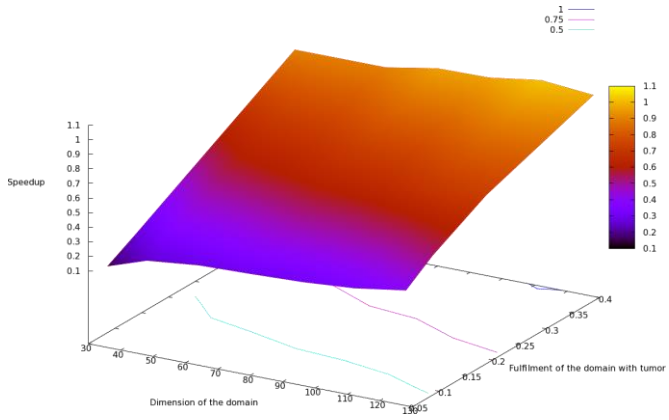


Figure 4. The speedup of the GPU computations over the legacy sequential code (with reshaping).

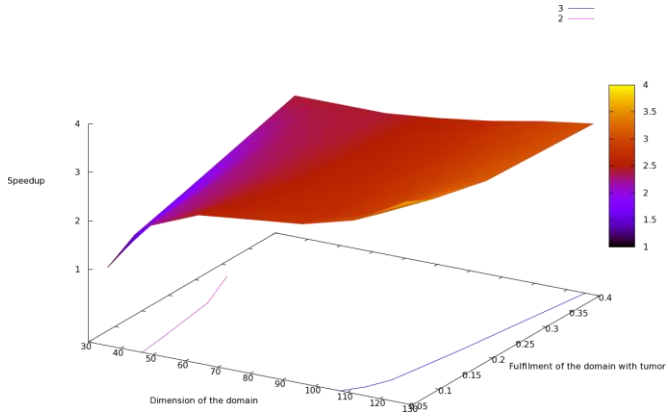


Figure 5. The speedup of the GPU computations over the legacy sequential code (with hybrid reshaping).

A. Tests on the parallelization of first mesh scan.

The parallelization of the first mesh scan computations is not a high-demanding task. The first mesh scan routine consists of independent operations over GCs. Thus, the computations over all GCs could be performed by different threads simultaneously. The parallelization of the CPU code has been performed with the use of the OpenMP dividing the domain evenly for all four cores. On the GPU architecture, one thread was deployed for each of the GCs. Unlike the CPU threads, the GPU threads are quick and lightweight. That is why there was no additional overhead introduced by launching such a huge number of threads.

In Fig. 1 and 2 the speedup of GPU and OpenMP computations over the legacy code is presented. The plots present the actual speedup in respect to the size of the domain (x axis depicts the side of the cubic domain) and its fulfillment with the tumor (y axis depicts the fulfillment of the domain). As one may observe the GPU architecture is performing the computations in the fastest manner. The speedup is increasing proportionally to the amount of the actual computations, achieving almost 40 fold speedup while performing computations over the largest domain (1283) with the highest tumor's fulfillment (80%). Such performance is an outcome of a very good parallelization process.

The multithreaded code achieved the performance varying between 1.2 and 3 fold speedup over the legacy code. The fluctuations of the actual speedup probably are caused by the inhomogeneous layout of the tumor in the computational domain and the incapability of the OpenMP interface to fully utilize the computational capabilities in such environment. Considering the fact that the processor consisted of 4 cores, the speedup was sufficient.

B. Tests on the parallelization of first mesh scan with standard sequential second mesh scan

In this section the benchmark of the tumor evolution of the tumor cell populations in comparison with the spatial evolution of the tumor is presented. The first mesh scan routine was performed on one of the three aforementioned architectures (CPU, multi-CPU, GPU), whereas in all cases the second mesh scan was performed in the sequential manner on CPU. As mentioned in the previous section, performing the computations on the GPU architecture in this test implied that the whole computational domain had to be copied from the memory of the GPU (accelerator) to the host memory after the first mesh scan in order to perform the second mesh scan and finally copied back to GPU. It resulted in a large performance loss of the simulation, since the data exchange between CPU and GPU is relatively slow. The overall performance of the Oncosimulator has been presented in Fig. 3 and 4 in the same manner as in previous section.

As it comes to multithreaded code the speedup of the simulation is slightly worse than the one without the second mesh scan. The small drop of performance was caused by the fact that still some part of the computations was performed in a sequential manner (the same as in legacy code), which

decreased the positive performance impact of the parallel computations. In result, the speedup was between 1.2 and 1.8 fold.

When it comes to the simulation performed on the GPU architecture one may observe that the influence of the superfluous data transfer made the computations inefficient in comparison to other architectures. Depending on the properties of the computational domain (size of the mesh and shape of the tumor), the GPU simulation was from 10 times slower up to similar performance in comparison to the sequential legacy code. Because of inefficient (in most cases of computational domain) computations, the hybrid method has been designed and implemented on the second mesh scan as described below.

C. Tests on the parallelization of the first mesh scan with the second mesh scan performed in a hybrid manner (CPU+GPU)

The main considered constraints of the parallelization of the second mesh scan process were: the consistency of the simulation and the deterministic results. As mentioned in previous paragraphs, the process of reshaping the tumor involves the whole tumor at once and aims to balance the over-loaded and under-loaded GCs within the tumor (growing and shrinking of the tumor). However, even though the process seams trivial, since it is just moving biological cells among GCs, it is highly data intensive. It involves all registers, which in a typical simulation are equal to 90 per GC. Moreover, when performing computations on the GPU, it requires that all of the registers would be copied from the GPU before each second scan mesh, and back on the GPU afterwards. In many cases it means that the amount of transferred data is equal to 1, 2 or even more gigabytes at each cycle. Knowing that the theoretical peak bandwidth of the PCI Express bus is equal to 8GB/s (in practice it is approximately 6GB/s), transferring the data back and forth may consume almost one second of a simulation time, whereas the first mesh scan routine performed on the GPU architecture, for data size 1.2GB would consume between 7 and 30ms. One can see that the overhead of the data transfer is consuming the whole performance benefits gained by the use of the GPU architecture.

During the parallelization of the second mesh scan routine, we have limited the number of registers transferred to the minimum required to perform deterministic and consistent reshaping. In particular, the total number of cells in each GC and the flags describing GCs (like for instance the actual location of the tumor). These two registers are then sequentially processed on the CPU. Unfortunately, this is the only possibility to perform the deterministic second mesh scan, since the process uses function generating pseudorandom numbers, and any variations in the domain traversal would influence the outcome of the simulation.

The sequential process traverses the computational domain seeking for over and under-loaded GCs. Any actual operations are performed only on the two mentioned registers, which results in a major speedup of data transfer. All of the performed transpositions within the computational domain are recorded by the process into a separate vector which enables redoing them on GPU. After the sequential part of the second mesh scan routine is finished, the vector with the transitions is copied back to the GPU, and all of the operations are repeated on the GPU on the rest of the Grid Functions. In order to preserve the consistency of the computational domain, the vector with transitions is analyzed sequentially, i.e. each transition is performed if all the previous transitions have been already performed. Despite the second mesh scan routine limitations, the possible parallelism of the process has been observed in two separate places:

- multiple registers; each of the registers is processed by a separate CUDA block, independently of each other;
- performing a shift operations on GCs within recorded direction at once by multiple threads; a vector of CUDA threads within a block reads all the shifted GCs, perform the intra-block synchronization (to avoid the read-after-write synchronization issue), and store the GCs into the new locations.

Despite of the observed parallelism and the high bandwidth of the GPU memory, the second mesh scan performed on the GPU is only a few (approximately up to 4; Fig. 5) times faster than performed on the CPU, whereas the ratio of GPU and CPU memory bandwidth is more than 10. The reason of such a moderate performance speedup is that the second mesh scan routine consists of multiple small transitions in random places and directions and the GPU is much more sensitive to inhomogeneous memory accesses than CPU. Despite this fact, the speedup of this process is still noticeable.

IV. CONCLUSION

In this research, the Oncosimulator has been ported onto the new architectures, including the multithreaded CPU and the newest GPU. The simulation has been split into the two most time consuming routines of the Oncosimulator: the first and the second mesh scan. In both routines possible parallelism has been tracked and utilized. The first routine gives a very high speedup on the GPU in comparison with the legacy application, in contrast to the second one that requires portions of sequential operations in order to preserve consistent and deterministic simulation results. Finally, a hybrid method has been designed, developed and implemented in the second mesh scan routine which performs required sequential logic on CPU, limiting the requirement for time consuming data transfers and leaving the data intensive operations on GPU, to be evaluated as a speedup over the legacy Oncosimulator. The simulation performed on GPU without the second mesh scan resulted in almost 40 fold speedup in comparison to 4 fold speedup of simulation performed with the second mesh scan. The difference was caused by the fact that the second mesh scan consists of multiple random memory accesses that are not efficiently performed on the GPU architecture. For the multithreaded version, the achieved speedup was 1.8 and 3.0 fold for the simulation with and without the second mesh

scan.

REFERENCES

[1] S. C. Kaste, J. S. Dome, P. S Babyn, N. M. Graf, P. Grundy, J. Godzinski, G. A Levitt, and H. Jenkinson, "Wilms tumour: prognostic factors, staging, therapy and late effects," *Pediatric Radiology*, vol. 38, no.1, pp. 2–17, Jan. 2008.

[2] G. S. Stamatakos, D. D. Dionysiou, A. Lunzer, R. G. Belleman, E. A. Kolokotroni, E. Ch. Georgiadi, M. Erdt, J. Pukacki, S. Rüping, S. G. Giatili, A. d' Onofrio, S. Sfakianakis, K. Marias, C. Desmedt, M. Tsiknakis, and N. M. Graf, "The Technologically Integrated Oncosimulator: Combining Multiscale Cancer Modeling With Information Technology in the *In Silico* Oncology Context," *IEEE J. Biomedical and Health Informatics*, vol. 18, no. 3, pp. 840–854, May 2014.

[3] G. S. Stamatakos, E. Ch. Georgiadi, N. M. 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*, vol. 6, no. 3, e17594, Mar. 2011.

[4] E. Ch. Georgiadi, D. D. Dionysiou, N. M. Graf, and G. S. Stamatakos, "Towards *in silico* oncology: Adapting a four dimensional nephroblastoma treatment model to a clinical trial case based on multi-method sensitivity analysis," *Comp. Biol. Med*., vol. 42, no. 11, pp. 1064–1078, Nov. 2012.

[5] N. M. Graf, A. Hoppe, E. Ch. Georgiadi, R. Belleman, C. Desmedt, D. D. Dionysiou, M. Erdt, J. Jacques, E. Kolokotroni, A. Lunzer, M. Tsiknakis, and G. S. Stamatakos, " 'In silico' oncology for clinical decision making in the context of nephroblastoma," *Klinische Paediatrie*, vol. 221, no.3, pp.141–149, May-Jun 2009.

[6] M. Christen, O. Schenk, and H. Burkhart, "PATUS: A Code Generation and Autotuning Framework for Parallel Iterative Stencil Computations on Modern Microarchitectures," in *Proc. 25th IEEE Int. Parallel & Distributed Processing Symposium*, Anchorage, 2011, pp. 676–687.

[7] K. Datta, M. Murphy, V. Volkov, S. Williams, J. Carter, L. Oliker, D. Patterson, J. Shalf, and K. Yelick, "Stencil computation optimization and auto-tuning on state-of-the-art multicore architectures," in *Proc. 2008 ACM/IEEE conference on Supercomputing*, Austin, TX. pp. 4:1–4:12.

[8] S. M. Faizur Rahman, Q. Yi, and A. Qasem, "Understanding stencil code performance on multicore architectures," in *Proc. 8th ACM International Conference on Computing Frontiers*, New York, 2011, pp. 30:1–30:10.

[9] T. Goodale, G. Allen, G. Lanfermann, J. Massó, T. Radke, E. Seidel, and J. Shalf, "The Cactus framework and toolkit: Design and applications," in *Vector and Parallel Processing – VECPAR'2002, 5th International Conference, Lecture Notes in Computer Science*, Berlin, 2003.

[10] M. Blazewicz, S. R. Brandt, P. Diener, D. M. Koppelman, K. Kurowski, F. Löffler, E. Schnetter, and J. Tao, "A massive data parallel computational framework on petascale/exascale hybrid computer systems," International Conference on Parallel Computing, Ghent, Belgium, 2011.

[11] M. Blazewicz, S. R. Brandt, M. Kierzynka, K. Kurowski, B. Ludwiczak, J. Tao, and J. Weglarz, "Cakernel - a parallel application programming framework for heterogenous computing architectures," *Scientific Programming*, vol. 19, no. 4, pp. 185–197, Dec. 2011.

[12] M. Blazewicz, I. Hinder, D. M. Koppelman, S. R. Brandt, M. Ciznicki, M. Kierzynka, F. Löffler, E. Schnetter, and J. Tao, "From physics model to results: An optimizing framework for cross-architecture code generation," *Scientific Programming*, vol. 21, pp. 1–16, July 2013.

The VPH Hypermodelling Framework for Cancer Multiscale Models in the Clinical Practice*

D. Tartarini, K. Duan, N. Gruel, D. Testi, D. Walker, and M. Viceconti

***Abstract*— The VPH Hypermodelling framework is a collaborative computational platform providing a complete Problem Solving Environment to execute, on distributed computational architectures, sophisticated predictive models involving patient medical data or specialized repositories. In the CHIC[1] project, it will be enhanced to support clinicians in providing prompt personalised cancer treatments. It supports several computational architectures with strict security policies.**

I. Introduction

Cancer is a complex disease that behaves differently from patient to patient. It can affect several organs of the human body requiring specific treatments. It challenges clinicians in making reliable diagnosis, prognosis and choosing the tailored treatment. The oncological research is further complicated since cancer manifestation spans several spatio-temporal scales, from molecular and cellular level to organ level, from nanosecond molecular reactions to tumour evolution in years [1]. Researchers have proposed numerous mathematical models describing cancer progression and treatments from a biochemical and biophysical perspective [1],[2],[3]. Generally, due to computational resource limitations and the complexity of bridging scales, these models are focused on a single phenomenon at a particular space-time scale. It is evident that a synergic cross-discipline collaboration among researchers (clinicians, biomedical scientists/engineers and computer scientists) would improve the chance of success. In fact, merging knowledge and expertise of researchers would allow the development of more accurate integrative cancer models that, with the help of computer simulations, can support the clinicians in their decisions [2]. This is the philosophy underpinning the Virtual Physiological Human (VPH)[4] agenda. In alignment with this, we aim to provide a hypermodelling environment where experts can contribute with their knowledge to develop sophisticated integrative models.

A. The VPH Hypermodelling framework

The VPH Hypermodelling Framework (VPH-HF) is a collaborative computational platform providing a complete Problem Solving Environment (PSE)[5] to execute, on distributed computational architectures, sophisticated predictive models involving patient medical data or specialized repositories. It is based on a fully-fledged prototype developed in a previous VPH project, the Osteoporotic VPH (VPH-OP)[2], which addressed the estimation of bone fracture risk due to osteoporosis. The aim of VPH-HF is to improve the effectiveness of diagnosis, prognosis and treatment of specific diseases in clinical practice with the ultimate objective to foster the personalized medicine paradigm and perform *in silico* clinical trials [6].

In particular, in the frame of the EC project Computational Horizons in Cancer (CHIC)[1] the VPH-HF is customized for the oncological needs targeting two primary users: the clinician and the researcher. The former needs easy out-of-the-box software tools to analyse patient medical data and simulate cancer behaviour to address specific clinical questions. The latter has a broader profile that includes the creation and validation of complex integrative models [7]. Researchers can populate the CHIC data and model repositories with their experimental data and provide the integrative/predictive models implemented in the computational format of their convenience. Furthermore, in CHIC a semantic annotation system is provided to support the development of *ad hoc* ontologies and folksonomies.

The VPH-HF framework is very flexible: almost any model can be integrated and run in the PSE, whether it is an executable binary file, an interpreted script (e.g. written in Python, Perl, Matlab, Octave) or a more complex case involving external licensed software like Ansys® or Abaqus®. The underlying assumption is that an hypermodel (i.e. an integrative model or a composition/orchestration of models) can be described as a workflow where its composing hypomodels (i.e. models) are connected to produce an output result from a given input and data from repositories and/or patient specific data. Therefore a workflow can be represented as a graph where the nodes are models or data repositories while connections are data or control flows. Two models are connected when an output of the first is an input of the second, while data repositories can be connected to any of the models. In order to build a workflow, hypermodels and hypomodels can be considered as black boxes with a standardized abstract interface exposing input and output ports and control data flow (Figure 1). This interface is well defined within the CHIC project and is called *Component Model Generic Stub*. It ensures the

*Research supported by the "Computational Horizons in Cancer" project, funded by the European Union, EC Seventh Programme for research, technological development and demonstration under grant agreement No [600841] (CHIC Project).

D. Tartarini, D. Walker, K. Duan, N. Gruel, M. Viceconti are with the INSIGNEO Institute for in silico Medicine, Sheffield, S13JD UK (corresponding author phone: +44 (0) 114 222 6173; e-mail: d.tartarini@sheffield.ac.uk).

D. Testi is with Consorzio inter-universitario CINECA, Bologna, Italy

[1] http:// http://chic-vph.eu/

[2] http://www.vphop.eu

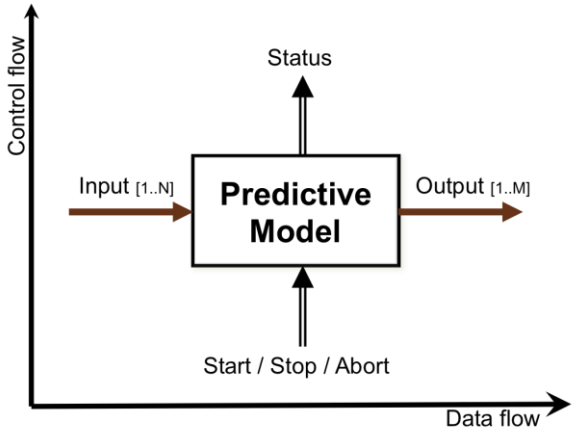


Figure 1. Predictive model as black box

interoperability between all the provided hyper- and hypo-models.

