PROCEEDINGS
2ND IARWISO

2ND INTERNATIONALADVANCED
RESEARCH WORKSHOP ON IN SILICO
ONCOLOGY, KOLYMPARI, CHANIA,
GREECE, 25TH AND 26TH
SEPTEMBER 2006

EDITORS: KOSTAS MARIAS AND GEORGIOS STAMATAKOS
2nd International Advanced Research Workshop on In Silico Oncology: Advances and Challenges

September 25-26, 2006

Orthodox Academy of Crete
Kolympari, Chania, Greece

Organizing Committee
K. Marias, D. Kafetzopoulos, I. G. Tollis (FORTH)
G. Stamatakos, D. Dionysiou, N. Uzunoglu (ICCS-NTUA)

Enquiries: kmarias@ics.forth.gr, gestam@central.ntua.gr

The 2nd International Advanced Research Workshop on In Silico Oncology: Advances and Challenges aims to contribute to the shaping of the emerging discipline of In Silico Oncology. The workshop focuses on the different levels of cancer modelling from molecular and cellular to tissue, organ, organism and population approaches.

At the same time it offers an international medium for the exchange of novel, computationally-oriented ideas and clinical/experimental experience related to Oncology.

Organizers:
- Foundation for Research and Technology - Hellas (FORTH) - Institute of Computer Science (ICS)
- National Technical University of Athens - Institute of Communication and Computer Systems (ICCS-NTUA)

SPONSORED BY:
ACGT - Advancing Clinico-Genomic Clinical Trials on Cancer
ICCS-NTUA
FORTH-ICS

PROPONDIS
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table of contents</td>
<td>3</td>
</tr>
<tr>
<td>Preface</td>
<td>5</td>
</tr>
<tr>
<td>Workshop Program</td>
<td>6</td>
</tr>
<tr>
<td><strong>PRESENTED PAPERS</strong></td>
<td></td>
</tr>
<tr>
<td><em>Modelling and Simulation of Tumour Growth and Response to Therapeutic Schemes I</em></td>
<td></td>
</tr>
<tr>
<td>Mathematics and computation for systems biology</td>
<td>8</td>
</tr>
<tr>
<td>Joke Blom</td>
<td></td>
</tr>
<tr>
<td>Established sciences and technologies as sources of inspiration and guidance for the emerging discipline of in silico oncology. The ISOG/ICCS/NTUA approach</td>
<td>10</td>
</tr>
<tr>
<td>N. Uzunoglu</td>
<td></td>
</tr>
<tr>
<td>Towards a collaborative formulation of the Mathematical Principles of Oncology. An outline of the ISOG/ICCS/NTUA tumor growth and response to therapies models</td>
<td>12</td>
</tr>
<tr>
<td>G. Stamatakos</td>
<td></td>
</tr>
<tr>
<td>The ISOG/ICCS/NTUA simulation model of in vivo tumour response to radiotherapeutic schemes. The case of imageable glioblastoma multiforme</td>
<td>14</td>
</tr>
<tr>
<td>D. Dionysiou</td>
<td></td>
</tr>
<tr>
<td>A four dimensional simulation model of the in vivo response of nephroblastoma to vincristine</td>
<td>17</td>
</tr>
<tr>
<td>N. Sofra</td>
<td></td>
</tr>
<tr>
<td><em>Bioinformatics and Cancer</em></td>
<td></td>
</tr>
<tr>
<td>Inferencing in In Silico Oncology: Exploiting Expressions, Biomarkers and Clinical Data for Clinical Decision Support</td>
<td>20</td>
</tr>
<tr>
<td>D. I. Fotiadis</td>
<td></td>
</tr>
<tr>
<td>Prediction of novel miRNAs and their gene targets with implications in tumourigenesis</td>
<td>22</td>
</tr>
<tr>
<td>Anastasis Oulas</td>
<td></td>
</tr>
<tr>
<td>On the Overriding Importance of Good Study Design</td>
<td>24</td>
</tr>
<tr>
<td>Emmanuel N. Lazaridis</td>
<td></td>
</tr>
<tr>
<td>Microarrays: Quality counts</td>
<td>26</td>
</tr>
<tr>
<td>Th. Margaritiis</td>
<td></td>
</tr>
</tbody>
</table>
Modeling and Simulation of Tumour Growth and Response to Therapeutic Schemes II

Simulating the response of a solid tumour to radiotherapy and hyperthermia
V. Agababov

A. The Center for the Development of a Virtual Tumor, CViT – An NCI Integrative Cancer Biology Program”. B. “A 3D multilevel agent based tumor model including gene-protein interaction profiles, cell phenotypes and multicellular patterns in brain cancer
T.S. Deisboeck

Investigating the interaction between Oncogene and Tumor Suppressor Protein
M.Akay

What are the expectations of a Clinician from in Silico Oncology
Norbert Graf

Novel Imaging technologies and Cancer

3D In-vivo Optical Imaging of Fluorescent Proteins In the Visible
Jorge Ripoll

3D Optical Diffuse Mammography for Breast Cancer Diagnostics (2 papers)
Olga V. Kravtsenyuk

Imaging, modelling and data analysis

Statistical Modelling of Shape and Biomechanical Properties – Application to Computer-Assisted Orthopaedic Surgery and Population-Based Orthopaedic Implant Shape Optimisation
M.A.Ballester

Optimized MR Imaging methology for tumor characterization
F.Zacharopoulou

Multi-level analysis and information extraction considerations for validating 4D models of human function
K. Marias
Preface

Modelling of tumour growth and response to various treatment modalities (surgery, radiation, chemotherapy, immunotherapy) has been a great challenge that motivated many efforts in the past four decades. Experimental and mathematical models have been developed but the complexity and stochasticity of biological mechanisms dictated the need of using algorithms and computer simulations. Tumours behave as complex, self-organizing, opportunistic dynamic systems. In an attempt to better understand and describe the highly complicated tumor behavior, this workshop focuses on the different complexity levels of cancer modelling from molecular and cellular to tissue/organ as well as relevant clinical issues.

The purpose of the 2nd International Advanced Research Workshop on In Silico Oncology, is to bring together researchers from around the world to exchange computationally-oriented ideas as well as experimental and clinical experience in the vast area of Oncology. The main goal of the workshop is to contribute to the shaping of the emerging discipline of In Silico Oncology and set the directions for future improvements.

We are particularly grateful to all the invited speakers, authors and researchers/staff from FORTH and ICCS-NTUA for their invaluable help to organize this workshop.

The Organizing Committee,

Georgios S. Stamatakos  
Research Associate Professor  
In Silico Oncology Group  
Laboratory of Microwaves and Fiber Optics  
Institute of Communication and Computer Systems  
School of Electrical and Computer Engineering  
National Technical University of Athens  
Phone: +30 210 772 2288, Fax: +30 210 772 3557  
E-Mail: gestam@central.ntua.gr  
URL: http://www.in-silico-oncology.iccs.ntua.gr/

Kostas Marias  
Associated Researcher  
Biomedical Informatics Laboratory  
Institute of Computer Science  
FORTH  
Phone: +30 2810 391696  
Fax: +30 2810 391428  
E-mail: kmarias@ics.forth.gr  
URL: http://www.ics.forth.gr/eHealth/marias.html

Dimitra Dionisiou  
Researcher  
In Silico Oncology Group  
Laboratory of Microwaves and Fiber Optics  
Institute of Communication and Computer Systems  
School of Electrical and Computer Engineering  
National Technical University of Athens  
Phone: +30 210 772 3556, Fax: +30 210 772 3557  
E-Mail: dimdio@esd.ece.ntua.gr  
URL: www.in-silico-oncology.iccs.ntua.gr

Dimitris Kafetzopoulos  
Group Leader Post - Genomic Research  
Institute of Molecular Biology and Biotechnology  
FORTH  
Phone: +30 2810 391594 (office)  
Phone: +30 2810 391927 (lab)  
Fax: +30 2810 391101  
e-mail: kafetzo@imbb.forth.gr  
URL: http://www.imbb.forth.gr/people/kafetzo/dimitris.htm

Ioannis Tollis  
Head of Biomedical Informatics Laboratory  
Professor of Computer Science  
Institute of Computer Science  
FORTH  
Phone: +30 2810 391671, 2810 393579  
Fax: +30 2810 391428  
E-mail: tollis@ics.forth.gr  
URL: http://www.ics.forth.gr/bmi/tollis/index.html

Nikolaos Uzunoglu  
Professor  
School of Electrical and Computer Engineering  
Institute of Communication and Computer Systems  
Laboratory of Microwaves and Fiber Optics  
In Silico Oncology Group  
National Technical University of Athens  
Phone: +30 210 772 3556, Fax: +30 210 772 3557  
E-Mail: nuzu@cc.ece.ntua.gr  
URL: www.in-silico-oncology.iccs.ntua.gr
## 2nd IARWISO Conference Program

### 25th September

#### 8:45-9:00  Conference Welcome (K. Marias, G. Stamatakos)

<table>
<thead>
<tr>
<th>Time</th>
<th>Session</th>
</tr>
</thead>
<tbody>
<tr>
<td>9:00 - 13:00</td>
<td><strong>Modeling and Simulation of Tumour Growth and Response to Therapeutic Schemes I</strong></td>
</tr>
<tr>
<td></td>
<td>Joke Blom, CWI, Netherlands, “Mathematics and computation for systems biology”</td>
</tr>
<tr>
<td></td>
<td>N. Uzunoglou, ICCS-NTUA, Athens, Greece, “Established sciences and technologies as sources of inspiration and guidance for the emerging discipline of in silico oncology. The ISOG/ICCS/NTUA approach”</td>
</tr>
<tr>
<td>11:00 - 11:30</td>
<td><strong>Coffee Break</strong></td>
</tr>
<tr>
<td>13:00 - 14:30</td>
<td><strong>Lunch</strong></td>
</tr>
</tbody>
</table>

#### 14:30-18:00  Bioinformatics and Cancer

<table>
<thead>
<tr>
<th>Time</th>
<th>Session</th>
</tr>
</thead>
<tbody>
<tr>
<td>14:30 - 18:00</td>
<td><strong>Bioinformatics and Cancer</strong></td>
</tr>
<tr>
<td></td>
<td>Anastasis Oulas¹, Martin Reczko², Panayota Poirazi¹, ¹Computational Biology Laboratory, IMBB-FORTH, Heraklion, Crete, Greece, ²ICS-FORTH, Heraklion, Crete, Greece. “Prediction of novel miRNAs and their gene targets with implications in tumourigenesis”</td>
</tr>
<tr>
<td>16 - 16:30</td>
<td><strong>Coffee Break</strong></td>
</tr>
<tr>
<td>19:30</td>
<td><strong>Conference Dinner</strong></td>
</tr>
</tbody>
</table>
26th September

9:00 -13:00  Modeling and Simulation of Tumour Growth and Response to Therapeutic Schemes II
V. Agababov1, G. Stamatakis2, N. Uzunoglu2, R. Aghgashyan1, 1State Engineering University of Armenia, Yerevan, Armenia, 2ICCS-NTUA, Athens Greece “Simulating the response of a solid tumour to radiotherapy and hyperthermia”
T.S. Deisboeck1, L. Zhang1, Z. Wang1, 1Complex Biosystems Modeling Laboratory, Harvard-MIT (HST) Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Charlestown, MA, USA; “A. The Center for the Development of a Virtual Tumor, CVIT – An NCI Integrative Cancer Biology Program”. B. “A 3D multilevel agent based tumor model including gene-protein interaction profiles, cell phenotypes and multicellular patterns in brain cancer”
E. Pirogova1, M.Akay2, I. Cosic1, 1School of Electrical and Computer Engineering, RMIT, Melbourne, Australia, 2Harrington Department of Bioengineering, Fulton School of Engineering, Arizona State University Tempe Arizona, USA, “Investigating the interaction between Oncogene and Tumor Suppressor Protein”

11:30 -11:30  Coffee Break
Norbert Graf, University Hospital of the Saarland, Germany, “What are the expectations of a Clinician from in Silico Oncology”

12:10-15:10  Novel Imaging technologies and Cancer
Jorge Ripoll, IESL FORTH, Heraklion, Crete, Greece, “3D In-vivo Optical Imaging of Fluorescent Proteins In the Visible”
13:00-14:30  Lunch
Olga V. Kravtsenyuk, IESL FORTH, Heraklion, Crete, Greece, “3D Optical Diffuse Mammography for Breast Cancer Diagnostics”

15:10-18:00  Imaging, modelling and data analysis
M.A.Ballester, MEM Research Center, University of Bern, Switzerland, “Statistical Modelling of Shape and Biomechanical Properties – Application to Computer-Assisted Orthopaedic Surgery and Population-Based Orthopaedic Implant Shape Optimisation”
16:16:30  Coffee Break
F.Zacharopoulou, K. Marias, E.Georgiadi, G. Tollis and T.G.Maris, Radiology Department, Medical School University of Crete, Greece and ICS-FORTH, Crete Greece, “Optimized MR Imaging methodology for tumor characterization”
K. Marias, Th. Margaritis, F. Zacharopoulou, E. Georgiadi, T.G. Maris, G. Tollis, C.P. Behrenbruch, ICS FORTH, Crete, Greece, IMBB FORTH, Crete, Greece, Medical School University of Crete, Siemens, “Multi-level analysis and information extraction considerations for validating 4D models of human function”

18:00-18:30  Conclusions (K. Marias, G. Stamatakos)
Abstract— The Holy Grail of systems biology is the computation of Life at the level of cells or organisms on the basis of the complete genomic, transcriptomic, proteomic, metabolomic, and cell-physiomic information that will become available in the forthcoming years, with the idea that if the cell can be approached in a rational and integrated way, it can be, e.g., utilized as a 'cell factory' to produce chemicals or pharmaceutical components. In practice, systems biology is a systematic approach to understand the functioning of a living cell or organism, in which experiments and theory go hand in hand. It involves (i) building biological hypotheses based on quantitative experiments (high-throughput, genome-wide, living cell, in silico); (ii) transforming these hypotheses into mathematical and computational models; (iii) formulating new hypotheses and predictions based on the model; (iv) testing these model predictions with new experiments; and not unimportantly (v) integrating these computational models in a hierarchical manner into models that describe a larger part of the functioning. We focus on the development, implementation, and validation of mathematical and computational techniques for the systems biology of the cell. Together with theoretical and experimental biologists we formulate realistic - first phenomenological, later on quantitative - mathematical models of metabolic and regulatory networks including intrinsic stochasticity and spatial non-homogeneity. Depending on the cellular phenomenon considered, models and methods of appropriate temporal and spatial scales are developed and then applied: models in the form of ordinary differential equations and methods for system reduction; multi-adaptive computational methods for partial differential equations for moderate spatial and temporal variability within a cell or an organelle; particle models describing the interaction of individual molecules and computational methods for the evaluation of the dynamic behavior; and methods for integration of these different approaches into a single simulation.

Mathematics and Computation for Systems Biology

Joke Blom

Dept. Modelling Analysis and Simulation (MAS)
Center for Mathematics and Computer Science (CWI)
P.O. Box 94079, NL-1090 GB Amsterdam
Email: Joke.Blom@cwi.nl

Modelling regimes in biochemistry

- **PTS - Glucose uptake and metabolism**
  Aim: Investigate the role of diffusive transport in controlling cellular function [5, 6, 9].
  Flux and control
  Gradients

- **Cardiomyocyte mitochondria**
  Modelling ROS-Induced ROS Release, preconditioning, apoptosis and necrosis.
  Experimental front observation (Brady[3, 2])

Red: polarization, Green: cytosolic ROS
Comparison of mesoscopic methods for biochemical systems

**Motivation**
Different mesoscopic methods can predict different fluctuations although averages are the same.

**Method regimes**

- **GFRD** [10]
  - event-driven scheme (variable time-step),
  - tabulated analytical solutions of Smoluchowski diffusion equation,
  - 2 molecules can react per $\Delta t$,
  - not a package.

