



## **Deliverable No. 2.3**

### **Requirements for enhancing hypermodels beyond the domain of cancer**

|                              |  |
|------------------------------|--|
| Grant Agreement No.:         | 600841   |
| Deliverable No.:             | D2.3   |
| Deliverable Name:            | Requirements for enhancing hypermodels beyond the domain of cancer |
| Contractual Submission Date: | 30/09/2014   |
| Actual Submission Date:      | 01/12/2014   |



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

| <b>COVER AND CONTROL PAGE OF DOCUMENT</b>         |   |
|---|---|
| Project Acronym:                                  | <b>CHIC</b>   |
| Project Full Name:                                | Computational Horizons In Cancer (CHIC): Developing Meta- and Hyper-Multiscale Models and Repositories for In Silico Oncology |
| Deliverable No.:                                  | D2.3  |
| Document name:                                    | Requirements for enhancing hypermodels beyond the domain of cancer  |
| Nature (R, P, D, O) <sup>1</sup>                  | R   |
| Dissemination Level (PU, PP, RE, CO) <sup>2</sup> | CO  |
| Version:  | 1.0   |
| Actual Submission Date:                           | 01/12/2014  |
| Editor:   | Quoc-Chinh Bui  |
| Institution:                                      | Philips Research  |
| E-Mail:   | quoc-chinh.bui@philips.com  |

#### ABSTRACT:

In this deliverable we describe the requirements for enhancing hypermodels beyond the domain of cancer. Since the CHIC platform consists of multiple levels ranging from fundamental levels such as IT infrastructure and security framework to domain specific levels, there are many aspects need to be taken into account in order to reuse or extend hypermodels for other domains. However, as fundamental levels are independent from the hypermodels, we only focus on requirements that directly relate to the hypermodels which are biological models, hypermodelling infrastructure, ethical and legal framework, and validation of hypermodels in clinical trials.

#### KEYWORD LIST:

Requirements, model, hypermodels, infrastructure, multiscale, interactions, information, data, deployment, technical resources, legal framework, clinical trials.

<sup>1</sup> R=Report, P=Prototype, D=Demonstrator, O=Other

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

*The research leading to these results has received funding from the European Union Seventh Framework Programme under grant agreement no 600841.*

*The author is solely responsible for its content, it does not represent the opinion of the European Community and the Community is not responsible for any use that might be made of data appearing therein.*

| <b>MODIFICATION CONTROL</b> |             |               |   |
|-----------------------------|-------------|---------------|---|
| <b>Version</b>              | <b>Date</b> | <b>Status</b> | <b>Author</b>   |
| 0.1                         | 25/09/2014  | Draft         | Chinh Bui, Philips Research   |
| 0.2                         | 28/10/2014  | Draft         | Daniele Tartarini, Dawn C. Walker, Marco Viceconti, UFD   |
| 0.3                         | 29/10/2014  | Draft         | Roman Niklaus, Michael Kistler, Philippe Büchler, UBERN   |
| 0.4                         | 29/10/2014  | Draft         | Mauricio Reyes, UBERN   |
| 0.5                         | 30/10/2014  | Draft         | Fay Misichroni, Georgios Stamatakos, ICCS   |
| 0.6                         | 30/10/2014  | Draft         | Norbert Graf, USAAR   |
| 0.7                         | 07/11/2014  | Draft         | Marc Stauch, Iryna Lishchuk, Iheanyi Nwankwo, LUH   |
| 0.8                         | 26/11/2014  | Draft         | Eleftherios Ouzounoglou, Katerina Argyri, Eleni Kolokotroni, Eleni Georgiadi, Markos Antonopoulos, Dimitra Dionysiou, Georgios Stamatakos, ICCS<br>Caterina Guiot, Ilaria Stura, UNITO<br>Philippe Buechler, UBERN<br>Ravi Radhakrishnan, UPENN |
| 0.9                         | 28/11/2014  | Pre-Final     | Chinh Bui, Philips Research   |
| 1.0                         | 1/12/2014   | Final         | Fay Misichroni, ICCS  |

#### List of contributors

- Chinh Bui, PHILIPS
- Anca Bucur, PHILIPS
- Norbert Graf, USAAR
- Mauricio Reyes, UBERN
- Roman Niklaus, UBERN
- Michael Kistler, UBERN
- Philippe Büchler, UBERN

- Daniele Tartarini, USFD
- Dawn C. Walker, USFD
- Marco Viceconti, UFD
- Marc Stauch, LUH
- Iryna Lishchuk, LUH
- Iheanyi Nwankwo, LUH
- Eleftherios Ouzounoglou, ICCS
- Katerina Argyri, ICCS
- Eleni Kolokotroni, ICCS
- Eleni Georgiadi, ICCS
- Markos Antonopoulos, ICCS
- Dimitra Dionysiou, ICCS
- Georgios Stamatakis, ICCS
- Fay Misichroni, ICCS
- Caterina Guiot, UNITO
- Ilaria Stura, UNITO
- Ravi Radhakrishnan, UPENN

## Contents

|       |   |    |
|-------|---|----|
| 1     | EXECUTIVE SUMMARY.....  | 8  |
| 2     | INTRODUCTION .....  | 9  |
| 2.1   | OVERVIEW .....  | 9  |
| 2.2   | REQUIREMENTS FOR ENHANCING HYPERMODELS FOR OTHER DOMAINS .....                                      | 12 |
| 2.2.1 | <i>Requirements for biological models</i> .....   | 12 |
| 2.2.2 | <i>Requirements for IT infrastructure</i> .....   | 12 |
| 2.2.3 | <i>Requirements for ethical and legal framework</i> .....   | 13 |
| 2.2.4 | <i>Validation of hypermodels in clinical trials</i> .....   | 13 |
| 3     | REQUIREMENTS FOR BIOLOGICAL MODELS.....   | 14 |
| 3.1   | BIOMECHANICAL MODELLING.....  | 15 |
| 3.2   | ATOMIC AND MOLECULAR LEVEL.....   | 16 |
| 3.3   | SPECIFIC EXAMPLES .....   | 18 |
| 3.3.1 | <i>Angiogenesis models</i> .....  | 18 |
| 3.3.2 | <i>Tumour response to immunotherapy models</i> .....  | 20 |
| 3.3.3 | <i>Phenomenological Universalities (PUN) approach applied in different biological context</i> ..... | 22 |
| 3.3.4 | <i>Nephroblastoma tumour growth and response to chemotherapy model</i> .....                        | 23 |
| 3.3.5 | <i>Acute lymphoblastic leukemia evolution and response to therapy model</i> .....                   | 24 |
| 4     | REQUIREMENTS FOR IT INFRASTRUCTURE .....  | 26 |
| 4.1   | IMAGE PROCESSING AND VISUALISATION .....  | 26 |
| 4.2   | HYPERMODELLING INFRASTRUCTURE .....   | 28 |
| 4.3   | MODEL REPOSITORY, IN SILICO TRIAL REPOSITORY AND CLINICAL DATA REPOSITORY.....                      | 28 |
| 4.3.1 | <i>Model repository and in silico trial repository</i> .....  | 28 |
| 4.3.2 | <i>Clinical Data Repository</i> .....   | 31 |
| 4.3.3 | <i>Potential exploitation of clinical data repository</i> .....                                     | 33 |
| 5     | REQUIREMENTS FOR ETHICAL AND LEGAL FRAMEWORK.....   | 35 |
| 6     | VALIDATION OF HYPERMODELS IN CLINICAL TRIALS .....  | 39 |
| 7     | REFERENCES .....  | 43 |

## 1 Executive Summary

The main objective of this deliverable is to describe the requirements for enhancing hypermodels beyond the domain of cancer. First we start with an overview of the CHIC platform, which consists of multiple levels ranging from fundamental levels such as IT infrastructure and security framework to domain specific levels, that supports accessibility, reusability and extensibility of VPH mathematical and computational hypermodels. Next we describe general requirements that need to be taken into account in order to reuse or extend hypermodels for other domains. As fundamental levels are independent from the hypermodels, we only focus on requirements that directly relate to the hypermodels which are biological models, hypermodelling infrastructure, ethical and legal framework, and validation of hypermodels in clinical trials.



## 2 Introduction

### 2.1 Overview

In this document we describe requirements for enhancing hypermodels beyond the cancer domain. The document starts with a landscape of hypermodels by recalling some general concepts and requirements for hypermodels documented in the deliverable D2.2 “Scenario based user needs and requirements”. As described in the deliverable D2.2, the hypermodels are built on three layers (see Fig. 2.1):

- A multiscale biology layer
- A engineering layer
- A software layer

The multiscale biology layer forms the biological basis for the models and hypermodels starting from the molecular level, to the cellular, to the tissue, to the organ, to the body, and up to the epidemiological level taking into account the biological time constraints. The engineering layer is concerned with the interaction between different models. All models or components are simple and basic. Based these models or components, more complex hypermodels are built. The software layer is where the models are programmed.

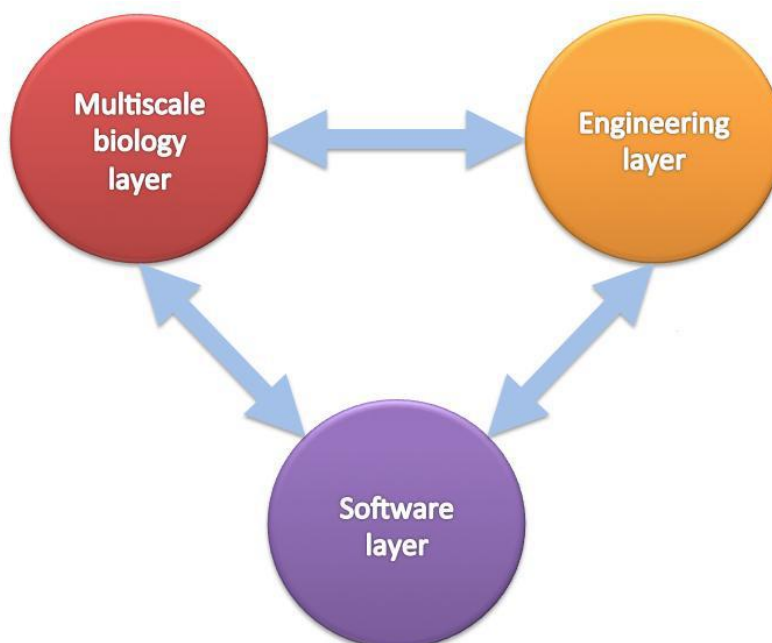


Figure 2-1. Interactions between different layers

The implementation requirements proposed below are designed to assure the development of a functional and state-of-the art system:

- Software as a Service (SaaS)
- Interoperability
- Reusability
- Modularity
- Security and granular access for end users
- Social networking frames

The most important for the architecture is the modularity of the system. All developed software, tools and services should be as granular and modular as possible and provide standardized, open interfaces and functionality descriptions so that a user can easily build new models as a composition of existing granular tools. For example, one needs only to develop a tool that links gene expression data of a tumour with the KEGG database. If this tool is as generic as possible and the interface between the gene expression data and the KEGG database is standardized then this tool will be able to use in different settings and models, independence of the underlying tumour or disease. It is important that for each granular tool, a standardized interface to data be defined so that the different research groups can establish what preconditions are needed to run such composed models.