The VPH-HF is compliant with this interface and provides a software implementation that follows the Wrapper pattern [8]. It allows the actual integration of any of the computational instances of the models in a workflow including data and control flow in an appropriate standard format. The *model wrapper* (Fig. 2) provides the following functionalities: adapts the parameters from the format used in the actual model to the standard one of the Component Model Generic Stub interface, retrieves input from the storage services, produces execution logs, runs the model and stores the results in the Storage service. In order to ensure modularity the wrapper exposes an XML-RPC/SOAP interface that is used by the Taverna Server [9] to run the model.

B. *The VPH-HF architecture*

The architecture of VPH-HF is inspired by the concept of modularity: each component can be used in isolation or ensemble with others to offer more sophisticated functionalities. This approach ensured an effective extension of the VPH-HF prototype developed in the VPH-OP project to the new requirements and scenarios of the CHIC project. The whole back-end VPH-HF architecture is hidden to the users since the target goal is to allow the adoption in the clinical practice through an easy and secure interface. In particular, as shown in (Figure 3), the clinician interacts only with the *CHIC portal* that embeds the key functionalities: uploading and (pseudo)anonymisation of patient data, visualization of processed patient data and images (CT, MRI etc), semantic searching for data and models using metadata information, configuring and running existing workflows using the *Hypermodelling Editor*, retrieval and validation of results. On the other side, VPH-HF also provides applications with an interface conceived for an expert user to fully control the execution of a workflow in the framework (i.e. the Hypermonitor).

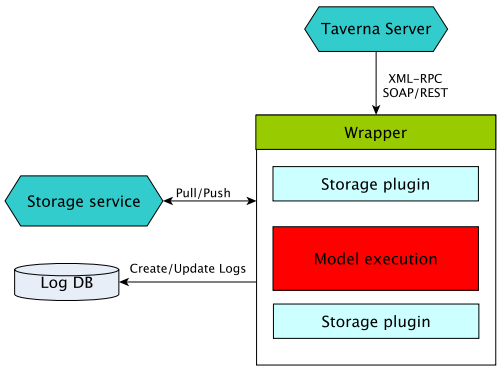


Figure 2. The model wrapper for the Generic Stub interface

In summary, the user interacts with the whole system through the following tools:

1. *The CHIC Portal*: a Liferay[3] based portal offering tools to anonymise patient data, to upload/download data, search model and data through their metadata, execute workflows.
2. *The CHIC Hypermodelling Editor*: is a Java portlet allowing the user to *compose* hypomodels into hypermodels and execute new or pre-built workflows on patients data or data repositories.
3. *The Hypermonitor*: a VPH-HF stand-alone application enabled to launch existing workflows on specific data, monitoring the execution status, and visualize the execution logs.
4. *The VPH-HF portlet*: provides the Hypermonitor functionalities into a web portal.

In particular the first two tools are developed by partners of the CHIC project and integrated in VPH-HF, while the last two are applications developed within the VPH-HF software stack. Nevertheless the VPH-HF adopted the security and authentication mechanisms to fully integrate with the CHIC requirements and services. In fact the whole computational platform has to obey to strict security policies given the involvement of patient data.

The VPH-HF architecture is designed to be flexible, modular, easy to maintain and be customized to mutable needs of the users, as well as sophisticated computational architectures. All the software components (Figure 3) expose

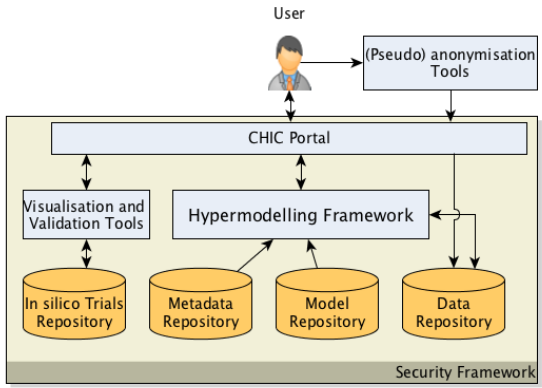


Figure 3. User perspective of the CHIC architecture

[3] http://www.liferay.com/

a standard interface (e.g. XML-RPC, SOAP, REST) that potentially allows them to be used in isolation. In the most common scenario the authenticated user prepares a workflow (using the available hypermodels or the CHIC workflows), submits it and waits for results. The end-user applications interact with the *Workflow Manager* that interprets their instructions and, through the *Communication bus (i.e. MAF3 Event Bus)*, initialises the necessary services, transfers data, sets up the execution environment, runs the workflow on the computational infrastructure and retrieves the results. A key role is played by the *Workflow Orchestrator* (an instance of the *Taverna Server* [9]*)*, which actually orchestrates the execution of all the models composing the workflow and the data flow between them. Each model exposes a standardized interface through the *model wrapper* that takes care of adapting the input/output to the particular model interface. The framework can address more sophisticated scenarios, thanks to the functionalities offered by its components (Figure 4):

- *Workflow Manager or Director*: orchestrates the entire process from the authentication of the user, the setup of the execution environment, the control of execution of the workflow, storage and retrieval of the results.
- *Communication bus*: provides the communication services between the software components of the framework even in a distributed computational architecture.
- *Authentication service*: manages the user authentication, permissions, role and a single sign on service provided by partners in the CHIC consortium.
- *Workflow Orchestrator*: it is based on the Taverna server that actually interprets the workflow and coordinates the model execution and data transfer.
- *Registry service*: provides a registry for the services available on the framework and the models that can be executed with their respective status.
- *Storage services*: provide the storage for the workflow inputs and results through a REST interface.
- *Log management*: stores the execution log of all the models composing a workflow.
- *Probabilistic engine*: allows users to create probabilistic variables when required in the workflow execution.
- *Wrapper*: implements a standardised interface to interact with every model instance, executes models, performs the push/pull operations from the data storage services and creates execution logs
- *Mechanical Turk*: an application programming interface, implemented in the *Director*, that integrates functionalities that require human intervention or manual operations during the execution of hypermodels (e.g. request for the user to validate data from partial workflow execution).

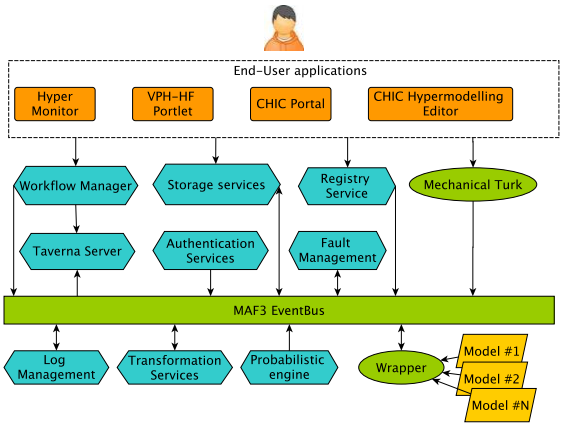


Figure 4. The VPH-HF architecture has a modular design where components (in hexagonal shapes) communicate each other through the MAF3 Event Bus. User can design, execute and monitor workflow execution through Web or stand alone applications (shown in rectangular boxes).

- *Transformation services:* allow the data from one format to be transformed to another according to the model specifications.

C. *The computational scenarios*

The VPH-HF architecture is designed to support several computational architectures since the candidate scenarios are diverse. In fact, some models are computationally intensive, may require high bandwidth, or specific hardware or software with some machine specific license. In particular, we identified and support the following scenarios:

- Parallel systems: multiprocessor systems and clusters for computationally intensive models or models implemented through parallel computing libraries.
- Distributed systems: systems including remote data repositories and storage services, High Performance Computing facilities, architectures with specialised hardware as attached accelerators (GPU, Xeon Phi), systems with specific licensed software.
- Private Cloud infrastructure.

In the CHIC project the target architecture adopted is a private cloud based on OpenStack [10]. The VPH-HF has been deployed on this platform in a configuration where all the VPH-HF services, the CHIC repositories and storage services are installed in the same system. This solution enforces the security requirements and the execution performance especially for those models requiring substantial bandwidth and a high degree of coupling.

D. *The CHIC clinical use case*

In the field of computational cancer research, one typical workflow could, for example, address the growth of a generic tumour embedded in a particular tissue. It can combine models describing the tumour growth, nutrient and oxygen diffusion, vascularization, and actions of any chemotherapeutic agents. Specifically the CHIC project is addressing different cancer types, according to the expertise of the groups involved in the project. These include

glioblastoma, nephroblastoma, lung, prostate and colon cancer. Clinicians are offered visualisation tools (e.g. Dr Eye[4] [11]) to analyse patient data and the hypermodelling framework to predict cancer evolution and the impact of different treatments. The aim of the project is to identify a personalised treatment in a short time to effectively help the clinician to minimise the patient suffering while awaiting treatment and surgery. Simulations based on specific patient data (e.g. imaging data) can better capture the cancer evolution in time and the best treatment strategy to adopt to increase the chances of success.

REFERENCES

[1] T. S. Deisboeck, Z. Wang, P. Macklin, and V. Cristini, "Multiscale Cancer Modeling," *Annu. Rev. Biomed. Eng.*, vol. 13, no. 1, pp. 127–155, Aug. 2011.

[2] H. M. Byrne, "Dissecting cancer through mathematics: from the cell to the animal model," Mar. 2010.

[3] T. S. Deisboeck and G. S. Stamatakos, *Multiscale Cancer Modeling*. CRC Press, 2010.

[4] P. Kohl and M. Viceconti, "The virtual physiological human: computer simulation for integrative biomedicine II," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 368, no. 1921, pp. 2837–2839, Jun. 2010.

[5] "Computer as thinker/doer: problem-solving environments for computational science," *IEEE Computational Science and Engineering*. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=326669. [Accessed: 29-Oct-2014].

[6] G. Clermont, J. Bartels, R. Kumar, G. Constantine, Y. Vodovotz, and C. Chow, "In silico design of clinical trials: A method coming of age," *Critical Care Medicine*, vol. 32, no. 10, pp. 2061–2070, Oct. 2004.

[7] A. Anderson and V. Quaranta, "Integrative mathematical oncology," 2008.

[8] E. Gamma, R. Helm, R. Johnson, and J. Vlissides, *Design patterns: elements of reusable object-oriented software*. 1994.

[9] K. Wolstencroft, R. Haines, D. Fellows, A. Williams, D. Withers, S. Owen, S. Soiland-Reyes, I. Dunlop, A. Nenadic, P. Fisher, J. Bhagat, K. Belhajjame, F. Bacall, A. Hardisty, A. Nieva de la Hidalga, M. P. Balcazar Vargas, S. Sufi, and C. Goble, "The Taverna workflow suite: designing and executing workflows of Web Services on the desktop, web or in the cloud.," *Nucleic Acids Res.*, vol. 41, no. Web Server issue, pp. W557–61, Jul. 2013.

[10] O. Sefraoui, M. Aissaoui, and M. Eleuldj, *OpenStack: toward an open-source solution for Cloud Computing Platform Using OpenStack*. International Journal of Computer …, 2012.

[11] E. Skounakis, V. Sakkalis, K. Marias, K. Banitsas, and N. Graf, "DoctorEye: A multifunctional open platform for fast annotation and visualization of tumors in medical images," presented at the 2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 3759–3762.

[4] http://biomodeling.ics.forth.gr/

Incorporating Data Protection in *In Silico* Research: A Case of the CHIC Project*

Elias Neri and Wouter Dhaeze

***Abstract*— This is a case study of the solution provided by the CHIC project (http://chic-vph.eu) for the processing of sensitive retrospective and prospective patient data in a research environment. The case study focuses on the de-identification aspects of the CHIC data protection solution.**

I. Introduction

One of the aims of the CHIC project is to develop cutting edge ICT tools, services and infrastructure to foster the development of elaborate and reusable integrative models (hypermodels) in the field of cancer diagnosis and treatment. During the development of these tools both retrospective and prospective patient data will be used to develop, test and validate these models. The processing of this sensitive data does have legal and ethical requirements and consequences.

This resulted in a concrete data protection framework which comprises a number of legal and organisational measures (such as contracts) as well as technical measures allowing patient data to be used to develop, test and validate the CHIC *In Silico* models.

This document does not aim to describe all these measures but instead, as a case study, focuses on the measures taken and tools used to de-identify patient data.

II. CHIC Data Protection Framework

Health data of a patient is collected by the treating physicians and analysed and stored within the treating hospital. If a patient agrees to participate in CHIC, the treating physician triggers the transmission of respective medical data.

When exporting data from a hospital for research purposes, anonymisation is the best way to protect a patient's privacy [1]. If research results reveal that a certain therapy would be highly effective for a given patient, CHIC envisages the re-identification towards the hospital of that patient. Therefor the CHIC data protection framework cannot fully anonymise data but instead it is based on de facto anonymous data [2].

The data is first pseudonymised at the hospital. Afterwards this pseudonymised data is uploaded through the CHIC pseudonymisation services in the CHIC data repository (Figure 1). These pseudonymisation services will encrypt all pseudonyms with a key held by a Trusted Third Party (TTP) [5]. This implies that the involvement of the TTP is required to go back from the encrypted to the original pseudonym, resulting in de facto anonymous data. Accordingly, the TTP also serves as vault for the link back to the patient [3].

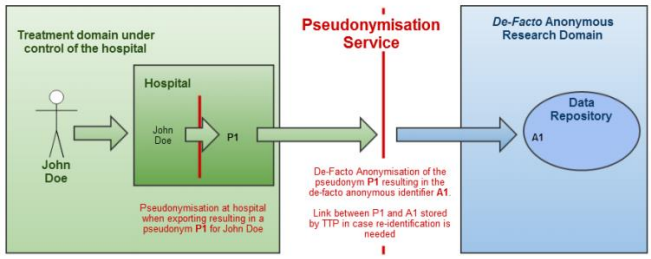


Figure 1. Data Transfer Flow

In order to be able to consider the de-identified data as de facto anonymous; various legal, organisational and technical security measures are required in addition to the involvement of a TTP (such as contracts between all parties, authentication, authorization and auditing) [3][6]. This paper does not describe all these required measures but focuses on the tools and operations available to de-identify the data.

III. CHIC Data Transfer Scenario

A. Create Privacy Profiles

Before the data upload can start, privacy profiles, which define how the data should be pseudonymised, need to be created (Figure 2). As a first step the data uploader/physician needs to create two sets of privacy profiles. One set defines the pseudonymisation processing at the hospital and the other defines the second pseudonymisation round executed by the CHIC pseudonymisation services.

CAT (further described in section 0) is a tool that can be used to create the privacy profiles. Once created, they should be uploaded to the privacy profile store.

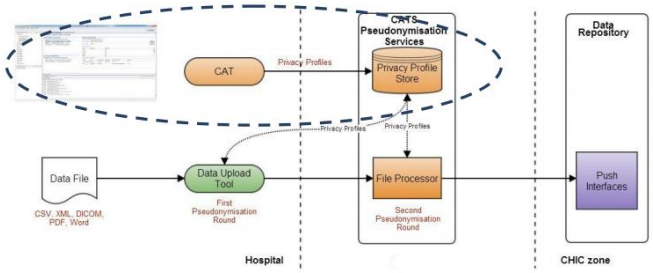


Figure 2. Step 1: Create Privacy Profiles

*The work reported here has been partially funded by the EU-FP7 grant CHIC, Grant Agreement No 600841 (CHIC project). The views expressed are those of the authors and not necessarily those of the Commission.

Elias Neri and Wouter Dhaeze are with Custodix NV, Kortrijksesteenweg 214 b3, 9830 Sint-Martens-Latem, Belgium (corresponding author phone: +3292107890; fax: +3292110666; email: elias.neri@custodix.com; wouter.dhaeze@custodix.com)

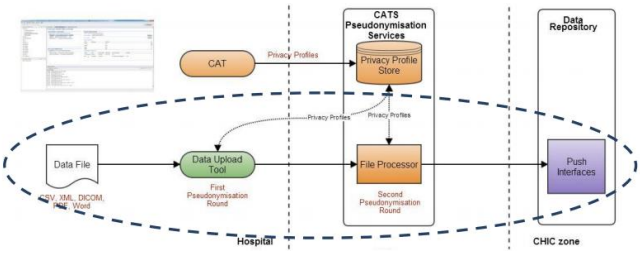


Figure 3. Step 2: Process and Upload Data

B. *Process and Upload Data*

The physician exports the data form the hospital database or information system to his local drive. By means of the CHIC data upload tool (Figure 3), the physician can select these exported files. The data upload tool wraps the pseudonymisation engine, which is responsible for the de facto anonymisation. Before the data upload tool (i.e. the Pseudonymisation engine) starts processing a given file, the correct privacy profile must be selected. The privacy profile store holds multiple privacy profiles, each of them designated to a given file format or given file content. The anonymisation engine will check the media type of the input file, select the privacy profile, and start processing. The resulting pseudonymised file is rendered on screen by the data upload tool, for verification.

Once the physician confirms that the data is correctly de-identified, the data is uploaded to the CHIC pseudonymisation services implemented by CATS (Custodix Anonymisation Tool services). By selecting relevant second round privacy profiles, CATS (i.e. the Pseudonymisation engine embedded in CATS) will encrypt all pseudonyms with a key held by the TTP. Optionally the data may be put on hold until confirmed that it is indeed de facto anonymous. Once approved the data is uploaded to the CHIC data repository and available for use by the CHIC partners to develop, test and validate their models.

IV. CHIC Pseudonymisation Services & Tools

A. *Pseudonymisation Engine*

At the core of the CHIC Pseudonymisation Services lies the Custodix Pseudonymisation Language and Engine. The Pseudonymisation Engine de-identifies data files by executing a privacy profile written in the Pseudonymisation Language. This language allows source data files (such as CSV, XML, Text, DICOM, microarray data and relational databases through SQL) through the privacy profile to be mapped to a generic data model (Figure 4). The profile in addition defines the operations to be executed on the data model. This approach allows the uniform processing of data in different formats (i.e. the same operations can be executed on different data formats such as CSV or XML). Not only is this more convenient in setting up a project, it also provides a higher assurance level with respect to compliance.