- **Smoldyn** [1]
  - every collision leads to reaction,
  - forward reaction rate = binding/unbinding radii,
  - usable package.

**Reaction-Diffusion Master Equation**

Variables: number of particles $\{U_{ij}\}$ of species $i$ in cells $j$.

$$\frac{d}{dt} p\left(U_{ij}; t\right) = \sum_{U_{ik}} W\left(U_{ij} \rightarrow U_{ik}\right) \cdot r\left(U_{ik}; t\right)$$

- in-flux from the neighbors to cell $j$,
- out-flux from cell $j$.

**Implementations**

- **SSA** [7] - solves CME (no space)
- **MesoRD** [8] - next reaction method
- **GMP** [4] - splitting diffusion and reactions

**Test case: gene expression**

- DNA + RNAp $\rightarrow$ Complex
- Complex $\rightarrow$ DNA + RNAp + P
- P $\rightarrow$ 0

**Protein noise vs lattice size**

**Isolated pair** - distribution of times between reactions.

Average time is the same, only moments differ.

**Conclusions**

- **GFRD**: properly models diffusion-limited reactions.
- **Smoldyn**: not able to reproduce higher moments.
- **RDME** (GMP, MesoRD):
  - properly accounts for spacial effects,
  - smaller lattice size yields correct higher moments.

**Thanks to**

Frank Bruggeman, Christof Franke, Klaas Krab, Hans Westerhoff (IMSW/VU)
Maciej Dobrzyński (CWI/MAS), Mark Peletier (now TUE/CASA)
Jordi Vidal Rodriguez, Jaap Kaandorp (UvA/SCS)
NWO - Netherlands Organization for Scientific Research Dutch BSIK/BRICKS project.

**References**

Established sciences and technologies as sources of inspiration and guidance for the emerging discipline of in silico oncology. The ISOG/ICCS/NTUA approach

Nikolaos K. Uzunoglu

Abstract—The high complexity of systems biology and especially in silico oncology dictates the use of several scientific and technological methods, techniques and tools. A combination of established sciences and technologies can serve as a theoretical foundation of these emerging fields. In this communication a short account of the domains that have been exploited during the development of the ICCS-NTUA tumour response to therapies simulation models is presented.

I. INTRODUCTION

SEQUENCING of the human genome and cataloging and analysis of every protein in the human body (proteomics) that is currently under way have shaped a completely new and promising environment in the vast area of biomedical sciences and technology. Detailed analytical understanding of a plethora of molecular mechanisms has already been successfully exploited for diagnostic and therapeutic purposes (e.g. computer drug design, gene therapy etc.).

Nevertheless, in many critical cases such as in the case of cancer, understanding disease at the molecular level, although imperative, is not generally a sufficient condition for a successful treatment. The astonishing complexity and degree of interdependence among the elementary biological mechanisms involved in tumor and normal tissue growth and response to therapeutic modalities as well as the partly stochastic character of cancer behavior dictate an extension of the analytical understanding of the disease to higher levels of biological complexity. Subcellular, cellular, tissue, organ, system, organism and population levels should also be addressed with rigour analogous to the one characterizing the molecular approach.

The challenge to mathematically describe cancer analytically and/or algorithmically might well be paralleled to the challenge of mathematically describing planetary motion as was posed millenia ago. Apparently Cancer is a natural phenomenon too and as such it must be amenable to mathematical description.

Analytical mathematics alone, despite the fact that it can be used to construct models of simple, mainly experimental tumor geometries such as tumor spheroids or non solid tumors, does not seem particularly adequate for the description of realistic tumors in vivo with complex geometries, complex metabolic activity and complex spatial proliferation distribution. On the contrary a combination of discrete state algorithmic descriptions of the system under consideration in conjunction with analytical mathematics and probability theory have been shown to be a quite efficient approach. Therefore, it is postulated that appropriate combinations of discrete and analytical mathematics with probability theory can describe critical aspects of malignant tumor behavior at the molecular and cellular levels of biological complexity.

Evidently, experimental and clinical validation of such methodological principles in conjunction with the determination of their predictability limits would play a central role is such an approach.

II. THE RADIATION THERAPY PARADIGM

In the radiation therapy case current treatment planning algorithms are based on the concept of physical optimization of the dose distribution and rely on rather crude biological models of tumour and normal tissue response. Such algorithms practically ignore the highly complicated dynamic behavior of malignant and normal cells and tissues.

The introduction of advanced biosimulation methods based on cell proliferation mechanisms and also on information drawn from the cellular and molecular properties of each individual malignancy and each individual patient are expected to substantially improve the radiation therapy efficiency.

This would be accomplished by using alternative fractionations, spatial dose distributions and even combination with other therapeutic modalities such as chemotherapy, hyperthermia etc. Therefore, efficient modelling, simulation and visualization of the biological phenomena taking place before, during and after irradiation is of paramount importance.

Discrete time algorithmic descriptions (simulations) of the various phenomena offer the possibility of taking into account a large number of involved mechanisms and interactions. The same philosophy has already been extensively applied to purely technological problems and the emerged numerical methods (e.g. the Finite Difference Time Domain (FDTD) technique) have proved to be very efficient and reliable.

A further prominent characteristic of the biological phenomena under consideration is stochasticity. The fate of a single irradiated cell cannot be accurately predicted for
example. Only survival probabilities can be assigned to the cell based on the accumulated experimental and clinical observations made on large cell populations. Furthermore, the exact spatiotemporal distribution of the various cell cycle phases within the tumor volume is generally unknown, although some plausible macroscopic hypotheses can be made.

Therefore, stochastic techniques such as the generic Monte Carlo method seem to be particularly appropriate for the prediction of tumor growth and response to radiation therapy. The practical usefulness of such methods is both to improve understanding of the cancer behavior and to optimize the spatiotemporal treatment plan by performing in silico (= on the computer) experiments before the actual delivery of radiation to the patient.

The clinician would be able to perform computer simulations of the likely tumor and adjacent normal tissue response to different irradiation scenarios based on the patient’s individual imaging, histologic and genetic data. The simulation predictions would support him or her in selecting the most appropriate fighting strategy. To this end a substantial number of experimental and analytical models have been developed. On the contrary a rather small number of actual three-(or four-) dimensional computer simulation models have appeared in the literature. Exploitation of the potential of current visualization techniques is even more limited. The basic philosophy of our group is to develop detailed four dimensional simulation models of the biological systems under consideration whereas at the same time to make use of advanced technology (e.g. visualization systems, client-server architectures, parallelization etc.)


The following events motivated the development of the ISOG/ICCS/NTUA oncological simulation model:

a. Historical advances in Biology (Human Genome Project etc.)

b. Massive development and applications of numerical computer methods to technological problems that has led to accurate and effective solutions (e.g. Finite Difference Time Domain Method (FDTD) for the solution of complex electromagnetic problems)

c. Intense Involvement of the Microwave Laboratory in Oncological Research (development of hyperthermia systems etc.)

d. Considerable experience with the solution of complex electromagnetic problems using numerical techniques

REFERENCES


Towards a collaborative formulation of the Mathematical Principles of Oncology. An outline of the ISOG/ICCS/NTUA tumour growth and response to therapies multiscale models

Georgios S. Stamatakos

Abstract—Inspired by the history of science and approaching biology as the physical science of living matter, we propose the development of a parsimonious mathematical formulation of the multiscale biological phenomena. Obviously such a long term endeavour has to be highly collaborative on a worldwide scale. Computational cancer biology / in silico oncology can serve as a valuable paradigm of such a process. Within this frame the tumour growth and therapy response models developed by the In Silico Oncology Group of the National Technical University of Athens during the last ten years are outlined. Current activities are also mentioned.

I. INTRODUCTION

The impressive rate of accumulation of both experimental and observational (clinical) knowledge pertaining to living matter dictates the formulation of a parsimonious system of "laws" in analogy to Newton's Mathematical Principles of Natural Philosophy. This seems to be a necessary step if a rational, coherent and transparent understanding of the biological phenomena is to be sought. Such a system would consist of a finite, yet considerable number of principles and refer to all levels of biocomplexity (Fig.1), as in accordance to D. Noble there is no privileged level of causality. The experimental, observational and theoretical study of cancer, a markedly multiscale biological phenomenon of obvious clinical importance, may be viewed as an excellent ground for the establishment of a number of such multiscale laws. Their formulation might well be achieved in a combination of algorithmic - discrete and continuous mathematics terms.

II. IN SILICO ONCOLOGY AND THE PROBLEM OF CANCER PREDICTABILITY

Within this frame the emerging field of in silico (computational) oncology has already provided some quite reliable descriptions of several biological mechanisms characterizing cancer, of both continuous and discrete nature. Obviously cancer is far from being a purely deterministic phenomenon. Instead it seems to behave like a mixture of deterministic (e.g. sequence of cell cycle phases) and stochastic (e.g. radiation cell kill probability) processes. Subsequently, stochasticity aspects should always be taken into account. Nevertheless, the more critical knowledge becomes available, the more deterministic the cancer phenomenon appears to become. An illustrative example supporting this hypothesis is that more detailed knowledge of the genetic status of a tumour may lead to a better prediction of its response to therapeutic interventions, thus to an apparently more deterministic tumour behaviour.

III. AN OUTLINE OF THE ISOG/ICCS/NTUA TUMOR GROWTH AND RESPONSE TO THERAPIES MODELS

Based on the previous thoughts, the In Silico Oncology Group, National Technical University of Athens, has developed a number of hybrid discrete Monte Carlo / cellular automata and continuous differential equation simulation models of tumour growth and response to therapeutic modalities. The models range from tumour growth and radiotherapy response in vitro to the clinical tumour response to radiotherapeutic and chemotherapeutic schemes in vivo, based i.a. on actual imaging data.

ECOSYSTEM LEVEL
POPULATION LEVEL
ORGANISM LEVEL
SYSTEM LEVEL
ORGAN LEVEL
TISSUE LEVEL
CELLULAR LEVEL
SUBCELLULAR
MOLECULAR LEVEL
ATOMIC LEVEL

Fig. 1. Fundamental levels of biocomplexity
Processed molecular data is used in order to perturb the radiobiological or pharmacodynamic cell kill parameters about their population based mean values. A prototype system of quantizing cell clusters included within each geometrical cell of a discretizing mesh covering the anatomic area of interest lies at the heart of the proposed simulation approach. Cell cycle phase durations and imaging based metabolism distribution define i.a. the quantization equivalence classes considered. Several algorithms have been developed so as to simulate i.a. various macroscopic mechanisms such as tumour expansion/shrinkage and mechanical boundary conditions as well as the effects of particular drugs (e.g. temozolomide) and radiation on the tumor under consideration.

A number of the models developed, mainly referring to imageable glioblastomas, have already been clinically validated to a substantial degree. Long term clinical testing and adaptation procedures are in process. The response of treatment affected normal tissues by radiotherapeutic schemes has also been addressed for certain cases. Currently, a substantial extension of the simulation models to the nephroblastoma (Wilm's tumor) and breast cancer cases is being performed within the frame of the European Commission funded project ACGT (Advancing Clinico-Genomic Trials on cancer), in collaboration with several clinical and technological institutions across Europe (Fig.2).

The whole effort profits considerably from the US NIH-NCI supported Center for the Development of a Virtual Tumor (CViT). It is worth noting the remarkably collaborative character of this and other complementary research efforts on a global scale.

The expected practical usefulness of the type of models mentioned would be the possibility of virtually experimenting in silico (on the computer) with the intention of optimizing the cancer treatment strategy based on the specific molecular, histopathologic, imaging and historical data of each individual patient. Deeper understanding of the cancer disease and at the same time of the related natural phenomenon is a further intermediate goal of considerable importance.

REFERENCES


Fig. 2. An oversimplified diagram of the “Oncosimulator” to be clinically validated, adapted and optimized within the frame of the ACGT project.
The ISOG/ICCS/NTUA simulation model of in vivo tumour response to radiotherapeutic schemes. The case of imageable glioblastoma multiforme

Dimitra D. Dionysiou

In Silico Oncology Group, Institute of Communication and Computer Systems, School of Electrical and Computer Engineering, National Technical University of Athens, Iroon Polytechniou 9, GR-157 80 Zografos, Greece
email: dimdio@esd.ece.ntua.gr

Abstract—The aim of the present paper is to present comparative simulations performed with a recently developed simulation model of glioblastoma multiforme response to radiotherapy in vivo. The model is based on the available imaging, histopathologic and genetic data of the patient and numerous fundamental biological mechanisms are explicitly described. Of particular importance is the possibility of incorporating the effect of various genetic determinants on tumour response to radiotherapy.

I BRIEF OUTLINE OF THE MODEL

In the following paragraphs a brief outline of the simulation model is presented. For a detailed description refer to [1]-[4].

The clinician delineates the tumor and its metabolic subregions on the available imaging data by using a dedicated computer tool. A three-dimensional discretizing mesh covers the region of interest. The elementary cubic volume of the mesh is called “Geometrical Cell (GC)”. In each time step of the simulation the geometrical mesh is scanned and the updated state of a given GC is determined on the basis of a number of algorithms describing the behaviour of the cells constituting the tumour, which can be briefly described as follows:

Each GC of the mesh belonging to the tumour initially accommodates a Number of Biological Cells (NBC), which are distributed in a number of “classes” (compartments), each one corresponding to the phase in which its cells are found (within or out of the cell cycle: G1, S, G2, M, G0, Necrosis, Apoptosis).

The cytokinetic model presented in Figure 1 is adopted.

Cell killing by irradiation is described by the Linear Quadratic or LQ Model:

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

where \( S(D) \) is the surviving fraction after a (uniform) dose \( D \) (Gy) of radiation to a population of cells. The parameters \( \alpha \) (Gy\(^{-1}\)) and \( \beta \) (Gy\(^{-2}\)) are called the radiosensitivity parameters of the LQ model.

The simulation of tumour expansion or shrinkage is based on the following rules:

In case that the actual number of alive and dead (but still existing) tumour cells contained within a given GC is reduced to less than NBC/10, then a procedure which attempts to “unload” the remaining biological cells in the neighboring GCs takes place. In case that at the end of the unloading procedure the given GC becomes empty, it is assumed to disappear from the tumour. An appropriate shift of a chain of GCs, intended to fill the “vacuum”, leads to tumour shrinkage.

On the other hand, if the number of alive and dead cells within a given GC exceeds NBC+NBC/10, then a similar procedure attempting to unload the excess cells in the surrounding GCs takes place. In case that the unloading procedure fails to reduce the number of cells to less than NBC+NBC/10, then a new GC emerges. Its position relative to the “mother” GC is determined using a random number generator. An appropriate shifting of a chain of adjacent GCs leads to the expansion of the tumour.

This work was supported by the Operational Programme for Educational and Vocational Training II (EPEAEK II) Pythagoras II and cofunded by the European Social Fund (75%) and National Resources (25%).