The development of the architecture for CHIC needs to take interoperability issues into account. This also includes semantic interoperability of data. A model will process input data to produce a result. Such a result with its output data can be used as input data for another model thus going from simple models to hypermodels. For that purpose, we categorize models into four different levels as already proposed in p-medicine:

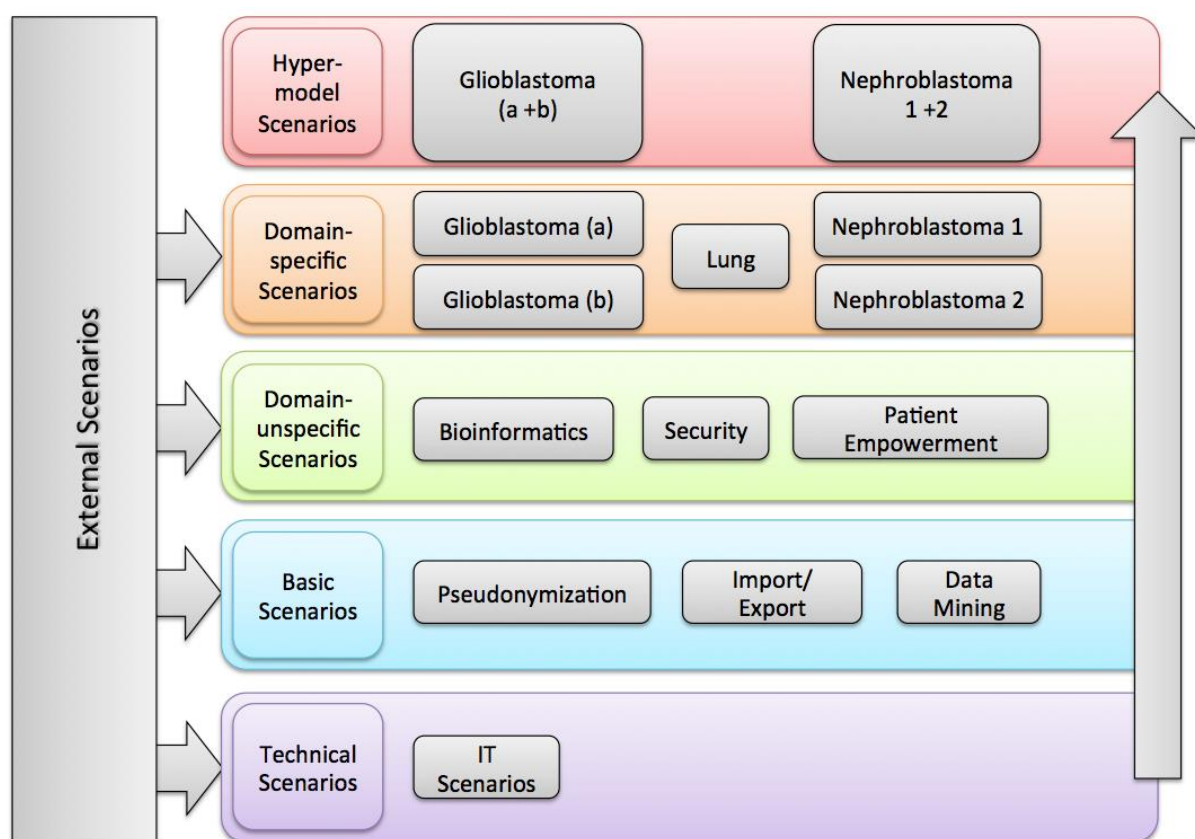
1. Level for fundamental tools
2. Level for basic tools
3. Level for modular tools
4. Level for domain specific tools

Fundamental tools are those that are fundamental for the architecture. This level includes mainly IT-tools that can be used in all models. The basic level will contain only such tools that are domain and scenario unspecific, e.g. pseudonymization tool, curation tool for data. Fundamental and basic tools can be re-used in different scenarios and domains. Modular tools are scenario specific but not domain specific, e.g. a tool for patient empowerment. At a higher level, tools, models and services are domain specific. For the purpose interoperability and standardization it is crucial to avoid building each tool repeatedly from scratch.

According to the classification of tools, scenarios will be classified in the following levels (Fig. 2.2):

1. Technical scenarios
2. Basic scenarios
3. Domain unspecific scenarios
4. Domain specific scenarios
5. Hypermodel scenarios

Technical scenarios are part of other scenarios including basic scenarios. Domain unspecific scenarios can be composed of basic scenarios and will be able to be used in domain specific scenarios. By doing so, scenarios do not need to be developed from scratch and it will foster the development of standards and interoperability. If interoperability and standards are taken into account then the integration of external scenarios into the CHIC framework is possible.



**Figure 2-2. Levels of scenarios giving examples on all levels. Interoperability and standards will allow the integration of external scenarios**

*The construction of such a modular architecture, with tools and models categorized into different levels, is a major factor contributing to the sustainability of the architecture beyond the project domain.*

Of high importance for the architecture of the CHIC platform is to facilitate data exchange with other health care systems in accordance with the legal framework. Otherwise it will become an “information island”, isolated from other information about the patients, with limited access and value. As a result, the CHIC platform should interoperate with other systems throughout the entire health and clinical studies information environment.

At a minimum, CHIC should export anonymized data to, and import-export data from other systems in a standardized (and interoperable) way. To provide interoperability, the CHIC platform should support from its inception communications, messaging, and content encoding standards as other health information systems (HIS) like HER.

At the basic level, IT tools are part of all hypermodels. These include tools for setting up an IT infrastructure, to store and share heterogeneous data and to process these data. In this setting tools for security and interoperability are also needed.

## ***2.2 Requirements for enhancing hypermodels for other domains***

Developing robust, reproducible, interoperable and collaborative hypermodels of diseases and normal physiology is necessity if rational, coherent and comprehensive exploitation of the invaluable information hidden within human multiscale biological data is envisaged. Responding to this imperative in the context of the broad Virtual Physiological Human (VPH) initiative, CHIC proposes the development of a suite of tools, services and secure infrastructure that will support accessibility, reusability and extensibility of VPH mathematical and computational hypermodels. The CHIC tools, services, infrastructure and repositories will provide the community with a collaborative interface for exchanging knowledge and sharing work in an effective and standardized way. Furthermore, to enhance the reusability, extensibility and accessibility of hypermodels for other domains, the following requirements should be taken into account.

### **2.2.1 Requirements for biological models**

It is important to understand that from an architectural perspective a hypermodel is always a composition of different component models. Each model needs input data for processing to output data that will be used as input data for other models. In that way hypermodels are composed. Furthermore, the hypermodels themselves can serve as component models for other hypermodels. To enhance the reusability and extensibility, modular and generality are important aspects when building the hypomodels. In addition, interoperability issues between these models are of utmost importance.

### **2.2.2 Requirements for IT infrastructure**

In addition to biological models, the hypermodelling infrastructure is needed to provide the orchestration and execution of hypermodel workflows on the CHIC computational infrastructure. This infrastructure also needs model repositories to store the models that describe and simulate different physiological processes. Therefore, the hypermodelling infrastructure should be designed in modular and be extensible in order to interact with model and data repositories through

standardised interfaces based on Web Services or Remote Procedure Calls. Finally appropriate image processing modules can enhance the scientific potential of the aforementioned IT infrastructure.

### **2.2.3 Requirements for ethical and legal framework**

In order to facilitate the extension of the CHIC models to other domains and to allow data exchange with other health care systems in accordance with the legal framework, a number of requirements must be implemented from the legal and ethical perspective. First and foremost, it remains imperative to maintain appropriate safeguards of the privacy rights of patients whose data are used in building, testing and applying the models.

### **2.2.4 Validation of hypermodels in clinical trials**

Another important aspect for enhancing hypermodels is that *in silico* models and hypermodels will be used in clinical settings such as “*in silico* trials”. The idea behind “*in silico* trials” is that a model is seen as a new drug that needs market approval. To get a drug on the market, preclinical testing and clinical testing within phase I to IV trials are required. Similarly, the same needs to be done with *in silico* models and hypermodels. This will help to validate the model to see if the model is indicating the right treatment. For example, a model indicates whether a preoperative chemotherapy scheme is better than primary surgery. One should look on tumour volume reduction predicted by the model. If the model predicts a reduction, chemotherapy is selected; if the model does not predict a reduction, go to primary surgery. In the standard arm all patients will receive preoperative chemotherapy and one could compare the reduction in tumour volume with the predicted reduction by the model as a validation means of the model. Comparing the results between both approaches would reveal whether the model is beneficial for the outcome of a patient.

In the following sections, we describe the requirements in more details.

### 3 Requirements for biological models

Some general requirements for enhancing hypermodels beyond the domain of cancer:

- **Modular character.**

Modular character is generally desirable so that specific modules dealing with elementary biomechanisms can be used beyond the domain of cancer.

Cancer models whenever possible should become modular to the extent that their constitutive hypomodels could be reused in the case of normal tissue development or pathologies other than cancer.

- **Enhancement with additional or modified restrictions/rules/checkpoints**, specific for the domain to which the models are to be applied.

For example, the distinct phases of the cell cycle in conjunction with several checkpoints could form a pretty generic cytokinetic model to be applicable to any kind of normal or pathogenic tissue model. The cell cycle checkpoints should be readily informed by the outputs of other elementary models (e.g. accounting for the homeostasis mechanisms in the case of non malignant tissues).

Care should also be taken so that the offspring of normal tissue cell mitosis are potentially channeled into different developmental branches, such as in the case of embryonic development, where both the epidermis and neuronal tissue emerge from the ectoderm.

- **Generality:** even some phenomenological, statistical characteristics of cell populations (e.g. collectively constituting a tumour) should be reformulated in a quite generic way so as to be applicable to normal cell populations and to different pathologies.

- In order to provide a more realistic simulating environment for tumour dynamics, not only tumour cells but **normal cells**, and in many cases their own functionalities or physiological activities, **should be modeled**, so as to recreate the tumour-host interactions. Such normal tissue cell models could obviously serve as the basis for modelling physiological activities or non neoplastic pathologies. An example of this is angiogenesis (blood vessel formation). Taking into account the fact that, at least during the early stages of angiogenesis, neoplastic cells stem from functional normal tissue cells, it would be advisable to extend tumour growth models by also considering some tissue cell functionality (e.g. biochemical functionality).

It should be also noted that there is a number of critical cancer modelling components which are characterized by **intrinsic generic applicability** to both physiological and pathological tissue. These include:

### **3.1 Biomechanical modelling**

The biomechanical component developed within CHIC aims at estimating the mechanical interactions between a growing/shrinking tumour and the surrounding healthy tissues. Biomechanical simulation provides insight on the direction of least pressure for the expansion/shrinking of the discrete tumour model and helps understanding tumour-induced angiogenesis. The mechanical models developed within this context can be used as a basis in different scenarios ranging from surgical planning to image processing.

Biomechanical finite element models of organs and tissues are application specific. Although the meshing approach used for the automatic creation of finite element meshes as well as the finite element solver are generic, the mechanical properties of the tissues as well as the boundary conditions are application- and patient- dependent. An application of the biomechanical modelling concerns the atlas-based image segmentation. Patient-specific segmentation of the tissue out of medical images can be achieved by the non-rigid registration of a reference atlas dataset that has been previously segmented. However, in the case of cancer, the atlas does not contain the tumour, which makes the registration process impossible. An alternative is to use a physiological model of tumour evolution and artificially grow a tumour in the atlas. Once the artificial tumour volume corresponds to the size of the tumour observed in the patient, a regular non-rigid registration step could be done to segment the healthy and pathological tissues. The quality of the segmentation depends on the accuracy of the prediction of the tumour growth, infiltration of the healthy tissues and simulation of the mechanical mass effect. Advanced simulations developed within CHIC could improve the accuracy of the segmentation approach and help planning radio-therapeutic treatments. Similarly, biomechanical simulations could be used to improve breathing compensation. Other applications of the biomechanical models for image-guided surgery include the prediction of craniotomy-induced brain-shift using patient-specific models. Numerical models could be used to update the MRI images used for pre-operative planning to account for the motion of the brain once the skull has been opened.

One of main challenges to extend the biomechanical models beyond the current applications concerns the mechanical description of the patient-specific behaviour of the tissue. This issue is not so critical for atlas-based segmentation, since the mechanical simulation will be followed by a final registration procedure. Therefore, generic properties could provide satisfactory outcomes. However, accurate predictions are required to predict intra-operative brain shift, which requires precise boundary condition and mechanical properties in order to avoid compromising the surgical navigation planning based on preoperative images. Another critical aspect to apply the biomechanical models developed within CHIC to clinical applications concerns the large calculation time required to solve the finite element problem. Calculation time needs to be significantly decreased to enable near real-time application, as it is the case for computer-assisted surgical planning. A solution could be to make use of GPU acceleration, which would require explicit time integration for good penalization. The approach seems suitable for soft tissues, which have a low Young's modulus, which would lead to relatively large stable time increments. [Related model in deliverable D6.1: C+Te33]

### 3.2 Atomic and molecular level

Techniques dealing with molecular dynamics, molecular networking and molecular pathways. These are also characterized by a high degree of generality.

In order to consider the potential requirements for enhancing beyond the domain of cancer the component models that refer to the atomic and the molecular scales of bio-complexity, planned to be developed and/or used in the context of the CHIC project, the following categorization could be adopted:

1. **Molecular Dynamics (or Structural Bioinformatics) (hypo)models** (i.e. Ae1, in deliverable D6.1), which focus on the analysis and prediction of the three-dimensional structure and molecular properties of biological macromolecules and **Molecular Interaction (or Docking Simulation) (hypo)models** (i.e. Ae1, Ae3, Ae4 in deliverable D6.1), aiming to simulate the interactions between (macro)molecules of known 3D structure.
2. **Statistical (or Machine Learning) Predictive (hypo)models** (i.e. Ad1, Md4, Md5, Md6, Md7 in deliverable D6.1), where computational algorithms are used in order for large-scale patient data (usually adopted using high-throughput experiments) to be analysed and correlated with a patient's disease characteristic (e.g. a specific cancer sub-type). An example of such an analysis planned to be done in the context of CHIC, for the nephroblastoma case, was presented in deliverable D2.2 (pp.30-33).
3. **Mechanistic (hypo)models of biochemical networks (pathways)** (i.e. Me1, Me2, Md1, Md3, Md8 in deliverable D6.1), which focus on the simulation of biomolecular interactions that regulate the diverse biological sub-cellular machineries. For example the model Me1, aims to simulate the biochemical regulation of the cell cycle using Ordinary Differential Equations (ODEs).

Important aspects to be considered include the modeling methodologies, principles and frameworks, the adjustment of the models in a problem-specific manner and the possible need for creation of new models.