The engine does not only allow source variables to directly be mapped to corresponding generic variables, but in additions variables can also be mapped to privacy types such as identifier, person and date. Operations can then be executed on all variables which are a date, a person …

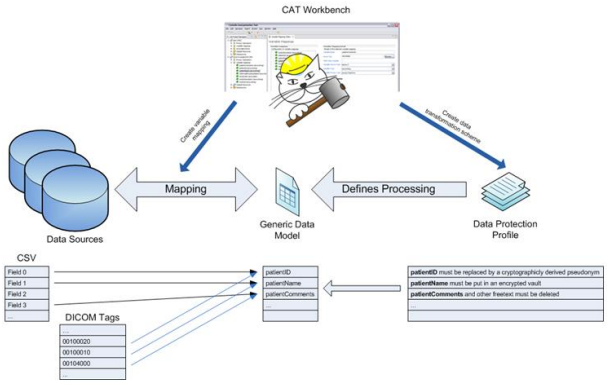


Figure 4. CATS mapping & operations definition

The engine contains out of the box privacy instructions such as:

- Remove patient names and identifiers
- Calculate a pseudo-ID based on a patient-ID
- Make all patient visit-dates relative to the patient data of birth and randomise that date.
- Process all free text to remove identifying information through named entity recognition
- Aggregate the data
- Etc.

The Pseudonymisation Engine does not provide a human interface. It is designed as an API to be included in applications and tools that need file de-identification capabilities.

B. *CHIC Upload Tool*

The CHIC data upload tool is a Java client application that allows a physician to locally anonymise files by integrating the Pseudonymisation Engine. The result of the processing is rendered on screen for review by the physician. Upon approval the application uploads the pseudonymised file to the CATS web application for further processing.

C. *CATS*

CATS [4] is a web-based application that serves two major goals:

1. Act as a privacy profile store.
2. Perform the second pseudonymisation round.

As described in section [III.A] a physician can create a privacy profile intended for a given file type (and file content). Privacy profiles must be uploaded to the central CATS instance. This way privacy profiles can be shared by multiple physicians uploading files in similar formats.

CATS main goal remains processing data files. Consequently CATS integrates the Pseudonymisation Engine as well. It uses its local store as privacy profile provider.

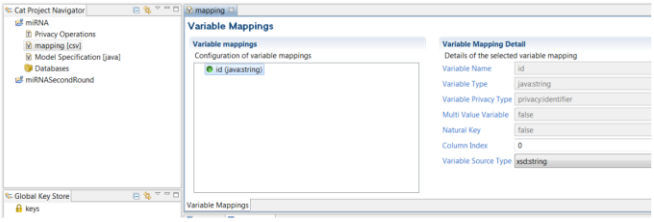


Figure 5. Cat Workbench through which Privacy Profiles can be created

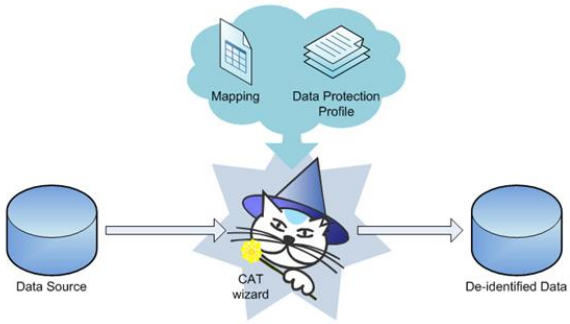


Figure 6. The CAT wizard aids in testing profile definitions on locally available test data sources

D. CAT

Writing a privacy profile in XML format can be a cumbersome and painstaking task. CAT [7] provides a graphical interface (Figure 5) to create a profile. By means of CAT a data model is created, operations on that model are defined and a mapping between (structured) data files and the data model is defined. CAT can upload a privacy profile to the central store.

In addition, CAT can be used to test privacy profiles locally (Figure 6). Creating and optimizing a privacy profile is a process of trial and error. In turn, CAT integrates the Pseudonymisation Engine in order to test profile definitions on test data.

V. EXAMPLE

As example we will define a privacy profile for a CSV data file with the headers as specified in Table .

Table I Headers of example CSV File

Patient ID	Name	DOB	Visit Date	Diagnosis

- Patient ID is a string that refers to the local centre identifier.
- Name is a string representing the patient's name
- DOB is the patient's date of birth.
- Visit Date is the date at which the patient visited the centre.
- Diagnosis is a free text field containing a description of the diagnoses. This field could potentially contain identifying information such as names, identifiers and dates.

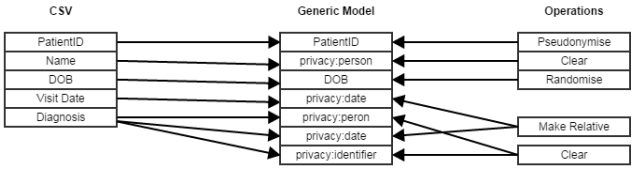


Figure 7. Privacy Profile Schematic Overview

```
<simpleVariable index="0" name="patientID" sourceType="xsd:string" type="java:string"/>
<simpleVariable index="1" name="name" privacyType="privacy:person" sourceType="xsd:string" type="ja
<simpleVariable index="2" name="DOB" sourceType="xsd:string" type="java:date" conversionParams="com
<simpleVariable index="3" name="visitDate" privacyType="privacy:date" sourceType="xsd:string" type="
<simpleVariable index="4" name="diagnosis" privacyType="privacy:freeText" sourceType="xsd:string" ty
```

Figure 8. CATS Model Mapping

Figure 7 gives a schematic overview of a privacy profile. Take particular note of the Diagnosis field. The pseudonymisation engine can automatically detect privacy sensitive text in free text fields. These will then be assigned a privacy concept so that they can be processed by operations that are defined on those corresponding concepts.

As a first step in defining the privacy profile the CSV schema should be mapped to a generic CATS data model (Figure 8).

All fields with as privacy type "freetext" will be parsed by the free text engine and consequently split up in sub variables with privacy types such as "privacy:identifier", "privacy:date" and "privacy:person".

Now that all variables are mapped the operations should be created. Figure 9 defines the replacement of the patient identifier with a random generated anonymous identifier. Figure 10 anonymises the date of birth and Figure 12 moves all other dates accordingly. Figure 11 defines the removal of all person names and identifiers.

```
<function ref="core:randomAnonId">
    <argument name="outputLength">
        <constant type="java:int" value="20" valueType="xsd:string"/>
    </argument>
    <returnValue name="patientID"/>
</function>
```

Figure 9. Anonymise Patient Identifier

```
<function ref="core:randomDate">
    <returnValue name="DOB"/>
</function>
```

Figure 10. Anonymise Date of Birth

```
<forEachVariable itemName="var" privacyType="privacy:person">
    <function ref="core:clear">
        <returnValue name="var"/>
    </function>
</forEachVariable>
<forEachVariable itemName="var" privacyType="privacy:identifier">
    <function ref="core:clear">
        <returnValue name="var"/>
    </function>
</forEachVariable>
```

Figure 11. Clear all person names and identifiers

```
<forEachVariable itemName="var" privacyType="privacy:date">
    <function ref="core:makeDateRelative">
        <argument name="inputDate">
            <variable ref="var"/>
        </argument>
        <argument name="referenceDate">
            <!-- take as reference original input date of birth-->
            <variable ref="DOB"  original="true"/>
        </argument>
        <argument name="newReferenceDate">
            <!-- take as new reference the randomised date of birth-->
            <variable ref="DOB" />
        </argument>
        <returnValue name="var"/>
    </function>
</forEachVariable>
```

Figure 12 Recalculate all dates to the new date of birth as reference date

VI. Conclusion

This paper described the CHIC Pseudonymisation Services and Tools as a case study of incorporating data protection in In Silico Research. CHIC makes use of tools such as an upload and Custodix anonymisation tool in combination with legal, organizational and technical security measures. This results in a data transfer protocol which ensures that imported data is de facto anonymous and can thus be used to develop, test and validate the models.

References

[1] N. Forgó, A. N. Haidar and H. Gerhartinger. “Security and Privacy in Sharing Patient Data,” in Computational Biomedicine, Modelling the Human Body, P. Coveney, V. Diaz-Zuccarini, P. Hunter and M. Viceconti (editors), Oxford, 2014. 207-231.

[2] N. Forgó. “Ethical and Legal Requirements for Transnational Genetic Research,” Hart Publishing, 2010.

[3] I. Nwankwo, M. Stauch, I. Lishchuk, E. Neri and N. Forgó, “Development of the data protection and copyright framework for CHIC first iteration,” 2014.

[4] “The CATS de-identification platform,” Available: https://www.custodix.com/index.php/cats

[5] Jefferys BR, Nwankwo I, Neri E, Chang DCW, Shamardin L, Hänold S, Graf N, Forgo´ N and Coveney P., “Navigating legal constraints in clinical data warehousing: a case study in personalized medicine,” 2013, Interface Focus 3: 20120088.

[6] I Nwankwo, S. Hänold and N. Forgó, “Legal and ethical issues in integrating and sharing databases for translational medical research within the EU,” IEEE 12th International Conference on BioInformatics & BioEngineering, 2012, S. 428-433; doi: 10.1109/BIBE.2012.6399764.

[7] Brecht Claerhout, “The CAT Platform,” in P-Medicine Newsletter Issue no 3, 2014.

MyHealthAvatar Survey: Scenario Based User Needs and Requirements*

Ruslan David, Feng Dong, Yvonne Braun, and Norbert Graf, *Member, IEEE*

The MyHealthAvatar (MHA) is an EC FP7 project aimed to focus on research and demonstration actions, through which the achievability of an innovative representation of the health status of citizens is explored. The aim of MHA survey was to enhance our understanding of patient and public views about the MHA platform. First, we investigated the background (sociodemographic characteristics) of our respondents, their understating of the basic terms like EHR and PHR. Second, we addressed complex questions about the expected and proposed for implementation MHA platform's functionalities and of special interest were the questions about the security and privacy concerns the end users might have.

In general terms the survey allowed us to examine individuals' specific hopes and concerns about MHA platform and now we have a deeper understanding of patient and public views about further MHA platform's functionalities.

I. Introduction

The MyHealthAvatar (MHA) EC FP7 project [1] is focused on research and demonstration actions, through which the achievability of an innovative representation of the health status of citizens is explored. The Avatar is anticipated as an interface that will allow data access, collection, sharing and analysis by utilizing modern ICT technology. It is expected to become the citizen's lifelong companion, providing long-term and consistent health status information of the individual citizen along a timeline representing the citizen's life, starting from birth. Data sharing will be encouraged, which will potentially provide to an extensive collection of population data to offer extremely valuable support to clinical research. The avatar will be equipped with a toolbox to facilitate clinical data analysis and knowledge discovery.

MHA can be described as a personal bag carried by individual citizens throughout their lifetime. It is a companion that will continually follow the citizen and will empower them to look after their own health records. This fits very well into the recent trend of developing patient-centred healthcare systems.

The initial purpose of MHA project was to describe and to underline the specific user needs by being focused on comprehensive and detailed MHA Scenarios / Use Cases. This purpose has been successfully extended to cover as well the general public and patient's views and their expectations for MHA platform. The achieved results are presented below.

II. Survey Design

A. Survey tool

The survey has been elaborated by using a web-based, open source, survey management system, named LimeSurvey [2]. All questions with related answers have been in details discussed with all project partners. The received suggestions related to survey structure improvements have been implemented. As result, a survey with minimum questions but with a wide coverage has been elaborated and disseminated to the available news channels and audiences.

MHA Survey has been designed with the requirements to keep the questions and answers as much as possible understandable for general public, simple to follow and with minimum possible amount of items (questions and predefined answers). As a performance check all survey respondents have been allowed to share their feedback on the survey. Additionally, we allowed all survey respondents to provide us with their contact e-mail in case of interest to receive the final survey report.

The final version of MHA Survey (English version) has been translated in two additional languages (German and Hungarian).

B. Survey questions

MHA survey questions have been divided in four major sections:

- General section with questions to underline the background of the survey respondents. First, simple sex and age related questions, and, afterwards, more complex information have been requested (e.g. health status, computer skills, social networks awareness, healthcare related job).
- Medical / Health Records section with questions to underline the knowledge and the experience of the survey respondents with EHR / PHR systems. Despite only two visible questions on this section, a number of additional questions had the interactive

The research leading to these results has received funding from the European Union's Seventh Programme for research, technological development and demonstration under grant agreement No [600929] (MyHealthAvatar)

R. David is with the Saarland University, Dep. Pediatric Oncology and Hematology, 66421 Homburg, Germany (e-mail: ruslan.david@uks.eu).

F. Dong. Author is with the Department of Computer Science and Technology, University of Bedfordshire, Luton, UK (e-mail: Feng.Dong@beds.ac.uk).

Y. Braun is with the Saarland University, Dep. Pediatric Oncology and Hematology, 66421 Homburg, Germany (e-mail: yvonne.braun@uks.eu).

N. Graf is with the Saarland University, Dep. Pediatric Oncology and Hematology, 66421 Homburg, Germany (corresponding author phone: 0049 6841 1628397; fax: 0049 6841 1628302; e-mail: graf@uks.eu).

status (e.g. If 'Yes' answer, 'Please specify' question has been addressed).

- MHA functionalities section with questions to underline the most expected functionalities from MHA platform. Additionally, it allowed us to refresh the elaborated Scenarios / Use Cases and to priorities MHA functionalities in line with end-users expectations and views.
- Security and Privacy section with questions in line with MHA project' Work Package 11. Our main goal was to identify the end-users' perspectives, views and concerns on security and privacy of the proposed for implementation MHA platform.
- Feedback section had only two questions, the first question allowed all survey respondents to share any feedback on our survey and the second question allowed all interested respondents to provide their e-mail in order to receive the final survey report.

III. Survey Data Analysis

Survey responses have been in details presented in the frames of the MHA project Deliverable No. 2.2 "Scenario Based User Needs and Requirements" document [3]. The survey has collected in total 270 responses (with 161 full responses and 109 incomplete responses).

A. General questions

General section with its related questions has been elaborated in order to underline the background of the survey respondents. Survey started with simple sex and age related questions, and, afterwards, more complex information have been requested (e.g. health status, computer skills, social networks awareness, healthcare job). Some selected questions with related answers and collected data are presented in Table I, Table II, Table III, and Table IV.

TABLE I. Please select your gender

Answer	Count	Percentage (%)
Female (F)	97	42,73
Male (M)	88	38,77
No answer	42	18,50

TABLE II. Please select your age range

Answer	Count	Percentage (%)
< 20	5	2,20
20 – 35	78	34,36
36 – 45	69	30,40
46 – 55	31	13,66
56 – 65	20	8,81
> 65	4	1,76
No answer	20	8,81

TABLE III. What is your highest level of education?

Answer	Count	Percentage (%)
No academic qualification	4	1,76
Elementary school	2	0,88
Vocational Qualification (e.g. technical college)	23	10,13
Higher degree	156	68,72
No answer	42	18,50

TABLE IV. How would you rate yourself computer skills?

Answer	Count	Percentage (%)
A super user	42	18,50
Advanced user	95	41,85
Intermediate user	49	21,59
Basic user	5	2,20
New user	1	0,44
No answer	35	15,42

B. Medical / Health Records questions

With 'Medical / Health Records' related questions we had the goal to underline the knowledge and the experience of the survey respondents with EHR / PHR systems. Despite only two visible questions on this section, a number of additional questions had the interactive status (i.e. If 'Yes' answer, 'Please specify' question has been addressed). The selected questions with related answers and collected data are presented in Table V and Table VI.

C. MHA Functionalities questions

The selection of the collected result are presented in Table VII, Table VII, Table IX, Table X, Table XI, and Table XII.

TABLE V. Have you ever heard anything about Electronic Health Records (EHR)?

Answer	Count	Percentage (%)
Yes (Y)	140	71,07
No (N)	56	28,43
No answer	1	0,51

TABLE VI. Do you use the electronic Personal Health Records (PHRs)?

Answer	Count	Percentage (%)
Yes (Y)	27	13,71
No (N)	169	85,79
No answer	1	0,51

TABLE VII. Enter, import, store and export personal medical data

Answer	Count	Percentage (%)
Very interested	95	52,78
Somewhat interested	46	25,56

Unsure	17	9,44
Somewhat uninterested	10	5,56
Very uninterested	9	5,00
No answer	3	1,67

TABLE VIII. SCHEDULE AND COORDINATE MEDICAL APPOINTMENTS

Answer	Count	Percentage (%)
Very interested	79	43,89
Somewhat interested	57	31,67
Unsure	17	9,44
Somewhat uninterested	12	6,67
Very uninterested	11	6,11
No answer	4	2,22

TABLE IX. FIND PERSONALISED INFORMATION ABOUT THE DISEASES, DRUGS, VITAMINS, FOOD, ETC.

Answer	Count	Percentage (%)
Very interested	91	50,56
Somewhat interested	54	30,00
Unsure	15	8,33
Somewhat uninterested	15	8,33
Very uninterested	2	1,11
No answer	3	1,67

TABLE X. ENTER, IMPORT, STORE AND EXPORT YOUR AVATAR PERSONAL MEDICATION LIST

Answer	Count	Percentage (%)
Very interested	77	42,78
Somewhat interested	53	29,44
Unsure	26	14,44
Somewhat uninterested	6	3,33
Very uninterested	14	7,78
No answer	4	2,22

TABLE XI. ABILITY TO MANAGE WHO HAS ACCESS TO YOUR INFORMATION (AVATAR)

Answer	Count	Percentage (%)
Very interested	139	77,22
Somewhat interested	22	12,22
Unsure	7	3,89
Somewhat uninterested	3	1,67
Very uninterested	4	2,22
No answer	5	2,78

TABLE XII. TRACK WHO HAS ACCESSED YOUR DATA/INFORMATION (AVATAR)

Answer	Count	Percentage (%)
Very interested	144	80,00
Somewhat interested	15	8,33
Unsure	6	3,33
Somewhat uninterested	4	2,22
Very uninterested	6	3,33
No answer	5	2,78

D. Security and Privacy questions

Some selected questions with related answers from 'Security and Privacy' section are presented in Table XIII and Table XIV.

TABLE XIII. DO YOU HAVE ANY SECURITY CONCERNS ABOUT MYHEALTHAVATAR PLATFORM?