II COMPARATIVE SIMULATIONS

The molecular basis of glioblastoma multiforme (GBM) radiosensitivity has been extensively studied during the last decades. In general, an increased radioresistance has been observed for GBM cells lacking functional wt p53, and this was manifested by a relatively lower $\alpha$ parameter in the LQ model. Based on one such study [5], we considered two hypothetical GBM tumors otherwise identical except for p53 gene status:

1) a GBM tumor with wild type p53:
   $\alpha_p = 0.61\text{Gy}^{-1}$, $\beta_p = 0.02\text{Gy}^{-2}$
2) a GBM tumor with mutant p53:
   $\alpha_p = 0.17\text{Gy}^{-1}$, $\beta_p = 0.022\text{Gy}^{-2}$

In consistence with experimental biology [6]-[7], we assumed $\alpha_{G0} = \alpha_p / \text{OER}$ and $\beta_{G0} = \beta_p / \text{OER}^2$ where OER: the Oxygen Enhancement Ratio, taken equal to 3 [7], and $\alpha_S = 0.6 \alpha_p + 0.4 \alpha_{G0}$, $\beta_S = 0.6 \beta_p + 0.4 \beta_{G0}$. The meaning of the above symbols is the following: $\alpha_p, \beta_p$: the LQ Model parameters for G1, G2, M phases, $\alpha_S, \beta_S$: the LQ Model parameters for S phase, $\alpha_{G0}, \beta_{G0}$: the LQ Model parameters for G0 phase.

The boundary of the tumour and its necrotic area has been delineated based on MRI imaging data. The dimensions of each GC are 1mm × 1mm × 1mm. Such a volume contains roughly $10^6$ biological cells (NBC=10$^6$) [8]. Since GBM is generally considered a poorly differentiated tumour, as a first approximation all non-clonogenic cells are considered to be necrotic (sterile cells are not taken into account). Other simulation parameters of importance are the cell cycle duration (TC=40h) and the cell loss factor (CLF=0.3) [9].

The following radiotherapy schemes have been simulated: a standard fractionation scheme (2 Gy once a day, 5 days a week, 60 Gy in total), a hyperfractionation scheme (1.2 Gy twice a day, 5 days per week, 72 Gy in total), an accelerated hyperfractionation scheme (HART) (1.5Gy three times a day, 5 days per week, 54 Gy in total), a hypofractionation scheme (6Gy once a week, 60 Gy in total) and the Continuous Hyperfractionated Accelerated Radiotherapy (CHART) scheme (dose fraction 1.5Gy, three fractions per day, 7 days per week, 54Gy in total). The distribution of the absorbed dose in the tumour region is assumed to be uniform. After the completion of a radiotherapy schedule, if the tumour cells have not been killed, they begin to repopulate the tumour.

III RESULTS

Fig. 2 and Fig. 3 depict the number of alive tumor cells, as a function of time, for the simulated radiotherapy schemes, and for the two hypothetical GBM tumours differing in their p53 status.

![Figure 2](image1.png)  
**Figure 2.** The number of surviving tumour cells as a function of time for a hypothetical GBM tumour with wild-type p53, irradiated according to several different radiotherapeutic schemes.

![Figure 3](image2.png)  
**Figure 3.** The number of surviving tumour cells as a function of time for a hypothetical GBM tumour with wild-type p53, irradiated according to several different radiotherapeutic schemes.

Fig. 4 depicts 3D reconstructions of the hypothetical tumours with wt and mt p53 right after having been irradiated according to the standard radiotherapy schedule. As expected, 3D visualization offers improved insight into the macroscopic geometry and structure of the tumour.

The results of the comparative simulations are biologically reasonable. In the case of the tumor with wt p53 the trend for reduction of the number of alive tumor cells is clearly pronounced; all radiotherapy schemes, apart from hypofractionation, seem to kill all the clonogenic cells we have initially assumed, so the tumour does not regrow after the end of the treatment. The tumour with mt p53 is so radioresistant that all radiotherapy schedules fail to hinder clonogenic cells from rapidly proliferating during therapy.

It should be noted that in clinical practice the choice of the appropriate radiotherapy schedule requires taking into account the effect of irradiation on the surrounding normal tissues. Nevertheless, by examining the simulation results, useful remarks can be made concerning tumor cell kill.
Figure 4. 3D reconstruction of the hypothetical tumours with (a) wild type p53 and (b) mutant p53 at the end of a standard fractionation radiotherapy treatment. Colour code → red: proliferating cell region, green: G0 cell region, blue: dead cell region.

Particularly interesting is the comparison of HART and CHART radiotherapy schedules. Both schemes employ the same total dose and fraction dose, but in the case of CHART there is no treatment gap during the weekend; the irradiation of the tumor takes place twelve consecutive days. Therefore, CHART seems to be advantageous in terms of tumor cell repopulation restrain during therapy. Nevertheless, since CHART’s duration is shorter, if it does not succeed in eliminating all clonogenic tumor cells – as in the case of tumor with mt p53- the repopulation of the tumor begins earlier.

IV CONCLUSIONS

The simulation results are in agreement with experimental observations and clinical experience. The model satisfactorily simulates characteristics of tumor behavior such as tumor shrinkage, repopulation and expansion and offers the advantage of readily adapting the parameters that take into account the influence of genetic determinants such as the p53 gene status.

Obviously, experimental and clinical feedback should always be used in order to improve the reliability of the model. The software system is currently undergoing a clinical adaptation procedure, by comparing the model “predictions” with clinical data before, during and after a radiotherapy course. In parallel, all the involved phenomena are constantly being studied, in order to keep pace with the ever-accumulating scientific knowledge.

REFERENCES

A four dimensional simulation model of the \textit{in vivo} response of nephroblastoma to vincristine

Nikoletta Sofra, Georgios Stamatakos, Norbert Graf, Nikolaos Uzunoglu

Abstract—Computational tumor modeling is fast evolving and focusing on real clinical problems is expected to prove its practical usefulness as a patient individualized decision support platform. In this context a novel spatiotemporal model simulating the behaviour of the nephroblastoma neoplasm (Wilm’s tumour) and its response to vincristine dose administration schemes is presented. The input of the model can be the real patient’s imaging data. The model is based on a number of algorithms previously developed by the In Silico Oncology Group, National Technical University of Athens, most of which have nevertheless been substantially modified and extended. A careful study of vincristine pharmacology has been conducted in order to realistically simulate both the drug pharmacokinetics and pharmacodynamics. Computer implementation of the model has been carried out using object oriented programming, thus facilitating code re-use and eventual future extensions. Reliability checks and pertinent parametric studies have been performed. Preliminary results agree at least qualitatively with clinical observations. A systematic clinical validation, adaptation and optimization of the model has been planned to take place within the frame of the EC funded project “ACGT: Advancing Clinicogenomic Trials on Cancer” (FP6-2005-IST-026996).

I INTRODUCTION

DURING the last decades several efforts to mathematically simulate the behaviour of neoplasms and their response to various therapeutic schemes have been made. The ultimate goal is to enable the clinician to perform \textit{in silico} (on the computer) experiments so that decisions concerning therapeutic schemes can be substantially rationalized, supported and accelerated. The field of \textit{in silico} oncology has thus emerged. The \textit{In Silico} Oncology Group, ICCS, NTUA has been engaged in developing simulation models towards this direction, by fully exploiting the insight gained by molecular biology and other disciplines as well as the individual patient’s data.

The special case of the nephroblastoma neoplasm treated with vincristine is given special consideration and a coordinated effort to maximize the clinical aspect of the model is taking place especially within the frame of the ACGT research program.

Vincristine pharmacology, both on the systemic (pharmacokinetics) and on the cellular (pharmacodynamics) scale provides critical information for the simulation of the cytotoxic action of the drug and subsequently the response of the tumour to chemotherapy.

The spatiotemporal model proposed has been based on a number of algorithms previously developed by the In Silico Oncology Group, National Technical University of Athens [1]-[4], most of which have nevertheless been substantially modified and extended.

As a first realistic approximation, the individual patient’s imaging data are used to describe the tumour geometry in terms of a triaxial ellipsoid.

Technical issues such as computation time or the use of object oriented programming and validation issues are also addressed.

The simulation predictions concerning the development of the nephroblastoma tumour and its response to existing chemotherapeutic schemes of vincristine are in at least qualitative agreement with clinical observations.

II THE CLINICAL PROBLEM

The nephroblastoma neoplasm (Wilm’s tumour) is a common pediatric solid kidney tumour mainly affecting children younger than 5 years. Although the genetic causes that lead to this neoplasm are not yet known, nephroblastoma has been associated with gene WT1 [5]. It is normally treated with chemotherapy and surgery. Vincristine is widely used in nephroblastoma treatment, mainly in combination with other drugs (dactinomycin and/or doxorubicin).

There are two different treatment approaches, one requiring chemotherapy to take place prior to surgery and the other one dictating the use of chemotherapeutic agents after removal of the tumour. The model under discussion...
might contribute to the selection of the optimal treatment for any individual patient taking also into account specific serum biochemical data.

The scheme simulated is the following: intra venous (i.v.) bolus injection of 1.5mg/m² every week for four successive weeks. This scheme is widely used in clinical practice. The model can however be readily adapted to simulate different dose administration schemes.

III VINCristine

A. Vincristine Pharmacokinetics

Following an extensive comparative bibliographical study of vincristine pharmacokinetics, reference [6] has been selected as the most suitable source of pharmacokinetic parameters of vincristine for the model developed. A complex algebraic analysis, using the corresponding parameters and appropriate pharmacology equations resulted in the following equation:

\[
C = 0.117e^{-6.300t} + 3.868 \times 10^{-3}e^{-0.0145t} \text{mg/L},
\]

(1)

where \(C\) is the concentration of vincristine in plasma and \(t\) is the time elapsed after injection.

Equation 1 is used by the model to determine the concentration of vincristine in the plasma of the patient. This information is also exploited by the corresponding pharmacodynamics equation.

B. Vincristine Pharmacodynamics

Both a qualitative and a quantitative analysis of vincristine pharmacodynamics are vital for the proper simulation of the drug’s cytotoxic effect.

Vincristine’s antineoplastic effect is usually attributed to the drug’s ability to destroy the cell’s microtubules’ functionality by binding to the protein tubulin [7]. The microtubules form the mitotic spindle. Failure of the mitotic spindle causes the cell cycle to stop during mitosis, inducing programmed cell death (apoptosis) [7]. This is the reason why vincristine is characterised as an M-phase specific drug.

In order to quantitatively express the drug’s antitumour effect the survival curve presented in [8] has been used.

IV Fundamental Algorithms

The concept of cellular automaton is adapted in order to model cell cycling and transitions to and from other possible cell states (cytokinetic diagram). The cytokinetic diagram constructed to integrate vincristine’s cytotoxic action is depicted in Fig.1. It is clear that vincristine can enter the cell and bind to tubulin at any point of the cell cycle, but its cytotoxic effect is observed only when the cell enters mitosis.

The tumour is computationally placed within a virtual discretizing mesh, each geometrical cell of which is considered to occupy a certain size of space and can therefore contain a specific number of biological cells. The growth or shrinkage of the tumour is modeled with the shift of the contents of the discretizing mesh cells, so that some geometrical cells can be added to the tumour or can no longer belong to it, depending on the local cell population at any given time.

V Technical Issues and Validation

Object oriented programming has been introduced into the computer implementation of the model. The code developed can be easily adapted and most of the classes implemented can be reused in the case of simulating different drugs for nephroblastoma or even different kinds of neoplasms.

A typical execution time is 2 min for the simulation of the response of a 600 mm² tumour to the above mentioned therapeutic scheme during 28 days on a Pentium 4, 3 GHz, 512 MB RAM.

As far as validation is concerned, various numerical, qualitative and quantitative tests have been devised and implemented so as to check the integrity and the stability of the code.

VI Results

The simulation of the development of a nephroblastoma neoplasm of initial size (0.6 cm, 1.2 cm, 1.8 cm) in terms of the three axes of a triaxial ellipsoid is depicted in Fig.2. It is worth noting that the doubling time derived from the curve i.e. 16 days, is in good agreement with clinical observations.
Ellipsoid tumour of initial size of axes (1cm, 2cm, 3cm). It is chemotherapeutic scheme are depicted in Fig.3 for an of the tumour after a few days is apparent.

Days after the last chemotherapy session, which is in interesting to note that following each therapeutic session removal takes place roughly at that time. Finally, regression that minimum tumour size is reached approximately two simulated and depicted. The results of the simulation show number of cells falls. An interval of 50 days has been simulated and depicted. The results of the simulation show that minimum tumour size is reached approximately two days after the last chemotherapy session, which is in agreement with the clinical observation that surgical removal takes place roughly at that time. Finally, regression of the tumour after a few days is apparent.

Fig.2 Predictions of the simulation of the development of a nephroblastoma neoplasm of initial axes (0.6 cm, 1.2 cm, 1.8 cm). Tumour size in terms of total number of cells vs. time (days).

The results of the simulation of the above mentioned chemotherapeutic scheme are depicted in Fig.3 for an ellipsoid tumour of initial size of axes (1cm, 2cm, 3cm). It is interesting to note that following each therapeutic session there is a temporary rise of dead cells. After some time, which is the time needed for the immune system to trace and eliminate the products of the apoptotic cell death, the total number of cells falls. An interval of 50 days has been simulated and depicted. The results of the simulation show that minimum tumour size is reached approximately two days after the last chemotherapy session, which is in agreement with the clinical observation that surgical removal takes place roughly at that time. Finally, regression of the tumour after a few days is apparent.

Fig.3 Predictions of the simulation of the neoplasm’s response to a specific chemotherapeutic scheme (4 weekly sessions). The total number of cells (→ tumour size), the total number of dead cells (taking into account the space that is occupied by the apoptotic death products) and the total number of actually living cells (dormant and proliferating) are depicted. The point when minimum tumour size is reached is shown.

VII CONCLUSIONS – DISCUSSION

The model presented, after being extensively clinically tested and adapted, is expected to be able to support clinicians’ decisions concerning various candidate cancer treatment schemes and thus facilitate patient individualized treatment optimization. Suggestion of new therapeutic strategies as well as contribution to the training of doctors, life scientists, researchers or interested patients by demonstrations of the likely tumour response to different therapeutic schemes are further expected uses of the model. A long term validation of the model is to take place within the frame of the ACGT project.

Adaptation of the model to include co-administration of vincristine with dactinomycin is in progress. Further adaptations include the use of the actual shape and size of the tumour instead of the ellipsoid assumption and the integration of serum protein data which are expected to considerably enhance patient individualization of the model.

REFERENCES


Abstract—The large number of bioinformatics applications during the recent years offers an abundance of experimental data related to oncology. This leads to the need for efficient algorithms and computational techniques that integrate different types of data (coming from different sources) and derive knowledge out of a the evolving volume of data. In this paper we demonstrate some of our recent research work towards inferencing in In Silico Oncology.

Keywords: gene expressions, genetic networks, genetic sequence pattern analysis, data mining.

Bioinformatics, i.e. the creation and advancement of algorithms, computational and statistical techniques, and theory to solve formal and practical problems posed by or inspired from the management and analysis of biological data, during the last years offers a large number of approaches applicable to oncology. However, due to the abundance of experimental data, efficient algorithms and techniques which integrate different types of data (coming from different sources) and derive knowledge out of a significant and continuously evolving volume of data are needed in order to support clinicians/biologists in their medical practice. Inferencing out of these data include the exploitation of gene expressions, demographic, clinical data including biomarkers as well as sequence data.

Processing of microarray images typically usually consists of gridding and spot finding, segmentation and intensity extraction [1,2]. Several difficulties appear in all above steps, such as variations of block and spot positions, existence of non-expressed spots that have zero intensity, existence of dust or other contamination on the slide that generates artefacts in the image. For this reason efficient computational analysis tools and techniques are required.