Starting from the molecular dynamics and molecular interaction category, the related modeling techniques are characterized by an intrinsic generic applicability to both physiological and pathological tissue. For the former, it could be argued that the prediction of the structure of macromolecules either un-changed or changed due to mutations in protein coding sequence could be informative for numerous pathophysiological and biological topics. Similarly, for the latter, the simulation of interactions between (macro)molecules could extract useful information for a wide spectrum of biological modeling applications such as the binding properties of a ligand (or a drug) to a receptor or the probability for the creation of heterodimers between proteins. Regarding the methodologies used, the tools planned to be used in the context of CHIC (NAMD (Phillips et al., 2005), VMD and Carma (Glykos, 2006) for Molecular Dynamics and AutoDock (Morris et al., 2009) and Glide (<http://www.schrodinger.com/Glide>) for Molecular Interactions) are widely accepted and can be used for the modeling of any macro-molecule with known 3D structure. However, in the case where the macro-molecules (e.g. specific receptors) thought to be important for the pathophysiological/biological phenomenon for which a new hyper-model would be created are



different from those studied in the context of CHIC, a de novo analysis should be performed using the available information for them. In the context of CHIC, the properties of the Epidermal Growth Factor (EGF) and Anaplastic lymphoma kinase (ALK) tyrosine kinase receptors and the consequences of sequence mutations on them are studied using the aforementioned methods. Similarly, the properties of another tyrosine kinase receptor could be studied, for example the Insulin receptor, which is known to be correlated with leprechaunism and Rabson-Mendenhall syndrome when mutated (Longo et al., 2002)

As far as statistical predictive models are concerned, the methods used to identify, for example, the differentially expressed genes or microRNAs (statistical tests like t-test) and the machine learning algorithms used to classify patients based on these expressions, are commonly used even in applications irrelevant to biology and medicine due to their high degree of generality (“black box” methods). Therefore, the model creation framework could be clearly adopted for any other pathological or physiological modeling topic in which large scale data should be correlated with a specific state or characteristic. However, it is evident that since these methods are data-driven and in general no manually curated information is included into the derived models, the entire analysis should be repeated using the data acquired for the specific case. A randomly chosen study from literature where gene expression data are analyzed in a similar way to identify indicators of heart failure etiologies is given in (Tan et al., 2002).

Finally, regarding mechanistic models of biochemical networks, again, as far as the modeling methodologies, principles and frameworks are concerned, it is believed that they could be directly adopted in many different applications. More specifically, the ODE-based modelling framework for the mathematical representation of biochemical reactions, has been already used in a great number of heterogeneous models hosted in BIOMODELS database (Li et al., 2010) ([www.ebi.ac.uk/biomodels/](http://www.ebi.ac.uk/biomodels/)) using the SBML descriptive language (Hucka et al., 2003). Moreover, the majority of these models are created by integrating knowledge from canonical pathways (e.g. those presented in KEGG, Panther or Reactome pathway databases), which represent general sub-cellular biochemical machineries that could be appointed to any type of cell, but in a different extent of activation due to the different functionalities of the cells (different tissue types) and the different extracellular environmental context. However, due to the heterogeneity that characterizes different pathologies, some pathways deregulated in cancer (like those of cell cycle or the Epidermal Growth Factor Receptor pathways), may not present a significant malfunction in other diseases and, therefore, focus should be given to those pathways that their deregulation majorly contributes to the pathogenesis to be modeled. For example, in the case where a model for diabetes is to be created, a component model that addresses the insulin secretion of pancreatic cells (Jiang et al., 2007) may be of significant interest. Such a model is hosted in BIOMODELS database (BIOMD0000000239). Moreover, even in the case where a model developed to be a component of a cancer hypermodel is considered to be re-usable for another domain, any cancer specific modeling choices (e.g. over-expression or under-expression of a protein) as well as the model’s parameterization should be reviewed to ensure that they fit the needs of the new application area, taking into account the observed in nature cell type specificity in signaling phenotypes (Sung and Hager, 2012).

### **3.3 Specific examples**

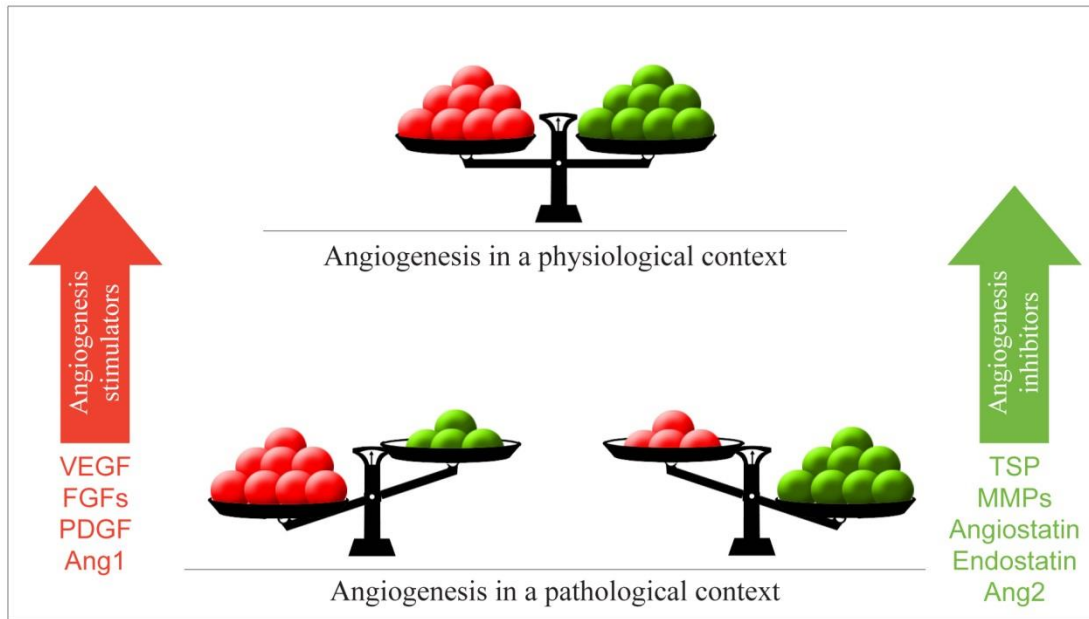
Provided that the above-mentioned general requirements are satisfied, numerous non cancer diseases and physiological functions can be simulated. The provided examples are of an indicative nature:

#### **3.3.1 Angiogenesis models**

This section elaborates on the potential of modeling angiogenesis in a way as independent of the underlying tumour or disease as possible. The goal is to end up with modeling approaches that are easily applicable to domains beyond cancer.

Angiogenesis is a fundamental bioprocess occurring both in a physiological and in a pathological context. In the latter case, numerous pathologies like cancer, ischemic diseases, reproductive disorders, inflammatory diseases etc. can occur, while under physiological circumstances angiogenesis can be met during early embryogenesis (Riseau, 1997), in the female reproductive system in response to ovulation or gestation (Reynolds et al. 1992) etc.

It is a well-recognized fact that there exist many qualitative and quantitative differences between physiologically and pathologically induced angiogenesis as well as among the vascular patterns exhibited by different angiogenesis-dependent pathologies. Nevertheless, there is a common modulator in the sense that it is the precise balance of pro-angiogenic and anti-angiogenic factors (or the lack of it) that determines whether an angiogenic response is physiological or not. Specifically, while in the physiological setting the action of endogenous angiogenic stimulators is counterbalanced by the action of the endogenous angiogenic inhibitors, in the pathological setting the effect of the one group of mediators overtakes the effect of the other one resulting either in excessive [case of cancer (Prager et al. 2012), endometriosis (Rocha et al. 2013), Crohn's disease (Danese et al. 2006), blinding retinopathy (Gariano and Gardner, 2005), psoriasis (Heidenreich et al. 2009), rheumatoid arthritis (Marrelli et al. 2011), atherosclerosis (Khurana et al. 2005), osteomyelitis (Ross, 2006), asthma (Ribatti et al. 2009) etc.] or in impaired angiogenesis [case of cardiac failure (Hilfiker-Kleiner et al. 2006), pre-eclampsia (Maynard and Karumachi, 2011), age-related nephropathy (Kang et al. 2001), pulmonary fibrosis (Antoniou et al. 2006), emphysema (Voelkel et al. 2007) etc.] (Figure 3-1).



**Figure 3-1: Regulated and dysregulated balance between endogenous angiogenesis stimulators and endogenous angiogenesis inhibitors. The latter case results either to excessive angiogenesis (left) or to insufficient angiogenesis (right).**

The first step prior to an attempt to extend the use of a model to another domain would be to re-examine the validity of underlying assumptions. Several of them may need to be relaxed depending on the nature of the new problem or, if this is not feasible, the existing model may be not applicable at all. Nevertheless, the minimum requirement for one to model angiogenesis in a generic setting would be to equip the model with explicit mathematical formulation of both key-players of angiogenic bioprocess (pro-angiogenic and anti-angiogenic factors) so as to capture the distinct kinetics and the interplay of growth regulatory proteins which act either to stimulate (such as vascular endothelial growth factor, fibroblast growth factors, platelet-derived growth factor, angiopoietin-1 etc.) or to inhibit (such as thrombospondin, angiostatin, endostatin, angiopoietin-2, etc.) blood vessel growth (Figure 3-1).

*The paradigm of a population dynamics related vascular tumour growth model: a few thoughts on a potential extension to wound healing and endometriosis*

The dynamical system suggested by (Poleszczuk et al. 2011) and extended by (Argyri et al. 2012) is governed by a pair of ordinary differential equations (ODEs) which describe the interaction between tumour volume and a time-dependent carrying capacity reflecting maximal tumour volume that can be supported by the given vasculature. This two-population model could be applied to wound healing in order to monitor the time evolution of the healthy cell population in the wound site under angiogenic control.

Wound healing is divided in three temporally overlapping phases, i.e. the inflammatory, proliferative and remodeling phases (Mendonca and Coutinho-Netto 2009). The cascade of biological events that constitute angiogenesis occurs practically during all three of the aforementioned phases and exhibits crucial similarities to angiogenesis occurring in the context of tumour growth. Similarly to the case of

cancer where angiogenesis is mainly hypoxia-induced, once the tissue is injured, the locally increased metabolic needs prompt the initiation of angiogenesis with the mediation of proangiogenic factors (Polverini, 2012). Specifically in the wound healing case, macrophages are induced to secrete VEGF and other cell signaling related proteins that directly stimulate endothelial proliferation and migration (Pollard, 2009). Additionally, macrophages also act as suppressors of wound angiogenesis through the production of angiogenesis inhibitors (DiPietro and Polverini, 1993). During the later stages of the remodeling phase where the normal tissue structure is recovered, most vessels disappear from the wound site through migration, apoptosis or other mechanisms of cell death due to the eventual overtake of angiogenesis inhibitors signaling. This phenomenon resembles to tumoural post-vascular dormancy where angiogenesis is eventually restrained and tumour growth ceases due to the balance eventually achieved between pro-angiogenic and antiangiogenic factors, a growth pattern which is addressed by the specific model (theoretical tumour plateau corresponding to the lethal burden of the disease).

Similarly, the model could also be applied for the pathological case of chronic wounds, like those occurring to diabetic patients (Tellechea et al. 2010), by selecting suitable parameter values that would reproduce the healing process to its full temporal extent spanning from months to years.

The development and maintenance of endometriosis, which is defined by the presence of endometrial tissue outside of the uterine cavity, also depends on the recruitment of blood vessels to the sites of endometriotic lesions. Thus, angiogenesis is stimulated in order for the lesions to acquire adequate oxygen and nutrient supply. Extending the same line of reasoning for the specific problem, the two – population model could be exploitable for the simulation of the interplay of the cell population that constitutes a lesion with the time-dependent vasculature as reflected by the carrying capacity of the population. [Related model in deliverable D6.1: C+Te20]

Another potential for extending hypermodels related to cancer angiogenesis and neovasculature *would be* to the domain of the **cardiovascular system**. Since the same fluid mechanics and deformable solid mechanics laws apply to both cases and neovasculature constitutes part of the overall body vasculature, it is obvious that with limited adaptation efforts (hyper)models developed for the cancer angiogenesis and neovasculature phenomena can be extended to simulate blood flow through small vessels and related phenomena throughout the human body.

### 3.3.2 Tumour response to immunotherapy models

At the core of the mechanistic model developed within CHIC to describe the immune system-Glioblastoma Multiforme (GBM) interactions after vaccination, is the time evolution of the various cell populations in the blood. Mature dendritic cells (DCs) present tumour antigens to naïve T cells, and drive their activation and proliferation. Regulatory T-cells are believed to get activated at tumour site, and then spread their immunosuppressive actions systemically. The cytokine milieu within lymph nodes, driven by dendritic cells, T helper cells, regulatory T cells and cytotoxic T cells propagates either tumour specific-immune responses or tolerance.

Being able to describe quantitatively the evolution of immune cells populations and their interactions after an antigen-specific stimulation, opens the door to applying similar mathematical descriptions, most probably with different parameters, to several self-immune diseases.