Answer	Count	Percentage (%)
Yes	114	67,46
No	24	14,20
I'm not sure	30	17,75
No answer	1	0,59

TABLE XIV. DO YOU HAVE ANY PRIVACY CONCERNS ABOUT MYHEALTHAVATAR PLATFORM?

Answer	Count	Percentage (%)
Yes	115	68,05
No	26	15,38
I'm not sure	27	15,98
No answer	1	0,59

E. Feedback questions

42 responses to the question 'Please provide any feedback on this survey' have been received and analyzed. We are proud to conclude that almost all received comments were positive and related to good wishes in our further project activities.

As response to the possibility to leave the e-mail in order to receive the final survey results, 75 e-mail addresses have been provided.

IV. SURVEY CONCLUSIONS

One of the major conclusions is related to the high interest of all survey respondents in further MHA platform.

MHA survey has been on-line for only two months (June 2013 and July 3013). Nevertheless, we managed to collect in total 270 responses (with 161 full responses and 109 incomplete responses).

The received responses allowed us to refresh the elaborated MHA project Scenarios / Use Cases. Additionally, new Scenarios / Use Cases will be elaborated by taking into account the collected survey results.

Of special interest is the advanced profile of our survey respondents. Here we would like to mention that according to the collected responses:

- 42,73% of survey respondents are 'Female' and 38,77% are 'Male' (Table I);
- The top selected age ranges are 20-35 Years and 36-45 Years (Table II);

- 68,72% of survey respondents reported ‘Higher Degree’ as the highest level of education (Table III);
- Most of the survey respondents have advanced (41,85%) and intermediate(21,59%) computer skills (Table IV).

In the terms of the survey respondents’ knowledge and information about Medical / Health Records we would like to underline that according to the collected responses:

- 71,07% of survey respondents are aware about HER (Table V);
- 58,93% of survey respondents (with no awareness about EHR) would be interested in finding more information about EHR in the frames of MHA project’s web page;
- 85,79% of survey respondents are not using PHR (Table VI).
- 66,86% of respondents (which are not using any PHR) reported their interest in a new, secure, advanced and personal health platform that lets gather, store, analyse and visualise health information online.

MHA functionalities section from MHA survey allowed us to highlight the most expected functionalities from further MHA platform. Additionally, it allowed us to refresh the elaborated Scenarios / Use Cases and to priorities MHA functionalities in line with end-users expectations and views. Of special interest are:

- 52,78% of respondents are ‘Very’ interested and 25,56% are ‘Somewhat’ interested in 'Enter, import, store and export personal medical data (e.g. Electronic Health Records)' (Table VII)
- 43,89% of respondents are ‘Very’ interested and 31,67% are ‘Somewhat’ interested in 'Schedule and coordinate medical appointments' (Table VIII)
- 50,56% of respondents are ‘Very’ interested and 30,00% are ‘Somewhat’ interested in 'Find personalised information about the diseases, drugs, vitamins, food, etc.' (Table IX)
- Most of the respondents (27,22%) are 'Very uninterested' and 25,56% are ‘Unsure’ about the function/functionality to 'Provide/Share your data/information (Avatar) to other Avatar(s)'
- 41,67% of respondents are ‘Very’ interested and 35,00% are ‘Somewhat’ interested in 'Provide/Share your personal Avatar to your doctor (e.g. GP)'
- 27,78% of respondents are ‘Very’ interested and 24,44% are ‘Unsure’ about the 'Ability to manage your personal Avatar participation in health research (e.g. Clinical Trials)'
- 42,78% of respondents are ‘Very’ interested and 29,44% are ‘Somewhat’ interested in 'Enter, import, store and export your Avatar personal medication list' (Table X)
- 48,33% of respondents are ‘Very’ interested and 30,00% are ‘Somewhat’ interested in 'Enter, import, store and export your Avatar personal laboratory results'
- 77,22% of respondents are ‘Very’ interested and 12,22% are ‘Somewhat’ interested in 'Ability to manage who has access to your information (Avatar)' (Table XI)
- 80,00% of respondents are ‘Very’ interested and 8,33% are ‘Somewhat’ interested in 'Track who has accessed your data/information (Avatar)' (Table XII)
- 47,22% of respondents are ‘Very’ interested and 21,67% are ‘Somewhat’ interested in 'Ability to access and manage your Avatar by using a mobile application (App)'
- 51,67% of respondents are ‘Very’ interested and 24,44% are ‘Somewhat’ interested in 'Enter, import, store and export your personal medical images (e.g. DICOM files)'

In the frames of the Security and Privacy section of MHA Survey our main goal was to identify the end-users’ perspectives, views and concerns on security and privacy of the proposed for implementation MHA platform, the top collected results are:

- 67,46% of respondents have security concerns (Table XIII) and 39,05% of respondents selected the ‘High risks’ option, and 43,79% opted for ‘Moderate risks’ of the level of possible risks to the security of MHA platform
- 68,05% of respondents have privacy concerns (Table XIV) and 40,24% of respondents selected the ‘High risks’ option, and 42,01% opted for ‘Moderate risks’ of the level of possible risks to the privacy of MHA platform

In general terms the survey allowed us to examine individuals’ specific hopes and concerns about MHA platform and now we have a deeper understanding of patient and public views about further MHA platform’s functionalities.

ACKNOWLEDGMENT

The authors would like to thank all MHA project partners for their support and assistance by conducting the survey.

REFERENCES

[1] MyHealthAvatar Project, http://www.myhealthavatar.eu (October 2014)

[2] LimeSurvey is a free and open source on-line survey application written in PHP based on a MySQL, PostgreSQL or MSSQL database, distributed under the GNU General Public License, http://www.limesurvey.org

[3] MHA project Deliverable No. 2.2, “Scenario Based User Needs and Requirements”, Chapter IV “MHA Survey”, Deliverables section from MHA project web site http://www.myhealthavatar.eu/?page_id=1519 (October 2014).

Multi-Modal Medical Data Analysis Platform (3MDAP) for Analysis and Predictive Modelling of Cancer Trial Data*

Georgios C. Manikis, Evangelia Maniadi, Manolis Tsiknakis, *Member IEEE*, and Kostas Marias, *Member, IEEE*

***Abstract*— This paper presents a user-friendly web-based collaborative environment for analyzing, assessing the quality of large multi-level clinical datasets and deriving predictive models. The Multi-Modal Medical Data Analysis Platform (3MDAP) follows two main objectives: a) to empower the user to analyze with ease clinic-genomic data in order to get simple statistics on selected parameters, perform survival analyses, compare regiments in selected cohort of patient and obtain genomic analysis results, and b) to perform heterogeneous clinical data modeling for deriving and cross-validating in multiple datasets predictive clinic-genomic models of patient response, and assessing the value of candidate biomarkers.**

3MDAP's enhanced functionality is coupled with a security framework for enabling user authentication and authorization, a set of services that facilitate the process of loading and retrieving data from a data-warehouse (either locally based or in a cloud), and a widget-based front-end environment for assisting users in interacting with the platform's functionality in a user friendly manner. For each running analysis, 3MDAP supports an engine to create dynamically analysis reports. Last, the framework provides an internal database where a full analysis record of an executed analysis is stored, including metadata information (i.e. timestamp information, the examined data, any memory constraints, the dynamically generated reports in both .pdf and .html format, and etc.) in order to be used for future reference.

I. Introduction

The proposed platform for cancer clinic-genomic data analysis is centered in empowering the user (e.g. clinical researcher or bioinformatician) to obtain simple descriptive statistics and compute with simple, high-level operations predictive models that can be seamlessly validated in multiple datasets within a single platform in the same session. The main design objective is to allow any user to use the platform even without expertise in computational tools such as the R software environment for statistical computing [1]. The developed predictive analysis functionality featuring a comprehensive clinical trial data-viewer has been largely driven by the clinical scenarios for the INTEGRATE VPH project [2] as well as by extensive discussions with experts bioinformaticians/clinicians involved in the project.

This paper presents in detail the platform explaining how it is capable to enable scientists from diverse backgrounds to employ with ease (at the push of a button) a) sophisticated statistical analysis tools that play and important role in deeply understanding and preparing the available multi-level data for further analysis, and, b) to derive predictive models (again at the push of the button) from cancer clinical trial data.

*Research supported from the INTEGRATE project funded by the European Commission under the 7th Framework Programme.

G.C. Manikis is with the Computational Medicine Laboratory (CML) of the Institute of Computer Science (ICS) in the Foundation for Research & Technology - Hellas (FORTH), Vassilika Vouton, P.O Box 1385, GR-71110 Heraklion, Crete, Greece (corresponding author phone: +30-2811-391593; fax: +30-2810-391428; e-mail: gmanikis@ics.forth.gr).

E. Maniadi (maniadi@ics.forth.gr), M. Tsiknakis (tsiknaki@ics.forth.gr), and K. Marias (kmarias@ics.forth.gr) are with the Computational Medicine Laboratory (CML) of the Institute of Computer Science (ICS) in the Foundation for Research & Technology - Hellas (FORTH).

G.C. Manikis and E. Maniadi contributed equally to this work.

II. 3MDAP Platform

A. System Architecture

The idea behind 3MDAP is to provide users with a web-based interface that supports user authentication and authorization, data handling, execution of the tools and models, and visualization and storage of the analysis reports. To achieve this goal, the programming aspects of the different environments and languages adopted for implementing the framework's facilities, and the connectivity process which allows the interaction between these components are kept at the back-end of the framework, hiding the complexities of the computational infrastructure. The architecture and specifications of the developed framework are divided into the following fields:

- The core functionality of the platform
- The authentication and authorization process
- The data retrieval system
- The web services infrastructure

1) The core functionality

From the technical perspective, the core functionality is composed of the front-end and the back-end component of the platform. The front-end, hiding the complex infrastructure is based on the Liferay Portal [3]. Liferay Portal is an enterprise web framework based on Java technologies. Our Liferay based front-end is enhanced with JavaServer Faces (JSF), a Java technology for building component-based user interfaces for web applications. An Ajax-based JSF framework named as PrimeFaces [4] was chosen to be used in 3MDAP, offering over 100 individual components, covering a diverse range of widgets including Ajax, input fields, buttons, data display controls, panels, overlays, menus, charts, dialogs, multimedia presentations, drag/drop and other controls.

The back-end (Fig.1) consists of a complex heterogeneous environment of several software components. The main part of it is the statistical and predictive modelling analysis software scripts, implemented in R language [1] and using publicly available libraries from its large repository. To facilitate embedding R functionality in our Java-based interface, a client/server concept using TCP/IP protocol [5] is used for the communication between the R system and the front-end allowing interaction between the analysis platform and the execution environment. At the same time connections between multiple clients-users and the R system are established using their own data space and working directory without interfering with other connections.

The platform supports an engine [6] to create dynamically statistical and predictive analysis reports by enabling integration of R code and Latex documentation [7]. On-the-fly reporting in both .pdf and .html format is generated by combining the programming source code and the corresponding documentation into a single file. 3MDAP is equipped with an internal database that stores all the metadata information for every executed analysis. In other words, all users have a private space were all analyses are stored and can be used at a later stage for further analysis (e.g. an already built model can be used to predict the clinical response of new cancer trial data). Each analysis record contains metadata information such as timestamp information, type of the analysis (e.g. descriptive statistics), selected variables used for the analysis, execution time, the analysis cohort in a tabular format, the status of the analysis (e.g. in progress or completed), etc. This metadata information is stored in the platform's database where allowed users can navigate through the database for a) viewing the generated analysis report in either .pdf or .html format, b) edit the .html report(s) using a basic editing toolbar and save the changes back to the server, c) compare the results from different executed analyses by vertically aligning their html reports in the browser, and d) view the selected cohort that was used as input data for the analysis in a tabular format.

2) Authentication and authorization

3MDAP relies on a security framework for enabling authentication and (basic) authorization. For enabling the authentication, the Liferay standard authentication modules are extended and connected to a central Identity Provider (IdP). This IdP provides an implementation of the Single-Sign-On (SSO) browser based on a Security Assertion Markup Language (SAML) profile. SSO is a property of access control that permits user to log in once and gain access to multiple applications concurrently. If a user tries to access the platform, he is redirected to this IdP and a security token is issued. This security token will then be validated by the local authentication module of the security framework and the obtained validation result is used to make an access decision for that specific user.

3) Retrieving the analysis data

The trial data in the 3MDAP platform can be either retrieved by a local database or from external repositories. To achieve interoperability between the 3MDAP platform and an external data repository, web-services are deployed at the basis of security proxy servers. Using them the data is retrieved and saved locally in order to perform the multiple statistical and predictive analyses. More specifically, during data retrieval the necessary queries are built by the framework and sent to the repositories over the web-service. Then the queries are executed and the information is returned back to the platform. Once the data is retrieved, the user can execute the provided tools and models.

4) Web services and interactivity with other platforms

The integration between the 3MDAP platform and external tools allows to query and filter large sets of patient data available in order to compose cohorts for further analysis. This is also available via a secure web-service and interoperability is achieved by having as a prerequisite that both 3MDAP and

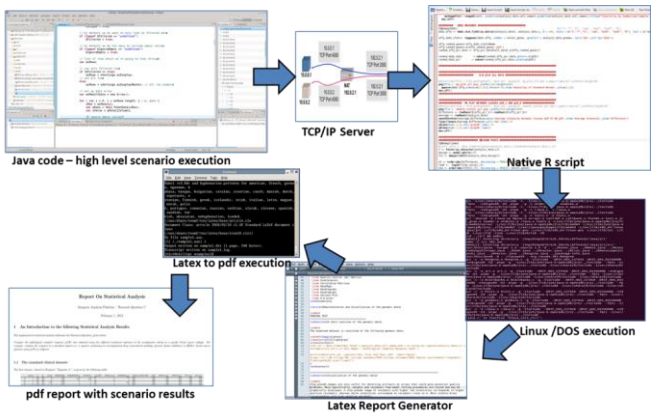


Figure 1. The back-end functionality of the platform

the cohort selection tools achieve connectivity to the same data repository. Using the cohort selection tool the user gets access to the database, and defines cohorts on the fly, by using SNAQL queries. These queries can be very complex, allowing the user to find highly specific patient cohorts in the datasets. The web-service has been deployed using the Apache CFX 2.7 framework [8] while the communication between the two platforms is over secured http. The web-service contains information about the selected cohort, the selected statistical analysis (e.g. apply descriptive statistics to tumor grading size of the selected population) and the analysis results (figures, tables, etc.).

B. The analysis scenarios covered by the platform

In the functional view of 3MDAP, the platform assists the statistical analysis and predictive modelling regarding patient response through a semi-automatic strategy that involves specific scenarios tailored for the needs of analyzing either homogeneous or heterogeneous multi-modal data. These scenarios include:

- Descriptive statistics for rapidly assessing the variability, dependency and the distribution of certain clinical characteristics across patient population.
- Comparison tests and evaluation of the response rate of different examined regimens when applied to a certain patient population.
- Defining if specific clinical parameters are surrogate markers for survival, involving the modeling of time to event data in survival analysis.
- Performing quality control tests to the genomic data, identifying statistically significant genomic information that discriminates subpopulations (i.e. patients achieved pathological complete response VS patients who didn't), and apply unsupervised learning techniques to the entire genomic information.
- Assisted predictive analysis model when homogeneous data (i.e. gene expression) is used for building, running and evaluating the predictive efficacy of the model.
- Heterogeneous integration modeling framework where multi-modal data are fused for the

development of multi-scale models for predicting drug response, and assessing candidate biomarkers.

C. *Workflow*

Each component at the front-end of the platform plays a specific role, starting from the "user authentication" for allowing accessibility to 3MDAP, the "data sources" for interacting and retrieving the analysis data from a data repository, the "analytical tools" for performing the statistical analysis, the "predictive models" for doing the predictive modelling, and finally the "history" component for accessing the internal database of the framework where the metadata information of every executed analysis is stored. A general pipeline workflow using the platform is as follows:

- The platform authenticates the user with the provided credentials and interacts with the data-warehouse to retrieve the data.
- Data are then displayed in a widget-based table and a filtering functionality allows the user to constrain a request by obtaining subpopulations and build cohorts based on specific ranges of values.
- For the selected cohort multiple tools or models can be scheduled for execution in a single step.
- The layout of the platform communicates with the back-end functionality and the required software, and the overall analysis workflow is presented in a functional diagram format.
- A table with metadata information for each completed or pending analysis is displayed to the user. Additionally the user can view, edit or compare the reports of completed analyses.

Fig.2 shows a screen shot of the platform's workflow while executing selected tasks and Fig.4 and 5 some indicative genomic and descriptive statistics result outputs.

III. Indicative Analysis Results

The integration of heterogeneous multi-scale biomedical data for predicting drug response is one of the major challenges in 3MDAP. Different data streams like clinical, demographic, genomic, etc. are represented in a unified framework, overcoming differences in scale and dimensionality. Therefore, aside of a predictive model implemented within the platform for analyzing homogeneous data (e.g. genomic data), the platform formulates the data integration task in machine learning terms, relying on kernel-based methods [9] in order to construct integrated meta-datasets for prediction analysis. Fig. 3 shows the selection process of heterogeneous parameters for predictive modelling by the user. By implementing a Multiple Kernel Learning (MKL) model [10] for data integration, the heterogeneities of the multi-modal data are resolved by transforming the different data into kernel matrices. The MKL model is then extended in feature selection techniques applied to kernel space, where data that contribute to the highest

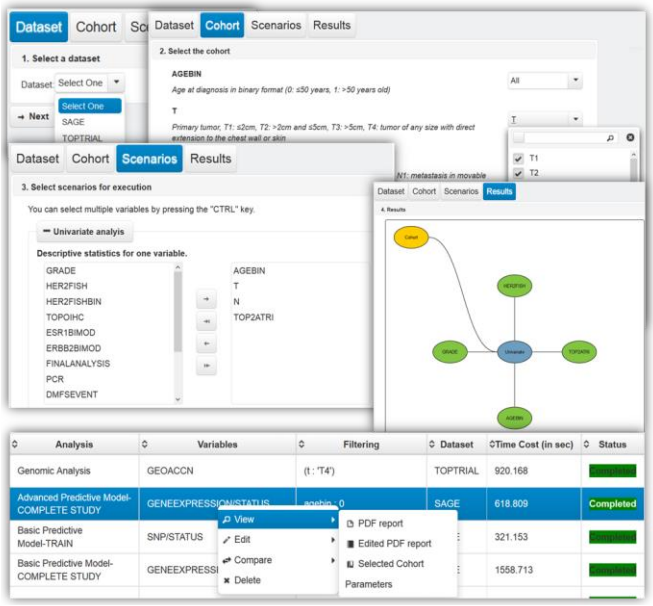


Figure 2. An indicative workflow of the platform

discrimination between the classes are chosen as the most significant for classification [11].