In [3] we introduced a novel five-step method to deal with all these difficulties. Initially, the raw microarray image is preprocessed with a template matching technique. In the second step, the blocks of the image are located. The third step is the spot finding in each block. In this step outlier detection is applied on each row and column of spots in order to remove the artefacts. The next step is the detection of the non-expressed spots. Finally a grid is fit on the image using a Voronoi diagram. The results of the above method are promising and this will lead us in the development of an effective segmentation technique in order to accurately extract the gene’s expression.

Gene Expression data can be exploited for the reconstruction of the genetic networks (see Fig.1). Probabilistic methods can by used for inferring complex relations between genes. In this way, regulatory mechanisms will be inferred and protein functions can be revealed. Bayesian networks have been long examined as a prominent approach for deriving network structure [4,5]. Still, there are a number of challenges that remain open. The first challenge is how to handle missing values in the procedure of learning Bayesian networks. For this reason, in one of our recent works, we employed the Structural EM [6], to handle missing values in the learning of the Bayesian network’s structure. The Bayesian network recovers the structure of regulatory interactions between the different genes.

Fig. 1: Reconstruction of genetic networks from microarray data.

A major drawback of the above procedure is the need for high computational effort, even for relatively small networks. The solution we proposed is a parallel implementation of the whole procedure. Following several tests, our results are quite satisfactory demonstrating the efficiency of our approach in genetic network reconstruction. Our results also demonstrate the need for incorporation of GRID technologies in expression data analysis.

Besides gene expression data, sequence data could be useful for In Silico Oncology applications. An innovative
project, MATCH [7], investigates how data mining and pattern recognition techniques can combine clinical and sequence data for profiling and diagnostic purposes in colon cancer. MATCH is developing a web based multi functional platform that integrates medicine and molecular biology to provide more effective treatment and enhance pharmaceutical research and drug discovery. In MATCH, clinical and biological data are integrated in order to (i) discover correlations between SNPs and colon cancer and (ii) allow for patient diagnosis, staging and treatment selection.

Data integration is the key to MATCH platform. Clinical data derived from the electronic healthcare records and biological information derived using data mining techniques from patient genomic and proteomic sources, are analyzed in order to provide patient profiles. A specially built Decision Support System matches new patient data (demographic, clinical, genetic) against these profiles. The relationship between sequence profiles under different experimental conditions and biological processes can be drawn through pattern analysis. This newly designed data mining model provides an efficient way to translate the large collection of existing profiles so as to be a handy reference for clinicians who face cancer early detection, clinical diagnosis and treatment decisions.

In MATCH architecture, a special ontology undertakes the role of facilitating knowledge sharing and reuse, while it will be experimentally integrated in the decision making process, the latter being a quite innovative feature [8]. MATCH ontology is built on top of already existing and proved ontologies (i.e. GeneOntology [9], National Cancer Institute ontology [10], Sequence Ontology [11]).

The architecture of the MATCH project is shown in Figures 2 and 3.

Fig. 2: MATCH architecture – components participating in training phase.

Fig. 3: MATCH architecture – components participating in decision support phase.

REFERENCES

Prediction of novel miRNAs and their gene targets with implications in tumourigenesis

Anastasis Oulas¹, Martin Reczko², Panayiota Poirazi¹

¹Computational Biology Laboratory, Institute for Molecular Biology and Biotechnology-FORTH, Heraklion, Crete, Greece. ²Institute of Computer Science-FORTH, Heraklion, Crete, Greece.

Background

Micro-RNAs (miRNAs) belong to a recently identified group of the large family of non-coding RNAs (1). A high percentage of miRNA genes have been located in cancer-associated genomic regions (CAGRs) and fragile sites (FRA) (3), thus implicating miRNAs in tumourgenic events. An example of such a case is the report of two known miRNAs (miR-15a and miR-16a) located in chromosome 13q14, a region deleted in more than half of B cell chronic lymphocytic leukemias (4) (B-CLLs). About 30% of miRNAs are found in the introns of other genes. It is generally thought that in most cases where the miRNA lies in the same orientation; the miRNA is co-transcribed with the host gene (5). This can have obvious implications when investigating genes that are highly expressed in certain cancers, whereby the careful investigation of their intronic sequences can lead to the identification of putative, intron-residing miRNAs, with a fundamental role in cancer development.

In this study we aim to implement efficient miRNA precursor and target prediction, computational tools, using Hidden Markov Models (HMM) trained to recognise existing biological features of miRNA biogenesis, and conservation. Previous prediction approaches adopt a pipeline architecture, whereby sequences are eliminated as the pipeline proceeds (6, 7). The drawback of these approaches is that they lose numerous true miRNAs along the line due to stringent cut-offs that often do not apply to all miRNAs. Other approaches use homology to detect novel miRNAs with similarity to previously identified miRNAs (8-10). This method obviously fails when scanning distant related sequences and when novel miRNAs lack detectable homologs. A recent study (11) also used HMMs to simultaneously consider sequence and structure for pre-miRNAs, however; did not incorporate conservation information, a very profound characteristic of miRNAs and one of the main differences to siRNA. In our approach we use HMM to simultaneously incorporate sequence, structure as well as conservation information into trained models, ultimately concluding to a very sensitive prediction algorithm, which is capable of predicting novel pre-miRNAs.

Thereafter, we will use our prediction tool(s) to scan cancer associated genomic regions (CAGR), fragile sites (FRA), and introns of genes involved in tumourigenesis, leading to the prediction of novel miRNAs; residing within these areas of the genome. We further aim to predict gene targets using existing target site prediction algorithms, for these cancer-related miRNAs. This may lead to the identification of novel functional gene modules and additional pathways involved in tumourigenesis

Methods and Algorithms

Hidden Markov Models (HMM)

We used the already implemented software called HMMER (12) for building a HMM capable of separating true miRNA precursors from other hairpin-like structures.

Combining Sequence, Structure and Conservation

In order to devise an approach that simultaneously considers multiple significant biological features; we created a 16 character code that encapsulates information regarding sequence, structure and conservation for every nucleotide position in a given miRNA

Results and Discussion

MiRNA Identification using Supervised Learning

Preliminary results from our HMM pre-miRNA prediction program reveal the effect of using different combinations of biological features: sequence, structure and conservation, on the classification of 260 true miRNA and over 35,000 negative miRNA sequences, used as a test set. The negative sequences are hairpin structures, derived from 3’UTR (these regions have not yet been documented to contain miRNA genes).

Results show an additive effect of the three features used to train the HMM. There is a clear improvement in classification performance as models are trained using increasing number of features, highlighting the importance of incorporating more biological information during the training procedure. The best results are obtained when all three features are used to train the HMM, achieving a ~85% sensitivity and specificity at a score of 3 given by the HMM program.

Future work entails the improvement of sensitivity and specificity using additional features and the implementation of novel miRNA target prediction tools for prediction of interactions between miRNA and genes with possible implications in tumourigenesis.

Prediction of novel miRNAs and their gene targets with implications in tumourigenesis

Anastasis Oulas¹, Martin Reczko², Panayiota Poirazi¹

¹Computational Biology Laboratory, Institute for Molecular Biology and Biotechnology-FORTH, Heraklion, Crete, Greece. ²Institute of Computer Science-FORTH, Heraklion, Crete, Greece.
References

On the Overriding Importance of Good Study Design

Emmanuel N. Lazaridis

Abstract—Based on the positive results of an exploratory study which had discovered a gene expression signature that could distinguish between different cancer tissues, investigators collected data for a confirmatory study. Whereas the first study used membrane-based microarrays spotted with radioactively-tagged cDNA targets, the second study measured gene expression using high-throughput RT-PCR. Some samples were analysed in both studies; however, the followup study lacked formal statistical design. The consequences are highlighted.

I. INTRODUCTION

Data derived from a few dozen images of microarray instances were analysed by this statistician a few years ago in an exploratory study. Recently, the same investigators requested analysis of a new dataset which they had collected with the intention to confirm the earlier results. The specifics of the biology underlying this report and the origins of the data studied are not described here for reasons that will shortly become evident to the reader.

The earlier study (S1) sought to identify a set of expressed genes that could be used to discriminate between a normal tissue (N) and two kinds of cancerous tissues that are often difficult to distinguish based on histology (X1 and X2). The data analysed in S1 were derived from membrane-based microarrays spotted with radioactively-tagged cDNA targets. Each microarray instance corresponded to a single tissue sample that had been previously characterised as one of N, X1 or X2 according to independent methods. As with all microarray studies, a substantial amount of “missing data” resulted from the quantification procedure, much of which was due to consideration of the faint spots corresponding to gene products with low expression in certain tissue samples. Data augmentation techniques based on a missing-at-random assumption were used to complete the data in S1. These were then subjected to partial least squares analysis in a multiple imputation context. Summary results indicated that linear projections of the data exist which, in two dimensions, could be used to clearly differentiate among N, X1 and X2 tissues.

Following analysis of S1, this statistician did not again participate in the study until, after some time, the primary investigators requested assistance in analysing new data collected with the intention to confirm the earlier results.

The second study (S2) employed robotics for high-throughput, 40-cycle RT-PCR using an array of approximately 100 expressed genes including most of those identified in S1, 5 “housekeeping” genes, and some genes that had been reported elsewhere to be differentially expressed between X1 and X2. Quantities for analysis were derived using the comparative threshold method. The S2 samples consisted of dozens of new cases of unknown pedigree, and 4 cases each of known N, X1 and X2 tissues that had been analysed in S1.

II. METHODS AND RESULTS

The S2 dataset was missing observations for a small set of combinations of gene product and biological sample (~4.7% of the measurements). Gene product that is not observed after 40 cycles of PCR is of very low prevalence in the corresponding sample. Because statistical models prefer that all conditions be observed, multiple imputation data augmentation was again employed; however, since the missing values in S2 are informative, a different imputation model was needed. A nonparametric model of the upper tail of the distribution of PCR measurements was constructed. Estimates were derived using the pooled PCR data. Imputation then consisted of random draws from the fitted PCR data model. Analysis was performed on the augmented data, as well as on another completed dataset wherein the missing cells were all set to the value 40.

A partial least squares (PLS) [1] model was fit to the (augmented) data of nearly 100 gene products measured on the 12 known samples (four each of N, X1 and X2). The fitted model was then used to classify the uncharacterised cases in addition to the 12 known samples.

A biplot of the first two components derived from the PLS model is given in Figure 1. The biplot shows good separation in two dimensions.
among the three tissue types despite reliance on a paltry 12 samples in the training set.

Prediction probability estimates for a subset of the S2 tissue samples is given in Table 1. As expected, each of the 12 known samples is assigned with highest probability to the correct group. Of the remaining tissue samples, most were estimated either as more likely to be of type X2, or about equally likely to be of type N or X2 but not likely to be of type X1. One of the uncharacterised tissue samples was predicted to be undifferentiable from normal tissue.

III. DISCUSSION

Perhaps the most amazing thing about the above analysis is that a suggestive result could be obtained despite serious structural issues with S2, indicating that a strong gene expression signature indeed appears to differentiate among these tissues. Thus, the conclusion must be that the investigators have paid a heavy price for trying to do too much with too little.

As frequently happens when better technology becomes available to investigators, the choice was made not to switch measurement modalities mid-stream. It has been observed in other studies that the correlation between measurements of gene expression arising from different experimental techniques is often low [2][3]. In this case, an attempt was made to map the gene expression signature from S1 to S2, however, the overall correlation between the microarray and PCR data was so low that this approach was precluded. Instead, the distinguishing signature had to be re-estimated from the S2 data alone. The relatively low estimates of membership probability for the 12 known samples are a direct result of the fact that, by itself, S2 is underpowered for purposes of estimating gene expression signatures.

The investigators selected the unknown tissues for study in S2 because their pedigree could not be identified by histology alone; however, a confirmatory study needs a standard by which to judge success of the classifier. While the unknown samples may represent real-world conditions for the application of an expression signature, they are of questionable use in S2. Moreover, a classification model based on histologically-distinguishable tissues may be entirely inadequate for developing knowledge about intermediate types.

Thus, S2 does not serve its intended purpose. This report re-emphasises the important notion that in silico and experimental techniques must correspond if they are to address the purposes for which they are employed.

![Fig. 1. Biplot of the first two components derived from the PLS model, demonstrating good separation in two dimensions among the three tissue types despite reliance on a paltry 12 samples in the training set.](image)

<table>
<thead>
<tr>
<th>Sample ID</th>
<th>Group</th>
<th>Pr(N)</th>
<th>Pr(X1)</th>
<th>Pr(X2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>N</td>
<td>61.4</td>
<td>15.3</td>
<td>23.3</td>
</tr>
<tr>
<td>2</td>
<td>N</td>
<td>44.9</td>
<td>27.6</td>
<td>27.5</td>
</tr>
<tr>
<td>3</td>
<td>N</td>
<td>46.6</td>
<td>32.5</td>
<td>21.0</td>
</tr>
<tr>
<td>4</td>
<td>N</td>
<td>67.6</td>
<td>15.3</td>
<td>17.1</td>
</tr>
<tr>
<td>5</td>
<td>X1</td>
<td>21.5</td>
<td>60.9</td>
<td>17.5</td>
</tr>
<tr>
<td>6</td>
<td>X1</td>
<td>24.9</td>
<td>55.2</td>
<td>20.0</td>
</tr>
<tr>
<td>7</td>
<td>X1</td>
<td>32.6</td>
<td>51.3</td>
<td>16.1</td>
</tr>
<tr>
<td>8</td>
<td>X1</td>
<td>25.1</td>
<td>54.2</td>
<td>20.7</td>
</tr>
<tr>
<td>43</td>
<td>X2</td>
<td>21.2</td>
<td>27.3</td>
<td>51.4</td>
</tr>
<tr>
<td>57</td>
<td>X2</td>
<td>17.4</td>
<td>28.2</td>
<td>54.3</td>
</tr>
<tr>
<td>58</td>
<td>X2</td>
<td>19.5</td>
<td>14.0</td>
<td>66.5</td>
</tr>
<tr>
<td>59</td>
<td>X2</td>
<td>17.4</td>
<td>19.0</td>
<td>63.6</td>
</tr>
<tr>
<td>42</td>
<td>--</td>
<td>36.8</td>
<td>21.8</td>
<td>41.4</td>
</tr>
<tr>
<td>46</td>
<td>--</td>
<td>38.9</td>
<td>16.3</td>
<td>44.8</td>
</tr>
<tr>
<td>48</td>
<td>--</td>
<td>37.5</td>
<td>18.9</td>
<td>43.6</td>
</tr>
<tr>
<td>67</td>
<td>--</td>
<td>56.6</td>
<td>22.9</td>
<td>20.5</td>
</tr>
<tr>
<td>69</td>
<td>--</td>
<td>22.6</td>
<td>18.7</td>
<td>57.7</td>
</tr>
<tr>
<td>71</td>
<td>--</td>
<td>21.9</td>
<td>21.3</td>
<td>56.8</td>
</tr>
<tr>
<td>73</td>
<td>--</td>
<td>22.4</td>
<td>34.4</td>
<td>43.2</td>
</tr>
</tbody>
</table>

Probability estimates in each row may not add up to exactly 100% due to rounding error.

ACKNOWLEDGMENT

The author would like to thank colleagues who provided the data on which this report is based, and the Foundation for Research and Technology Hellas for its support.