### Systemic Lupus Erythromatosus (SLE).

Although the exact etiology of SLE is unknown, aberrant apoptosis or insufficient clearance of apoptotic cells are believed to be a starting cause of the subsequent self-immune responses. According to the model proposed in (Fransen et al., 2010), apoptotic blebs and nucleosomes, which are modified during apoptosis, are taken up by immature DCs. DCs subsequently undergo maturation, as indicated by the increased expression of co-stimulatory molecules on their surface and their production of proinflammatory cytokines. Mature DCs can induce activation of Th1, Th2 helper cells, support the development of Th17 helper cells and inhibit the development of regulatory T cells. After activation by T helper cells, autoreactive B cells produce autoantibodies which can form immune complexes with apoptotic material. Plasmacytoid DCs can take up these immune complexes and secrete IFN- $\alpha$ , which further enhances autoantibody production. This results in increasing concentrations of immune complexes, which can associate with the (glomerular) basement membrane. The resulting influx of immune cells ultimately gives rise to multiple disease manifestations, and local tissue damage, thereby inducing an increase in apoptotic cells, which feeds back immature DCs with immunogenic material.

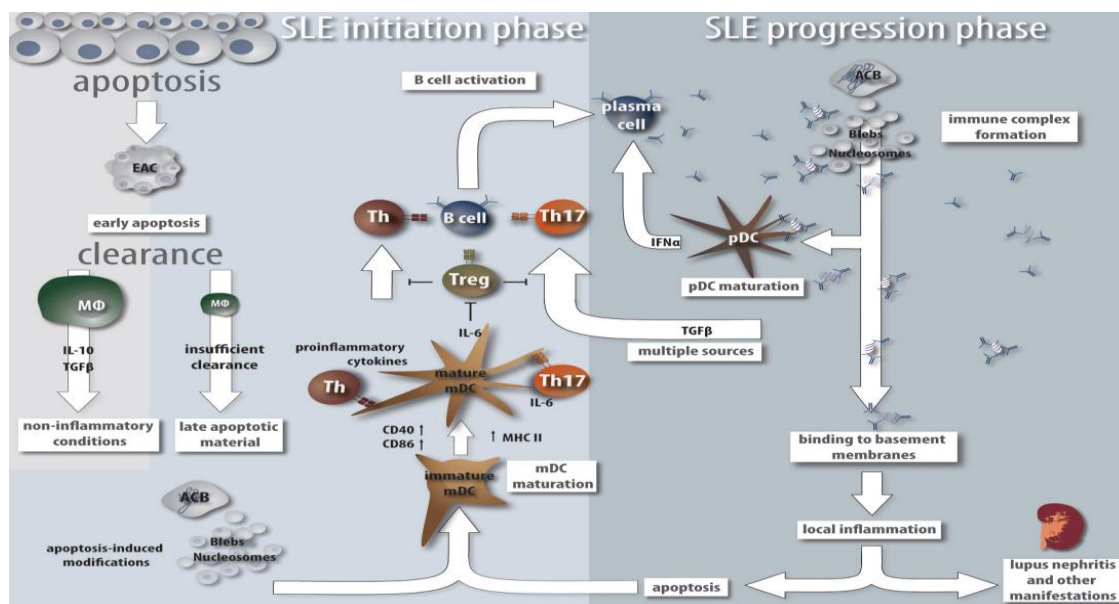


Figure 3-2: (Adopted from Fransen et al, 2010). ACB, apoptotic cell body; Blebs, apoptotic blebs; EAC, early apoptotic cell; MΦ, macrophage; mDC, myeloid dendritic cell; Nucleosomes, apoptosis-induced hyper acetylated nucleosomes; pDC, plasmacytoid dendritic cell; Th, T helper cell; Treg, regulatory T cell.

The evolution of several immune cell populations, stratified by type and/or antigen specificity (as in the GBM case) can be utilized as a first step in quantifying the self-immune response. An additional module seems to be needed, quantifying the concentrations/populations of B-cells and auto-antibodies.

#### Multiple Sclerosis

Multiple sclerosis is a T-cell mediated disease, which consists in the development of neuroantigen-specific T cells. It is not known what causes the development of these cells. Nevertheless, autoreactive T cells are activated by antigen presenting cells, and dendritic cells are the primary antigen presenting cells directing T-cell actions. The role of dendritic cells in multiple sclerosis has been addressed in various studies (Wu and Laufer, 2007).

Again, the approach of quantifying the immune cell response by addressing the evolution of various immune populations in the blood as well as the Central Nervous System (CNS) could serve as a first approach in modelling the immune system functions in multiple sclerosis.

#### Rheumatoid Arthritis

Rheumatoid arthritis (RA) is a chronic inflammatory disease associated with the destruction of affected joints and represents one of the most common autoimmune-related diseases. Clinically, RA is associated with swelling and pain in multiple joints, and is characterized by synovial hyperplasia and progressive joint destruction (Firestein, 2003). In this respect, there is a vast migration of T cells, B cells, fibroblast-like synoviocytes, macrophages and DC into the synovial tissue (Tak, 2006; Lebre and Tak, 2008).

By extending the model in the GBM case, time-evolutions of immune cells populations their circulation and localization can provide a first approach to modelling phenomena involved in this disease.

### **3.3.3 Phenomenological Universalities (PUN) approach applied in different biological context**

This approach was already successfully used to model human growth (Gliozzi et al., 2012; Galletto et al., 2014).

Moreover, the two populations model, that we use to simulate two cancer cell populations, could be adapted to ecological models in which the two populations are vegetal species in different environment or animal species in competition in the same territory (e.g. two wolf packs in the same hunting or sheep and goats in the same pasture); other authors studied this possibility, e.g. (Barberis and Condat, 2012). [Related model in deliverable D6.1: C+Te27]



### 3.3.4 Nephroblastoma tumour growth and response to chemotherapy model

"Nephro" means kidney, and a "blastoma" is a tumour made of embryonic tissue that has not yet fully developed. Wilms' tumour, or nephroblastoma, is the most common renal malignancy of childhood and is thought to arise from alterations in the coordinated differentiation of nephrogenic progenitor cells within the developing kidney. Subsequently, Wilms' tumour is closely linked to early kidney development and is often associated with remaining embryonic kidney precursor cells and other renal abnormalities (Bennington, 1975; Beckwith, 1990). From a biological perspective, this provides a window for future use of Wilms tumour models for simulating mechanisms involved in early kidney development and possibly for studying the properties of embryonic kidney stem cells.

Wilms' tumour is classified as a primitive malignancy of embryonic renal precursors (Bennington, 1975). Normally, nephrogenesis occurs as a result of reciprocal tissue interactions between the ureteric bud (UB) epithelium, the condensed metanephric mesenchyme surrounding UB tips, and stromal mesenchyme (Saxen, 1987; Dressler 2006). These interactions coordinate the differentiation of the condensed metanephric mesenchyme into epithelial elements that give rise to functional adult nephrons. This process of mesenchymal-to-epithelial transition occurs reiteratively throughout nephrogenesis and is normally completed before birth in humans. Wilms' tumours typically comprise three cell types that are histologically reminiscent of those found in the developing kidney but that show no functionally organized tissue architecture. The theory that these tumours have multi potential is supported by a characteristic 'triphasic' histology, which includes undifferentiated blastema, primitive nephronic epithelia, and stroma elements (Beckwith, 1997; Schmidt, 1995). Identical changes have been detected in these different components by microdissection analysis, indicating that they are all derived from a pluripotent malignant cell (Green, 1997). The blastemal, or primitive component, consists of tightly packed round-to-oval undifferentiated mesenchymal cells with little cytoplasm. The epithelial component might show different degrees of differentiation, ranging from poorly defined glands to structures resembling glomeruli. Stromal cells often resemble fibroblasts, but they can also differentiate into smooth muscle, skeletal muscle or neural elements. All three elements can be found in most tumors although the predominance of one histological component over another is variable.

In addition to its apparent derivation from pluripotent precursors, a significant subset of kidneys from children who develop Wilms' tumours contains putative precursor lesions, known as nephrogenic rests, that also manifest histologic features of the condensed metanephric mesenchyme and also points to a connection with early kidney development (Beckwith, 1990). Nephrogenic rests consist of blastemal cells with varying degrees of differentiation. Wilms' tumours themselves share histological features with the developing kidney, and are frequently cited as an example of impaired differentiation in tumorigenesis. Of all the stages of kidney development, Wilms' tumour most closely resembles the developing nephrogenic mesenchyme. The epithelial components of the tumour resemble comma-shaped bodies, S-shaped bodies and glomeruli, and the blastemal components are similar to the condensing nephrogenic mesenchyme (Rivera, 2005). At a molecular level, studies have shown that Wilms' tumours express markers of early kidney development and, more recently, microarray gene-expression studies have provided additional evidence for the similarities that are suggested by morphology. Genes that are overexpressed in Wilms' tumours tend to be expressed at the time of the first contact between the ureteric bud and the metanephric mesenchyme. Similarly, genes that are under-represented in Wilms' tumours are usually expressed at

later stages of normal kidney development (Li, 2002; Li, 2005). *[Related model in deliverable D6.1: C+Te12]*

### 3.3.5 *Acute lymphoblastic* leukemia evolution and response to therapy model

In the hematopoietic system, the presence of a developmental hierarchy has been well established. All cells of the blood arise from a common origin, the hematopoietic stem cells (HSCs), located in the bone marrow (Janeway et al., 2001). These pluripotent stem cells generate cells with progressively more limited potential (e.g., multipotent progenitor cells, myeloid and erythroid progenitor cells, early lymphoid progenitor cells etc.) and ultimately differentiate into three major types: the leukocytes, the platelets, and the erythrocytes.

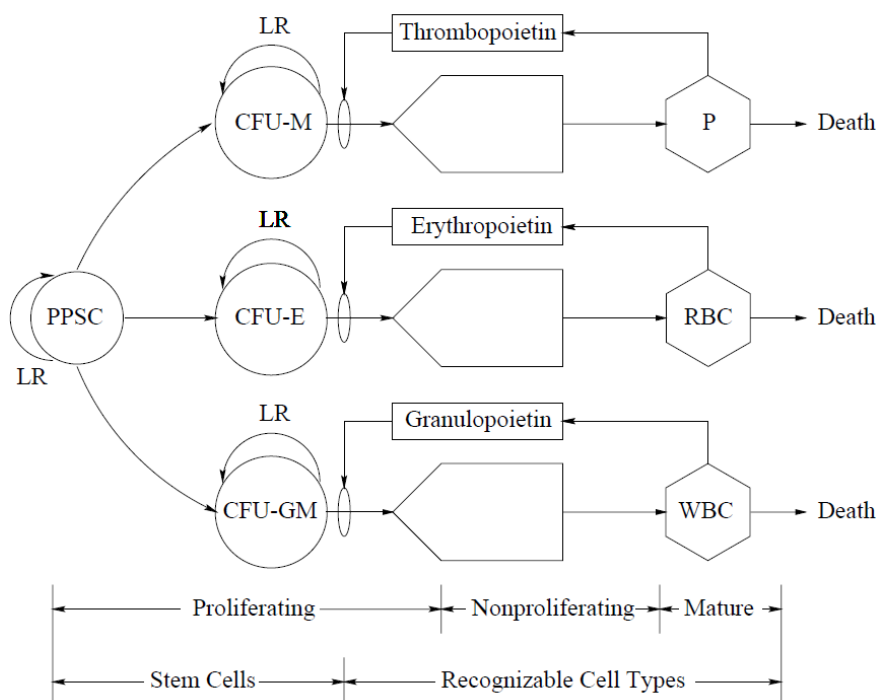
The simulation model of (Kolokotroni *et al.*, 2014) for acute lymphoblastic leukemia (ALL) evolution and response to therapy is able to describe hematopoiesis process. The modelling approach adopts a developmental hierarchy structure between the considered cell compartments, thus allowing for a realistic and detailed representation of the hematopoietic system. It incorporates a refined description of the cell cycle, considering actively proliferating cells and quiescent cells in the so-called G0 phase. Cells with self-renewal capacity and cells found at various levels of maturation are also incorporated in the model. The dynamics of such a cell community is regulated by the balance between active cell cycle, quiescence, self-renewal, differentiation and death. Each considered biological mechanism is represented by an adequate number of input model parameters, thereby enabling the independent handling of each phenomenon. Via the proper tuning of the input parameters, the model could comprise the basis for studying a wide spectrum of physiological and pathological situations. Indicative examples are given below:

- The model could be used to describe normal hematopoiesis, as a stable steady state, or under conditions of increasing demand e.g. in the case of bleeding, infection or injury (Peixoto et al., 2011). The specialized microenvironment of the bone marrow provides signals both for the development of progenitors from hematopoietic stem cells and for the subsequent differentiation (Janeway et al., 2001). For example the production of erythrocytes (erythropoiesis) appears to be regulated by specific cytokines (e.g. erythropoietin) via a negative feedback mechanism such that a decrease in the number of erythrocytes leads to an increase in erythrocyte production and vice versa (Foley and Mackey, 2009). Such interactions can be easily modelled and incorporated into the ALL model through the most relevant input parameter(s) e.g. by increasing the self-renewal rate of stem cells in the presence of a stimulating signal. Care should be taken for the definition and incorporation of the control loops (Figure 3-3).
- Periodic hematological disorders are of particular modelling interest due to their dynamical behaviors. These disorders include cyclical neutropenia, periodic chronic myelogenous leukemia (PCML), cyclical thrombocytopenia, and auto-immune hemolytic anemia. They are characterized by oscillating patterns in one or more of the circulating blood cells (Colijn and Mackey, 2005). The period of oscillations can be on the order of days to months. When all cells oscillate with the same



period, a dynamic destabilization at the stem cell level may be involved (Foley and Mackey, 2009). Such dynamic behaviours can be caught by the ALL model under the requirement of time variant free-growth parameters modeled e.g. through feedback loops. Because of the detailed modelling of the cell cycle and the cell maturation processes, delays that need to be included in the transition of feedbacks due to the above biological processes are inherent characteristics of the ALL modelling approach, offering a unique advantage in the description of such disorders.

