



SIGBio Record

Newsletter of the SIGBio
ACM Special Interest Group

TABLE OF CONTENTS

3 Editor's notes, *Pietro Hiram Guzzi, Young-Rae Cho, Pierangelo Veltri*

4 Chair's message, *Aidong Zhang*

Contributed Articles

5 Connecting digital cancer model repositories with markup:

Introducing TumorML version 1.0, by *David Johnson, Steve McKeever, Thomas S. Deisboeck, Zhihui Wang*

12 Computational Biologist Profile:

Ruth Nussinov, by *Amarda Shehu*

Conference, Workshop and Symposium Reports

15 ACB Bioinformatics, Computational Biology and Biomedical Informatics Conference Report (ACM BCB 2013), by *Cathy Wu and Sridhar Hannenhalli*

16 ACB BCB 2013 Paper and Poster awards, by *Dongxiao Zhu and Srinivas Aluru*

About the newsletter

18 SigBio Record - Submission Guidelines

Editor's Notes

The SIGBio community is growing and the research interests are broadening. We presented the new structure of the SIGBio Record during the SIGBio meeting at ACM BCB Conference. Consequently, we hope the SIGBio Record will contain in future more articles and perspective covering novel arguments. We also propose a new structure for the Record with contribute articles and discussions, workshop and conference reports, scientist profiles, all topics interesting for SIGBio community. Moreover, the SIGBio community will host health informatics topics: we welcome Pierangelo Veltri as novel associate editor in charge of gathering contributions in such area.

The current issue presents a contributed article describing an XML based repository for digital cancer data by A. Joh, and a scientist profile description of Ruth Russinov, by Amarda Shehu.

Also, a short report of the ACM Bioinformatics, Computational Biology and Biomedical Informatics (ACM BCB 2013) held in Washington D.C. (September 22-25). The conference had high number of participants, hosted included workshops and a Symposium on Health Informatics. The conference report also include best papers and best poster awards assignement.

We thank contributors for this issue and hope that readers will find interesting references to their work in Bioinformatics and Health Informatics area.

Pietro Hiram Guzzi, Young-Rae Cho, Pierangelo Veltri - SIGBio Record Editors

Notice to Contributing Authors to SIG Newsletters

By submitting your article for distribution in this Special Interest Group publication, you hereby grant to ACM the following non-exclusive, perpetual, worldwide rights: (i) to publish in print on condition of acceptance by the editor (ii) to digitize and post your article in the electronic version of this publication (iii) to include the article in the ACM Digital Library and in any Digital Library related services (iv) to allow users to make a personal copy of the article for noncommercial, educational or research purposes. However, as a contributing author, you retain copyright to your article and ACM will refer requests for republication directly to you.

Chair's message

SIGBIO was established in January 2011 and was chartered as transitional and had up to a 2 year period to perform to expectations before moving to full SIG status. On October 1, 2013, ACM Special Interest Groups (SIG) Governing Board met and with great satisfaction of SIGBio's program performance, the board granted SIGBIO the full SIG status for the next 4 years.

Aidong Zhang

Connecting digital cancer model repositories with markup: Introducing TumorML version 1.0

David Johnson^{*}
Discovery Sciences Group
Department of Computing
Imperial College London
Huxley Building, South
Kensington Campus
London, SW7 2AZ, UK
david.johnson@imperial
.ac.uk

Steve McKeever
Department of Informatics &
Media
Uppsala University
Box 513, 751 20 Uppsala
Sweden
steve.mckeever@im.uu.se

Thomas S. Deisboeck[†]
Harvard-MIT (HST) Athinoula
A. Martinos Center for
Biomedical Imaging
Massachusetts General
Hospital
Charlestown, MA 02129
deisboec@helix.mgh
.harvard.edu

Zhihui Wang
Department of Pathology
The University of New Mexico
Albuquerque, NM 87131
ZWang@salud.unm.edu

ABSTRACT

The cancer research community requires a standardized way of describing mathematical and computational models to enable interoperability between systems, repositories, and between the models themselves. In this paper we describe a new markup language, TumorML¹, for describing computational models that fall within the domain of cancer. TumorML is an XML-based markup language that wraps existing cancer model implementations with metadata for model curation, parametric interface description, implementation description, and compound model linking.

In this paper we first introduce the rationale for a new markup language for computational cancer model description based on our experiences and requirements from the European Commission's 'Transatlantic Tumor Model Repositories' project. The aim of the project is to develop a European-based digital cancer model repository to link and interoperate with a similar established repository based in the United States. TumorML was developed to enable this interoperability between repositories. We introduce the language and describe the main features of the specification and go on to describe a real application of TumorML where

a molecular pathway model has been packaged using the new markup language.