REFERENCES

[1]
[2]
[3]
INTRODUCTION

Microarrays have extensively been used the last seven years to address complex biological questions and have produced a vast amount of data, but have yet to realize their full potential. The interpretation and merging of microarray data remains a major challenge, because of the inherent noise and the complexity of the underlying biological networks. To produce biologically meaningful expression data, it is crucial to develop standardized approaches for manufacturing, controlled and updated annotation and careful quality control of experiments to achieve better statistics and reproducible results. In the next sections, we present the key steps needed for quality assurance of microarray experiments, so that global gene expression data can become more reliable and accurate to facilitate information sharing.

I. MICROARRAYS

To understand gene function, it is helpful to know where, when, how much a gene is expressed and what other genes are coexpressed with it. Since the expression of each gene results in increased concentration of the corresponding mRNA, DNA microarrays estimate this concentration using reporters, that each matches a particular mRNA in the cells. The extracted mRNA is converted to cDNA and then every sample is labeled. Microarrays consist of arrays of immobilized DNA reporters, where hybridization of labeled samples takes place to study differential expression or patterns of gene expression [1].

The synergy of novel chemistries, robotics and computational tools, as well as the completions of many sequencing projects, including the Human Genome Project, gave microarrays a dramatic boost, allowing the researchers to track the expression levels of all known genes simultaneously. Microarrays are a platform technology and has a wide range of applications, including expression profiling (functional genomics, molecular markers, drug discovery, pharmacogenomics), gene structure (mutations, polymorphisms, chromosomal aberrations, amplifications, short tandem repeats) molecular interactions (DNA-protein interactions, antisense oligo scanning, drug target validation) and environmental studies (monitoring microbial populations).

Although initially exciting, microarray work soon became highly frustrating for the researcher, as the amount of data has grown enormously and there is an absence of standards in both the annotation of reporters and the quality of the results found in the public databases. In a recent study, a consortium of ten highly-accredited labs analyzed identical RNA samples, using three widely-used platforms. There were relatively large differences in data obtained in labs using the same platform, but the results from the best-performing labs agreed rather well [2].

<p>| Table 1: Measures of accuracy and reproducibility in different lab using the same samples (from Irizarry et al (2005) Nature Methods 2, 345-350) |
|-------------|----------------|------------------------|-----------------|-----------------|</p>
<table>
<thead>
<tr>
<th>Platform</th>
<th>Lab ID</th>
<th>Correlation</th>
<th>SD</th>
<th>Signal (SE)</th>
<th>Proportion of Agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Affy oligo</td>
<td>1</td>
<td>0.48</td>
<td>0.32</td>
<td>0.65 (0.19)</td>
<td>0.72 0.56 0.54</td>
</tr>
<tr>
<td>Affy oligo</td>
<td>2</td>
<td>0.56</td>
<td>0.17</td>
<td>0.29 (0.03)</td>
<td>0.80 0.70 0.70</td>
</tr>
<tr>
<td>Affy oligo</td>
<td>3</td>
<td>0.67</td>
<td>0.24</td>
<td>0.70 (0.14)</td>
<td>0.68 0.66 0.60</td>
</tr>
<tr>
<td>Affy oligo</td>
<td>4</td>
<td>0.79</td>
<td>0.15</td>
<td>0.51 (0.08)</td>
<td>0.80 0.70 0.65</td>
</tr>
<tr>
<td>Affy oligo</td>
<td>5</td>
<td>0.89</td>
<td>0.25</td>
<td>0.36 (0.09)</td>
<td>0.64 0.68 0.55</td>
</tr>
<tr>
<td>two-color cDNA</td>
<td>1</td>
<td>0.65</td>
<td>0.23</td>
<td>0.59 (0.13)</td>
<td>0.68 0.64 0.65</td>
</tr>
<tr>
<td>two-color cDNA</td>
<td>2</td>
<td>0.68</td>
<td>0.21</td>
<td>0.06 (0.15)</td>
<td>0.28 0.30 0.38</td>
</tr>
<tr>
<td>two-color cDNA</td>
<td>3</td>
<td>0.46</td>
<td>0.23</td>
<td>-0.08 (0.13)</td>
<td>0.72 0.68 0.50</td>
</tr>
<tr>
<td>two-color oligo</td>
<td>1</td>
<td>0.68</td>
<td>0.51</td>
<td>0.03 (0.16)</td>
<td>0.40 0.36 0.33</td>
</tr>
<tr>
<td>two-color oligo</td>
<td>2</td>
<td>0.50</td>
<td>0.10</td>
<td>0.87 (0.17)</td>
<td>0.44 0.72 0.81</td>
</tr>
</tbody>
</table>

II. QUALITY / NOISE REMOVAL

As in most studies, the target of the microarray researcher is to increase the precision / reproducibility of the measurements without sacrificing accuracy. A typical microarray experiment consists of many successive steps (Figure 1). The point is not that these steps are ‘not perfect’; it is that they may vary from array to array and experiment to experiment. The noise that is causing this variability can be attributed to the array manufacturing, the initial samples (biological variability, purity, integrity (Figure2)) and the protocols / people to get the labeled samples, the dye properties, the hybridization and scanning, the image analysis, the data transformations (‘background’ correction, intra- and inter-array normalization) and so on.
Researchers have to be aware of these issues and avoid it, if possible. Else, the bad quality data have to be recognized and corrected or even discarded and repeated, as even one bad array can damage the final results. A big part of the noise is systematic and normalization procedures have been utilized to deal with it [3]. On the other hand, normalization can ruin data, if wrong assumptions about the ‘average’ data are made [4]. Well-placed external controls (spike-in controls) with the addition of of corresponding RNA in known quantities to sample RNA, that span the range of both ratios and intensities can help the researchers estimate the accuracy and reproducibility of the measurements, as well as the validity of the chosen normalization procedure (Figure 3).

We have recently identified the microarray raw images as a major source of noise, which makes them an ideal target for employing statistical and image analysis procedures to correct it. We have employed two parallel procedures. The first one treats the image as a homogeneous entity and denoises it using various techniques. Initial results have shown a significant increase in the signal to noise ratio and a big improvement in the subsequent image analysis procedure (Figure 4, unpublished data).

Accurate spot segmentation is an essential analysis step in spotted arrays technologies, as different segmentation algorithms lead to different results [5]. The aim is to reduce the image to single gene-expression values per spot, i.e. the log ratio of the fluorescent intensities. Background pixels can underestimate the true expression value of each channel, leading to potentially false negative calls in differential expression. On the other hand, outlier pixels, representing hybridization defects, nearby spots, dust, etc., may overestimate the expression value and create potential false positive calls. We developed a two-channel segmentation framework that aims to provide a more robust and intuitive segmentation, based on the “hybridization ground truth” [6]. Initial results indicate that our method can achieve the smallest range of estimated log ratios in replicate spots, used as a measure of reproducibility, compared to widely-used state-of-the-art segmentation methods. The range of log ratios in replicates spots can become a standard quality control measure providing uncertainty weights for individual differential expression values.
Fig. 5 Comparison of different segmentation methods by using the range of log ratios in replicate spots as a measure of reproducibility. (SegA and SegAM come from the 1st and the 2nd step of our segmentation method respectively, Imagenes from a commercial image analysis software, GOGAC and SRG from Spot and SpotSeg from spotSegmentation Bioconductor package. The range equals zero in the ideal case.

REFERENCES

Simulating the response of a solid tumour to combined radiotherapy and hyperthermia

Victor Agababov, Georgios Stamatakos, Nikolaos Uzunoglu, Ruben Aghgashyan

Abstract—A brief outline of a spatiotemporal simulation model of the response of solid tumours to combined radiotherapy and hyperthermia is presented. The model is based on previous work done by the In Silico Oncology Group, National Technical University of Athens, Since it is currently under refinement, the paper focuses on crucial technical issues rather than on clinical validation. The latter is to follow after completion of the model development and numerical behaviour checks. Indicative examples of the new data structures and algorithms introduced are given.

I INTRODUCTION

Radiation therapy plays a critical role in the treatment of several types of cancer. Radiotherapy aims at delivering the highest possible radiation dose to the tumour, keeping damage to the surrounding normal tissues as low as possible. Cancer biology is a central factor determining radiotherapy outcome. Therefore, efficient modeling, simulation and visualization of the biological phenomena taking place before, during and after treatment are of high importance. During the past four decades understanding of tumour growth and response to radiation therapy has been enhanced by means of computer simulation models. Continuous analytical mathematical models are valuable tools to study certain aspects of tumour radiobiology. However, their main limitation lies in the modest number of biological mechanisms they can incorporate without becoming prohibitively complex and impractical. On the contrary, discrete time algorithmic descriptions (computer simulations) have the advantage of high adaptability in treating complex situations. They offer the possibility of accounting for a large number of biomechanisms and interactions and they are therefore particularly suitable to describe in vivo tumour growth and response to irradiation. The stochastic nature of the involved biological phenomena favours the choice of stochastic techniques such as the generic Monte Carlo method appropriately adapted for oncological simulations. Most of the previously developed computer models employ single cell discretization lattices (each geometrical cell of the lattice contains at most one biological cell). Such discretization lattices allow for small tumours (in vitro tumour spheroids or in vivo tumours of the early avascular stage) to be simulated. Models [1] –[9], however, refer mainly to actual clinical macroscopic tumours of arbitrary shapes. Finite state cytokinetic models aiming at the simulation of the dynamics of the cell cycle have been recently incorporated into tumour behaviour models. The latter modules are of particular importance when considering the effect of radiation therapy on the ultimate fate of cells, since this is cell cycle phase-dependent. Cells residing in the G0 and S phases tend to be considerably radio-resistant, which in turn lessens the effect of radiation therapy. On the other hand hyperthermia, renders cells in these phases more radiosensitive, thus considerably improving the results of radiation treatment.

II A BRIEF OUTLINE OF THE SIMULATION MODEL

A 3D discretizing mesh is virtually superimposed on the anatomic region of interest. The elementary cubic volume of the mesh, called “Geometrical Cell” (GC) at the current version can accommodate one tumor cell. Below a number of indicative assumptions are listed.

i. Each tumour GC is assumed to contain a cell characterized by the phase in which it is found (within or out of the cell cycle, i.e.: G1, S, G2, M, G0, Necrosis, Apoptosis).

ii. Throughout the simulation procedure, the transitions of the cells from M to G1 or G0 phase and from G0 to G1 or N phase take place with probabilities depending on the metabolic sub-region to which the GC belongs.

iii. At each time step the geometrical mesh is scanned and the new “state” of a given GC is determined as follows:

iv. At time instants corresponding to the delivery of radiation to the tumour, the probability of cell kill in a particular GC is calculated based on the Linear Quadratic (LQ) Model. The fraction of cells surviving from a uniform radiation dose D is given by

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

(1)

where \( \alpha \) (Gy−1) and \( \beta \) (Gy−2) characterize the initial slope and the curvature, respectively, of the survival curve and relate to cell radiosensitivity. Cell radiosensitivity varies considerably throughout the cell cycle. The S phase is regarded as the most resistant. Cells in any proliferating phase are more radiosensitive than hypoxic cells residing in.
If the instant corresponds to the combined hyperthermia delivery then according to [10] the cell radiosensitivity will be approximately doubled. In case of various tumor types the multiplier may change. Thus, if the hyperthermia is applied along with radiation therapy equation (1) becomes

\[ S_{CT}(D) = \frac{S(D)}{2} \]  

(2)

At each time step the time registers of all GCs increase by 1 h. Cell loss due to apoptosis and necrosis is computed based on the values of the cell loss factor due to apoptosis and the cell loss factor due to necrosis. Cell cycle phase transitions take place according to the cell cycle model presented in [1].

V. Several growth algorithms have been applied such as the following:

1. Random differential growth direction. In that case when a new cell is created a cell chain shift is induced in a random direction.
2. Minimum cell shift direction. In this case the cell is placed in the direction which leads to minimum cell shifting.
3. Direction opposite to the one “looking” at the centre of mass of the tumour.

VI. Shrinkage algorithms select directions essentially opposite to those of the previous growth algorithms.

III. OBJECT ORIENTED CODE DEVELOPMENT

A UML diagram of the internal data structures is shown on Fig.1. Flexibility of the system is achieved using the following design solution: the class which is responsible for the tumor growth and shrinkage simulation is implemented using standard C++ language and is platform and compiler independent. This allows easy porting to different operating systems and compilers.

The GeometricalCellContainer class is easily extensible for the needed GUI interface. In Fig.1 it is extended for usage within the Microsoft MFC framework.

The GeometricalCellContainer is basically an aggregation of GeometricalCell class objects, each one having its own timer, cell phase and other properties. The GeometricalCellContainer class provides interface for setting the schedule of radiation and hyperthermia, shrinkage/growth algorithms, cell parameters etc.

The simulation algorithm is grossly depicted in Fig.2, using a UML sequence diagram.

REFERENCES

http://ieeexplore.ieee.org/search/wrapper.jsp?arnumber=1046955  
[http://dx.doi.org/doi:10.1109/JPROC.2002.804685 ]

[http://dx.doi.org/10.1016/j.jtbi.2004.03.024 ]


A. “The Center for the Development of a Virtual Tumor, CViT (An NCI Integrative Cancer Biology Program)”.  
B. “A 3D multilevel agent based brain tumor model that enables simulation and analysis of gene-protein interaction profiles, cell phenotypes and multicellular patterns”.

T.S. Deisboeck, L. Zhang, Z. Wang

Abstract—A. The Center for the Development of a Virtual Tumor (CViT) is an international community of investigators who focus on modeling and simulation of tumors. Charged with facilitating cancer systems biology research, CViT is dedicated to providing access to networked expertise, sharing data, software and resources securely, rapid dissemination of information, a digital model depository and a semantic layered research platform.

B. Evidence suggests that epidermal growth factor receptor (EGFR)-mediated activation of the signalling protein phospholipase C γ plays a critical role in a cancer cell’s phenotypic decision to either proliferate or to migrate at a given point in time. We have developed a novel three-dimensional multiscale agent-based model to simulate this cellular decision process in the context of a virtual brain tumor. Each tumor cell is equipped with an EGFR gene-protein interaction network module that also connects to a simplified cell cycle description. The simulation results show that over time proliferative and migratory cell populations not only oscillate but also directly impact the spatio-temporal expansion patterns of the entire cancer system. The percentage change in the concentration of the sub-cellular interaction network’s molecular components fluctuates, and, for the ‘proliferation to migration’ switch we find that the phenotype triggering molecular profile to some degree varies as the tumor system grows and the microenvironment changes. We discuss potential implications of these findings for experimental and clinical cancer research.

A. A Brief View of CViT

The URL of CViT is http://www.cvit.org. Two representative webpages are shown in Fig. 1 and Fig. 2.

T.S. Deisboeck is with Complex Biosystems Modeling Laboratory, Harvard-MIT (HST) Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Charlestown, MA, USA (Tel: 617-724-1845, Fax: 617-726-7422, Email: deisboeck@helix.mgh.harvard.edu)

L. Zhang1, is with Complex Biosystems Modeling Laboratory, Harvard-MIT (HST) Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Charlestown, MA, USA

Z. Wang1, is with Complex Biosystems Modeling Laboratory, Harvard-MIT (HST) Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Charlestown, MA, USA
B. ASPECTS OF THE 3D MULTILEVEL AGENT BASED TUMOR MODEL

Figure 3: The diagram displays the sub-cellular EGFR gene-protein interaction network that combines nucleus, cytoplasm and membrane compartments.