- The ALL modelling approach could also be applied, following proper calibration, for the modelling of the hematopoietic reconstitution after stem cell transplantation or for the modelling of the side effect of drugs on the immune system.



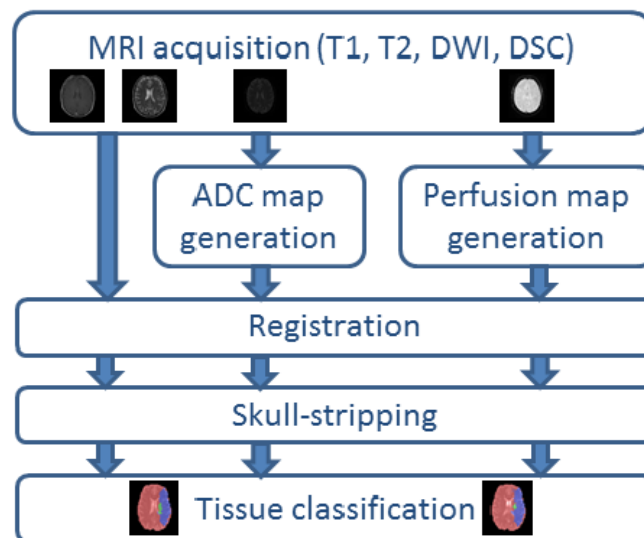
**Figure 3-3: Architecture of Hematopoietic regulation.** A schematic representation of the control of platelet (P), erythrocyte (RBC) and white blood cell (WBC) production, showing loops mediated by the various poietins, as well as local regulatory (LR) loops within the various stem cell compartments. CFU refers to the various colony forming units (M=megakaryocytic, E=erythroid, GM=granulocyte/macrophage) which are the in vitro analogs of in vivo committed stem cell populations, all of which arise from the pluripotential stem cells (PPSCs). (Figure taken from Mackey (1997))

## 4 Requirements for IT infrastructure

### 4.1 Image processing and visualisation

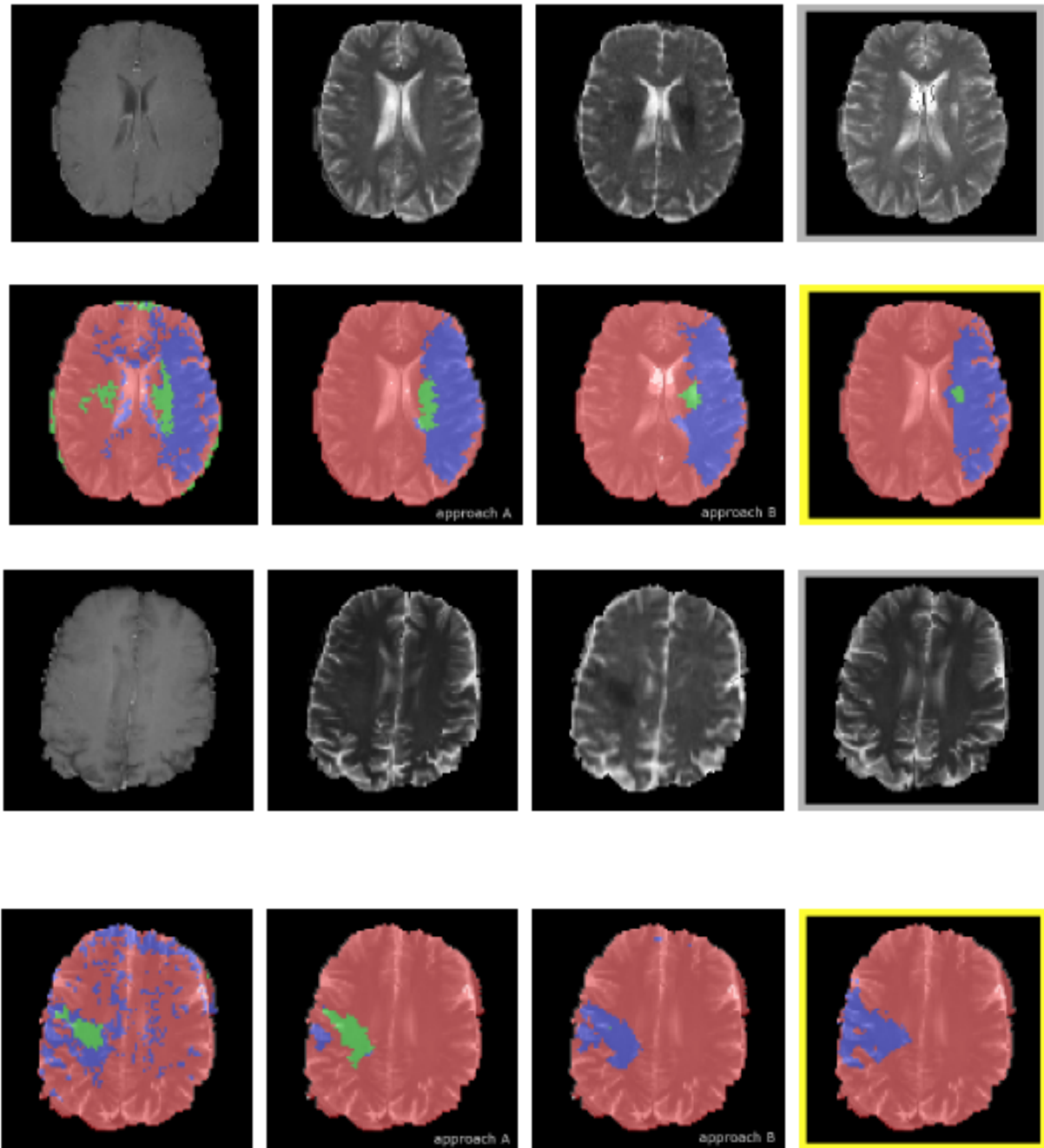
#### *Towards automatic volumetry for treatment selection in acute ischemic stroke patients.*

The methodologies and algorithms produced for the automatic segmentation of brain tumors (high-grade and low-grade gliomas) have been extended to consider other brain lesions, namely acute ischemic stroke patients. In this clinical scenario it is crucial for an appropriate treatment selection to quantify the ratio between the infarct core and the penumbra region (tissues in potential risk). We developed an automatic volumetry algorithm based on a supervised segmentation methodology, in principle, following the same rationale used for the segmentation of brain tumors. The algorithm employs the common clinical image sequences (T1, T2, Diffusion Weighted and Perfusion maps), and manually segmented images of the core and penumbra regions, to learn the mapping between imaging features and tissue labels. On given new unseen images the algorithm performs tissue classification, which enables volumetric analysis of tissues at risk. Figure 4-1 describes the image processing pipeline used in this application. The whole application requires in average approximately five minutes to perform a full tissue classification.



**Figure 4-1: Processing pipeline for automated tissue classification into infarct core and penumbra region. Image modalities including structural (T1-weighted, T2-weighted) and functional (diffusion and perfusion) are pre-processed by rigid registration and skull-stripping, followed by tissue classification.**

Figure 4-1 presents two exemplary patient cases where the approach was tested and compared against the state of the art approach RAPID. Two different approaches were designed to consider as well 3-month postoperative images (approach B in Figure 4-2). In this scenario, the tissue classification corresponds to a prediction engine capable of estimating the 3-month scenario. First results using the proposed approach are very promising for an improved quantification of stroke patient images.



**Figure 4-2: Results of the proposed approach on two exemplary cases (rows 1 and 3, from top to bottom). In rows 2 and 4 and from left to right: RAPID result, approach A (no 3-month follow used for training, pure classification approach), approach B (including 3-month follow-up, prediction model), ground truth segmentation. Note that for approach B, no 3-month follow-up is necessary to perform the tissue classification.**

The use of supervised learning approaches in combination with advanced image processing techniques and domain-knowledge methodologies have great potential for the modelling of complex disease processes. The example presented above on ischemic stroke patients is one out of many other applications where these technologies can leverage medical treatment.

## **4.2 Hypermodelling infrastructure**

The hypermodelling infrastructure is based on the VPH Hypermodelling Framework that provides the orchestration and execution of hypermodel workflows on the CHIC computational infrastructure. Its design is modular and extensible since it interacts with model and data repositories through standardised interfaces based on Web Services or Remote Procedure Calls. The hypermodelling infrastructure is therefore not tied to any specific model or data, and it can serve domains beyond cancer.

### ***Requirements for enhancing hypermodels beyond the domain of cancer***

The initial specifications of hypermodelling in the cancer domain have been captured in the deliverable D7.1 Hypermodelling Specifications. They take not only into account the needs of the cancer domain but also the future exploitation of the CHIC platform in other domains of the VPH initiative. In fact, the design of the VPH Hypermodelling Framework (VPH-HF) is inspired by the principles of extensibility, reusability, modularity, interoperability, security and robustness. The partners involved in WP7 are already obeying these principles, in fact they are heavily exploiting outcomes from previous EC projects like TUMOR and VPHOP. In the latter, the VPH Hypermodelling Framework was developed for the osteoporotic domain and is currently enhanced to serve the needs of the cancer community.

In order to decouple the execution framework from the actual model implementation, the Component Model Generic Stub has been introduced. It is a standardised interface that allows any model to be considered as a black box with input, output and control flow ports. In the current version of the VPH Hypermodelling Framework, it is implemented through a Model Wrapper software component that hides the details of the models and their implementation behind the same interface. Therefore, the Hypermodelling Framework itself is agnostic of the actual wrapped models and it can be easily adopted beyond the cancer domain by simply wrapping any required model. It is the responsibility of the researcher to compose the hypomodels into hypermodels to suit their modelling needs. Therefore the CHIC-HF is ready to be exploited in any domain beyond cancer.

## **4.3 Model repository, in silico trial repository and clinical data repository**

### **4.3.1 Model repository and in silico trial repository**

*Important note: The model repository can store hypomodels, hypermodels, data transformation tools, linkers and other tools. The term “module” will be used throw-out the following section in order to represent more correctly the nature of what it is actually stored in the model repository.*

The model repository is the technological platform where the modules used in the construction and execution of hypermodels are stored. End users and other technological components may access the stored modules, or store new ones, using the graphical or programmatic interfaces provided by the model repository.

The *in silico* trial repository is the technological platform where the data related to *in silico* trials are stored. Similar to the model repository, end users and other technological components may access the *in silico* trial repository in order to prepare new *in silico* trials or to store the results of the conducted *in silico* experiments.

The information model adopted in the representation of key entities of the model repository and the *in silico* trial repository is generic enough to be used in non-cancer related domains.

#### **4.3.1.1 Model Repository**

The modules that can be stored in the model repository are hypomodels, hypermodels, data transformation tools, linkers and other tools. The basic descriptive information that differentiates one module from another is stored in the entity “Tool”. We have also introduced the entity “Reference” which provides direct or indirect links to additional material, extending in this way the knowledge base related to the specific module. Both “Tool” and “Reference” entities are domain agnostic.

The additional information that further characterizes or classifies the module is included in the entity “Property”. It must be noted that properties can only be used in correlation with “Tool” entity as they may only supplement the basic descriptive information of a module. In the model repository of CHIC there are predefined properties that are specific for the cancer domain and are related to the 13 perspectives described in “D6.1: Cancer hypomodelling and hypermodelling strategies and initial component models”. Nevertheless, the design choice to store the default properties as data, and not as fields, in a separate structure makes the use of CHIC model repository in other medical domains effortless.

The aforementioned three entities assist the user in understanding the nature of the module and facilitate the abstract construction of hypermodels. However, the ultimate goal of CHIC is to provide concrete specifications of modules (hypomodels, hypermodels etc) that can be joined together programmatically. In order to succeed this, the “Parameter” entity is introduced. This entity facilitates the transition from an abstract representation to a concrete one. The modules are treated as generic stubs, as described in “D7.1: Hypermodelling specifications”, which have entry and exit points. Logical compatibilities between connected parameters are taken into account along with the aspect of units, in order to avoid inconsistencies between the connected modules. For usability reasons, the units used in CHIC are restricted to the ones that are used in the cancer domain. Minimum effort is required in order to enhance the application layer of the model repository with additional units, making the transition from cancer to other medical domains very easy.