MKL uses each input variable to generate its corresponding kernel and aims to select the relevant variables of the corresponding kernels according to their relevance to the task of classification. In this way, the variable weights and the classification boundary are trained simultaneously and the most relevant variables (variables with the highest weighted value) that leading to the best classification performance are selected. During the process, cross-validation techniques for estimating the generalization performance in this context in a way to protect the classification model against over-fitting are applied. Finally, metrics for evaluating the classification performance such as the sensitivity, specificity, accuracy, precision and area under the curve (AUC) are computed and reported over the total number of the iterative procedure. Fig. 6 shows indicative predictive analysis results using the MKL model.

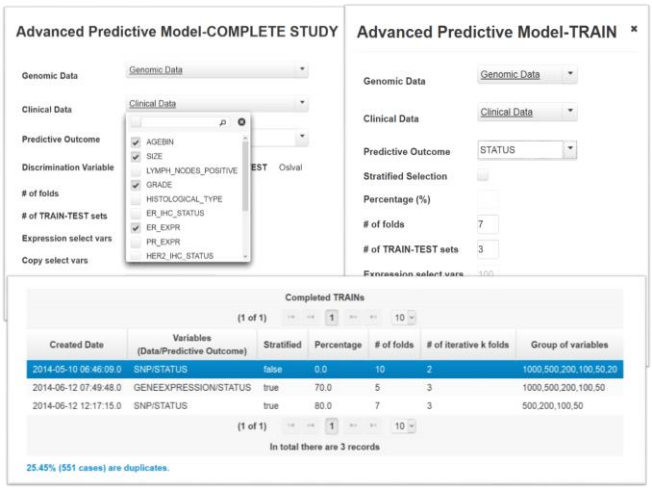


Figure 3. Setting up a predictive model

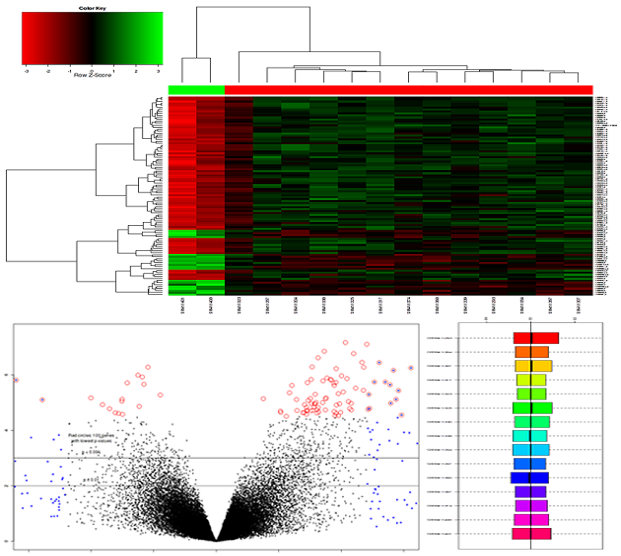


Figure 4. Indicative genomic analysis results showing a heatmap (on top), a volcano plot (bottom left), and a quality assessment plot (bottom right)

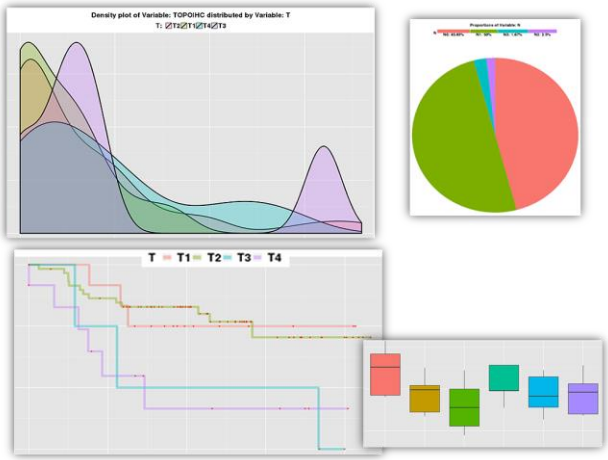


Figure 5. Indicative descriptive statistics results showing density plots (top left), a pie chart (top right), survival analysis plots (bottom left), and boxplots (bottom right)

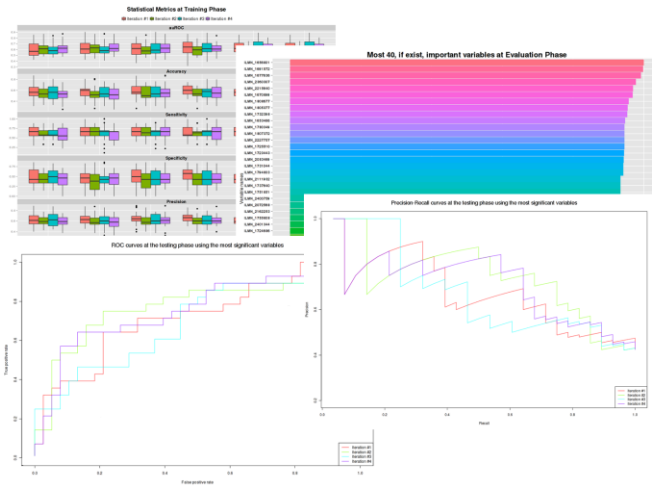


Figure 6. Indicative predictive analysis results using the MKL model showing boxplots (top left), significant degree of each variable (top right), ROC curves (bottom left), and precision-recall curves (bottom right)

IV. Conclusion

This paper outlined the implementation of a platform assisting users in deriving statistical and predictive analysis tools and models based on multi-level heterogeneous data provided by clinical trials. The main focus was on explaining 3MDAP by first presenting the architecture and specifications and explaining the core functionality. Particular focus is given on presenting the components developed for the analytical tools and the predictive models as well as the integration with independent tools for cohort selection.

The main goal of this platform is to empower scientists from diverse backgrounds to employ with ease (at the push of a button) sophisticated statistical analysis tools and to derive predictive models (again at the push of the button) from clinical trial data. It assists users in employing the statistical analysis tools implemented within the framework, addressing specific analysis scenarios and enables them to construct and validate their own predictive models. Although 3MDAP involves various scenarios for statistical and predictive analyses of homogenous or heterogeneous multi-modal data, it does not currently offer the possibility of editing the related R scripts. This function would facilitate the users to configure the output of the analyses (.pdf reports) according to their needs. We plan in future releases version of the tool to implement the possibility of uploading a new analysis R script plus an .xml file with all the meta-data information (analysis description, mandatory and optional variables, etc.).

Acknowledgment

The authors would like to thank all the collaborators of the project and specially Alexandre Irrthum from the Breast International Group (BIG).

References

[1] The R project for Statistical Computing (www.r-project.org).
[2] The INTEGRATE project funded by the European Commission under the 7th Framework Programme (http://www.fp7-integrate.eu/).
[3] Liferay (www.liferay.com).
[4] Primefaces (www.primefaces.org).
[5] Rserve, a binary R server (www.rforge.net/Rserve).
[6] Leisch, F. (2002), Sweave: Dynamic Generation of Statistical Reports Using Literate Data Analysis, in Wolfgang Härdle & Bernd Rönz, ed., 'Compstat 2002 - Proceedings in computational statistics', Physica Verlag, Heidelberg, pp. 575-580.
[7] Latex, a document preparation system (www.latex-project.org).
[8] Apache CXF: An Open-Source Services Framework (http://cxf.apache.org/).
[9] Shawe-Taylor, J. & Christianini, N. (2004), Kernel Methods for Pattern Analysis, Cambridge University Press.
[10] Lanckriet, G. R. G.; Cristianini, N.; Bartlett, P. L.; Ghaoui, L. E. & Jordan, M. I. (2002), Learning the Kernel Matrix with Semi-Definite Programming., in Claude Sammut & Achim G. Hoffmann, ed., 'ICML', Morgan Kaufmann, pp. 323-330.
[11] Chen, Z.; Li, J. & Wei, L. (2007), 'A multiple kernel support vector machine scheme for feature selection and rule extraction from gene expression data of cancer tissue.', Artificial Intelligence in Medicine 41 (2) , 161-175.

Intellectual Property Rights Issues in Multiscale Cancer Modeling*

Iryna V. Lishchuk, Marc S. Stauch, and Nikolaus P. Forgó

***Abstract*— *In silico* hyper-modeling is a complex process which requires interdisciplinary effort. Scientists from biology, medicine, bio-informatics, mathematics, engineering and other fields collaborate and contribute their knowledge and expertise. Researchers deserve recognition, intellectual input deserves protection and investments deserve reward. This paper investigates several IP regimes which may apply to cancer models and seeks to find solutions which would guarantee protection and reward.**

I. Introduction

In silico oncology requires an interdisciplinary approach to enable simulation of cancer progression in space and time using multi scale cancer modeling. "*A model is considered to be "multiscale" if it spans two or more different spatial scales and/or includes processes that occur at two or more temporal scales"* [1]. Multiscaling is achieved *in silico* by constructing hyper-models from component models which correspond to elementary biological models and relation models which reflect relations across them.

Modeling work is a complex process and encompasses several phases – it starts with the creation of elementary models and adoption of modeling techniques, goes through complex mathematical computations and simulations and leads to structuring models into specific choreographies. Modeling requires significant intellectual input and expertise. Researchers who invest their expertise and research justifiably expect such investment to be rewarded, and to receive protection against marauding interests. Protecting Intellectual Property (IP) by patents, copyright and/or legal regime of know-how aims to provide creators with recognition and possibly financial benefit. However, just as multiscale cancer modeling is a challenge for scientists, its novelty makes the choice of appropriate legal protection a challenge for lawyers.

In this paper we analyze different IP regimes potentially applicable to in silico modeling, investigate their pros and cons and seek to find appropriate solutions. Copyright, know-how protection and patent law all come into question.

* Research supported by the European Union Seventh Framework Programme FP7/2007-2013 under grant agreement No 600841 (CHIC Project).

I. V. Lishchuk is with the Institute for Legal Informatics, Leibniz University Hannover, Königsworther Platz 1, 30167 Hannover, Germany (corresponding author phone: +49 0511 762 8283; fax: +49 0511 762 8290; e-mail: lishchuk@iri.uni-hannover.de).

M. S. Stauch and N. P. Forgo are with the Institute for Legal Informatics, Leibniz University Hannover, Königsworther Platz 1, 30167 Hannover, Germany (e-mails: stauch@iri.uni-hannover.de; forgo@iri.uni-hannover.de).

II. Substance and Nature of Models

What legal regime may apply to cancer models depends first on the manner in which these are implemented and defined.

A. Scientific Models and Computer Models

There are two types of models that are of interest here: scientific models and computer models. Scientific models are defined as: "*finalized cognitive constructs of finite complexity that idealize an infinitely complex portion of reality through idealizations that contribute to the achievement of knowledge on that portion of reality that is objective, shareable, reliable and verifiable*" [2]. Scientific models are implemented in silico via computer models. In the context of cancer modeling a computer model is defined as: "*a computer program that implements a scientific model, so that when executed according to a given set of control instructions (control inputs) computes certain quantities (data outputs) on the basis of a set of initial quantities (data inputs), and a set of execution logs (control outputs)"* [3].

B. Modeling Process

In moving from the scientific model to its instantiation as a computer model, complex modeling work needs to be done. First, the biological process of tumor growth is analyzed. Then it is broken down into elementary processes, such as cell cycling, the angiogenesis process, probability of a cell to apoptosis after a particular treatment, etc. Further, modeling techniques - *discrete, continuum,* or *hybrid - which would represent the progression best are identified* [1]. Finally, computer codes corresponding to biological processes are developed. At this stage the scientific model is transformed into an executable form, i.e. encoded in a computer program. It is the program that then instructs the computer as to what steps to follow to simulate the biological process captured by the scientific model.

In the above scheme, the core element of a computer model, which provides an initial candidate for legal protection as intellectual property, is the model code, i.e. computer program. Here the law of copyright will come into play.

III. COPYRIGHT IN COMPUTER MODELS

A. Protection of Software under Copyright

Copyright is a traditional type of protection enjoyed by software under European and International Law. Article 4 WIPO Copyright Treaty [4], and Article 10 TRIPS Agreement [5] afford such protection to computer programs as literary works in the meaning of Berne Convention (1886). The same principle is followed by the European law, Article 1 Directive 2009/24/EC on the legal protection of computer programs which states: "*Member States shall protect computer programs, by copyright, as literary works*

*within the meaning of the Berne Convention for the Protection of Literary and Artistic Wo*rks" [6].

B. Prerequisites for Copyright Protection

Copyright will apply to the extent that a computer program constitutes an original expression of the author´s own intellectual creation. Copyright, if granted, would normally subsist in the source and object code, as established by Article 10 TRIPS Agreement. In broad terms, protection would cover any form of expression of a program which "*permits reproduction in different computer languages, such as the source code and the object code*" [7].

Nonetheless, not all aspects of the efforts that go into computer modeling will attract such protection. Most notably, the ideas and principles which lie beneath any element of a computer program are exempted from protection by copyright as Recital 11 Directive 2009/24/EC clarifies. In this regard a computer program which is too abstract and resembles rather a mere idea than an expression may lack the necessary dependence on creativity to be protected by copyright [8]. The same might be true for computer models.

Originality is another essential attribute of copyright work. A model code, as any computer program, is supposed to be written in a programming language, such as: C, C++, Python, etc. Programming languages and computer programs usually consist of symbols, figures, mathematical elements, concepts, etc. These items, considered in isolation, do not constitute the intellectual creation of the author where copyright could subsist. "*It is only through the choice, sequence and combination of those words, figures or mathematical concepts that the author may express his creativity in an original manner and achieve a result… which is an intellectual creation"* [9]. Hence, only models written by their authors in an original way may be eligible for copyright protection. Consequently, if a model code is generated automatically, it will lack sufficient intellectual input to attract copyright.

C. Drawbacks of Copyright

A limitation of copyright as a means of protecting software is that by focusing so much on expression it leaves open various possibilities of exploiting the works in circumvention of copyright protection. Thus, it may still be possible to achieve the same computer performance by rewriting the instructions without infringing software copyright as long as substantial copying of the original is not proven [10].

It may be noted that some common law countries tend to go beyond protection of programs from literal copying of the code and extend software copyright to the program non-literal expression, i.e. "look and feel", its structure, organization and managing of input and output routines [11]. Thus according to UK case law, it is not only literal expression which needs to be protected from copying, but rather the skill and labor which went into the design and coding of the program "… *United Kingdom cannot prevent the copying of a mere idea but can protect the copying of a detailed ´idea´. It is a question of degree where a good guide is the notion of over-borrowing of the skill, labor and judgment which went into the copyright work*" [11]. However, as long as this approach is not recognized by the ECJ or becomes a well established practice in all EU Member States, software copyright as a means of protecting original ideas embodied in the code may not be the best option. Here, as discussed later, patent law, may have advantages:

IV. COPYRIGHT IN MODELING WORK

A. Copyright in Preparatory Design Material

Yet an advantage of copyright is that copyright covers not only the program itself, but extends also to the preparatory design material. According to Article 1 Directive 2009/24/EC, "*the term 'computer programs' shall include their preparatory design material*". Preparatory design material is defined as: "*work leading to the development of a computer program provided that the nature of the preparatory work is such that a computer program can result from it at a later stage"* [12]. This extended scope of copyright in a program might also be relevant for the modeling work.

As was considered above, a computer model evolves in the course of modeling work, so a computer program is the end product of an extended software development process usually comprising several stages. *First, a problem to be solved by a computer is analyzed, then methods of solving the problem are adopted and stages of running the program are identified. Subsequently, detailed further instructions for a computer to perform operations necessary for the execution of the program are developed* [13].

*For its part, the modeling work also falls into a number of separate stages. Initially modeling comprises analysis of a tumor growth to be modeled, then models of elementary biological processes are identified, and modeling techniques and stages of executing models are defined. Finally, when all steps of implementing a model are clear, instructions for a computer to execute the model (model codes) are defined. With this comparison in mind, the developing documentation leading to the creation of a model code has a good chance to be considered as p*reparatory design material within the meaning of Article 1 Directive 2009/24/EC.

B. Prerequisites for Copyright Protection

A prerequisite for protection is that the preparatory work should have original creative substance and lead to the reproduction or the subsequent creation of a computer program [7]. There are no specific requirements as to how the preparatory design work must be expressed. Development documentation set down in writing, inter alia, data flow plans, designs of commands and information cycles, exhibits of scientific or technical art, expressed in any form, including mathematical, technical or graphic symbols, and which enable the production of a program, should suffice [14]. Provided the preparatory work is recorded, leads

to reproduction of a computer model and can be related to relevant aspects of software development, it will stand a good chance to be covered under the umbrella of copyright together with the model code. Hence, copyright seems to be an optimal option to protect modeling materials - if not under extended protection of a computer model, then as a separate copyright protected work on its own.

Modeling work constitutes a substantial piece of research and may well deserve individual protection as a copyright protected work in scientific and literary domain in its own right [15].

V. COPYRIGHT IN HYPER-MODELS

So far we have been considering copyright potentially applicable in all cases of modeling (including where the resulting model remains a simple model); however, further possibilities of protection arise where discrete models are integrated into composite hyper-models, or, more exactly, hyper-model structures. Here the integrative process itself may be protected by copyright, if the hyper-model is designed in an original creative way.