Categories and Subject Descriptors

[**Applied computing**]: Life and medical sciences—*Bioinformatics*; [**Computing methodologies**]: Modeling and simulation—*Model development and analysis*

General Terms

Design, Languages, Standardization

1. INTRODUCTION

The 'Transatlantic Tumor Model Repositories' project (TUMOR) aims to develop a European digital cancer model repository [14], building on past and existing European Commission (EC) projects, to interoperate with the US National Institute of Health/National Cancer Institute (NIH/NCI) semantic-layered cancer research platform, CViT (Center for the Development of a Virtual Tumor) [20]. The ultimate goal is to bridge cancer research communities across the Atlantic through the provision of internationally available data and computing services for cancer modelers, researchers and, ultimately, clinicians to support both basic scientific research in cancer and to develop personalized computer-aided cancer diagnosis and treatment.

Biological model repositories are not novel, as demonstrated by existing open source software such as the Physiome Model Repository [19], and open data services such as the European Bioinformatics Institute's BioModels Database [12]. However, one of the key aims of TUMOR is to enable a European-based cancer model repository to seamlessly interoperate with its US counterpart, the CViT Digital Model Repository (DMR). As an enabling technological component of the project, a simulation markup language specifically targeted at the cancer modeling domain was developed to act as the standard communication format between elements of

^{*}Corresponding author.

[†]On leave of absence; current address: ThinkMotu LLC, Wellesley, MA 02481. Email: ts.deisboeck@thinkmotu.com

¹<http://www.github.com/tumorml>

the TUMOR infrastructure [15], and for exporting models externally.

1.1 Rationale

We have discussed extensively the need for dealing with diversity in computational cancer modeling [10, 8, 9, 7]. Our requirements are based on the premise that existing biological markup languages are not suited to the immediate needs for state-of-the-art cancer model description. For example, the ‘oncosimulator’ set of *in silico* cancer models [16] developed by our colleagues in precursor projects to TUMOR produce simulations of tumor growth based on a combination of techniques including: nondeterministic finite-state automata, Monte Carlo methods, differential equations to simulate population dynamics of a tumor mass, and pharmacokinetic/pharmacodynamic (PK/PD) interactions with candidate chemical or radiation treatments. Existing markup languages include CellML [5] and SBML (Systems Biology Markup Language) [4] however we deemed both unsuitable for our needs due to lack of expressive notation for statistical models, incorporating random factors necessary in Monte Carlo simulations, or for procedural operations. Oncosimulator models also consider multiple scales in aggregate, covering molecular scale interactions and the clinical perspective using patient histories, description of which cannot be integrated into either CellML or SBML models. Further arguments are described in our 2013 *Cancer Informatics* commentary [10].

2. THE TUMOR MODEL REPOSITORIES MARKUP LANGUAGE

TumorML (Tumor model repositories Markup Language) was developed to overcome the limitations of existing markup languages, not as a competitor to either of CellML and SBML, but to deal with storing and transmitting existing cancer models among research communities. We have previously described the concept and requirements for TumorML [8] and in this paper we present an introduction and highlights of the first published specification of the markup language developed by TUMOR. TumorML has been developed as an XML-based domain-specific vocabulary that includes elements from existing vocabularies to avoid ‘reinventing the wheel’. The vocabularies, reused in part, in TumorML version 1.0 include the following:

1. Dublin Core - For basic curation of model description documents and model implementations, we reuse a subset of the Dublin Core Metadata Element Set [11].
2. xCard - To describe entities that create the model descriptions themselves we use an XML representation of vCards [13].
3. BibTeXXML - Whilst not a standardized XML vocabulary, BibTeXXML is a representation of the BibTeX format for bibliographic references [6]. In this first version of TumorML we reuse the document categories with a view to implementing the full BibTeX reference structure in future TumorML versions.
4. JSDL - The Job Description Markup Language (JSDL) provides an established vocabulary for specifying Grid

execution jobs, including job execution requirements [1]. A subset of the JSDL execution requirements description is used in TumorML for describing the basic hardware and software requirements in order to run TumorML packaged implementations with a computational framework or batch execution system.

5. xMML - The Multiscale Modeling Language, MML, and its XML version, xMML, proposes a standard way of describing multiscale models and, most significantly, how to couple models of different simulated scales [2]. TumorML version 1.0 reuses a simplified part of the xMML coupling markup to represent the coupling of computational models.