To facilitate the automatic execution of modules, the model repository is able to store the module’s computational representation. The modules can be expressed in any form that a computer can understand, provided that the associated software is installed on it. The modules may be provided in a static (e.g. a web service), configurable (e.g. an executable) or migrating (e.g. a virtual machines) form as described in the “D7.1: Hypermodelling specifications”. Configurable modules are stored as files (or set of files) in the model repository. On the other side, only the access specifications of static modules are stored in the model repository. More specifically, the modules may be provided as:

- compiled, interpreted or scripting code, ready to be executed from the platform (eg. as an application) or the corresponding engine (eg. java vm, matlab, copasi for SBML etc). In this case all the files needed for the execution of the module can be retrieved from the model repository.
- web-service, accessed using standardized protocols and representation forms, such as REST, SOAP, JSON, XML etc. In this case only the access specifications are provided from the model repository.
- virtual machines. This approach is quite complicated. Eventually a combination of configurable (for the actual execution) and static (for accessing) form is applied.

From the description above it is straight forward that our approach is not limited to cancer domain, but it can cover a wide range of applications that demands computational representation of elements. The only limitation is that the computational platform to be used for the execution of the modules, either the user's PC or a centralized computational framework, is equipped with the extra programmes that are needed.

#### **4.3.1.2 In silico trial repository**

The design of the *in silico* trial repository is based on the classical and standardized principles of the real clinical trials. Similar to the model repository, one of the fundamental requirements of the *in silico* trial repository is the reusability in other medical domains.

The *in silico* trial repository consists of three main entities, the subjects, the experiments and the trials. The subject can be a citizen (patient or healthy), an animal etc. The experiment is the process of performing a simulation on a subject. The result of the experiment is another state of the original subject. The experiments are organized in trials, where the same simulation is performed on different subjects. By the description above it is obvious that the *in silico trial* repository is domain independent.

Moreover one of the main purpose of the *in silico* trial repository is to test the repeatability and reproducibility of the experiments conducted in the context of any *in silico* medicine domain.

Repeatability is the ability for an individual to show that an experiment, repeated using the same material and equipment, yields the same result. In *in silico* medicine this means that if we run the same module multiple times on the same computer using the same software the same result would be yielded.

Reproducibility is the ability for different individuals to show that an experiment repeated using different but similar material and different equipment yields the same statistical result. In *in silico* medicine this means that we are able to recreate a simulation without necessarily using the same software or computer that was used in the original simulation. Reproducing an experiment is one important approach that scientists use to gain confidence in their conclusions.

Recently, the scientific community was shaken by reports that a troubling proportion of peer-reviewed preclinical studies are not reproducible. Because confidence in results is of paramount

importance to the broad scientific community, new initiatives are announced to increase confidence in the studies.

The *in silico* trial repository can serve perfectly the aforementioned initiatives. By storing in one place the complete information concerning the input data, the output data and the modules which participate in the *in silico* experiments and the *in silico* trials, the *in silico* trial repository can advance *in silico* medicine in general, by facilitating the validation of the current *in silico* medicine discoveries.

### 4.3.2 Clinical Data Repository

The clinical data repository will permanently host all the related medical data produced or collected by the CHIC project. Standardized interfaces allow to import and export the contents of the clinical data repository. In this way the data can be sustained after the expiration of the project's lifetime and reused and exploited continuously within the limits allowed by the legal framework of the project.

The clinical data repository has been developed with an approach that ensures its flexible integration into other infrastructures. Open implementations schemes have been selected to interact with external services as well as to store the clinical data as generic objects. As results, few adaptations are required to extend the application of this system beyond the domain of cancer. Adaptation required for some key components are outlined in the next sections.

This section covers topics such as; potential exploitations, how the access to the repository is handled, how the data can be stored and retrieved, and what data can be stored.

#### 4.3.2.1 Authentication

The access to the clinical data repository is managed by the CHIC security framework. This framework makes use of the brokered authentication approach, which differentiates between the Identity Provider (IdP) and Service Provider (SP). The clinical data repository acts as an SP and accepts identity assertions issued by the IdP of the CHIC security framework. All human users, including patients, physicians, researchers and administrator are managed by the IdP. With this level of abstraction, it is straightforward to replace the complete user base of the clinical data repository by switching the accepted IdP. It is also possible to support multiple user bases by accepting multiple IdPs.

#### 4.3.2.2 Interaction/data access

The clinical data repository makes use of the REST (Representational State Transfer) architectural style. Therefore, the interactions accepted by the clinical data repository can be reduced to the set of common methods for HTTP (Hypertext Transfer Protocol). The flexibility, extensibility, performance, security, and ease of use facilitated by the REST architecture allows developers to quickly implement our API (Application Programming Interface) into their applications.

Versioning is applied by embedding the version number within the identifier of the resources, to maintain the flexibility and extensibility of the API. The following list shows which types of changes require a new version and which don't.



Changes that don't require a new version:

- New resources
- New methods on existing resources
- New data formats
- New attributes or elements on existing data types

Changes that require a new version:

- Removed or renamed resources
- Removed or renamed attributes or elements on existing data types
- Removal of support for methods on existing resources

#### **4.3.2.3 Supported data formats**

The data model is centered on the concept of an object. An object can be any kind of medical image format, segmentation, 3D models or other clinical data. Each time an object is updated (for example, with an improved segmentation), a new version of this object is created. The mechanism enables other users of the system to continue working with the previous version, without being affected by the modification of the files. Modification on the core infrastructure considers the situation where different users provide their own processing to the same data. A typical case corresponds to different segmentations of the same DICOM series with different tools or focused on different structures present in the images.

Besides the generic object approach, the clinical data repository relies on data based on a standardized format. This ensures to some extent that the data can be further processed by other applications. The not exhaustive list of accepted standardized formats includes:

- DICOM (Digital Imaging and Communications in Medicine), MetaImage, Analyze, Nifti
- HDF (Hierarchical Data Format) mostly for statistical shape models
- CDISC ODM (Clinical Data Interchange Standards Consortium / Operational Data Model)

However, the concept of generic objects used to store clinical data can be easily extended to store any kind of data. However, the internal processing of the dataset and the amount of information/metadata that can be obtained from the system will depend on the standard selected by the users to upload their files. Therefore, accepted standard file format are preferred for the new implementations.

#### **4.3.2.4 Metadata**

Data usually contains general information describing the content. For most standardized formats the general information can be defined in advance and extracted for discovery and identification purposes.



Besides general information, it is also possible to use predefined terms to annotate the content of the data. Those predefined terms are managed by external ontology or folksonomy services. Within the CHIC project, the ontology and folksonomy services are provided by RICORDO. However, the clinical data repository will be designed, so that the dependency to RICORDO could be replaced by another service.

### **4.3.3 Potential exploitation of clinical data repository**

#### **4.3.3.1 Centralized medical data storage solution**

The solutions developed within CHIC for the clinical data repository are flexible and scalable. Not only various type of data can be stored on the system, which could be used to build large open database for research purposes. One of the key features of the system developed in CHIC is the large variety of image file format that can be interpreted, annotated and stored on the system. These functionalities could be directly transferred to the development of large repositories of medical images, including basic processing such as segmentation. An application topic with receives a growing attention from the biomedical field is the generation of statistical shape models. These models are routinely implemented for automatic image segmentation or object identification in medical images. In these fields, however, the acquisition of the large training datasets, required to develop these models, is usually a time-consuming process. Even after this effort, the collections of datasets are often lost or mishandled resulting in replication of work.

To solve these problems, the clinical data repository could be proposed as a centralized storage system where the data necessary to build statistical shape models can be stored and shared. The clinical data repository system could be made available online tailored to the needs of the medical research community. The processing of the most common image file types, a statistical shape model framework, and an ontology-based search provide the generic tools to store, exchange, and retrieve digital medical datasets. Such a centralized and open approach, would favour the exchange of dataset and help researcher to build models which are more representative of the targeted population. The hosted data are accessible to the community, and collaborative research catalyses their productivity.

#### **4.3.3.2 Challenges**

An important topic for research team developing algorithms is the ability to test their development on a set of reference datasets. This enables the objective comparison of various technical solutions proposed by the research teams. The clinical data repository could be used to propose an infrastructure to host and run such open competitions. Only small adaptations would be required to enable the CHIC data repository to host challenges in image segmentation. However, the concept is generic and could be applied to many areas in the biomedical field such as biomechanics modelling or prediction of biological processes. The challenge data would be publicly available and any team around the world to develop and train their segmentation algorithms. In addition, the clinical data repository would be able to evaluate their submissions against the challenge reference. For this purpose the existing data upload tool could be used to allow users to upload their segmentations and the existing website could be extended to allow users to start the evaluation process. The evaluation

of the different quality metrics would be calculated as a background process to evaluate the quality of the submission (such as Dice coefficients segmentation challenges). Individual and overall results of the evaluation could be automatically published on the website and downloadable as a comma separated values (CSV) file for further statistical calculations.

## 5 Requirements for ethical and legal framework

In order to facilitate the extension of the CHIC models to other domains, a number of requirements must be implemented from the legal and ethical perspective. First and foremost, it remains imperative to maintain appropriate safeguards of the privacy rights of patients whose data are used in building, testing and applying the models. As described in Deliverables D4.1 and 4.3.1, the main piece of legislation in the EU on data protection is Directive 95/46/EC. Its aim is to protect the right to privacy individuals enjoy by regulating, inter alia, the “processing of personal data wholly or partly by automatic means”. Data concerning health or sex life belong to the special categories of data (“sensitive personal data”) that may not be processed unless specific exemptions apply. As applied to medical hypermodelling in general, the relevant exemptions are: explicit consent of the data subject; where the processing is required for certain medical and health-care purposes for the direct benefit of the individual patient; and where a member state, subject to safeguards, lays down exemptions for reasons of substantial public interest (under art 8(4) of the Directive).

It is apparent that building hypermodels will require large amounts of data. In general the option of first choice (in best allowing patient control over information) would be for explicit consent to be obtained for the use of the data for the specific purpose of building a model. However, this may raise practical or scientific problems. Frequently, the data at issue will be retrospective, perhaps collected years before, and contacting patients to agree the new use will be an immensely difficult or even impossible task. Where research is cutting edge or new questions arise, it might even be difficult to obtain valid explicit consent, which requires a voluntary and considered decision and a detailed understanding of how the data will be used. In such cases new contributors of data from other domains shall be required to ensure the lawfulness and ethicality of data use is underpinned by the informed consent of the patient and/or appropriate ethics approval obtained by the contributor in question.

As further discussed in the relevant WP4 deliverables, a separate requirement of data protection law is that data so far as possible (consistent with the achievability of the purposes for which it is being processed) be de-identified prior to use. This serves to protect the interests of patient in not suffering unnecessary harm, in the form of discrimination or stigmatisation, as a result of sensitive medical data being individually linked to him. As also described in detail in D4.3.1, within CHIC a dedicated security framework will take care of de-identification aspects using a secure process of double pseudonymisation backed up by the use of a trusted third party and advanced state of the art authentication and encryption services in order to secure the data from unauthorised access and misuse. As noted earlier in the deliverable, the relevant pseudonymisation tool is part of the fundamental architectural level of the project, which comprises IT tools that are domain and scenario unspecific: this means it has in-built flexibility, and can automatically be deployed in domains, beyond the cancer domain.

At the same time, while technical measures to de-identify and control access to the sensitive patient data are an extremely important and necessary part of safeguarding the patients’ privacy interests, the rise of ever better algorithms and increased processing power means that truly de-identifying (in effect anonymizing) data is in some circumstances very difficult, particularly when the data is longitudinal and large amounts of variables are available in it, as may be the case with data needed in hypermodelling scenarios. This is because combining such data elements with data from other

external sources (data-matching) may well permits re-identification of the data subject. In this respect, privacy cannot be guaranteed by technical means alone. Rather there is a need to combine technical, organizational and legal controls so together these ensure that, as per the legal test in Directive 95/46/EC, data cannot be linked back to the data subject using ‘reasonably likely means’. It is this combination of safeguards that has been established, for the ongoing Cancer domain model development work of the project through the adoption of the CHIC data protection framework set out in Deliverable D4.3.1, with its set of binding agreements in which the consortium partners commit themselves to treat all CHIC data confidentially and securely.

An important implication of this is that, in terms of extending the CHIC modeling processs to new domains, particularly where data will be contributed and mined by parties outside the existing project consortium, this requires an extension of the existing framework of agreements, in which new users also bind themselves to all relevant privacy and security obligations envisaged by those agreements. The legal framework is indeed scalable in this manner so, assuming the new users are prepared to sign the relevant agreements as a precondition for participating in the enlarged modeling domain, this should not in itself present an obstacle.

However, an updating of the existing framework will be necessary to the extent that patient data, once contributed and stored in the CHIC data repository, is to be processed for new purposes (research in other domains than cancer) to those for which it was originally collected and provided to the repository by the existing consortium partners: this will necessitate a reappraisal of the scope of the original consents and where this is limited to specific oncological research, re-consent and/or ethics committee approval would be required before making the data available for use in other domains. A related requirement arises insofar as the repository becomes a permanent resource, in which the medical data produced or collected by the CHIC project is sustained after the expiration of the project’s lifetime for ongoing reuse and exploitation. Here it will be important to comply with relevant member state law that stipulates the necessary safeguards to be taken (pursuant to art. 6(e) of Directive 95/46/EC), where personal data is retained for research purposes in this way. These matters will receive close attention in Deliverable D4.3.2.