Figure 4: Shown is a 3D snapshot of the tumor system at time step $t = 50$

REFERENCES

Investigating the interaction between Oncogene and Tumor Suppressor Protein

E. Pirogova, M. Akay, Senior Member IEEE, and I. Cosic, Senior Member, IEEE

Abstract—Cancer develops when cells in a part of the body begin growing uncontrollably. Because cancer cells continue growing and dividing disorderly, they never mature into the specific tissue and thus, are functionally different from normal cells. Instead of dying, they outlive normal cells and persist developing new abnormal cells. The cell becomes cancerous with the alteration of the right gene combinations. However, there are some genes, known as tumor-suppressor genes that prevent malignant cell behavior. We have investigated the structure and function relationships of oncogene and p53 tumor suppressor proteins using the Resonant Recognition Model (RRM), a physico-mathematical approach based on digital signal processing methods.

I. INTRODUCTION

Bioengineering plays a significant role in current cancer research, focusing on understanding and interpreting this disease using gene identification, protein, and DNA modeling. This might facilitate evaluating the biochemical processes in cells/tissues, understanding the grounds of disease, and developing novel diagnostics and drugs using engineering, instrumentation, and computer science methodologies.

p53 genes, known as tumor-suppressor genes, help prevent malignant cell behavior. Although detected in the human genome, these genes are difficult to isolate and analyze. Other genes, known as proto-oncogenes, promote cancer if they acquire new properties due to mutations whereupon they are called oncogenes. Most common cancers involve both inactivation of specific tumor-suppressor genes and activation of certain proto-oncogenes. Proto-oncogene proteins are the products of proto-oncogenes. Although lacking oncogenic or transforming properties, they are involved in cell growth regulation or differentiation.

Analyzing the gene encoding p53 proteins could serve to evaluate the effectiveness of a cancer treatment. Mutations in this gene occur in half of all human cancers, and protein regulation is defective in a variety of other tumors. Novel strategies exploiting the knowledge of the function and regulation of p53 are being actively investigated [1]. Strategies directed at treating tumors containing p53 mutations include gene therapy, viruses replicating in p53 deficient cells, and searching for small molecules that reactivate mutant p53. An analysis of the mutual relationships between oncogene and p53 tumor suppressor proteins is extremely important for developing both new methodologies and drug design for cancer treatment.

II. METHODOLOGY

The Resonant Recognition Model (RRM) [2-4] employed in this study is a physico-mathematical model that interprets linear protein sequence information using digital signal processing methods. The RRM was developed to analyze protein-protein and protein-DNA interactions.

The RRM procedure first converts amino acid sequences into numerical sequences, and then analyzes this sequence by appropriate digital signal processing methods (Fourier Transform, Wavelets etc.) [5-10].

Assigning a particular number for each amino acid in the protein sequence is crucial in all these calculations. These numbers must have a physical meaning commensurate to the protein’s biological function. We used the energy of delocalized electrons (calculated as the electron-ion interaction pseudo-potential, EIIP [11]) of each amino acid residue. Thus, the resulting numerical series represents the distribution of the free electron energies along the protein. Then these numerical series are analyzed by Discrete Fourier Transform (DFT) to facilitate extracting information pertinent to the protein biological function.

A multiple cross-spectral function is defined and calculated to obtain the common frequency components from the spectra of a group of proteins. Peaks denote common frequency components for all sequences analyzed. Signal-to-noise ratio (S/N) for each peak described a similarity between the analyzed sequences is calculated as the ratio between the signal intensity at the particular peak frequency and the mean value over the whole spectrum. Through extensive study, the RRM reached a fundamental conclusion: one characteristic frequency represents a crucial parameter of the recognition between the interacting bio-molecules [2-4]. Knowing the characteristic frequency for a particular biological function/interaction makes it possible to identify individual “hot spot” amino acids contributing to this observed biological function, and design bioactive peptides having both the determined characteristic frequency and desired biological function [2-10]. Here the RRM approach has been employed in structure-function analysis of the relationships between oncogene, interleukin, and p53 tumor suppressor proteins.

III. RESULTS

In this study, we investigated 44-oncogene and 13-p53 protein sequences. A multiple cross-spectral analysis was performed for each selected protein group in addition to their mutual combination using the EIIP values. Fig.1-Fig.3 shows the multiple cross-spectral functions of analyzed proteins and their peak frequencies.
For a group of 44 oncogene sequences, the characteristic frequency was identified at $f=0.0322 \pm 0.023$, $S/N=408.77$ (Fig.1). This frequency component, common in all analyzed sequences within the group, is the consensus characteristic of their common bioactivity: promoting uncontrolled cell growth and proliferation. In Fig.1, we observe another less significant peak identified at the frequency $f=0.0536 \pm 0.023$ that represents a characteristic feature of proto-oncogene proteins found in our previous research [4]. The prominent characteristic frequency of p53 proteins was identified at $f=0.4326 \pm 0.077$, $S/N=159.97$ (Fig.2). As mentioned above, each specific protein function represents a single frequency. Thus, the frequency $f=0.4326$ identified is considered a characteristic feature of the specific biological activity of p53 proteins: preventing tumor formation [4]. Less significant peaks shown in Fig.2 correspond to other functions of p53 proteins, i.e. their involvement in other biological processes.

IV. DISCUSSION

Fig.3 shows one dominant peak corresponding to the common frequency component for the combined group of oncogene and p53 proteins identified at $f=0.0537 \pm 0.018$, $S/N=406.38$, and $f=0.0322 \pm 0.018$ that is a characteristic frequency of oncogenes. This prominent peak at $f=0.0537$ presents a characteristic frequency of proto-oncogene proteins [4]. Thus, the interaction between oncogene and p53 protein leads to changes in oncogene biological function and to the suppression of uncontrolled cell growth and tumor development. Because the characteristic frequencies represent the common feature of the interacting molecules [7-19], the RRM characteristic frequency is a relevant parameter for mutual recognition between bio-molecules. Thus, the RRM characteristic frequency may dictate the specificity of the protein interactions.

V. CONCLUSION

We have previously shown that the digital signal processing methods can be used to analyze linear sequences of amino acids to reveal the informational content of proteins [2-10]. This study extends the utility of the RRM procedures to oncogene and p53 tumor suppressor proteins. The results of our computational analysis clearly indicate that the RRM presents an engineering approach based on digital signal processing that is able to efficiently identify the protein characteristic patterns in protein sequences related to the common biological function/interaction of different protein families, allocate protein’s biological active site(s), and design new peptides with the desired biological function [8-10].

This novel prediction scheme can be used to facilitate the structure-function studies of different proteins and thus, could result in significant cost saving and improvement in current biotechnology and quality of new biomaterials.

References

What are the expectations of a Clinician from *In Silico Oncology*?

Norbert Graf* and Alexander Hoppe

Abstract—Treatment of most patients with cancer is still based on surgery, radiotherapy, chemotherapy and psychosocial support. Although most patients with cancer respond to therapy, not all are cured. Despite the fact that our ability to characterize and understand the various forms of cancer is growing exponentially, there is a demand for further individualization of cancer treatments to reduce mortality by improving therapies. Targeted therapies are hopefully one of three major approaches to gain this objective. These drugs block the growth of cancer cells by interfering with specific targeted molecules produced by the patient’s tumour and involved in carcinogenesis and tumour growth. A second attempt will be the conduct of more and better clinicogenomic trials as proposed by the EC funded project “ACGT: Advancing Clinicogenomic Trials on Cancer”. These trials will provide the medical and scientific community with new insights, answers and capabilities regarding all aspects of cancer. Furthermore it is known, that the enrolment of patients in a clinical trial - conducted according to GCP criteria - guarantees them the best available treatment and thus a better outcome. *In Silico Oncology* is anticipated to be the third approach in gaining a more individualized treatment for patients with cancer. From a clinical point of view 6 different simulation experiments have to be developed from *In Silico Oncology*. These models should answer the following questions for an individual patient: 1. What is the natural local tumour growth over time in size and shape? 2. When and where is it a tumour metastasizing? 3. Can the response of the local tumour and the metastases to a given treatment be predicted in size and shape over time? 4. What is the best treatment schedule for a patient regarding drugs, surgery, irradiation and their combination, dosage, time schedule and duration? 5. Is it possible to predict severe adverse events (SAE) of a treatment and to propose an alternative treatment to avoid them without deteriorating the outcome? 6. Is it possible to predict a cancer before occurring and to recommend a treatment, that will prevent the occurrence or a recurrence of a cancer in an individual patient? All approaches together will give doctors a better way to tailor cancer treatment, thus holding the promise of applying a more individualized treatment with increasing survival, reducing side effects, and improving the quality of life. (FP6-2005-IST-026996)

I. INTRODUCTION

The goal of cancer research is to find better ways to treat patients with cancer. During the last decades basic research and clinical trials have gathered a lot of new insights in the molecular biology of cancer providing new drugs and treatment approaches. This can be clearly shown for Wilm’s Tumour [2]. Nevertheless the outcome for most cancers is still dismal demanding new and better treatments for patients.

A clinical trial is one of the final stages of a long and careful cancer research process. The search for new treatments begins in the laboratory. Molecular biology did help to better understand carcinogenesis, thus finding new targets for interfering with agents, that are able to reverse the process of developing cancer cells. New methods and technologies in molecular biology will result in an exponential increase of information in future that can be handled by the advances of high-computing and informatics. It is of paramount importance to gather this information with clinical data to gain new knowledge for developing better treatments for cancer patients. This approach will result in clinicogenomic trials, as ACGT is proposing and running. Such trials will lead to more individualized cancer treatments and providing better chances of cure for patients with less side effects.

With the help of *In Silico Oncology* it is expected that cancer growth and response to different treatments can be simulated. Such *in silico* experiments might help clinicians in future to find the best way of treating an individual patient by simulating different treatments in the computer before starting the treatment in reality. The *In Silico Oncology* Group, ICCS, NTUA has been engaged in developing such simulation models, by fully exploiting the insight gained by molecular biology and other disciplines as well as the individual patient’s data.

II. PRECONDITIONS FOR TRUSTING IN IN SILICO METHODS

From a clinical point of view two preconditions are of utmost importance, if one can trust predictions of *in silico* methods:

1. *Every in silico* method has to be part of a clinicogenomic trial
2. *Every prediction of an in silico* method has to be compared with the reality

In the process of developing *in silico* methods it is necessary to define the needed data in a first step, including data from the tumor (molecular biology, pathology, imaging), from the patient (clinical data) and from the possible treatment (pharmacokinetics of drugs that will be used, the treatment schema). To make the simulation predictions as precise and realistic as possible it is crucial to get as much information from each of the different categories. The amount of data will be restricted by the availability of tumour material, imaging data and clinical data. Therefore *In Silico Oncology* has always to be integrated into or part of a clinicogenomic trial, where data management including data security and anonymisation or pseudonymisation of data as well as tumour banking is well established. In addition the trial is...
reviewed by an ethical committee and fulfils all other GCP criteria to get approval by regulatory authorities. The simulation prediction of each in silico method has always to be compared with the reality. The feedback given by the reality has to tune the in silico method to get better predictions. If treatment is based on predictions of in silico methods, this control loop has to be part of the method and automated. In that way the in silico experiments should work as a learning system. Only if there are no or minimal deviations between the prediction and the reality the in silico method is allowed to be used in a clinical setting. The clinician has to define, what can be accepted as a minimal deviation between prediction and reality. This definition should always be included in the biometrics part of a clinicogenomic trial protocol. For the safety of patients a stopping rule has to be defined, if clinical decisions are based on in silico experiments.

III. IMPORTANT IN SILICO EXPERIMENTS FOR CLINICIANS

For a clinician the following 6 questions should be addressed and answered precisely by in silico experiments:

1. What is the natural local tumour growth over time in size and shape?
2. When and whereto is a tumour metastasizing?
3. Can the response of the local tumour and the metastases to a given treatment be predicted in size and shape over time?
4. What is the best treatment schedule for a patient regarding drugs, surgery, irradiation and their combination, dosage, time schedule and duration?
5. Is it possible to predict severe adverse events (SAE) of a treatment and to propose an alternative treatment to avoid them without deteriorating the outcome?
6. Is it possible to predict a cancer before occurring and to recommend a treatment, that will prevent the occurrence or a recurrence of a cancer in an individual patient?

Comments to these questions from the clinical point of view are given in the following sections.

IV. WHAT IS THE NATURAL LOCAL TUMOUR GROWTH OVER TIME IN SIZE AND SHAPE?

In Paediatric Oncology tumours are known, that show spontaneous regression, as for example neuroblastoma (Fig.1) [5]. Even in stage IVs in infants, despite the fact that this tumour is metastasized to the liver and the bone marrow, a complete remission can occur. If there is no information on the natural course of a tumour the impact or benefit of a given treatment on the tumour can not be shown. For nephroblastoma these data are available by the documentation of individual cases refusing treatment [7] (Fig.2).

In children the growth of normal tissue has to be kept in mind. Only if the tumour growths faster than the normal tissue, there is a real tumour progression.

V. WHEN AND WHERETO IS A TUMOUR METASTASIZING?

Most important for the survival of a patient with cancer are metastases. If they occur prognosis is getting worse. The prevention and the treatment of metastatic disease has to be improved in modern oncology. If the development of metastases can be simulated accurately, it might be possible to run in silico experiments to test treatments or interventions that will prevent the occurrence of metastases. In case of overt metastatic disease a “pre-treatment” knowledge of response to different therapeutic options in an individual patient should be beneficial for this patient, as the physician will be able to start with the best available treatment.

VI. CAN THE RESPONSE OF THE LOCAL TUMOUR AND THE METASTASES TO A GIVEN TREATMENT BE PREDICTED IN SIZE AND SHAPE OVER TIME?

The evaluation of tumour response in the daily clinical practice of oncology is performed according to predefined criteria. It may, rather, be based on a subjective medical judgment that results from clinical and laboratory data that are used to assess the treatment benefit for the patient. It is necessary to define what is an “objective tumour response”. Besides this, a lot of patients do have more than one lesion. From a clinical point of view a combined assessment of all existing lesions has to be used to extrapolate an overall response to treatment. Such new guidelines to evaluate the response to treatment in solid tumors are still under discussion. For the in silico experiments the RECIST...
criteria and their further development should be used [1], [4], [6]. To what extend the kinetics of tumour shrinkage is of importance to the outcome of a patient should be analysed in the in silico experiments and compared with the reality.

Although most patients with cancer respond to therapy, not all of these are cured. Even objective clinical responses to a given treatment do not translate into substantial improvements in overall survival. The reason for this phenomenon can be explained by the fact that therapies successfully eliminating the vast majority of cancer cells may be ineffective against rare, biological distinct cancer stem cells. Therefore new methods for assessing treatment efficacy have to be developed, as traditional response criteria, as the RECIST criteria, measure tumour bulk and do not reflect changes in the rare cancer stem cells [3]. If in silico experiments are of more help for a clinician than to get a prediction of changes in tumour volume and shape the response of treatment to the small fraction of cancer stem cells would be of utmost importance. From a clinical point of view, cancer stem cells have to be taken into account, in developing in silico models for tumour response. It might be time to eliminate traditional measures of clinical response as trial end points and to evaluate activity on rare cancer stem cells. It seems obvious that treatment effective against the gross majority of differentiated cancer cells are ineffective for rare cancer stem cells. This implies a change of treatment after a patient is in clinical remission by destroying or removal of the bulky tumour burden. In silico experiments should focus on this topic. Data on cancer stem cells for each tumour have to be created by molecular biologists and clinicians have to provide them with tumour material. This again underlies the importance of enrolling patients into clinicogenomic trials, if in silico experiments are done and conclusive results are awaited.