2.1 Overview

The full specification of TumorML 1.0 is published as an EC deliverable by TUMOR. In this paper we will describe the main elements and concepts that make up the markup language used in TumorML model description documents. The root element of every TumorML document is `<tumorml>`, with a `<header>` and `<model>` as children. The `<header>` block contains metadata about the model, and the `<model>` block description of the model itself in terms of its parametric interface and implementation. A TumorML document follows the general pattern shown in Listing 1:

Listing 1: General pattern of TumorML documents.

```

1 <?xml version="1.0"?>
2 <tumorml
  xmlns="http://www.tumor-project.eu/tumorml/1.0"
  id="tp53_pathway" >
3   <!-- cancer model metadata -->
4   <header>
5     ...
6   </header>
7   <!-- cancer model description -->
8   <model>
9     <parameters>
10      <!-- model parameters list -->
11      ...
12    </parameters>
13    <implementation>
14      <!-- implementation description -->
15      ...
16    </implementation>
17    ...
18  </model>
19 </tumorml>

```

2.1.1 Header

The `<header>` element contains a list of metadata elements to aid in publishing, search and retrieval of TumorML models, and a list of external references to attribute. There may only be one `<header>` per TumorML document. The header consists of a set of general document curation elements taken from Dublin Core, a set of TUMOR repository specific elements, and a list of references where applicable.

The Dublin Core elements are inherited from the Dublin Core Metadata Element Set, Version 1.1. The TumorML specification does not use the element set exhaustively, but takes key elements including `<title>`, `<creator>`, `<description>`, `<publisher>`, `<contributor>` and `<date>`. TumorML

Table 1: Domain-specific taxonomic metadata for TumorML model descriptions.

Metadata element	Description	Values
<math>	The mathematical type that the cancer simulation model is based on.	discrete, continuous, hybrid
<biocomplexityDirection>	The biocomplexity direction of the model.	topDown, bottomUp, middleOut
<cancer>	The type of cancer described by the model.	glioma, nephroblastoma, lung, breast
<materialization>	The materialization of the cancer.	solid, liquid
<homogeneity>	The homogeneity of the tumor.	imageable, non-imageable
<imageBasedDetectability>	Whether or not the tumor can be detected in an image.	homogeneous, non-homogeneous
<freeGrowth>	Whether or not the model includes simulation of free growth.	true, false
<treatmentIncluded>	Whether or not the simulated tumor includes simulation of treatments.	true, false

also uses a domain-specific taxonomy of models relating to TUMOR that were designed as part of the database schema, summarized in Table 1, where the schema does not strictly enforce the enumerations, but also allows free text strings to be used. A relevant discussion on the different and diverse types of cancer models has been published by the authors in [10]. A list of <reference> elements describe bibliographic references of models and related components of model descriptions to attribute work done elsewhere that the model description is based on. References have a <title>, <source> containing a URL, <creator>, full text <citation>, and a <type> that categorizes the reference using BibTeXXML categories.

2.1.2 Model

The `model` element contains the description of a computational model, including its input and output interface, included implementations, and where applicable, compound sub-model topologies. A definition of a model serves as a description of the standard way to interface with the model via exposed parameters (both inputs and outputs), and as a container for any descriptions of model implementations. A model is treated like a ‘black box’ where the details of the model implementation are of no concern. What is of interest is how to communicate data to and from the model, as well as how to execute an implementation of a model, therefore enabling us to wrap up and publish existing implementations with TumorML.

As described in [8], models can be either ‘simple’ or ‘complex’, and this is reflected in the XML schema by having a choice of two patterns that can be enclosed within the <model> tag. To recap, a simple model description provides a wrapping up of a single computational cancer model, while a complex model describes a compound model; an aggregate of simple or complex models wrapped in TumorML.

A simple model consists of two key descriptors (an example of which shown in the general pattern in Listing 1): an input and output parameter specification, described with a <parameters> block with <in> and <out> elements to define inputs and outputs. Following this, at least one <implementation> block is used to describe a software implementation of the model. In each implementation specification, a description of the files that make up the model is included, as well

as a descriptor of how to run the entry point to the packaged files, along with the requirements for execution. A complex model contains the same header elements as a simple model, however under the model description section contained with <model> elements, multiple simple or complex models are referenced within <submodel> elements. Each sub-model is labelled with an identifier that is unique within the root <model> declaration. An additional section is used to define instances of models and a topology of model parameter connections. This is illustrated in Listing 2, where the example shown declares a single sub-model, `modelA`, with two parameters, and a topology consisting of two instances of `modelA` connected from instance A’s output parameter `p2` to instance B’s input parameter `p1`.

Listing 2: Complex model pattern fragment.

```

1 <model>
2   <sub-model name="modelA">
3     ...
4     <parameters>
5       <in name="p1">
6         <value type="double" />
7       </in>
8       <out name="p2">
9         <value type="double" />
10      </out>
11    </parameters>
12    ...
13  </sub-model>
14  <topology>
15    <instance id="A" sub-model="modelA" />
16    <instance id="B" sub-model="modelA" />
17    <coupling from="A.p2" to="B.p1" />
18  </topology>
19 </model>

```

3. AN EXAMPLE MODEL

Within TUMOR a number of models contributed by project partners from both the US and EU were wrapped in TumorML markup, including a TumorML complex model description developed using a model from each of the CViT DMR and the EU TUMOR repository, the details of which will be published in a forthcoming paper. We use one of the models taken from the CViT DMR here as an exemplar of how TumorML has been used to wrap up and package an existing model implementation. First described in Wang et

al [18], the US partners in TUMOR published the model implementation of an EGFR-ERK pathway in the CViT DMR. Listing 3 gives an example model wrapped up in TumorML markup.