A further issue concerns the technical means of storing and processing additional contributed data in the clinical data repository. In particular, where the volume of data processing requires the use of a cloud-based infrastructure, safeguards must be in place to ensure that authorized users retain exclusive control of the data; here the present private cloud solution within CHIC still needs to be retained as the default position, as a move to a public cloud solution, and – as recently suggested by the Article 29 Working Party (set up under Directive 95/46/EC) - the loss of control over sensitive data this implies, would simply be too risky in legal terms. Accordingly, it will be important for the private-cloud infrastructure to be able to handle the increased demands made upon it.

Moving beyond data security and privacy concerns, a second important legal requirement for enhancing the use of hypermodels beyond the cancer domain relates to their validation in order to ensure that they are fit for purpose, and that patients are not exposed to avoidable health risks as a result of their use. In this regard, validation of the models in the current cancer use cases will be necessary for generating trust in other domains (see: Deliverable D11.1 for validation requirements). It is also evident that, within a highly interdependent system in which the output of one model or hypermodel provides the input for another, the reliability of the overall results may be significantly

impaired by inaccuracies in just a single model. Here, the system layer of the models/hypermodels, ie, the programming functions, will need to be validated to ensure the quality of the system (see: FDA guidelines for medical software validation, IEC 62304, ISO 13485).

It indeed seems probable that a hypermodel would qualify as a medical device within the definition of the EU Medical Devices Directive 93/42/EC. The Commission in 2012 guidance on the application of Directive clarified that it extends to decision support software, widely defined as “computer based tools which combine medical knowledge databases and algorithms with patient specific data...[that] provide healthcare professionals and/or users with recommendations for diagnosis, prognosis, monitoring and treatment of individual patients”. If so the model would need to undergo a certification process, involving an application for approval from notified bodies at member state level, and potentially the need for sponsored ‘clinical investigations’ (similar to the clinical trials required for medicinal products). Here an issue for further exploration is whether, as applied in the context of personalized medical care, a hypermodel may qualify as a ‘custom-based device’ (under art. 11 para 6 of the Directive) exempting it from some aspects of the above regime.

A final set of legal issues arising from the enhancement of hyper-models beyond the domain of cancer concerns IPR. Computer models, as established in D.4.3.1, constitute copyright works and are protected by copyright as computer programs. Hence, enhancement of a model into other domain may require that model code be altered or extended in a way that a piece of new code is contributed or built on top of the existing one. Several points may arise and need to be address. The first relates to access to the source code, in cases where this is required for model enhancement. Here source code of a model would need to be provided to the model repository and access to the model source code authorised by the modelling party acting as a right holder. Should models be provided in compiled form only, enhancement of the code would be problematic. Source code might be attained by reverse engineering. However, reverse engineering, as exempted from authorization requirement by Article 6 Software Directive, is allowed for interoperability purposes only and may be performed in strict legal boundaries only. Therefore, the first precondition for enhancement would be provision of the model source code into the model repository.

Second, should any alteration or extension of the model code be required, authorisation of the right holder must be provided as well. Any adaptation, modification, alteration or re-arrangement of a model code is an exclusive right of the right holder. Modification of the code by any third party will be subject to authorisation by the right holder, i.e. the modelling party. Thus, the right to modify and distribute modified versions of the model code must be provided under the license under which a model will be released. Most “open source” licenses, such as Apache v2, BSD, GPL, etc., allow modifications and distribution of modified versions of the code subject, however, to the license conditions. On the other hand, should a model be released under a proprietary license, the right of modification for development purposes would usually not be provided and would require separate authorization by the modelling party.

A third issue which needs to be addressed is who will own the IP rights in enhanced models, if modification of the model code is allowed. In terms of copyright law, enhanced versions of the models would constitute derivative works. Under the first ownership rule in copyright (explained in D.4.3.1) the IP rights in derivative works would normally belong to the party who develops the derivative work, i.e. party who will make an enhanced version, subject that (a )making a derivative

work is authorised by the original right holder and (b) unless agreed otherwise. Consequently, should modification of the model code be allowed, ownership in the enhanced version must be considered. A first option would be to allow modifications and derivative works under grant back obligations, meaning that rights in enhanced or modified versions would pass to the original right holder, i.e. the modelling party in CHIC. Grant back obligations are seen sometimes in proprietary licenses, if derivative works are allowed at all. If grant back of the rights is not foreseen explicitly, the first ownership rule in copyright will apply and rights in the enhanced (modified) version would pass to the party who will make the enhancement (modification). This second option usually applies in “open source” licenses. As a rule, a party who contributes a piece of code or modifies the existing one owns copyright in his contribution – a piece of new code as contributed or the code as modified.

## 6 Validation of hypermodels in Clinical trials

Validation of hypermodels defined by CHIC will be done in the context of the three cancer domains: nephroblastoma, glioblastoma and lung cancer. Hypermodels beyond these three domains rely on the validation results for these three cancer types. The integration of these hypermodels in clinical practice requires standards for model validation. Methods are needed to ensure that models are consistent with their claims, accompanied by explicit statements concerning assumptions and limitations along with care in the acquisition of validation data and acknowledgement of errors and limitations.<sup>3</sup>

First of all the validation of software tools and/or clinical research computer systems like hypermodels is required by regulatory guidelines on Good Clinical Practice (GCP). The rationale for validation has a complex background but close to its complexity it makes as well a good business sense by ensuring quality, timeliness and efficiency.

According to the European Medicines Agency (EMA) and the World Health Organization (WHO), Good Clinical Practice (GCP) is a process, which incorporates international ethical and scientific quality standards for designing, recording and reporting trials that involve the participation of human subjects. Compliance with this standard provides public assurance that:

- 1) the rights, safety and wellbeing of trial subjects are protected;
- 2) the clinical trial data collected during the trial is credible.

The protection of clinical trial subjects is consistent with the principles set out in the Declaration of Helsinki with adoptions.<sup>4</sup> This is a statement of ethical principles developed by the World Medical Association.<sup>5</sup> Requirements for the conduct of clinical trials in the European Union (EU), including GCP and good manufacturing practice (GMP) and GCP or GMP inspections, are implemented in:

- the Clinical Trial Directive (Directive 2001/20/EC)<sup>6</sup>
- the GCP Directive (Directive 2005/28/EC).<sup>7</sup>

All hypermodels developed within CHIC are intended to be used in clinical care. To achieve this goal their validation within clinical trials is necessary and should be in strict conformance with the above EC directives.

Taking into consideration both the guidelines from the WHO and EMA, following are the basic principles of GCP<sup>8,9</sup>:

---

<sup>3</sup> Clapworthy GJ, Kohl P, Gregerson H, Thomas SR, Viceconti MD, Hose R, Pinney D, Fenner J, McCormack K and Lawford P, et al. Digital Human Modelling: A Global Vision and a European Perspective. Lecture Notes in Computer Science, 2007 (4561): 549-558.

<sup>4</sup> <http://www.wma.net/en/30publications/10policies/b3/> July 2011

<sup>5</sup> <http://www.wma.net/e/> July 2011

<sup>6</sup> DIRECTIVE 2001/20/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 4 April 2001, Official Journal of the European Communities, 2001

<sup>7</sup> COMMISSION DIRECTIVE 2005/28/EC of 8 April 2005, Official Journal of the European Union, 2005

<sup>8</sup> ICH Topic E 6 (R1) Guideline for Good Clinical Practice, Note for guidance on clinical practice, EMA, July 2002, CPMP/ICH/135/95

1. Clinical trials should be conducted in accordance with the ethical principles that have their origin in the Declaration of Helsinki and that are consistent with GCP and the applicable regulatory requirement(s).
2. Before a trial is initiated, foreseeable risks and inconveniences should be weighed against the anticipated benefit for the individual trial subject and society. A trial should be initiated and continued only if the anticipated benefits justify the risks.
3. The rights, safety and well-being of the trial subjects are the most important considerations and should prevail over interests of science and society.
4. The available non-clinical and clinical information on an investigational product should be adequate to support the proposed clinical trial.
5. Within GCP, clinical trials should be described in a clear, detailed protocol. The sponsor, often in consultation with one or more clinical investigators, generally designs the study protocol; clinical investigators may also design and initiate clinical studies, as sponsor-investigators. Integral to protocol development are the concepts of risk identification, study design and control groups, and statistical methodology. The sponsor and clinical investigator(s) should be aware of any national/ local laws or regulations pertaining to designing, initiating, and conducting the study.
6. A trial should be conducted in compliance with the protocol that has received prior Institutional Review Board (IRB)/Independent Ethics Committee (IEC) approval/favourable opinion.
7. The medical care given to, and medical decisions made on behalf of, subjects should always be the responsibility of a qualified physician or, when appropriate, of a qualified dentist.
8. Each individual involved in conducting a trial should be qualified by education, training, and experience to perform his or her respective task(s). Clinical investigators must be qualified and have sufficient resources and appropriately trained staff to conduct the investigation and be knowledgeable of the national setting and circumstances of the site and study population(s). Sponsors should review the requirements of the study protocol to determine the type(s) of expertise required and identify clinical investigators who have the particular medical expertise necessary to conduct the study and who have knowledge, training and experience in the conduct of clinical trials and human subject protection.
9. Freely given informed consent should be obtained from every subject prior to clinical trial participation. The clinical investigator has primary responsibility for recruiting subjects, ensuring that only eligible subjects are enrolled in the study, and obtaining and documenting the informed consent of each subject. Within GCP, informed consent must be obtained from each study subject prior to enrolment in the study or performing any specific study procedures.
10. All clinical trial information should be recorded, handled, and stored in a way that allows its accurate reporting, interpretation and verification.
11. The confidentiality of records that could identify subjects should be protected, respecting the privacy and confidentiality rules in accordance with the applicable regulatory requirement(s).
12. Investigational products should be manufactured, handled, and stored in accordance with applicable GMP. They should be used in accordance with the approved protocol. GCP requires that sponsors control access to the investigational product and also document the

---

<sup>9</sup> WHO-Handbook for good clinical research practice (GCP): guidance for Implementation; Clinical trials-methods. World Health Organization. ISBN 92 4 159392 X (NLM classification: W 20.5)



quantity(ies) produced, to whom the product is shipped, and disposition (e.g. return or destruction) of any unused supplies. GCP also requires investigators to control receipt, administration, and disposition of the investigational product.

13. Systems with procedures that assure the quality of every aspect of the trial should be implemented. Appropriate support systems and tools facilitate the conduct of the study and collection of data required by the protocol. Support systems and tools include, but are not limited to, trial-related information documents (e.g. investigator's brochure, case report forms [CRFs], checklists, study flow sheets, drug accountability logs, computer hardware and software, electronic patient diaries, and other specialized equipment. The sponsor is generally responsible for developing, maintaining, modifying, and ensuring the availability of support systems and tools for conducting the trial and collecting and reporting required data.

Hypermodels are not investigational drugs, but it seems worthwhile to regard them as such in the context of market approval. One of the reasons is, that physicians who should use them, are aware of the process for investigational drugs from preclinical testing to market approval.

As described in task 2.3 of the DoW the benefits of hypermodels in the domain of cancer and beyond needs to be shown. It is of utmost importance to demonstrate that *in silico* models and hypermodels will be used in clinical settings. The idea behind "*in silico* trials" is that a model is seen as a new drug that needs market approval.

To get a drug on the market, preclinical testing and clinical testing within phase I to IV trials are required. In comparison to the situation for drugs, the same needs to be done with *in silico* models and hypermodels. The preclinical phase will be done during this project. This preclinical phase needs to show that the tool or model delivers what is supposed to. This means that it gives correct answers without wrong calculations. This kind of validation refers to the correctness of the model from the mathematical viewpoint.

After this preclinical phase, clinical testing will start with phase I/II trials followed by phase III and phase IV trials. In phase I/II it is important to show that the logistics in running a hypermodel is according to the expectations of the clinicians: that the model can be used in due time, that it will deliver results in a timeframe that will support physicians in decision making, so that the model can be used as a decision supporting service. This is critical, as physicians cannot wait to start treatment in cancer or any other disease. If this problem is solved, a phase III trial can be conducted. According to phase III trials in drug development, such a trial will be a prospective randomized one, where the standard treatment is randomized against the treatment that is predicted by the model/hypermodel. At the end of the trial both treatment arms can be compared to see if patients treated according to the prediction of the model are doing better than those treated according to the standard conservative approach. In such a situation it is also possible to run the model in the group of patients treated according to the standard treatment. This will help to validate the model as one can see if the model is indicating the right treatment. E.g. A model indicates whether a preoperative chemotherapy scheme is better than primary surgery. One should look on tumour volume reduction predicted by the model. If the model predicts a reduction, chemotherapy is selected; if the model does not predict a reduction, go to primary surgery. In the standard arm all patients will receive preoperative chemotherapy and one could compare the reduction in tumour volume with the

predicted reduction by the model as a validation means of the model. Comparing the results between both approaches would reveal whether the model is beneficial for the outcome of a patient.