VII. WHAT IS THE BEST TREATMENT SCHEDULE FOR A PATIENT REGARDING DRUGS, SURGERY, IRRADIATION AND THEIR COMBINATION, DOSAGE, TIME SCHEDULE AND DURATION?

If in silico experiments aim to answer questions regarding the best available treatment for an individual patient and become part of the decision process in daily clinical practice, the treatment of individual patients is on a high level of individualisation. To attain this goal it is of utmost importance that the result of the in silico experiments are available in a short timeframe after diagnosis. This implies that all data that are necessary for running the in silico experiments have to be available in a short timely manner. This is especially important for molecular biologists and clinicians, who have to produce reliable data very fast. The in silico experiments itself should not be time consuming experiments.

VIII. IS IT POSSIBLE TO PREDICT SEVERE ADVERSE EVENTS (SAE) OF A TREATMENT AND TO PROPOSE AN ALTERNATIVE TREATMENT TO AVOID THEM WITHOUT DETERIORATING THE OUTCOME?

Chemotherapy might cause severe side effects in some patients and can even cause death. For a clinician is would be very helpful to know the individual risk of a patient regarding the occurrence of side effects. If they could be predicted treatment can be changed for an individual patient to reduce the risk of severe side effects. The in silico experiments should always provide the clinician with a result telling him what treatment will have the highest possibility of cure and the lowest risk of acute and late toxicity for an individual patient.

IX. IS IT POSSIBLE TO PREDICT A CANCER BEFORE OCCURRING AND TO RECOMMEND A TREATMENT, THAT WILL PREVENT THE OCCURRENCE OR A RECURRENT OF A CANCER IN AN INDIVIDUAL PATIENT?

We need to consider the possibility of earlier interventions in patients who have preneoplastic conditions and/or who are genetically predisposed to cancer development. There are tremendous opportunities to implement chemoprevention, dietary intervention, and lifestyle changes that could profoundly reduce the risk of cancer in these patients. With the help of in silico models such experiments can be run in defined cancer syndromes. It is the ultimate goal to prevent cancer or treat cancer before the disease is clinically overt.

X. CONCLUSIONS

Patients who take part in clinicogenomic trials may be helped personally by the treatment(s) they receive. They get up-to-date care from cancer experts, and they receive either a new treatment being tested or the best available standard treatment for their cancer. Of course, there is no guarantee that a new treatment being tested or a standard treatment will cure the patient. New treatments also may have unknown risks, but if a new treatment proves effective or more effective than standard treatment, study patients who receive it may be among the first to benefit. In Silico Oncology has to be tested now in the setting of clinicogenomic trials, to prove the expectations for getting better individualised cancer treatments with higher cure rates and less acute and late toxicity.

REFERENCES

Abstract- Optical imaging has emerged as a powerful tool for probing deep in tissues, where light propagates in a diffuse manner. While in most of the medical applications the source of contrast is inherent (e.g. tissue absorption or tissue scattering), recently a wide variety of fluorescent probes has been proposed as contrast agents for optical imaging studies 1. Fluorescence Molecular Tomography (FMT) incorporates the principles of optical tomography with the use of fluorescent probes as source of contrast 2, 3, in which case the subject that carries the fluorescent probe is exposed to light from different source positions or patterns and the emitted light is captured by detectors arranged in a spatially defined order. The information is then mathematically processed, resulting in a reconstructed tomographic image. Originally, the feasibility of this method was proven with NIR probes due to higher penetration in tissue of NIR light but in recent studies tomographic images of GFP expression from deep seated locations in vivo have also been presented 4.

The system used for performing Fluorescence Molecular Tomography is mainly composed of a CCD camera, a rotating stage, and a laser that scans over the mouse (see Fig 1).

Preliminary results of GFP-tagged expression of T-cells in the spleen are shown in Fig. 2. This image is an example of the the ability of the system to detect fluorescence signals in vivo from GFP-expressing T-cells deep in the body. The system and method described here are able to operate in different excitation and emission wavelengths using the appropriate filters, making feasible the detection of multiple fluorophores and fluorescing proteins, thus increasing the biological processes that could be targeted. Currently we are implementing algorithms for unmixing the fluorescence obtained from multiple fluorophores and reconstructing the fluorescence data.

In conclusion, the system presented could be used in vast number of different experimental models targeting different important biological processes and functions, as well as different types of diseases. These include developmental biology, cancer research, angiogenesis, immunology, drug development and others.

Fig. 1. Schematic representation of the FMT setup

Fig. 2. 3D reconstruction of GFP-tagged T-cells in the spleen. This figure reflects to the potential that FMT has on imaging biological processes in-vivo on the same subject.

Acknowledgements
This research was supported by E.U. Integrated Project "Molecular Imaging" LSHG-CT-2003-503259.

References
Fast Slice-to-slice 3D Method of Reconstruction for Optical Diffuse Mammography

Olga V. Kravtsenyuk, Vladimir V. Lyubimov, Alexander B. Konovalov, Igor I. Kutuzov

Abstract— We suggest a method that makes it possible to reduce the inverse problem of optical diffuse tomography (ODT) of strongly scattering media to the solution of the integral equation with integration along a curved photon average trajectory (PAT). Such an approach (PAT method) allows us to apply of standard fast algebraic algorithms as well as integral methods commonly used in the projection computer tomography, for the reconstruction of diffuse optical images. To demonstrate the capabilities of the PAT method, we consider the scheme for time-domain optical diffuse mammography, wherein sources and receivers lie on a conical surface. For slice-to-slice 3D reconstruction of absorbing inhomogeneities embedded in scattering media, we use the algebraic reconstruction technique and filtered backprojection algorithm. Numerical experiments show that the PAT method, which uses any of these algorithms, reconstructs diffuse optical images in real time and with good quality, and can be recommended for use in a computer programs specified for medical ODT applications.

Index Terms—optical diffuse tomography, reconstruction algorithm, trajectories.

I INTRODUCTION

OPTICAL diffuse tomography (ODT) is a very promising method for non-invasive clinical diagnostics providing both functional and metabolic imaging as well as structural one for different human organs and tissues. The principal goal of the medical DOT is the reconstruction of a spatial distribution of optical parameters of tissue from a set of data obtained during illumination of an object by a near-infrared radiation.

The main problem of medical DOT is connected with the dominance of multiple scattering, which causes light to propagate diffusely in a tissue. As a result there are no regular photon trajectories of the light migrating from a source to a receiver. In this case the multi-step algorithms based on exact models of light propagation through tissue are used [1]. These algorithms give a good resolution, but are not sufficiently fast for real time clinical explorations.

To save reconstruction time, we develop new DOT method based on the concept of the photon average trajectory (PAT) [2-4]. The essence of the PAT concept is in representing the process of the photon energy transport from a source to a receiver in a form admitting probabilistic interpretation. The opportunity for this approach comes from the physical nature of the problem. Indeed, if one fixes an instantaneous spatial distribution of photon density within the volume of the object, the input of each volume point to the signal registered will be a stochastic value. New method called by us the photon average trajectories method (PAT method) allows the inverse problem of DOT to be reduced to the solution of the integral equation with integration along a curved PAT. Therefore, it opens the opportunity to apply for reconstruction the fast trajectory algorithms traditionally used in projection tomography.

This paper shows the capability of algorithm based on the PAT method to reconstruct of optical inhomogeneities embedded in strongly scattering 3D object of a truncate cone shape. Such geometry is quite suitable for the case of optical diffuse mammography with sources and receivers located on a conical surface of an applicator, in which female breast is placed.

II RECONSTRUCTION PROBLEM STATEMENT

The problem of slice-to-slice 3D reconstruction using the PAT method has important specific feature. PATs connecting the sources and the receivers of one cross slice can not be located in a single plane and are turned out to be scattered in a relatively thick 3D layer. In such situation a 3D reconstruction approach is needed to reconstruct the internal structure of each slice.

To simplify the implementation of the PAT method, the statistical characteristics of photon trajectories are approximated by convenient functions. For an example, PAT with the good accuracy may be approximated by three step polygonal line [2]. The first and the last segments of such approximation have equal lengths and are normal to the object boundary in the source and receiver points correspondingly. If the length of the end segments for all PATs approximations of one slice is constant and greatly less than the length of the middle segments, the broken lines form a thin “dish”. To reconstruct a spatial area corresponding to...
the middle segments (a “dish” bottom), the standard algorithms for 2D reconstruction may be used. By reconstruction of each “dish” one can get the 3D image from summation of 2D reconstructions for all slices.

III RESULTS AND CONCLUSION

To demonstrate the capabilities of the PAT method, we carried out a numerical experiment on slice-to-slice 3D reconstruction of absorbing inhomogeneities embedded in scattering object in the form of a truncate cone. The height of object was equal to 7 cm, the diameters of cone bases are equal to 12 and 6 cm. The cone contains three absorbing inhomogeneities in the form of an ellipsoid, a cylinder, and a parallelepiped (Fig. 1). Semi-axes of the ellipsoid were 1, 1, and 1.5 cm. Semi-axes of the cylinder and its height were equal to 1 cm. Sides of the parallelepiped were also equal to 1 cm. Refraction index, diffusion and absorption coefficients were equal to 1.4, 0.0636 cm, and 0.042 cm\(^{-1}\), correspondingly. Absorption coefficient for the inhomogeneities was 0.17 cm\(^{-1}\). The number of the sources and the receivers were selected to be 16 for each slice. The sources and the receivers were located along a perimeter of the object cross-sections at equal step angles and alternated with each other so that the step angle between the nearest-neighbor source and receiver constituted 5.625\(^\circ\). The total number of slices was 35. The relative shadows caused by inhomogeneities were simulated via numerical solution of time-dependent diffusion equation for the instantaneous point source by the finite-element method. Reconstruction of each slice was realized onto basis 44×44 with the use of two algorithms: algebraic reconstruction technique (ART) \[3\] and filtered backprojection algorithm (FBP) \[4\]. The results of reconstruction of the object shown in the Fig 1 are presented in Fig. 2. The 3D images with localization of secant planes are given at the left of Fig. 2. At the right one can see the cross-sections corresponding to these planes. For a common personal computer (Pentium4-1700, 256-mb RAM) the time of reconstruction of one slice by ART and by FBP was 3.5 s and 0.5 s correspondingly.

It is obvious from Fig. 2 that inhomogeneities reconstructed by both algorithms are resolved perfectly clear without visible noise. Moreover, in the images one can recognized real geometrical contours of inhomogeneities: ellipsoid, cylinder, and parallelepiped. Using FBP it takes us around 18 s to obtain of 3D reconstruction that seems adequate for processing of images in real time operation mode. Thus, the PAT method can be recommended for medical ODT applications, for example, 3D optical mammography.

ACKNOWLEDGMENT

The authors thank their colleagues O. V. Lyamtsev, L. N. Soms, L. M. Yavorskaya, A. G. Kalintsev, A. G. Murzin, O. V. Golubkina, G. B. Mordvinov, and K. B. Domrachev for fruitful and helpful discussions during the research preparation.

REFERENCES

Multicolor Photon Density Wave Setup for Breast Cancer Diagnostics

Olga V. Kravtzenyuk, Ilya V. Turchin, Ekaterina A. Sergeeva, Vladimir I. Plekhanov and Vladislav A. Kamensky

Abstract—An experimental setup for multicolor frequency-domain optical diffuse tomography was designed to visualize soft biological tissue inhomogeneities of up to 5 mm in size at the distance of up to 6 cm. Scanning is performed by independent electronically controlled shift of source and detector placed in transmission mode. Employing illumination at three wavelengths (684, 794, and 1060 nm) which correspond to predominating absorption of oxygenated hemoglobin, deoxygenated hemoglobin and water provides determination of component composition of an inhomogeneity. Source power modulation at 140 MHz increases spatial resolution (compared to CW imaging) and improves quality of reconstruction procedure. Studies on model media and preliminary in vivo experiments were performed.

I INTRODUCTION

Experimental visualization of internal structure of biological tissue is limited by strong scattering of light; therefore, tissue acts as non-transparent medium in visible and near IR. Special techniques for turbid media imaging such as Optical Diffusion Tomography (ODT) [1-4] combined with reconstruction procedure allow for non-invasive recognition of biological tissue inhomogeneities.

ODT is based on acquiring information from multiply scattered light which penetrates into the tissue up to depths of several centimeters. This technique allows for imaging absorbing and scattering inclusions inside tissue and distinguishing between them after computer processing of an image. Typically ODT operates in transmission mode with spatially separated source and detector when an image is formed by light that is diffusively transmitted through scattering medium. Like in any transmission technique, the goal of ODT is to reconstruct distribution of scattering and absorption by a measured set of integrals over trajectories. Strong scattering blurs sharp edges of an inhomogeneity thus reducing spatial resolution down to 0.5 cm. Consequently, only large tumors can sized up properly. Frequency-Domain Optical Diffusion Tomography (FD ODT) employs illumination of the tissue by amplitude-modulated light [3,4]. This method significantly increases spatial resolution of the observed objects inside turbid media due to registration and processing of the amplitude and phase of signal envelope which both carry information about the tissue scattering and absorption inclusions. Some encouraging results were recently obtained by ODT technique in clinic for early diagnosis of deeply located tumors (mammary gland) and pathologic changes in cerebrum brain blood supply [5]. In the transparency window ($\lambda = 600$–1100 nm) absorption in biotissues is determined primarily by four components [6,7]: lipids, water, oxygenated hemoglobin and deoxygenated hemoglobin. The degree of blood oxygenation and content of water and lipid content are closely related to the tissue physiological features. In the present paper we report on recent experiments performed with the ODT device created at the Institute of Applied Physics (Russia). This device uses in parallel sources at three wavelengths, 684 nm, 794 nm and 1060 nm. The reason for choosing these wavelengths is the fact that at the vicinity of 700 nm deoxyhemoglobin makes the main contribution to absorption, wavelength of 800 nm corresponds to equal absorption of deoxyhemoglobin and oxyhemoglobin, and water and lipids are dominating absorbers at 1000 nm. As was shown in [3], mathematical processing of the images obtained at multiple wavelengths allows for determination of component composition of biotissue (oxy-, deoxyhemoglobin, adipose tissue).

II EXPERIMENTAL SET UP

Experiments on FD ODT were accomplished on the experimental setup created at IAP RAS (Fig. 1). The object of interest is placed between two parallel planes, source plane and receiving plane. Three laser fibers coupled in a single bundle with amplitude modulation at frequency of 140 MHz illuminate investigated volume at wavelengths 684 nm, 794 nm and 1060 nm. Independent scanning of source and detector in corresponding planes is performed by computer controlled stepping motors, scanning area is 15x15 square centimeters. Data reading at three wavelengths is realized by automated sequential switching from one laser to another at each source-detector position. Sensitivity of the receiving system is $2.4 \times 10^{-5}$ mW. Calculation of signal amplitude and phase is done by means of a computer.
The described FD ODT setup was applied for in vivo imaging of breast tissue. In the Fig. 2 images obtained from 49-year-old patient are shown. The image was acquired by simultaneous scanning of source and detector facing each other, with fixed source-detector separation (60 mm). In order to fit space between source and detector planes breast tissue was slightly compressed. The whole image is 35x35 pixels with a 2-mm pixel size. Amplitude and phase of the transmitted signal obtained at modulation frequency of 140 MHz are computed and presented in the figure.