Wang et al developed a multiscale model for investigating expansion dynamics of non-small cell lung cancer (NSCLC) within a small two-dimensional *in silico* mesh. At the molecular level, a specific EGFR-ERK intracellular signaling pathway was implemented. Alterations in these molecules are used to trigger phenotypic changes at the cellular level. As described fully in [18], the authors validated that increasing the amount of available growth factor leads to a spatially more aggressive cancer system, by computationally examining the relationship between extrinsic ligand concentrations, intrinsic molecular profiles and microscopic patterns. The kinetic model of the implemented NSCLC-specific molecular signaling pathway, which consists of 20 protein molecules, is illustrated in Figure 1. These proteins, including both receptor (epidermal growth factor receptor; EGFR) and non-receptor kinases, such as PLC γ , have previously been experimentally or clinically proven to play an important role in NSCLC tumorigenesis.

Functionally speaking as a computer-based implementation, the model takes as an input parameter an EGF concentration, simulates both the molecular signalling pathway and interactions between sub cellular molecules as a set of ordinary differential equations (ODEs) to produce as an output the cell cycle duration. This is then used as the basis for simulating the phenotypic changes and consequential proliferation or migration of cells in the two-dimensional mesh over a number of discrete time steps. In the following section we describe the markup used to wrap up the EGFR-ERK pathway model.

3.1 Model markup

In the EGFR-ERK pathway example, the metadata is extracted from the corresponding entry taken from the CViT DMR and inserted into a TumorML document under the `<header>` element. The `<description>` element is filled with the text of a short abstract, including references to the relevant published papers about the model itself. Next, stakeholders are described with the main author/developer of the model listed as the model `<creator>`, the publishing entity in the `<publisher>` element, in this case the Complex Biosystems Modeling Laboratory at Massachusetts General Hospital, Boston, MA, and finally other associated authors listed under `<contributor>`. Following a timestamp stored under `<date>`, we list the TUMOR repository taxonomic elements. In the case of the EGFR-ERK pathway model, the model uses `continuous` mathematics, is considered to take a `bottomUp` approach, concerns `lung cancer`, simulates the cancer as a `solid homogenous tumor` that is `imageable` and does not simulate `freeGrowth` nor treatment.

The model and implementation description is provided under the `<model>` element. First, the input and output parameters are described to enable the model to be run in an execution environment, like CViT's Computational Model Execution Framework [17], and linked with other TumorML model descriptions as a complex model. In the case of the EGFR-ERK pathway model here, it is a simple model de-

scription as it only concerns the one published implementation.

Under the `<parameters>` element an input parameter of `value type double` (double-precision floating point number) is used to represent the EGF concentration, labelled as `egf`. Two output parameters return the `cell cycle time` (an estimate average cell cycle time) and `PLC $\gamma$  concentration`, both also of type `double`. The implementation of the model has also been provided under the `<implementation>` element. In this case a LSID (Life Sciences Identifier [3]) has been used as the implementation `id`, the value of which has been taken directly from the CViT DMR. Under the `<package>` element a URL to the model source code has been provided. Under `<command>` the command-line interface instruction to execute the model is given. In this case, `EGFR_ODE_EC` is the executable file, while `$egf` is the input EGF concentration. Note that the command-line parameter `$egf` maps directly to `egf` as defined in the input parameters list earlier. It is assumed the output parameters, as specified in the `<parameters>` element are written out to a text file each corresponding to the parameter name, in this example i.e. `cell cycle time.txt` and `PLC $\gamma$ .txt`. Note that these input and output parameters map directly to the earlier definition in the `<parameters>` element, and in turn model the variable inputs and outputs of the actual model as shown in Figure 1. The internal processing implemented as the set of ODEs is treated as a 'black box' where it is of no concern to the functional description in TumorML. Finally, the minimum requirements for executing a model simulation are specified using standardized vocabulary taken from JSDL. In this case, `<operatingSystem>` has been specified as Linux, the `<CPUArchitecture>` as x86 64-bit, and the source code `<language>` as C++.

4. CONCLUSIONS

In developing TumorML our success criteria consisted of the following two goals. First, being able to demonstrate the import and export of models between the two repositories preserving as much metadata as possible, and translating between the CViT DMR and TumorML schemas where appropriate. Second, demonstrating a 'transatlantic' combination of models linked together via markup describing each component model's interfaces and their couplings.

Within TUMOR the exchange of models between the CViT DMR and the TUMOR model repository has been demonstrated, one of which, the EGFR-ERK pathway model, is described in this paper. Work towards the construction and execution of a compound model of modules taken from each repository has also been carried out and will be reported in a future publication. However, a greater validation of the markup's usage would be in its adoption by open and public repositories to broaden its currently small community profile. To this end an open-source reference database, supporting TumorML documents natively, is being developed to go alongside the published TumorML schemas to act as a publicly available demonstrator of the markup language.