Such phase I/II/III trials will be outside the scope of this project, as the time needed to develop, initiate, run and analyse such trials is much longer than what the timeframe of the project allows. But within CHIC the basis for such trials will be developed.

The same way of validation can be used for hypermodels beyond the three cancer domains of nephroblastoma, glioblastoma and lung cancer. The hypermodel for prostate cancer will serve as the subsequent hypermodel going through this proposed validation schema.

## 7 References

- Antoniou KM, Tzouvelekis A, Alexandrakis MG, et al. Different angiogenic activity in idiopathic pulmonary fibrosis and pulmonary sarcoidosis. *Chest*. 2006;130:982–8.
- Argyri K, Dionysiou D, Stamatakis G Modeling the interplay between pathological angiogenesis and solid tumor growth: The anti-angiogenic treatment effect. *Advanced Research Workshop on In Silico Oncology and Cancer Investigation - The TUMOR Project Workshop (IARWISOCI), 2012 5th International* , vol., no., pp.1,4, 22-23 Oct. 2012
- Barberis L, Condat CA, Describing interactive growth using vector universalities, *Ecol Model*, 227, 2012, 56-63.
- Beckwith JB. New developments in the pathology of Wilms tumor. *Cancer Invest*. 1997;15:153–162.
- Beckwith, J. B., Kiviat, N. B. & Bonadio, J. F. Nephrogenic rests, nephroblastomatosis, and the pathogenesis of Wilms' tumor. *Pediatr. Pathol*. 10, 1-36 (1990)
- Bennington, J. & Beckwith, J. Tumors of the kidney, renal pelvis and ureter (Armed Forces Institute of Pathology, Washington DC, 1975).
- Colijn C, Mackey MC. A mathematical model of hematopoiesis--I. Periodic chronic myelogenous leukemia. *J Theor Biol*. 2005 Nov 21;237(2):117-32. Epub 2005 Jun 21.
- Danese S, Sans M, de la MC et al (2006) Angiogenesis as a novel component of inflammatory bowel disease pathogenesis. *Gastroenterology* 130:2060–2073
- DiPietro LA, Polverini PJ. Angiogenic macrophages produce the angiogenic inhibitor thrombospondin 1. *Am J Pathol* 1993; 143: 678–684.
- Dressler GR. The cellular basis of kidney development. *Annu Rev Cell Dev Biol*. 2006;22:509–529.
- Firestein GS. Evolving concepts of rheumatoid arthritis. *Nature*. 2003 May 15; 423(6937):356-61.
- Foley C, Mackey MC. Dynamic hematological disease: a review. *J Math Biol*. 2009 Jan;58(1-2):285-322. doi: 10.1007/s00285-008-0165-3. Epub 2008 Mar 4.
- Fransen JH, van der Vlag J, Ruben J, Adema GJ, Berden JH, Hilbrands LB. The role of dendritic cells in the pathogenesis of systemic lupus erythematosus. *Arthritis Res Ther*. 2010;12(2):207
- Galletto C, Gliozzi A, Nucera D, Bertorello N, Biasin E, Corrias A, Chiabotto P, Fagioli F, Guiot C, Growth impairment after TBI of leukemia survivors children: a model- based investigation, *Theoretical Biology and Medical Modelling* 2014, 11:44 DOI: 10.1186/1742-4682-11-44.
- Gliozzi AS, Guiot C, Delsanto PP, Iordache DA. , A novel approach to the analysis of human growth, *Theor Biol Med Model*. 2012 May 17; 9:17. doi: 10.1186/1742-4682-9-17
- Glykos, N.M., 2006. Software news and updates. Carma: a molecular dynamics analysis program. *J. Comput. Chem*. 27, 1765–1768. doi:10.1002/jcc.20482
- Green, D. et al. in *Principles and Practice of Pediatric Oncology* (eds Pizzo, P. & Poplack, D.) 733-759 (Lippincott-Raven, Philadelphia, 1997).

- R. Heidenreich, M. Rocken, K. Ghoreschi Angiogenesis drives psoriasis pathogenesis *Int J Exp Pathol*, 90 (2009), pp. 232–248
- Hilfiker-Kleiner D, Landmesser U, Drexler H. Molecular Mechanisms in Heart Failure : Focus on Cardiac Hypertrophy, Inflammation, Angiogenesis, and Apoptosis, *J Am Coll Cardiol* 48: A56–66, 2006.
- Hucka, M., Finney, A., Sauro, H.M., Bolouri, H., Doyle, J.C., Kitano, H., Arkin, A.P., Bornstein, B.J., Bray, D., Cornish-Bowden, A., Cuellar, A.A., Dronov, S., Gilles, E.D., Ginkel, M., Gor, V., Goryanin, I.I., Hedley, W.J., Hodgman, T.C., Hofmeyr, J.-H., Hunter, P.J., Juty, N.S., Kasberger, J.L., Kremling, A., Kummer, U., Novère, N.L., Loew, L.M., Lucio, D., Mendes, P., Minch, E., Mjolsness, E.D., Nakayama, Y., Nelson, M.R., Nielsen, P.F., Sakurada, T., Schaff, J.C., Shapiro, B.E., Shimizu, T.S., Spence, H.D., Stelling, J., Takahashi, K., Tomita, M., Wagner, J., Wang, J., 2003. The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics* 19, 524–531. doi:10.1093/bioinformatics/btg015
- Janeway CA Jr, Travers P, Walport M, et al. (Eds) *Immunobiology: The Immune System in Health and Disease*. 5th edition. New York: Garland Science; 2001.
- Jiang, N., Cox, R.D., Hancock, J.M., 2007. A kinetic core model of the glucose-stimulated insulin secretion network of pancreatic beta cells. *Mamm. Genome Off. J. Int. Mamm. Genome Soc.* 18, 508–520. doi:10.1007/s00335-007-9011-y
- Kang DH, Anderson S, Kim YG, Mazzali M, Suga S, Jefferson JA, Gordon KL, Oyama T, Hughes J, Hugo C, Kerjaschki D, Schreiner GF, Johnson RJ: Impaired angiogenesis in the aging kidney: Vascular endothelial growth factor and thrombospondin-1 in renal disease. *Am J Kidney Dis* 37 : 601-611, 2001
- Khurana, R., et al., Role of Angiogenesis in Cardiovascular Disease—A Critical Appraisal, *Circulation* 112:1813-1824, Sep. 2005.
- Kolokotroni Eleni, Ouzounoglou Eleftherios, Stanulla Martin, Dionysiou Dimitra and Georgios S. Stamatakis. In *Silico Oncology: Developing and Clinically Adapting the Acute Lymphoblastic Leukemia (ALL) Oncosimulator by Exploiting Pathway Based Gene Expression Analysis in the Context of the ALL-BFM 2000 Clinical Study*. Virtual Physiological Human Conference 2014 (VPH 2014) Trondheim, Norway, 2014
- Lebre MC, Tak PP. Dendritic cell subsets: their roles in rheumatoid arthritis. *Acta Reumatol Port.* 2008 Jan-Mar;33(1): 35-45.
- Li, C. M. et al. Gene expression in Wilms' tumor mimics the earliest committed stage in the metanephric mesenchymal-epithelial transition. *Am. J. Pathol.* 160, 2181-2190 (2002)
- Li, W., Kessler, P. & Williams, B. R. Transcript profiling of Wilms tumors reveals connections to kidney morphogenesis and expression patterns associated with anaplasia. *Oncogene* 24, 457-468 (2005)
- Li, C., Donizelli, M., Rodriguez, N., Dharuri, H., Endler, L., Chelliah, V., Li, L., He, E., Henry, A., Stefan, M.I., Snoep, J.L., Hucka, M., Novère, N.L., Laibe, C., 2010. BioModels Database: An enhanced, curated and annotated resource for published quantitative kinetic models. *BMC Syst. Biol.* 4, 92. doi:10.1186/1752-0509-4-92
- Longo, N., Wang, Y., Smith, S.A., Langley, S.D., DiMeglio, L.A., Giannella-Neto, D., 2002. Genotype-phenotype correlation in inherited severe insulin resistance. *Hum. Mol. Genet.* 11, 1465–1475.

- Michael C. Mackey. Mathematical models of hematopoietic cell replication and control. In Case Studies in Mathematical Modeling---Ecology, Physiology, and Cell Biology Edited by Othmer H, Adler F, Lewis M, Dallon J. New Jersey: Prentice Hall; 1997
- A. Marrelli, P. Cipriani, V. Liakouli, F. Carubbi, C. Perricone, R. Perricone, et al. Angiogenesis in rheumatoid arthritis: a disease specific process or a common response to chronic inflammation? *Autoimmun Rev*, 10 (2011), pp. 595–598
- Maynard S.E., Karumanchi S.A. Angiogenic factors and preeclampsia. *Semin. Nephrol.* 2011;31:33–46.
- Mendonca, R. J. and J. Coutinho-Netto (2009). "Cellular aspects of wound healing." *An Bras Dermatol* 84(3): 257-62.
- Miguel N. Rivera & Daniel A. Haber, Wilms' tumour: connecting tumorigenesis and organ development in the kidney *Nature Reviews Cancer* 5, 699-712 (September 2005) | doi:10.1038/nrc1696
- Morris, G.M., Huey, R., Lindstrom, W., Sanner, M.F., Belew, R.K., Goodsell, D.S., Olson, A.J., 2009. AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility. *J. Comput. Chem.* 30, 2785–2791. doi:10.1002/jcc.21256
- Peixoto D., Dingli D., Pacheco J.M., Modelling hematopoiesis in health and disease. *Mathematical and Computer Modelling* Volume 53, Issues 7–8, April 2011, Pages 1546–1557
- Phillips, J.C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R.D., Kalé, L., Schulten, K., 2005. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* 26, 1781–1802. doi:10.1002/jcc.20289
- Pollard JW. Trophic macrophages in development and disease. *Nat Rev Immunol* 2009; 9: 259–270.
- Poleszczuk J, Bodnar M, Foryś U (2011) New approach to modelling of antiangiogenic treatment on the basis of Hahnfeldt et al. model. *Math Biosci Eng* 8(2):591-603. doi: 10.3934/mbe.2011.8.591
- J. P. Polverini, "Angiogenesis and wound healing: basic discoveries, clinical implications and therapeutic opportunities," in *Oral Wound Healing: Cell Biology and Clinical Management*, H. Larjava, Ed., pp. 175–195, John Wiley & Sons, 1st edition, 2012.
- Prager GW, Poettler M, Unseld M, Zielinski CC (2012) Angiogenesis in cancer: Anti-VEGF escapes mechanisms. *Transl Lung Cancer Res* 1: 14–25.
- Reynolds, L. P., Killilea, S. D. & Redmer, D. A. (1992) Angiogenesis in the female reproductive cycle. *FASEB J.*, 6, 886–892.
- D. Ribatti, I. Puxeddu, E. Crivellato, B. Nico, A. Vacca, and F. Levi-Schaffer, Angiogenesis in asthma, *Clinical and Experimental Allergy*, vol. 39, no. 12, pp. 1815–1821, 2009.
- Risau, V. (1997) Mechanisms of angiogenesis. *Nature*, 386, 671–674
- Rocha AL, Reis FM, Taylor RN. Angiogenesis and endometriosis. *Obstetrics and gynecology international*. 2013;2013:859619.
- Ross JJ: Angiogenic gene therapy as a potential therapeutic agent in chronic osteomyelitis. *Med Hypotheses* 2006, 67:161-163
- Saxen L, Sariola H. Early organogenesis of the kidney. *Pediatr Nephrol.* 1987;1:385–392.

- Schmidt D, Beckwith JB. Histopathology of childhood renal tumors. *Hematol/Oncol Clin North Am.* 1995;9:1179–1200.
- Sung, M.-H., Hager, G.L., 2012. Nonlinear Dependencies of Biochemical Reactions for Context-specific Signaling Dynamics. *Sci. Rep.* 2. doi:10.1038/srep00616
- Tak PP. Examination of the synovium and synovial fluid. In: Wollheim FA, Firestein GS, Panayi G, eds. *Rheumatoid arthritis: Frontiers in pathogenesis and treatment.* Oxford: Oxford University Press, Inc.; 2006:229-241
- Tan, F.-L., Moravec, C.S., Li, J., Apperson-Hansen, C., McCarthy, P.M., Young, J.B., Bond, M., 2002. The gene expression fingerprint of human heart failure. *Proc. Natl. Acad. Sci.* 99, 11387–11392. doi:10.1073/pnas.162370099
- L. E. Tellechea A, A. Veves, and E. Carvalho, “Inflammatory and angiogenic abnormalities in diabetic wound healing: role of neuropeptides and therapeutic perspectives,” *the Open Circulation and Vascular Journal*, vol. 3, pp. 43–55, 2010.
- Voelkel NF, Douglas IS, Nicolls M. Angiogenesis in chronic lung disease. *Chest* 2007;131:874–879.
- Wu GF, Laufer TM. The role of dendritic cells in multiple sclerosis. *Curr Neurol Neurosci Rep.* 2007 May;7(3):245-52.