A multiscale cancer model is an example of such a hyper-model, constituting "*a concrete instance of an integrative model, built as the orchestration of multiple computer models that might run on different computers at different locations using different simulation stacks*" [16]. Hyper-models seek to simulate complex biological processes and need to follow the laws of biology. More than that, computer simulations are based on sophisticated mathematical principles which also need to be observed. Structuring hyper-models in this way requires substantial intellectual input and inter-disciplinary expertise - a piece of work which may also need to be rewarded and protected.A hyper-model designed by a researcher according to the principles of mathematics and biology in an original way may here qualify for copyright protection as a compilation. "*Compilations of data or other material, in any form, which by reason of the selection or arrangement of their contents constitute intellectual creations*" are protected as such, both by Article 5 WIPO Copyright Treaty and Article 10 TRIPS Agreement. Copyright in a hyper-model structure would not affect the earlier copyright in constituent models. So copyright in those models and copyright in the integrated hyper-model may subsist together. A further possibility (and one indeed intended by the modeling community) is that the same models remain open to diverse re-assembly into different hyper-models. Once a new hyper-model is compiled in an original creative way, then in principle, independent copyright would arise in it. Copyright protection might admittedly be hindered, if a model design is not creative, but dictated by pure technical considerations.

VI. KNOW-HOW IN MODELING

A second type of protection which may be considered to protect model sources and modeling work from being disclosed is conferred by the legal regime of know-how.

A. Know-How in Legal Terms

The main legal instrument which affords protection to know-how is the TRIPS Agreement. Article 39 provides that information "which is secret in the sense that it is not generally known or accessible to circles of persons that normally deal with the kind of information in question, has commercial value because it is secret, and has been subject to reasonable steps by the person in lawful control of such information to keep it secret" may be eligible for legal protection as know-how. Any information, including, but not limited to, "technical or non-technical data, patterns, compilations, programs, devices, methods, techniques, financial data" may qualify for protection as long as the conditions for protection are fulfilled [17]. These are that the information must be secret and have economic value.

Here, the taking of measures to restrict the availability of information to a limited number of persons and subject to a contractual duty of confidentiality are considered as sufficient pre-requisites for proving secrecy [18]. Publication of information or disclosure of model codes in an "open source" format would by contrast destroy the regime of secrecy so that protection would be forfeited. As regards the requirement of economic value of modeling know-how, this will be considered to be present if through publication, the research investment and competitive standing of the entity doing the work would be undermined [18].

B. Protectable Subject Matter in Modeling

In the context of hyper-modeling for in silico oncology, a clear candidate for protection as know-how would be source codes, which have value for commercialization (unless released "open source").

Nonetheless, a problem for an entity wishing to assert know-how protection in the model source code is that, although software or models were released in compiled form only, i.e. object code, one has to count with a potential risk that a third party may uncover the source code by reverse engineering [19]. Indeed, such reverse engineering is sometimes exempted from the need for the right holder's approval by Article 6 Directive 2009/24/EC, namely when "indispensable to obtain the information necessary to achieve the interoperability of an independently created computer program with other programs" and executed by a person authorized to use a program. Although this exemption may appear tightly drawn, the argument that a third party disregarded the relevant boundaries in a given case can be hard to prove in practice.

Accordingly, while the legal regime of know-how may be appropriate to keep information undisclosed, modeling work once published or models released in source code cannot count as protectable know-how.

VII. PATENTABLE AND NON-PATENTABLE SUBJECT MATTER

If it is the model substance (or an idea how to simulate a

cancer model in silico) which is to be protected, then, as discussed earlier, protection by copyright runs into limitations. Here patent law would usually represent a better alternative.

In contrast to copyright, "patent does not protect expression of an idea, but the underlying substance of it", e.g. a patent protecting a micro-chip would not cover the expression, but the idea that circuits can be organized to operate in a particular way [20]. The same approach would apply to modeling. A patent would not protect the model expression, but the mode of implementation, namely that the models representing elementary biological processes can be organized in a specific way so that a more complex biological process can be simulated in silico. Patent protection is available for products or processes in all fields of technology which satisfy patentability requirements [21]. These are that the relevant claimed invention must be novel, involve an inventive step and be susceptible of industrial application [22]. The threshold for obtaining patent protection is rather high. Thus, it requires going through a stringent (and costly) application and registration process.

There are two further potential obstacles which might hinder patent protection in the particular case of cancer models. These are first that computer programs as such, methods for performing mental acts and mathematical methods are non-patentable in the EU [23]; the same is true, secondly, for "diagnostic methods practised on the human or animal body" [24]. The ambit of these exceptions has yet to be fully tested in case law, and their potential applicability to *in silico* modeling remains unclear.

However, in one case patent protection for a method for processing medical data for patient examinations with the use of artificial intelligence was denied in Germany on the ground of non-patentability of rules and methods for performing mental acts and presentations of information [25].

The claimed subject matter contained instructions related to the choice of examination modalities (e.g. X-ray, computer tomography, magnet resonance) and purpose-related application on the patient by means of a program using a symptom-specific and/or diagnosis specific database. The patent tribunal found that instructions on selecting one or more examination and measurement protocols as well as selection of examination modalities by a physician constitute non-technical elements which only aim to automate decisions already contemplated by the physician and solve no technical problem at hand. On that basis the claimed method was declared non-patentable as such [26]. For its part, the appeal court, while suggesting that in principle a method that is using a computer program to solve a technical problem could be patentable, agreed that here a technical problem which the claimed method solved could not be identified. Accordingly, patent protection was denied.

With these observations and case law in mind, even if patent protection for cancer models is an attractive option, because of high and stringent patentability requirements and a range of exceptions, patent protection would be hard to attain and needs a deep analysis.

VIII. Conclusion

From the above observations, it is clear that the decision which IP regime can confer optimum legal protection for multiscale cancer models depends on multiple factors. These include model substance, mode of implementation, the exploitation interests (or otherwise) of the researchers, disclosure of sources and intention to publish research results, etc. Considering the complex nature of multiscale cancer models as a whole and the collaborative nature of work in hyper-modeling, seeking protection of models under a bundle of IP rights seems to be the most appropriate solution. Copyright which protects original expression and is a recognized means of protecting software may be regarded as optimal for protecting computer models and underlying modeling work. Research work preserved as secret for publication may be subject to protection as know-how (although this might trigger a conflict with fundamental scientific standards on knowledge-sharing). Patent protection, while offering the best approach to protect model substance, has practical disadvantages due to the uncertain scope of exclusions under the EPC, as well as stringent patentability and registration requirements.

References

[1] T. Deisboeck, Z.Wang, P. Macklin, V. Cristini, "Multiscale Cancer Modeling," *Annual Review of Biomedical Engineering*, vol. 13, pp. 127-155, Aug. 2011.

[2] M. Viceconti, "A tentative taxonomy for predictive models in relation to their falsifiability", *Philos. Transact A Math Phys Eng Sci* vol. 369, no. 1954, pp. 4149-61, Nov. 2011.

[3] CHIC, Deliverable No. 7.1, "Hypermodelling specifications", submitted 30.06.2014 presented at the 6th IARWISOCI – The CHIC Project Workshop, Athens, Greece, 3-4.11.2014.

[4] WIPO Copyright Treaty, adopted in Geneva on December 20, 1996.

[5] Agreement on Trade-Related Aspects of Intellectual Property Rights, the TRIPS Agreement, Annex 1C of the Marrakesh Agreement Establishing the World Trade Organization, signed in Marrakesh, Morocco on 15 April 1994.

[6] Directive 2009/24/EC of the European Parliament and of the Council of 23 April 2009 on the legal protection of computer programs, Official Journal of the European Union, L 111/16, 5.5.2009.

[7] ECJ, Judgment of 22.12.2010, Case C 393/09, Bezpečnostní softwarová asociace – Svaz softwarové ochrany v Ministerstvo kultury.

[8] R.T. Nimmer, Legal Issues in Open Source and Free Software Distribution, adapted from Chapter 11 in Raymond T. Nimmer, The Law of Computer Technology, 1997, 2005 Supp.

[9] ECJ, Judgment of 02.05.2012, Case C 406/10, SAS Institute Inc v World Programming Ltd.

[10] M.S. McBride, "Bioinformatics and intellectual property protection," *Berkley Technology Law Journal*, vol. 17, pp. 1331-1363, 2002.

[11] S. Stokes, Digital Copyright, Law and Practice, 2014, p. 119.

[12] Directive 2009/24/EC of the European Parliament and of the Council of 23 April 2009 on the legal protection of computer programs, Official Journal of the European Union, L 111/16, 5.5.2009, Recital 7.

[13] WIPO expert group on legal protection of computer software, First Session Geneva, November 27 to 30, 1979, LPCS/I/2, 30.09.1979.

[14] Federal Court of Justice of Germany, Judgment of 09.05.1985 - I ZR 52/83, BGHZ 94, 276 – 292.

[15] Berne Convention for the Protection of Literary and Artistic Works of September 9, 1886, as amended on September 28, 1979, Article 2.

[16] P.A. Fishwick, "Hypermodelling: an intergrated approach to dynamic system modeling," *Journal of Simulation*, vol. 6, pp. 2-8, 2012.

[17] Hogan Lovells International LLP, “Study on Trade Secrets and Parasitic Copying (Look-alikes)”, MARKT/2010/20/D, LIB02/CM3SET/2743659.17.
[18] K. Lodigkeit, Intellectual Property Rights in Computer Programs in the USA and Germany, Peter Lang GmbH, 2006, pp. 98-101.
[19] R. Gopalan, “Bioinformatics: Scope of Intellectual Property Protection, *Journal of Intellectual Property Rights*, vol. 14.01.2009, pp. 36-51, 2009.
[20] A. M. St. Laurent, Understanding open source & free software licensing, O´Reilly, 1. Edition, 2004, p. 2.
[21] Agreement on Trade-Related Aspects of Intellectual Property Rights, the TRIPS Agreement, Annex 1C of the Marrakesh Agreement Establishing the World Trade Organization, signed in Marrakesh, Morocco on 15 April 1994., Article 27.
[22] Convention on the Grant of European Patents (European Patent Convention, EPC) of 5 October 1973, as revised, Article 52 Paragraph 1
[23] Convention on the Grant of European Patents (European Patent Convention, EPC) of 5 October 1973, as revised, Article 52 Paragraph 2 (c).
[24] Convention on the Grant of European Patents (European Patent Convention, EPC) of 5 October 1973, as revised, Article 53 (c).
[25] Case Law from the Contracting States to the EPC 2004 – 2011, Special edition 3, Official Journal EPO 2011, Federal Court of Justice of Germany, Judgment of 20 January 2009 - X ZB 22/07 - Equipment for selecting medical examination methods.
[26] Federal Court of Justice of Germany, Judgment of 20 January 2009 - X ZB 22/0, Rn.5.

Legal and Ethical Aspects of *In Silico* Medicine*

Iheanyi S. Nwankwo, Marc S. Stauch, Alan Dahi, and Nikolaus P. Forgó

***Abstract*— The following paper considers some of the novel ethical and legal issues that may arise in the context of *in silico*-based medicine, with particular reference to the development of hypermodels to optimize treatment decisions for specific diseases.**

I. Introduction

In silico medicine is a term denoting medical experiments performed on computer or via computer simulations [1]. A key aim of this area of medical research is to develop models representing different aspects of the human biological system, thus allowing biological processes to be simulated and studied virtually rather than in a laboratory (*in vitro*) or in the live subject (*in vivo*). Some positive outcomes have already been recorded. Thus, in one case researchers succeeded in identifying potential inhibitors to an enzyme associated with cancer activity *in silico*, of which fifty percent of the molecules were later shown to be active inhibitors *in vitro* [2]. Other papers such as [3] and [4] have also shown remarkable findings using *in silico* methods.

Although the idea of performing computer simulations is not entirely new in the sciences, its application in the field of medicine appears to be a novel starting point towards achieving the goals of personalized medicine, i.e. being far more tailored to the individual than is possible today. The knowledge captured in models at different scales of the biological process (molecular, cellular, tissue, organ, etc.) can be integrated into composite models (hypermodels) of increasing complexity, capable of simulating processes in relation to a given disease domain. These models could then be fed with data relating to an individual patient in order to answer specific disease treatment questions in relation to that patient [5].

The ultimate aim would be to fuse hypomodels covering discrete domains into an overall linked-up hypermodel covering the human patient as a whole. This could be used to create a digital patient, i.e. a virtual version (an avatar) of each living person, as well as to run simulations of health and disease processes on this virtual individual [6]. The results could then be used to make not only decisions in response to disease, but long range predictions about an individual's future health, allowing preventative strategies and interventions to be employed before disease appears.

*Research supported by European Union Seventh Framework Programme FP7/2007-2013 under grant agreement No 600841(CHIC Project).

I. Nwankwo, M. Stauch, A. Dahi, and N. Forgó are with the Institute for Legal Informatics, Leibniz University Hannover, Königsworther Platz 1, 30167 Hannover, Germany, (corresponding author phone: +495117628242; fax: +495117628290; e-mail: nwankwo@iri.uni-hannover.de); stauch@iri.uni-hannover.de; dahi@iri.uni-hannover.de; forgo@iri.uni-hannover.de

At present a number of research projects, including CHIC, are ongoing that focus upon the creation of hypermodels to assist treatment decisions in response to specific diseases [7]. However, an aspect of *in silico*-based medicine that has not generated much discussion in the research community is its legal and ethical landscape. Existing legal and ethical frameworks are largely in the context of traditional medical research directly using human subjects. Rules were made primarily to protect the human research subjects from harm. By contrast, *in silico* medical research relies on computational resources and data – using patient data to generate and validate models.

In this regard the big question is: should such research be subject to the current legal and ethical rules that apply in medical research or should new ones be developed to cater for the needs of the *in silico* community and facilitate their work? For example, should trial protocols and ethical committee approval be obtained for *in silico* trials? What validation procedure should be adopted for *in silico* outcomes? Similarly, as the digital patient represents a form of medical data [8], this implies that the rules governing the processing of sensitive data will apply in a number of scenarios involving the development of the models as well as their interaction with the EHR. This brings to the fore the privacy and data protection issues that may arise in this aspect of medicine. Furthermore, there may be other issues arising in the application of *in silico*-based medicine in treatment scenarios, such as the issue of liability when the models give inaccurate predictions, or how the whole system will affect the doctor-patient relationship [9].

This paper seeks to examine some of these issues and is divided into four parts. Section II considers the benefits that progress in *in silico* medicine could bring to society and patients at large. In terms of consequentialist ethical reasoning, this provides a prima facie justification for the promotion and pursuit of such research. However, it is also critically important that the rights and interests of individual patients receive due weight and respect. Section III looks at patient privacy interests implicated at the stage of building such models, section IV at those in relation to their validation, and section V those regarding their use in actual clinical practice.

II. Advantages Of *In Silico* Research

The global healthcare system is struggling with rising costs and uneven quality despite various policies aimed at improving the system. A particular challenge is the management of chronic diseases with an unpredictable nature, such as cancer or Parkinson's, that affect each individual differently and progress in very diverse ways. These require treatment tailored to the individual, and a need for models that are able to accurately predict each patient's condition and disease progression using his or her specific health data.

Advances in ICTs, including the use of powerful cloud computing, have enabled a lot of transformations in translational medicine, ranging from genomic sequencing to the availability of large bioinformatics databases. With these advancements, it is now possible to integrate clinical and molecular sciences with advanced engineering sciences so that physiological and pathological information from the living human body can be quantitatively described via biocomputing across multiple scales of time and size, and through diverse hierarchies of organization – from molecules to cells and organs to individuals [10].

Apart from the above, the construction or design of an *in silico* clinical trial could provide profound insight into the design of real life clinical trials, ranging from optimal patient selection to individualized dosage and duration of proposed therapeutic interventions. There may be other benefits of such prior checking *in silico*: the targeting of drugs based on individual patient profiling, reduced animal testing, identifying problematic side effects, creating tailored treatments, understanding costs and benefits at an individual level, etc. [11]. Trials can be run harmlessly on individual digital human models, as well as on entire virtual patient populations numbering hundreds or thousands, and which may reduce the number of subsequent *in vivo* tests. Where this trend is sustained, *in silico* clinical studies will reduce cost and error to a great extent, and aid in achieving the goals of personalized medicine.

The potential reduction of the risk of causing physical harm to real patients and volunteers compared to *in vivo* research is a good ethical reason to switch to *in silico* research, so far as the reliability of the results is acceptable. Elsewhere, *in silico* medicine has the potential to generate further positive benefits by unlocking new knowledge from patient data that can be used to optimize individual care and treatment. As noted above, a medium term application of this information gain lies in the development of disease-specific hypermodels.

III. Building Hypermodels: Privacy And Data Protection Issues

As regards implications for patient rights and interests that arise during the building and integration of hypermodels, the key issues relate to patient privacy and data protection. These are distinct but related concepts. Privacy, which the law and academia both struggle to define in a universally accepted manner, is generally recognized as a fundamental right in Europe [12].

Despite the problems of defining privacy, it can broadly be categorized into three spheres: physical privacy, informational privacy, and decisional privacy [13]. Physical privacy encompasses the freedom to have a space free from intrusion. Informational privacy allows one to control the communication of information about oneself to others. Decisional privacy is the freedom to make decisions without undue interference. Data protection, on the other hand, can be viewed as a tool to balance an individual's privacy rights with the interests of third parties in that individual's data [14]. In this regard, it is closely aligned with informational privacy. Indeed, for the purposes of *in silico* medicine in the context of this paper, data protection and informational privacy can be treated as one and the same.

The main piece of secondary legislation in the EU on data protection is Directive 95/46/EC [15]. Its aim is to protect the right to privacy individuals enjoy by regulating, inter alia, the "processing of personal data wholly or partly by automatic means" [16]. "Personal data" is defined as [17]

> Any information relating to an identified or identifiable natural person (data subject); an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity.

Data concerning health or sex life belong to the special categories of data ("sensitive personal data") that is prohibited from processing unless specific exemptions apply [18]. Regarding *in silico* medicine in general, the relevant exemptions are: explicit consent of the data subject [19]; where the processing is required for certain medical and health-care purposes, subject to some additional requirements [20]; and where a member state, subject to safeguards, lays down exemptions for reasons of substantial public interest [21].

Building models requires vast amounts of data. In general the option of first choice (in best allowing patient control over information) would be for explicit consent to be obtained for the use of the data for the specific purpose of building a model [22]. However, this may raise practical or scientific problems. Frequently, the data at issue will be retrospective, perhaps collected years before, and contacting patients to agree to the new use will be an immensely difficult or even impossible (or unethical) task. Where research is cutting edge or new questions arise, it might even be difficult to obtain valid explicit consent, which requires a voluntary and considered decision and a detailed understanding of how the data will be used [23].