Listing 3: Full listing of the EGFR-ERK pathway module wrapped up in TumorML 1.0 markup.

```

1 <?xml version="1.0"?>
2 <tumorml xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
3   xsi:schemaLocation="http://www.tumor-project.eu/tumorml/1.0" id="egfr_erk_pathway">
4   <header>
5     <title>EGFR-ERK Pathway</title>
6     <description>
7       This is a multiscale agent-based model for investigating expansion dynamics of epithelial cancers
8       (e.g., glioma, NSCLC) within a two-dimensional microenvironment. At the molecular level, we present a
9       specific EGFR-ERK intracellular signaling pathway. The goal of this work is to provide useful insights
10      into the quantitative understanding of the relationship between signaling properties of underlying
11      molecular changes and the multi-cellular responses they may trigger.
12
13      This particular version of the model has been first applied to non small cell lung cancer (NSCLC) and
14      has been published in Theoretical Biology and Medical Modelling 2007, 4:50.
15      (http://www.tbiomed.com/content/4/1/50). A follow-up work on cross scale sensitivity analysis of this
16      model has been published in BioSystems, 92(3): 249-258, 2008.
17    </description>
18    <creator>
19      <person id="zhiui_wang">
20        <fullname>Zhihui Wang</fullname>
21      </person>
22    </creator>
23    <publisher>
24      <person id="massachusetts_general_hospital">
25        <fullname>Complex Biosystems Modeling Laboratory (CBML) Mass. General Hospital</fullname>
26      </person>
27    </publisher>
28    <contributor>
29      <person id="thomas_s_deisboeck">
30        <fullname>Thomas S. Deisboeck, M.D.</fullname>
31      </person>
32    </contributor>
33    <date>2012-06-22T00:00:00+00:00</date>
34    <math>continuous</math>
35    <biocomplexityDirection>bottomUp</biocomplexityDirection>
36    <cancer>Lung Cancer</cancer>
37    <materialization>solid</materialization>
38    <homogeneity>homogeneous</homogeneity>
39    <imageBasedDetectability>imageable</imageBasedDetectability>
40    <freeGrowth>false</freeGrowth>
41    <treatmentIncluded>false</treatmentIncluded>
42  </header>
43  <model>
44    <parameters>
45      <in name="egf" optional="0">
46        <value type="double" />
47      </in>
48      <out name="cell cycle time" optional="0">
49        <value type="double" />
50      </out>
51      <out name="PLC_g" optional="0">
52        <value type="double" />
53      </out>
54    </parameters>
55    <implementation id="urn:lsid:cvit.org:cmef:0.919920521935164">
56      <title>
57        EGFR-ODE Model for EC revision #3 (6/25/2012) from Massachusetts General Hospital. Calculates
58        Cell Cycle Time for EGF concentration. (Updated for command line parameters)
59      </title>
60      <date>2012-06-25T00:00:00+00:00</date>
61      <package name="EGFR_ODE_EC" checksum="">
62        <file name="EGFR_ODE_EC"
63          source="http://mgh-cvit.infotechsoft.com/.../2012/5/25/EGFR_ODE-2012-06-25.zip" />
64      </package>
65      <command>EGFR_ODE_EC $egf</command>
66      <requirements>
67        <operatingSystem>linux</operatingSystem>
68        <CPUArchitecture>x86_64</CPUArchitecture>
69        <language>cpp</language>
70      </requirements>
71    </implementation>
72  </model>
73 </tumorml>