Dark spot observed in amplitude images taken at 684 nm and 794 nm may correspond to a region of higher absorption, probably tumor, since it does not appear at 1060 nm where blood absorption is low. This inhomogeneity can be seen in a phase image at 684 nm as well. However, the sensitivity of the receiving system is not sufficient to visualize the inclusion in phase image at 794 nm. At present, the setup is being updated in order to observe smaller inhomogeneities with better contrast. Computer processing of obtained images aimed to reconstruct profiles of scattering and absorption is on the way as well.

**IV CONCLUSION**

The described multicolor FD ODT device allows for detailed investigation of breast tissue. Illumination by multiple wavelengths provides determination of component composition of an inhomogeneity under appropriate image processing. However, insufficient sensitivity of the created system limits imaging depth by the value of 6 cm. We expect to increase imaging depth up to 8-9 cm by employing detectors with higher sensitivity (for example, cooled photomultiplier tube).

**ACKNOWLEDGMENT**

The authors are grateful to the Nizhny Novgorod State Medical Academy, Regional Hospital and Oncological Department for providing access to clinical material.

**REFERENCES**


Abstract— In-silico modelling of human anatomy and physiology should not only provide a realistic model of a “typical” human being, but it should also account for the variability in shape and function across the modelled population. Methods based on multivariate statistical analysis of a training set of data corresponding to a subset of the population have been proposed and are extensively used for medical image analysis. In this presentation, we will describe our work on modelling the shape of human bones, with an application to computer-assisted guidance on orthopaedic surgical procedures. We will show that it is possible to provide acceptable approximations of the 3D shape of bones without any pre-operative CT scan, and with minimal data acquired intra-operatively by means of a tracked pointer or an ultrasound probe. We will then show our work on modelling not only shape, but also bone quality, based on statistical models of CT images. Finally, we will show that it is possible to build statistical models of the biomechanical behaviour of anatomical structures, illustrated by the establishment of a statistical distribution of bone strength across the population. This will then be applied to the optimisation of the shape of orthopaedic implants, as to provide optimal anatomical fitting and biomechanical stability.

Fig. 1 Illustration of the concept of statistical modelling of anatomical variability. The aim is to establish a probability function representing the prevalence of particular anatomical features in a population.

Fig. 2 From a set of example shapes acquired by segmenting CT scans, it is possible to compute the average shape and a structured description of the main patterns of shape variability in the population.

Fig. 3 The same concept applies to images, so it is possible to compute the “average” image and the main patterns of variation.

Fig. 4 By using calibrated CT data sets, we are able to build a model of bone shape and quality at each location inside the bone.

M. A. Gonzalez Ballester is the Director of the Surgical Technologies Division and Head of the Medical Image Analysis Group at MEM Research Center, University of Bern.
Fig. 5 Finite element analysis is applied to compute the stress / strain at each location of the bone, based on the bone quality. Samples are drawn from the distribution to generate new “virtual bones”. The experiment in the illustration corresponds to a vertical load of approx. 2 times body weight, applied to the mean (centre) and ±2 std from the shape distribution.

REFERENCES


Optimized MR Imaging methodology for tumor characterization

F. Zacharopoulou, K. Marias, E. Georgiadi, G. Tollis and T. G. Maris
(1) Medical Physics Department, University of Crete, Faculty of Medicine, Heraklion, Greece,
(2) Institute of Computer science (ICS), Foundation for Research & Technology (FORTH), Crete, Greece

Abstract - In Silico oncology is anticipated to provide models that are capable of predicting the tumor growth and its response to the treatment. Therefore, this methodology is expected to become a valuable means for the clinicians to apply more individualized cancer treatments for their patients with increased probabilities of survival. However, every in silico method in order to be reliable has to be compared and validated by real clinical cases. The clinical data of interest should provide imaging data prior and post tumor treatment. In order to validate the proposed models it is crucial to be able to determine the tumor size and shape with high resolution, accuracy and precision, within a level of confidence, utilizing clinical image data. One of the imaging techniques used for tumor characterization is Magnetic Resonance Imaging (MRI). MRI can be used to characterize the size of the tumor and its microvasculature, providing information about tumor microvessel structure and function (1-5). MRI techniques can be divided into non-enhanced and contrast media-enhanced methods (6). Dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) is in most cases applied by the clinicians in order to decide upon the state of the tumor. This study proposes a methodology that can be applied along with the standard clinical practice associated with DCE. The aim of the proposed methodology is to optimize the imaging technique in order to provide more accurate and precise tumor information, thus the validation of in silico models can be performed within a higher level of confidence.

Introduction
Dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI)
Perfusion is a complex process that encompasses the delivery of nutrients (primarily oxygen and glucose), their diffusion and convection (if lymphatics are functional) into the tumor parenchyma, and the removal of waste products. Therefore, any perfusion information is crucial for characterizing the tumor microenvironment (Figure 1). The perfusion of tumors can be interrogated with time-dependent delivery of exogenous contrast, using contrast agents. These studies show a spatio-temporal heterogeneity of perfusion in tumors that is consistent with the chaotic vascular architecture.

Dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) is performed after the administration of intravenous contrast medium to noninvasively access tumor vascular characteristics. The success of this technique depends on its ability to demonstrate quantitative differences of contrast medium behaviour in a variety of tissues. Kinetic parameters can be correlated with immunohistochemical surrogates of tumor angiogenesis, including microvessel density, and with pathologic tumor grade. DCE-MRI is being applied to monitor the clinical effectiveness of a variety of treatments, including antiangiogenic drugs. Kinetic parameter changes following treatment have correlated with histopathological outcome and patient survival.

Other MRI techniques
All MRI techniques are based on the effect of the molecular motion on the MR related physical parameters (Figure 2). This effect results in discrimination between different types of human tissue according to their measured physical parameters (Figure 3). MRI techniques can actually be divided into two major categories: a) conventional techniques, providing Proton Density (PD), T1, T2, T2* weighted images and b) quantitative MR imaging techniques (qMRI) providing parametric image maps utilizing MR related physical parameters, such as PD, T1, T2, T2*. In addition, qMRI can calculate perfusion index maps utilizing contrast enhancement methods. This method is valuable when no DCE-MRI image data can be acquired.
Proposed MRI methodology

It has been observed that tumor tissue ‘liquifies’ after treatment due to destruction of the protein structure. qMRI techniques can actually quantify this alteration of the tissue structure by measuring its physical parameters (Figure 3, 4). In this study it is proposed that qMRI techniques should be included in the clinical protocols associated with the validation of in silico models. The alteration of the tissue structure, detected by MRI, could be an insightful parameter to trace the tumor especially in areas where DCE technique is unreliable (the periphery of the tumor). It is suggested that the collaboration DCE-MRI and parametric qMRI can provide more accurate and precise tumor information, thus the validation of in silico models can be performed within a higher level confidence.

In addition, when no DCE-MRI image data could be acquired a standard CE-MRI should be used to trace the tumor. However, this technique can be misleading on several conditions. It is common knowledge that the contrast media perfuse over time in tissue. There is no evidence that after a specific time interval the contrast agent perfuse by tumor tissue exclusively. Thus, a ‘perfusion index’ parametric slope map (Figure 5) is proposed in order to discriminate between the hyper vascularised and vascolarised tumor tissue.

References

Abstract— Modern advances in Medicine and Biology promise more efficient ways to optimally extract, analyse and describe biomedical information at all scales (from gene to organ). Recent research trends focus on how this information can be modelled and transformed into knowledge, in order to lead to a less interfering but also more individualised diagnosis and therapy. However, the integration of models from different levels remains an open issue. This paper addresses an important facet of this problem; the challenge of efficient and accurate extraction of biomedical information from different levels (from molecular to tissue/organ) in order to shed light in the computation of 4D maps of pathophysiological properties of tumours for model development and validation.

IV. INTRODUCTION

Multi-level data analysis and information extraction are necessary steps for developing and validating mathematical models of human physiology and pathology. Despite the recent advances in medical, molecular/genetic imaging, the robust extraction of physiological parameters remains an open issue since computing them from measurements (e.g. pixel values), isn’t a trivial task. Although, pathophysiological measurements are essential for modelling human processes; the physics of imaging modalities (e.g. MRI) does not absolutely generalise. It is the physics of MRI electromagnetism or the ionising radiation in CT that produces a representation of anatomy and physiology for interpretation. Also, the mechanism by which the image is formed inherently emphasises certain aspects and provides less sensitivity to others.

In addition, traditional diagnostic imaging, no matter how sophisticated, is only a representation of the underlying anatomical and (patho)physiological characteristics of disease. Even the most sophisticated MRI and CT techniques can only achieve an order of magnitude below millimetre spatial resolution [1,2]. Therefore the information captured in the image is only an average of what is contained within the sampling range of the instrument. Given the local complexity and “multi-scale” nature of human physiology, any attempt to use imaging to capture detailed physiological processes will understandably generalise the underlying phenomena. Thus, care must be taken when interpreting any kind of physiological parameter from imaging – that is not to say that image-based quantification of biological processes is not valid or useful, rather that the interpretation must give credence to the macroscopic scale at which phenomena is observed. In the case where diagnostic imaging is used to assess change in the conditions of a physiological mechanism, care must be taken to ensure that changes captured by the imaging technique are indeed changes reflective of the biology. In practical terms, this means a sensible clinical protocol and a better understanding of the imaging physics. This is particularly true in CE MRI analysis where the choice of a different analysis models may lead to a completely different result regarding the true spatial extent of a cancer.

V. PROBLEMS RELATED TO MULTI-LEVEL PHYSIOLOGICAL INFORMATION EXTRACTION

The above considerations are very important for the development of computational frameworks for multi-level (from molecular/genetic to tissue/organ) modelling and simulation of human pathophysiology. Multi-level measurements (e.g. molecular/genetic, diagnostic imaging, etc.), should be properly interpreted and combined in order to be used in simulation models of human function. Additionally, this process needs to be repeated in several time instances in order to assess the validity of each model temporally. This concept is schematically illustrated in Fig. 1 where information extraction of temporal multi-level data is driven into the corresponding multi-level model of breast cancer. To reach this goal, it is essential to be able to optimally extract robust temporal pathophysiological information from multi-level measurements (e.g. microarrays, CE MRI, etc.).
Medical Imaging has focused in providing anatomical information, mainly imaging human bones, dense tissue and arteries. Recent advances especially in PET and functional MRI allowed the study of various pathological processes via radio-labelled tracers (PET) or pharmaco-kinetic models in contrast enhanced MRI. The whole field of molecular medicine and molecular imaging is opening up new possibilities for targeted assessment of disease and disease mechanisms. Also, microarray imaging has created exiting possibilities for defining new disease biomarkers. Some of the most common problems related to extracting information in the abovementioned techniques can be classified in the following categories:

- Geometrical normalisation. It is necessary to ensure that a point correspondence is computed between different images. This is a typical problem in the case of breast imaging owing to the differences in breast shape/compression but also in newer applications such as molecular imaging and microarray imaging. Non-rigid alignment or registration is required to compensate for such differences. This problem does not only pertain to the multi-modal scenario, as a temporal acquisition of the same modality will still likely involve registration in order to facilitate comparison. Several registration frameworks have been proposed (see [3-5] for selected publications in the field), traditionally for medical imaging applications but more recently also for correcting time dependent geometries in 2D molecular optical imaging studies [6].

- Extraction of relevant information. Essentially, this implies the use of some kind of process (e.g. segmentation) to identify important structures and features in the images (e.g. tumours can be segmented using a pharmacokinetic model of gadolinium uptake with contrast-enhanced MRI, while microarray spots can be segmented by combining the two different information channels i.e. Cy3 and Cy 5 [7]). Another interesting application is the extraction of dense tissue from mammograms in order to assess density changes due to therapy [8]. Last, in CE MRI the dynamic behaviour of different structures within the breast can be monitored (functional imaging). In particular, malignant tumours exhibit an increased vascularity, since they begin to grow their own blood supply network. For this reason when the contrast agent is distributed, malignant masses enhance faster. This led to the development of models of contrast uptake as is illustrated in Figure 2.

- Quantification. Each imaging modality produces a representation that has different parametric properties. For example, the active volume of a tumour detected with X-ray mammography is less accurate than that using with MRI. Quantification is particularly important in microarray experiments for defining the actual quantitative relationship of two information channels (e.g., healthy tissue DNA labelled with Cy3 and a diseased tissue DNA labelled with Cy5). This way, differential expression of genes can be detected on the basis of ‘differing’ from the overall trend that characterizes the relationship of gene expressions in two tissues.

- Intensity normalization. As stated above each biomedical measurement in essence disguises the true physiological property due to the image formation process. The non-linearities introduced by varying imaging conditions may alter significantly the image-intensity profile and reduce the efficiency of generic analysis algorithms. An interesting example is the model of Highnam and Brady [9] for mammogram image normalisation that eliminates variations related to imaging conditions (e.g. tube voltage, time of exposure, etc). Highnam and Brady’s method estimates – in millimetres – the amount of interesting tissue in each pixel column, as illustrated in Figure 3. This effectively provides objective quantitative information about the breast anatomy. If, for example, the separation between the Lucite plates is 6.5cm, the amount of interesting tissue at a location (x,y) might be 4.75cm, implying 1.75cm of fat. This way, the algorithm estimates, and then eliminates the effects of, the particular parameters that were used to form the mammographic image providing true anatomical information.

Figure 1: Extraction of temporal pathophysiological information is essential for developing and validating multi-level models.

Figure 2: A two-compartment pharmacokinetic model with typical contrast curves for fat, parenchymal (glandular) tissue and enhancing regions of interest. $M_d$ is the mass of contrast injected into the blood stream with respect to time, $k_{12}$ and $k_{21}$ are inter-compartment exchange rates and $k_{out}$ is the leaving contrast rate.
This problem also exists in microarray imaging technologies where several non-linearities in the experimental process render the measured expression values prone to variability and often, to poor reproducibility. To achieve normalization, one has to adjust the sensitivity of detection (photomultiplier voltage with fluorescence or exposure time with radioactivity) so that the measurements occupy the same dynamic range in the detector and exploit the fact that the gene expression values (e.g. from the Cy3 and Cy5 matrices), should ideally follow a linear trend [10]. The later can be performed to all the genes or to a ‘ground truth’ subset that is known a priori to be the same in both channels (Cy3 and Cy5).

- Visualisation. This is a fundamental aspect of biomedical data information fusion that is typically less well addressed in the literature, but which can dramatically increase the clinical utility of a solution if implemented intelligently. The effectiveness of visualisation depends very strongly on how clearly different indicators can be extracted from data and therefore segmentation is of utmost importance. A great deal of effort has been made in this research to produce visualisations of the results of temporal and multi-modal image fusion that optimises the presentation of available clinical information (an example is illustrated in Figure 4).

VI. CONCLUSIONS

In order to approach the vision of the Virtual Physiological Human, it will be essential to develop and validate individualized, multi-level models taking into consideration pathophysiological information at all scales. As discussed in this paper, the extraction of useful anatomical and physiological information from biomedical measurements isn’t a trivial task due to the complex physical interactions involved in each acquisition as well as several systematic and random errors involved in the process. In many cases, the problems that arise in different scales are common (e.g. geometrical inconsistencies over time) and is therefore important to develop generic tools for multi-scale temporal analysis in order to robustly extract and visualize pathophysiological information over time. Such information is crucial for initializing (i.e. in the case of in silico models of cancer, 3D voxels should be classified as ‘proliferating’, ‘necrotic’, etc.), inspiring and validating 4D models of human function.

REFERENCES