Independently of the above, a further requirement of data protection law is for data, so far as possible (consistent with the purposes of processing it), to be de-identified prior to use [24]. This serves to protect the interest of patients in not suffering harm, e.g. discrimination by an insurance company, by preventing sensitive medical data from being linked to her. Indeed, because of the way the law is framed, if data is truly de-identified so there is no way to re-link it to an individual, it loses its status of personal data and will be released from the above-mentioned legal restrictions on use because it cannot do any harm to the individual it stems from.

While theoretically a clear and straightforward process, the rise of ever better algorithms and increased processing power means that truly de-identifying (in effect anonymizing) data is in some circumstances more or less impossible, particularly when the data is longitudinal with large amounts of variables. This is because combining such data elements with data from other sources (data-matching)

more often than not permits re-identification of the data subject.

A more practical approach is to make use of "pseudonymous data", a new category of personal data that is legislatively recognized in the Draft General Data Protection Regulation [25], but not in the Data Protection Directive. It refers to personal data that cannot be attributed to an individual without additional information. An example is the use of a key-code pseudonym that replaces direct identifiers. This also has the benefit that, exceptionally, it might be ethically appropriate to link back, e.g. where data-mining uncovers information of vital importance for the wellbeing of a particular patient.

Pseudonymous data can be protected by applying legal, organizational and technical safeguards that prevent the data subject from being re-identified, thereby leading to what may also be termed "secure de-identification". The legal safeguards encompass prohibitions on using the data for other than the strict research project purpose, on disclosing the data outside the researcher group or seeking to re-link the data to patient subjects either by the key-code (except in exceptional circumstances, as mentioned) or otherwise.

The organizational and technical safeguards should include strict data access policies and controls on individual data users, state of the art secure servers, and encryption of data during transit so as to make access by unauthorized persons virtually impossible. A specific issue here (and also later when deploying hypermodels) is that, where the volume of data processing requires the use of a cloud-based infrastructure, safeguards are in place to ensure authorized users retain exclusive control of the data; in normal circumstances this means a private rather than public cloud solution should be adopted [26].

IV. Validating Hypermodels: Safety And Quality Assurance Issues

Once a hypermodel has been built with the potential to guide a specific clinical decision (should the physician give treatment A or B to a given cancer patient with attributes p, q, r, s…?), the question of validating its accuracy arises. Some initial testing will certainly be possible virtually, by running the model on retrospective data available from other patient populations (not used to build the model). However, there will later be the need for legal compliance validation – in terms of ensuring that the model is fit for purpose for treating real prospective patients. Here, it seems likely that a hypermodel would qualify as a medical device within the definition of the EU Medical Devices Directive 93/42/EC. The Commission in its 2012 guidance on the application of the Directive clarified that it also extends to decision support software, widely defined as [27]:

> computer based tools which combine medical knowledge databases and algorithms with patient specific data. They are intended to provide healthcare professionals and/or users with recommendations for diagnosis, prognosis, monitoring and treatment of individual patients.

As a medical device the hypermodel would be required to undergo a certification process, involving an application for approval from notified bodies at member state level [28]. As foreseen in the Commission's draft Medical Devices Regulation, which is due to replace the Directive in 2015 [29], there will be the need for sponsored 'clinical investigations' of devices (similar to the clinical trials required in the field of medicinal products) [30].

The aim of the validation process, as in the case of medicinal products, is to assure the public that medical devices are safe and effective in practice. While sensible as a general safeguard prior to sanctioning new medical devices, it is interesting to explore a specific aspect of hypermodels that might argue for a lighter approach to approving their adoption. Indeed, given the early focus of hypermodels on assisting decisions in critical care, such as the treatment of life-threatening cancer conditions, there is a clear case for expediting their use where possible.

Here the most important feature to note is that the *in silico* simulations performed by hypermodels, augmented by image visualization functionality, aim to allow the clinician to view the patient's predicted progress in a chronologically contiguous manner, akin to repeat observations of that patient in real-time. It follows that there would be little or no delay in verifying a prediction: the clinician, by observing the patient, can see quickly if the actual course of events following treatment recommended by the hypermodel conforms to what the model indicated would happen. In the case of a divergence, the clinician could switch away from using the (inaccurate) model and adopt a different treatment.

Two caveats should be entered here: the first is that, to avoid the risk of observer bias, hospitals should ensure that observations of the real patient are not carried out by clinicians aware of the hypermodel's prediction. Secondly, a problem could arise in cases where adopting the treatment indicated by the hypermodel would involve a radical irreversible intervention, e.g. immediate surgery on a patient. In any case and in particular in such a situation, the clinician should arguably adopt the radical course only if this tallies with his own clinical judgment (perhaps informed over time by knowledge of the hypermodel's success in predicting the progress of patients for whom it favored a non-radical alternative).

Subject to the above, and also to the presence of informed consent from the patient or his legal representative (who should be told the clinician is – at least in part – basing the treatment on an automated algorithmic mechanism), hypermodel use by clinicians could plausibly be classed as a case of innovative therapy for the benefit of the individual patient. If that is so, there would also be an argument for exempting it (at least in parts) from the generic approval regime under EU medical devices legislation as a custom made device.

V. Deploying Hypermodels: Communication And Liability Issues

It is apparent that *in silico*-based medicine has the significant potential to improve healthcare delivery; however, it also poses some legal and ethical challenges when applied in treatment scenarios. This will require clarifying important ethical boundaries vis-à-vis to what extent reliance on *in silico* predictions may estrange or otherwise affect the physician-patient relationship. How much should one rely on leaving potentially vital decisions to an automated system that may not have the ability to appreciate the unique character and personality of every individual that doctors gain from physical interaction, training and years of experience?

Here, as noted, it will be important to secure the patient's informed consent to be treated with the aid of the hypermodel. As part of this, the patient should, to the extent that the clinician bases the decision which treatment to provide on the model's prediction, be made aware of this. However, it remains unclear how specific the consent would need to be (e.g. in explaining the logic underlying the decision) to be legally valid. The law would presumably have to take account of the practical difficulties clinicians and patients may have in giving and understanding a detailed explanation.

Challenges may arise too for the doctor's ethical duty of candour towards patients. Assuming for example that a model tells the doctor that any course of therapy will be hopeless for a certain patient, how should the doctor act on this information? [31] Such a scenario may also raise difficult distributive justice questions if models were to later include functionality for computing cost-effectiveness of different treatment options, or indeed determine, between patients, who would be the most efficient recipient of some resource-intensive therapy.

A further issue is that of legal liability in the event of adverse outcomes resulting from inaccurate or incorrectly interpreted models or data. Admittedly, the model can only give probabilistic information, but clearly if it gives a wildly wrong prediction the doctor may end up taking a decision he would not have taken otherwise. In this case, who should be liable – the doctor, the model-developers? Negligence and perhaps also statutory product liability need to be considered in this instance. For example, the STEP Consortium [32] outlines factors that may lead to an unforeseen adverse outcome in the VPH model such as: patient variability, databases populated with incorrect data, inappropriate use of data, the use of a flawed model, a misunderstanding of the assumptions associated with a model, etc.

VI. Conclusion

At the moment, it is not clear how safeguards to forestall these issues should be implemented, or how far this should influence the design of the models. At present, there is also no EU-harmonized legislation on medical liability, leaving a fragmentation of national laws and practices [33]. It is likely that, as in other areas of rapid technological advance, lawyers and ethicists will be busy for some time in assessing the legal and ethical implications of *in silico* medicine and devising appropriate solutions to emerging issues.

Acknowledgment

The views expressed are those of the authors and not necessarily those of the European Commission.

References

[1] See, http://insigneo.org.
[2] Wikipedia, In silico, http://en.wikipedia.org/wiki/In_silico.
[3] G. Stamatakos, E. Georgiadi, N. Graf, E. Kolokotroni and 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* vol 6, no. 3, pp. e17594, Mar. 2011.
[4] G. Clermont, J. Bartels, R. Kumar, G. Constantine, Y. Vodovotz and C. Chow, "In silico design of clinical trials: A method coming of age," *Crit. Care Med.* vol. 32, no. 10, Oct. 2004.
[5] J. Tester, "Personal health forecast: previewing our future self for decision making today," http://futureofcities.wikispaces.com/file/view/Tester_Personal+Health+Forecasts.pdf. See also, Personal health system foresight, http://www.phsforesight.eu/archives/2414.
[6] See the DISCIPULUS project roadmap for the digital patient, http://www.digital-patient.net/files/DP-Roadmap_FINAL_N.pdf.
[7] For example, Computational Horizons in Cancer (CHIC). See also, DISCIPULUS, AVICENNA.
[8] U. Charles, "The digital patient and the law", 16 October 2012, http://digitalpatientroadmap.blogspot.de/2012/10/the-digital-patient-and-law.html.
[9] P. Hynes, "Doctors, Devices and Defects: Product Liability for Defective Medical Expert Systems in Australia," *Journal of Law, Information and Science* vol. 15, 2004.
[10] D. Dimitrov, "Systems patientomics: the virtual in-silico patient," *New Horizons in Translational Medicine* vol. 2, pp.1 –4, Feb. 2014.
[11] http://avicenna-isct.org/projectinformation.html.
[12] Examples of privacy being recognized as fundamental right are: article 12 UDHR; article 17 International Covenant on Civil and Political Rights; article 8 European Convention on Human Rights.
[13] E. Chemerinsky, "Rediscovering Brandeis's right to privacy," *Brandeis Law Journal*, vol. 45, pp. 644-657, July 2006.
[14] D. Rowland, U. Kohl, A. Charlesworth, *Information Technology Law*. 4th ed. Abingdon, Oxon: Routledge, 2012, pp.150-152.
[15] See http://eur-lex.europa.eu/legal-content/DE/TXT/?uri=CELEX:31995L0046 .
[16] Article 3(1) Directive 95/46/EC.
[17] Article 2(a) Directive 95/46/EC.
[18] Article 8 Directive 95/46/EC.
[19] Article 8(2)(a) Directive 95/46/EC.
[20] Article 8(3) Directive 95/46/EC.
[21] Article 8(4) Directive 95/46/EC.
[22] D. Beyleveld, D.Townend, S. Rouille-Mirza and J. Wright Eds. *The Data Protection Directive and medical research across Europe*. Aldershot: Ashgate Publishing 2004, pp. 11–12.
[23] See the 2011 Opinion of the Art 29 Working Party (set up under Directive 95/46/EC) on the definition of consent, WP 187.
[24] Article 6(1)(c) and (e) Directive 95/46/EC.
[25] Draft General Data Protection Regulation, Parliament's First Reading (March 2014), Art 4(2).
[26] See the 20012 Opinion of the Art 29 WP on Cloud Computing, WP 197.
[27] See DG Health and Consumer, MEDDEV 2.1/6 (Jan 2012), p 20.
[28] Directive 93/42/EC, Annex III.
[29] Draft Medical Devices Regulation, COM(2012)542, Chapter VI.
[30] I.e. under the Clinical Trials Directive 2001/20/EC.

[31] I. Cohen, R. Amarasingham, A. Shah, B. Xie, and B. Lo , "The legal and ethical concerns that arise from using complex predictive analytics in health care," *Health affairs* 33, NO. 7: 1139–1147, July 2014.

[32] STEP Consortium, Seeding the EuroPhysiome: A Roadmap to the Virtual Physiological Human, p.80.

[33] Medical liability in Europe, http://www.bleedle.net/medical-liability-in-europe/.

A Brownian Motion Based Mathematical Analysis as a Potential Basis for Modeling the Extent of Infiltration of Glioma Cells into the Surrounding Normal Brain Tissue*

Markos Antonopoulos and Georgios Stamatakos

***Abstract*— Fast and extensive glioma tumour infiltration into the surrounding normal brain tissues is one of the most critical causes of glioma treatment failure. To quantitatively understand and mathematically simulate this phenomenon several diffusion based mathematical models have appeared in literature. The majority of them ignore the anisotropic character of diffusion of glioma cells since truly exploitable tomographic imaging data to this end is rare. Aiming at enriching the anisotropy enhanced glioma model weaponry so as to increase the potential of exploiting available tomographic imaging data, we propose a Brownian motion based mathematical analysis that could serve as the basis for a simulation model estimating the infiltration of glioblastoma cells into the surrounding brain tissue. The analysis is based on clinical observations and exploits diffusion tensor imaging data. Numerical simulations and suggestions for further elaboration are provided.**

I. Introduction

Glioblastoma multiforme (GBM) is the most malignant of all brain tumors. Apart from heterogeneity and highly invasive behavior, GBM cells tend to infiltrate the surrounding tissue by solely leaving the main tumor mass and traveling long distances inside the brain [1]. This diffusive behavior of GBM cells is one of the main causes of tumor relapse after resection. Since infiltrating cells are generally not visible by an MRI, and an extensive resection may damage surrounding tissue, small populations of such cells are almost always left at the resection margin of GBMs, and are widely believed to drive tumor relapse. Therefore, in order to provide quantitative insight into the non imageable phenomenon of tumor cell invasion, a number of mathematical models have been developed [2-8]. Most of the published mathematical models ignore the anisotropic character of diffusion of glioma cells since truly exploitable tomographic imaging data to this end is rare. In order to enrich the anisotropy enhanced glioma model weaponry so as to increase the potential of exploiting available tomographic imaging data, we propose a Brownian motion based mathematical analysis that could serve as the basis for a model estimating the infiltration of glioblastoma cells into the surrounding normal brain tissue. The analysis is based on clinical observations and exploits diffusion tensor imaging data. Numerical simulations and suggestions for further elaboration are provided. A realistic model based on the analysis presented could be useful for diagnosis as well as resection and radiotherapy planning.

* This work has been supported by the European Commission under the project Computational Horizons In Cancer (CHIC): Developing Meta- and Hyper-Multiscale Models and Repositories for *In Silico* Oncology (FP7-ICT-2011-9, Grant agreement no: 600841)

M. Antonopoulos is with the *In Silico* Oncology & *In Silico* Medicine Group, Institute of Communication and Computer Systems, National Technical University of Athens, Greece (e-mail: markos@central.ntua.gr)

G.S.Stamatakos is with the *In Silico* Oncology & *In Silico* Medicine Group, Institute of Communication and Computer Systems, National Technical University of Athens, Greece (corresponding author, project scientific coordinator: phone:+302107722287, fax:+302107723557, e-mail: gestam@central.ntua.gr).

II. Mathematical Modeling Of Diffusion

The reaction diffusion equation has been widely used for modeling the diffusion of tumor cells. In [5, 2] the authors assume that tumor cells move from regions of higher to lower densities and exploit the reaction-diffusion equation using an additional term corresponding to the net proliferation of tumor cells:

$$\frac{\partial c}{\partial t} = \nabla \cdot (D \nabla c) + \rho c \qquad (1)$$

where $c(x,t)$ is the concentration of tumor cells at time t and location x, and ρ reflects the net proliferation of tumor cells. D is the diffusion coefficient, assuming different scalar values on regions of white (D_w) and grey (D_g) matter, where $D_w > D_g$. For a review see [9]. Numerical treatments of (1) can be found in [7]. In [10, 11] the authors have expanded this model introducing in (1) the diffusion coefficient D in tensorial form, thereby including anisotropy in their models. Measurements of D in tensorial form where acquired through Diffusion Tensor Imaging techniques, which are described in the following section..

III. Diffusion Tensor Imaging

Diffusion Tensor Imaging (DTI) is a magnetic resonance imaging (MRI) technique measuring the diffusion properties of water molecules along specific directions. This is done by defining an ellipsoid in every voxel of a 3-dimensional space which mathematically corresponds to a 3x3 positive definite symmetric matrix:

$$D = \begin{bmatrix} D_{xx} & D_{xy} & D_{xz} \\ D_{xy} & D_{yy} & D_{yz} \\ D_{xz} & D_{yz} & D_{zz} \end{bmatrix}$$

This matrix can be decomposed in the following form:

$$D = [u_1 \quad u_2 \quad u_3] \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix} [u_1 \quad u_2 \quad u_3]^T$$

where $\lambda_1, \lambda_1, \lambda_3$ are the eigenvalues of D (positive, since D is positive definite) and $u_{1,}u_2, u_3$ are the corresponding orthonormal eigenvectors. The eigenvalues and eigenvectors of D define an ellipsoid with principal axes lying on the directions of $u_{1,}u_2, u_3$ each one having length $2\sqrt{\lambda_1}, 2\sqrt{\lambda_1}, 2\sqrt{\lambda_3}$ (Figure 1).

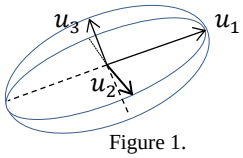


Figure 1.

This ellipsoid, called the diffusion ellipsoid describes the anisotropic diffusion of dyed water molecules in the specific voxel it refers to. If at the beginning of the observation period, a droplet of water molecules is placed at the center of the ellipsoid, after some time, the front of the diffusing water molecules will form an ellipsoid like the one in figure 1.This reflects the fact that at a certain location, water molecules do not move towards all directions at equal rates. Diffusion is faster in the larger axis (i.e. along the eigenvector corresponding to the larger eigenvalue).For each eigenvector of D, the larger the corresponding eigenvalue, the larger the diffusion along it's direction. DTI measurements provide this ellipsoid (actually, the principal axes directions and lengths) for each voxel (Figure 2).

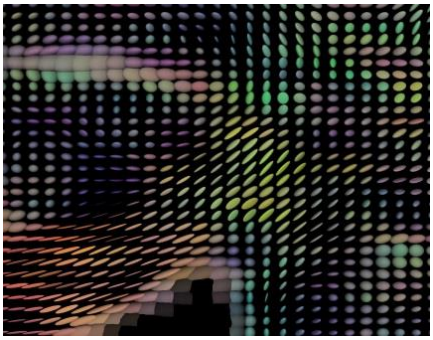


Figure 2. Picture by "DTI-axial-ellipsoids" by Thomas Schultz - Own work. Licensed under Creative Commons Attribution-Share Alike 3.0 via Wikimedia Commons - http://commons.wikimedia.org/wiki/File:DTI-axial-ellipsoids.jpg#mediaviewer/File:DTI-axial-ellipsoids.jpg

Diffusion tensor MRI has been used extensively for tractography in vivo. It is widely assumed that water molecules tend to move more easily along white matter tracts, therefore, the fiber bundle of local white matter tracts is assumed to be aligned with the largest axis of the diffusion ellipsoid [12].Furthermore, measuring the surface of the front of the diffusing water molecules (i.e. the surface area of the diffusion ellipsoid) provides a quantity known as apparent diffusion coefficient (ADC), which has been reported to inversely correlate with local cell density [13, 14].Taking into account that brain tumor cells are generally assumed to diffuse towards regions of lower cellular density [5, 2] and invade the surrounding tissue by moving along white matter tracts [1], the diffusion tensor acquired by DTI has been used to describe the diffusion of tumor cells as well [10, 11].