```

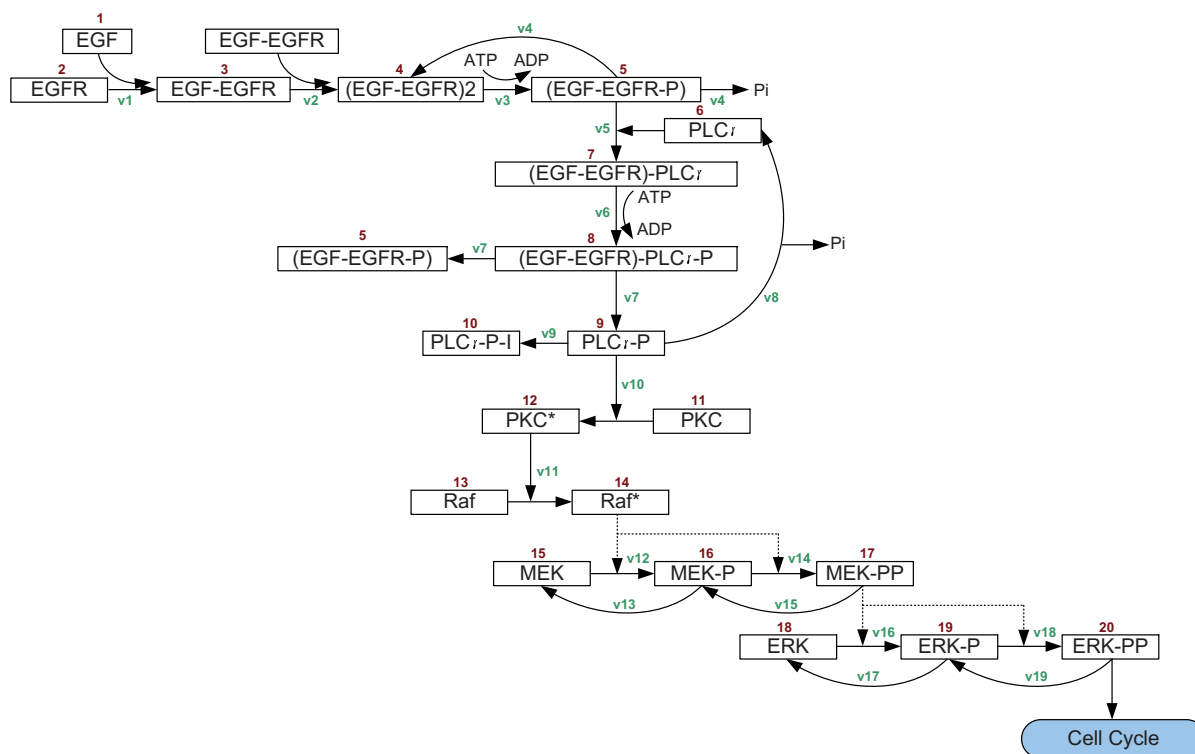


Figure 1: Implementation of the EGFR-ERK molecular pathway. Note that EGF concentration functions as the main variable input, and cell cycle time as a generated output. Diagram originally published by BioMed Central in Z. Wang et al in [18], reproduced with permission.

5. ACKNOWLEDGMENTS

This work was initially supported in part by the EC Seventh Framework Programme (FP7) under the ‘Transatlantic Tumor Model Repositories’ project (Contract #FP7-ICT-2009.5.4-247754) from 2010-2013. The authors gratefully acknowledge contributions from all members of the TUMOR consortium, in particular the following and their respective research & development teams: Konstantinos Marias and Vangelis Sakkalis (Foundation for Research & Technology - Hellas), Georgios Stamatakis (National Technical University of Athens), Norbert Graf (Saarland University Hospital), and Thomas Taylor (InfoTech Soft Inc., Miami, FL).

TumorML is now maintained as open-source software hosted on GitHub (<http://www.github.com/tumorml>) where we actively invite contributions to its development and community adoption.

6. REFERENCES

- [1] A. Anjomshoaa, F. Brisard, M. Drescher, D. Fellows, A. Ly, S. McGough, D. Pulsipher, and A. Savva. Job Submission Description Language (JSDL) specification, version 1.0. Technical report, Global Grid Forum, November 2005.
- [2] J. Borgdorff, E. Lorenz, A. Hoekstra, J. Falcone, and B. Chopard. A principled approach to distributed multiscale computing, from formalization to execution. In *Proc IEEE eScience Workshops*, pages 97–104, 2011.
- [3] T. Clark, S. Martin, and T. Liefeld. Globally distributed object identification for biological knowledgebases. *Brief Bioinform*, 5(1):59–70, March 2004.
- [4] A. Finney and M. Hucka. Systems biology markup language: Level 2 and beyond. *Biochem Soc Trans*, 31(Pt 6):1472–1473, Dec. 2003.
- [5] A. Garny, D. Nickerson, J. Cooper, R. W. dos Santos, S. McKeever, P. Nielsen, and P. Hunter. CellML and associated tools and techniques. *Philos T R Soc A*, 366(1878):3017–3043, 2008.
- [6] V. B. Gundersen and Z. W. Hendrikse. BibTeX as XML markup. <http://bibtexml.sf.net/>. Accessed online 2013-08-11.
- [7] D. Johnson, J. Cooper, and S. McKeever. Markup languages for in silico oncology. In *Proceedings of the 4th International Advanced Research Workshop on In Silico Oncology and Cancer Investigation (4th IARWISOCI) - The ContraCancrum Workshop*, pages 108–110, September 2010.
- [8] D. Johnson, J. Cooper, and S. McKeever. TumorML: Concept and requirements of an in silico cancer modelling markup language. In *Conf Proc IEEE Eng Med Biol Soc. IEEE EMBS*, September 2011.
- [9] D. Johnson and S. McKeever. The case for using markup for biomechanical modelling. In *Proceedings of the XXIII Congress of the International Society of Biomechanics*. ISB, July 2011.
- [10] D. Johnson, S. McKeever, G. Stamatakis, D. Dionysiou, N. Graf, V. Sakkalis, K. Marias,

- Z. Wang, and T. S. Deisboeck. Dealing with diversity in computational cancer modeling. *Cancer Informatics*, 12:115–124, 05 2013.
- [11] J. Kunze and T. Baker. The Dublin Core Metadata Element Set. RFC 5013 (Informational), Aug. 2007.
- [12] C. Li, M. Donizelli, N. Rodriguez, H. Dharuri, L. Endler, V. Chelliah, L. Li, E. He, A. Henry, M. I. Stefan, J. L. Snoep, M. Hucka, N. Le Novère, and C. Laibe. BioModels Database: An enhanced, curated and annotated resource for published quantitative kinetic models. *BMC Syst Biol*, 4:92, Jun 2010.
- [13] S. Perreault. xCard: vCard XML Representation. RFC 6351 (Proposed Standard), Aug. 2011. Updated by RFC 6868.
- [14] V. Sakkalis, K. Marias, G. Stamatakis, F. Misichroni, D. Dionysiou, S. McKeever, D. Johnson, T. S. Deisboeck, and N. Graf. The TUMOR project: Integrating cancer model repositories for supporting predictive oncology. In *Abstract booklet for VPH2012 Integrative Approaches to Computational Biomedicine*, September 2012.
- [15] S. Sfakianakis, V. Sakkalis, K. Marias, G. Stamatakis, S. McKeever, T. S. Deisboeck, and N. Graf. An architecture for integrating cancer model repositories. *Conf Proc IEEE Eng Med Biol Soc*, 2012:6628–31, 2012.
- [16] G. S. Stamatakis, D. D. Dionysiou, N. M. Graf, N. A. Sofra, C. Desmedt, A. Hoppe, N. K. Uzunoglu, and M. Tsiknakis. The ‘oncosimulator’: a multilevel, clinically oriented simulation system of tumor growth and organism response to therapeutic schemes. towards the clinical evaluation of in silico oncology. *Conf Proc IEEE Eng Med Biol Soc*, 2007:6629–32, 2007.
- [17] Z. Wang, J. Sagotsky, T. Taylor, P. Shironoshita, and T. S. Deisboeck. Accelerating cancer systems biology research through semantic web technology. *Wiley Interdisciplinary Reviews Systems Biology and Medicine*, 5(2):135–151, Mar-Apr 2013.
- [18] Z. Wang, L. Zhang, J. Sagotsky, and T. S. Deisboeck. Simulating non-small cell lung cancer with a multiscale agent-based model. *Theor Biol Med Model*, 4(1):50+, Dec. 2007.
- [19] T. Yu, C. M. Lloyd, D. P. Nickerson, M. T. Cooling, A. K. Miller, A. Garny, J. R. Terkildsen, J. R. Lawson, R. Britten, P. J. Hunter, and P. M. F. Nielsen. The physiome model repository 2. *Bioinformatics*, 27(5):743–744, 2011.
- [20] L. Zhang, S. Martin, and T. Deisboeck. Advancing cancer systems biology: Introducing the center for the development of a virtual tumor, CViT. *Cancer Informatics*, (5):1–8, 2007.

Computational Biologist in Profile:

Ruth Nussinov

Contributor: Amarda Shehu^{1,2} [amarda@gmu.edu]
Dept. of ¹Computer Science, ²Bioengineering
George Mason University, Fairfax, VA 22030
Tel: 703-993-4135 Fax: 703-993-1710



I first met Ruth during my graduate studies, when she was invited to give a seminar on her decade-long research in computational biology. I sat through her inspiring (and what has been etched on my memory as gentle) delivery of her work on macromolecular structure and interactions. Her informal meeting with graduate students after the talk left a deep impression on me of someone highly dedicated to providing students with guidance on what was important in biological research. Many years later, ending up near the National Cancer Institute (NCI), where one can find Ruth mostly these days as head of the Computational Structural Biology Group of the Center for Cancer Research Nanobiology Program at NCI, I simply could not give up the opportunity of obtaining a closer look at her journey as a scientist.

At ease and relaxed in her environment, though battling a cold, which she attributes to a recent visit with her grandson, Ruth naturally progresses to tell me about her work these days. Through the years, she has witnessed a rapid growth in computational power, an explosion

in the masses of biological data that need to be understood, and many advances in the biological sciences. Merging these with the fast pace and discoveries made by experimental studies these days, Ruth believes there is now an emergent need to re-evaluate directions. She turns to me and asks:

“Are we, as computational biologists, contributing as much as we can to the advancement of the sciences in key areas?”

Waiting for me to absorb her question, she then proceeds to tell me what guides her work these last few years. She tells me that research in the biological sciences is increasingly shifting towards understanding the causes and mechanisms of diseases and discovering therapeutics. This shift over the last years is fueled by the desire to alleviate human suffering, also a priority of most funding agencies these days. She asserts that computational biology research, if properly guided, can effectively contribute to the understanding of a broad range of biological processes, from the molecular and biochemical, to the organismal and population levels.