IV. Derivation Of the Model

Our model will use the diffusion tensor measurements provided by DTI to describe the stochastic movement of tumor cells within the brain tissue. Let us assume that we are given a DTI atlas, consisting of the diffusion tensor in each voxel like in [10]. Since the voxel size is of the order 0.5-1 mm3, we will assume that the diffusion tensor is piecewise constant: the diffusion tensor is constant within a voxel and is defined by the measured diffusion ellipsoid of the particular voxel. Thus, supposing that the observation period of the tensor measurement (see before) is Δt the movement of a particular water molecule within a voxel can be described by the equation

$$p(x, x_o, \Delta t) = \quad (2)$$

$$= \frac{1}{(2\pi)^{3/2}\det(ULU^T)^{1/2}} exp\left(-\frac{1}{2}(x - x_o)^T UL^{-1}U^T(x - x_o)\right)$$

Where $p(x, x_o, \Delta t)$ is the probability of a molecule starting at x_o to be at x after time Δt. U is the matrix having columns the orthonormal vectors $u_{1,}u_2, u_3$ of the diffusion tensor of the voxel, and L is a diagonal matrix with the respective eigenvalues $\lambda_1, \lambda_1, \lambda_3$ as elements of the main diagonal. Thus, given x_o, the probability density function of x is a multivariate Gaussian with mean x_o and covariance matrix ULU^T. The movement of tumor cells in after the same interval Δt can be described by the equation

$$p(x, x_o, \Delta t) = \quad (3)$$

$$= \frac{1}{(2\pi)^{3/2}\det(UL_aU^T)^{1/2}} exp\left(-\frac{1}{2}(x - x_o)^T UL_a^{-1}U^T(x - x_o)\right)$$

Where

$$L_{a,t} = \begin{bmatrix} \lambda_1\alpha & 0 & 0 \\ 0 & \lambda_2\alpha & 0 \\ 0 & 0 & \lambda_3\alpha \end{bmatrix} \quad (4) \quad L_{a,t}^{-1} = \begin{bmatrix} \frac{1}{\lambda_1\alpha} & 0 & 0 \\ 0 & \frac{1}{\lambda_2\alpha} & 0 \\ 0 & 0 & \frac{1}{\lambda_3\alpha} \end{bmatrix} \quad (5)$$

Using (4),(5) equation (3) can be written equivalently

$$p(x, x_o, \Delta t) = \quad (6)$$

$$\frac{1}{(2\pi)^{3/2}\det(UL_aU^T)^{1/2}} exp\left(-\frac{1}{2}(x - x_o)^T U \begin{bmatrix} \frac{1}{\lambda_1\alpha} & 0 & 0 \\ 0 & \frac{1}{\lambda_2\alpha} & 0 \\ 0 & 0 & \frac{1}{\lambda_3\alpha} \end{bmatrix} U^T(x - x_o)\right)$$

Parameter a rescales the eigenvalues of the tensor thereby rescaling conformally the axes of the diffusion ellipsoid. This reflects the fact that tumor cells may tend to move along the

axes of the ellipsoid, but do so with a different velocity than water molecules. Equations (3) and (6) again are Gaussian probability densities for the random variable x, i.e. the position of a tumor cell after time Δt given that the initial position of the cell is x_o.Using a standard linear transformation, the random variable x can be equivalently written in the form

$$x = U\begin{bmatrix}\sqrt{\alpha\lambda_1} & 0 & 0\\ 0 & \sqrt{\alpha\lambda_2} & 0\\ 0 & 0 & \sqrt{\alpha\lambda_3}\end{bmatrix}z$$

Where z is a normally distributed random vector with mean x_o and covariance matrix the identity matrix in R^3. Equivalently, we can write for the random variable $x - x_o$

$$x - x_o = U\begin{bmatrix}\sqrt{\alpha\lambda_1} & 0 & 0\\ 0 & \sqrt{\alpha\lambda_2} & 0\\ 0 & 0 & \sqrt{\alpha\lambda_3}\end{bmatrix}z'$$

Where z' is a normally distributed random vector with mean (0,0,0) and covariance matrix the identity matrix in R^3.This leads us to model the movement in continuous time by the equation

$$x_{t+\Delta\tau} - x_t = U\begin{bmatrix}\sqrt{\alpha\lambda_1} & 0 & 0\\ 0 & \sqrt{\alpha\lambda_2} & 0\\ 0 & 0 & \sqrt{\alpha\lambda_3}\end{bmatrix}b$$

Where

$$b \sim \frac{1}{(2\pi)^{3/2}(\Delta\tau)^{3/2}} exp\left(-\frac{b^T b}{2\Delta\tau}\right) = \frac{1}{(2\pi)^{3/2}(\Delta\tau)^{3/2}} exp\left(-\frac{\|b\|^2}{2\Delta\tau}\right)$$

i.e. the distribution of b is normal with zero mean and covariance matrix the identity matrix times $\Delta\tau$. This is equivalent to the stochastic differential equation

$$dx_t = \sqrt{a}\cdot U(x)L^{1/2}(x)dB_t \qquad (7)$$

Where B_t denotes standard Brownian motion in R^3. The matrix $U(x)$ depends on x and has columns the orthonormal eigenvectors $u_1(x), u_2(x), u_3(x)$ of the diffusion tensor at x.The matrix $L^{1/2}(x)$ is diagonal, and it's main diagonal entries are the square roots of the eigenvalues $\lambda_1(x), \lambda_2(x), \lambda_3(x)$ of the diffusion tensor at x. As mentioned before, the matrices $U(x)$, $L(x)$ are assumed piecewise constant. The parameter a is to be estimated from data. Equation (7) is the model we propose for describing the movement of tumor cells in the brain.

By assuming a twice differentiable diffusion tensor and denoting the probability density function of x_t by $p(x,t)$ the corresponding Fokker-Planck equation describing the evolution of $p(x,t)$ through time is

$$\frac{\partial p(x,t)}{\partial t} = \frac{1}{2}\sum_{i,j=1}^{3}\left[\frac{\partial^2}{\partial x_i \partial x_j}\left(\beta_{i,j}(x)p(x,t)\right)\right] \qquad (8)$$

where $\beta_{i,j}(x)$ are elements of the matrix

$$\beta(x) = U(x)M^{1/2}(x)\left(U(x)M^{1/2}(x)\right)^T$$

where M is the diagonal matrix with main diagonal entries $\sqrt{a\lambda_1(x)}$, $\sqrt{a\lambda_2(x)}$, $\sqrt{a\lambda_3(x)}$ [15]. Denoting by $p(x,0)$ the probability distribution of the position of a cell at time 0, we can estimate the probability that the particular cell will lie at a ball of center x at time t, by integrating the function $p(x,t)$ on that ball. Sampling from the distribution $p(x,t)$ can give us an estimate on how the cells have spread at the end of the time interval $[0,t]$. Assuming that the diffusion tensor is twice differentiable, solution of (8) could be approximated by numerical methods like finite differences. But DTI provides piecewise constant values for the diffusion tensor, so one should first approximate DTI measurements by a twice differentiable function.

We note that in the case of isotropic diffusion, i.e. $U(x)$ is constant and equals the identity matrix and $\lambda_1(x), \lambda_2(x), \lambda_3(x)$ are also constant and $\lambda_1(x) = \lambda_2(x) = \lambda_3(x) = \lambda$, equation (8) takes the form

$$\frac{\partial p(x,t)}{\partial t} = \frac{a\lambda}{2}\left(\frac{\partial^2 p(x,t)}{\partial x_1^2} + \frac{\partial^2 p(x,t)}{\partial x_2^2} + \frac{\partial^2 p(x,t)}{\partial x_3^2}\right)$$

which is the form of the diffusion term in equation (1) for isotropic diffusion. We will not try to solve equation (8) numerically in this work. Rather, by assuming a piecewise constant eigenvectors and eigenvalues of the diffusion tensor we are going to sample from the distribution $p(x,t)$ by numerically producing sample paths of (7) on the interval $[0,t]$ and keeping track of the random variable x_t, i.e. the position of the cell at time t.

V. Numerical simulations

To produce sample paths of (7) we will use the Euler-Maruyama [16] numerical scheme:

$$x_{n+1} = x_n + \sqrt{\Delta t}\sqrt{a}U(x_n)L^{1/2}(x_n)Z \qquad (9)$$

Where Z is a normally distributed random variable $Z \sim N(0, I_3)$, and Δt is the discretization step. In each iteration, the matrices $U(x_n), L(x_n)$ are defined by the eigenvalues and eigenvectors of the diffusion tensor of the voxel in which x_n lies. Scheme (9) has order of strong convergence ½. This means that if $\overline{x_T}$ is the solution of (7) at time T as calculated from (9) and x_T is the actual solution of (7) at time T then

$$E\|\overline{x_T} - x_T\| \le C\left(\Delta t^{1/2}\right)$$

for some positive constant C [16].Due to absence of an actual DTI atlas of the brain, simulations where performed using

various diffusion tensors. Parameter estimation is left for future work. Simulation results are shown in the following figures.

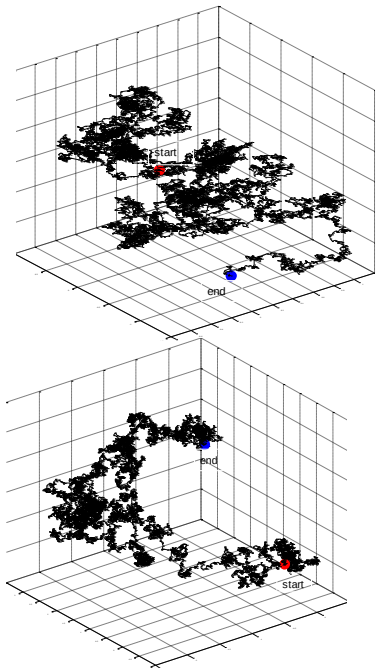
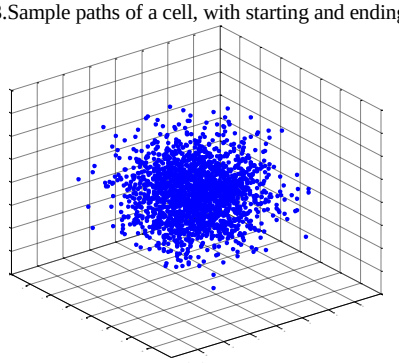
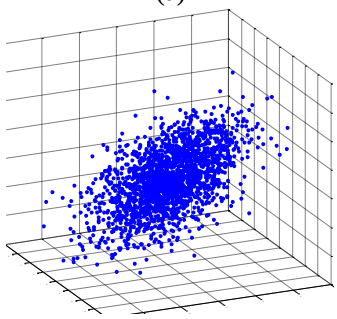


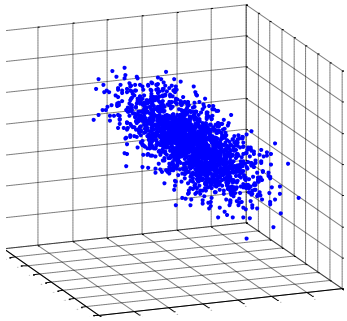
Figure 3.Sample paths of a cell, with starting and ending point.



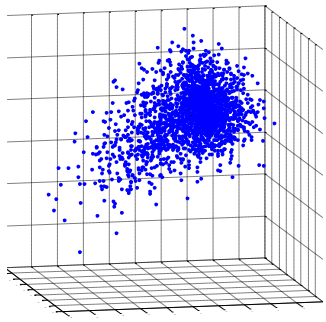
(a)



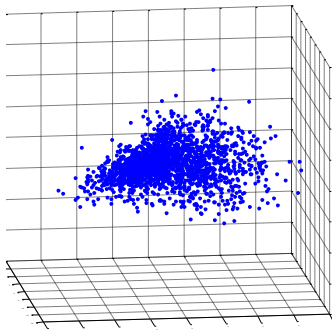
(b)



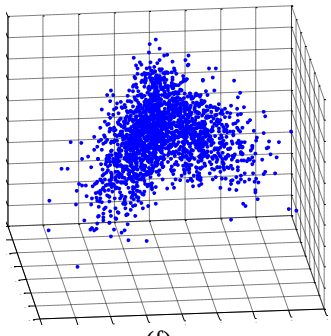
(c)



(d)



(e)



(f)

Figure 4. (a) isotropic diffusion. (b),(c),(d),(e),(f) anisotropic diffusion.

VI. CONCLUSIONS

We have proposed a mathematical treatment for simulating the diffusion of glioma tumor cells into the brain. The method consists of simulating probable paths that the tumor cells may follow, using a stochastic differential equation. We have not included proliferation of cells in the analysis as yet. The latter can be introduced by allowing the sample paths to branch during the time of the simulation in the following way: If a cell located at position y_o divides, sample paths can be produced from the equation (7) with initial position y_o thus producing probable paths for the cells resulting after mitosis.

We leave this for future work. An inspection of the results obtained so far shows that our approach satisfies basic characteristics of anisotropic glioma cell diffusion. Therefore, it will be further elaborated in order to finally serve as a tomographic data adaptable diffusion model of glioma invasion and thus eventually improve the simulation of the corresponding biological phenomena. In order to also address the response of glioma to treatment, a combination of such a model with a treatment focusing discrete-entity, discrete-event model [17] is envisaged.

REFERENCES

[1] A. Claes, A. J. Idema, and P. Wesseling, "Diffuse glioma growth: a guerilla war," *Acta Neuropathol*., vol. 114, no. 5, pp. 443-458, Nov. 2007.

[2] K. R. Swanson, E. C. Jr. 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.

[3] 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 tumors in MR images coupling diffusion with biomechanical deformation," *IEEE Trans. Medical Imag*. vol. 24, no. 10, pp. 1334-1346, 2000.

[4] O. Clatz, P.Y. Bondiau , H. Delingette, M. Sermesant, S. K. Warfield, G. Malandain, N. Ayache, "Brain Tumor Growth Simulation," Institut National de Recherche en Informatique et en Automatique (INRIA), France, Rapport de Recherche, Theme Bio, Systemes Biologiques, Projets Epidaure, No 5187, April 2004.

[5] J. D. Murray, *Mathematical biology II: Spatial Models and Biomedical Applications*, Chapter 11. Springer; 3rd edition 2011

[6] E. Konukoglou, O. Clatz, H. Delingette, and N. Ayache, "Personalization to Brain Gliomas Characterization and Radiotherapy Planning", in *Multiscale Cancer Modelling*, T. Deisboeck, G.Stamatakos Eds., Chapman & Hall/ CRC Press, Boca Raton, FL, 2011, pp. 385–406.

[7] S. G. Giatili and G. S. Stamatakos, "A detailed numerical treatment of the boundary conditions imposed by the skull on a diffusion–reaction model of glioma tumor growth. Clinical validation aspects," *Applied Mathematics and Computation*, vol. 218, no. 17, pp. 8779-8799, May 2012.

[8] A. Roniotis, G. Manikis, V. Sakkalis, M.E. Zervakis, I. Karatzanis, and K. Marias, "High grade glioma diffusive modeling using statistical tissue information and diffusion tensors extracted from atlases," *IEEE Trans Inform. Techn. Biomed.*, vol. 16, no. 2, pp. 255-263, 2012.

[9] H. L. Harpold, E. C. Jr. Alvord, and K. R. Swanson, "The evolution of mathematical modeling of glioma proliferation and invasion," *J Neuropathol Exp Neurol*. vol. 66, no. 1, pp. 1-9, 2007.

[10] S. Jbabdi, E. Mandonnet, H. Duffau, L. Capelle, K. R. Swanson, M. Pélégrini-Issac, and R. Guillevin, "Simulation of anisotropic growth of low-grade gliomas using diffusion tensor imaging," *Magn. Reson. Med*., vol. 54, no. 3, pp. 616-24, Sep. 2005.

[11] E. Stretton, E. Geremia, B. H. Menze, H. Delingette, and N. Ayache, "Importance of patient DTI's to accurately model glioma growth using the reaction diffusion equation", in *Proc. ISBI*, 2013, pp.1142-1145.

[12] R. Bammer, B. Acar, and M. E. Moseley, "*In vivo* MR tractography using diffusion imaging," *Eur J Radiol*., vol. 45, no. 3, pp. 223-34, Mar. 2003.

[13] L. Chen, M. Liu, J. Bao, Y. Xia, J. Zhang, L. Zhang, X. Huang, and J. Wang, "The correlation between apparent diffusion coefficient and tumor cellularity in patients: a meta-analysis," *PLoS One*, vol. 8, no. 11, Nov. 2013.

[14] K.M. Gauvain, R. C. McKinstry, P. Mukherjee, A. Perry, J. J. Neil, B. A. Kaufman, and R. J. Hayashi, "Evaluating pediatric brain tumor cellularity with diffusion-tensor imaging," *Am. J. Roentgenol*., vol. 177, no. 2, pp. 449-454, Aug. 2001.

[15] M. Grigoriu, *Stochastic Calculus: Applications in Science and Engineering*. Birkhäuser 2003, chapter 7.

[16] P. E. Kloeden, E. Platen, *Numerical Solution of Stochastic Differential Equations*. Springer 2013

[17] G. Stamatakos, "In Silico Oncology: PART I Clinically oriented cancer multilevel modeling based on discrete event simulation", in *Multiscale Cancer Modeling*, T.S. Deisboeck, G.S.Stamatakos Eds., Chapman & Hall, CRC, Boca Raton, Florida, USA, 2011, ISBN 9781439814406.

AUTHOR INDEX

Viceconti M.

Walker D.