Ruth's passion and energy are palpable and tireless. As if anticipating the next question, Ruth goes on to provide me a glimpse for her uncompromising work ethic. It all goes back to her father. Ruth Nussinov was born Ruth Hurwitz and raised in Rehovot Israel, where her father was one of the early researchers of agriculture and the recipient of the prestigious Israel Prize for his achievements. His working late at night after full days at the university and field trips on weekends, on which she often accompanied him, left a deep impression on her. Her work style emulates his to this very day.

Perhaps in this context, it is not hard to believe that Ruth is the author of close to 500 research articles, the recipient of the Biophysical Society Fellow Award “for her extraordinary contributions to advances in computational biology on both nucleic acids and proteins” and a Fellow of the International Society of Computational Biology (ISCB). Ruth was appointed the Editor-in-Chief of PLoS Computational Biology in 2012. She continues to be an editor in J. Biological chemistry (JBC), Physical Biology, Proteins, BMC Bioinformatics, and other journals. She is also a long term member on the NIH Study Section MSFD.

Despite the strong early exposure to science, Ruth's own scientific journey has been winding and, at-

times, hard, as well. Before starting her studies in microbiology at the Hebrew university in Jerusalem (largely under the influence of her microbiologist aunt), Ruth met and shortly thereafter married her loving husband, Shmuel Nussinov, then a physics graduate student at the Weizmann Institute in Rehovot. The unexpected departure of Shmuel's Ph.D. advisor caused the young couple to move to Seattle. With some head start, Shmuel finished his Ph.D. in two years. She recalls very well that she was not given full credit for the year at Hebrew University. Nonetheless, she managed to complete the 4-year equivalent undergraduate studies at the same time.

By the time the couple moved to Princeton University for Shmuel's postdoctoral studies, Ruth realized that she liked biochemistry. This was likely the outcome of fascinating courses in biochemistry and organic chemistry at the University of Washington. While Ruth was determined to start her work towards a Master's degree, at the time, Princeton still did not accept girls. So, Ruth went on to Rutgers, where she finished a non-thesis M.Sc. program in one year, just in time before their first son, Zohar (now a physics professor at Washington University in St. Louis and father of three) was born.

A several year break from science then followed, when two girls, Orna (now a medical doctor in Israel specializing in infectious diseases and a mother of three), and Osnat (now a litigation lawyer in Tel Aviv and a mother of three), were born. When Shmuel, who in the meantime received tenure at Tel Aviv university, went again for a one-year visit at the Institute for Advanced Studies in Princeton, and the kids were in day care, Ruth visited Rutgers and was fascinated when hearing from Dr. George Piczenick (who went on to become her Ph.D. advisor) about the sequencing of the first few-thousand unit long RNA bacteriophages.

Ruth then embarked on a project – with which she continued also after her Ph.D. – trying to devise algorithms to fold these single stranded RNA viruses into (hopefully biologically-significant) secondary structures. Upon realizing that she was making real progress, Shmuel asked for an extra year of leave of absence and was given just one more year from Tel Aviv University. Ruth completed her Ph.D. thesis work in the next half a year, where she presented the first dynamical algorithm for RNA folding after a fruitful collaboration with computer scientists - an experience she has repeated ever since.

Still, to formally obtain a Ph.D., Ruth was required to take various courses, and there was no time for that. So, it was agreed that the matter would be further discussed after she took the Qualifying exams. This seemed hopeless, since she had not studied any biology since her M.Sc. eight years earlier, and the field had exploded in the interim. Fortunately, she recalls, the extensive book by Lehninger had just come out. By reading and internalizing the book in three months of dedicated studies, she managed to do so well in the exams, that a few months later she defended her thesis just in time for the family's return to Israel and the beginning of an exciting postdoctoral fellowship at the Weizmann Institute.

It would have seemed that after three more years of postdoctoral experience at Weizmann and two one-semester visits at Berkeley and Harvard, Ruth should not have any difficulty in securing a tenured position. Unfortunately, things were not as simple. Having always worked independently and not being a protégé of an influential scientist, the Rector of Tel Aviv University vetoed putting her up for a National Fellowship as a Tel Aviv University candidate, arguing that it was enough to have one Tel Aviv University professor from the family – an attitude that fortunately has since been reversed there and everywhere else. As I was writing this down, I paused to realize how strange this all would seem now. Yet, Ruth continues to tell me in her calm demeanor of the next chapter.

Despite the headwind, Ruth continued at Tel Aviv University at the mathematics department and medical school with temporary (for a while, unsalaried) appointments. In the meantime, she started her affiliation with the NIH which has been ongoing ever since. Her independent work started to increasingly attract attention, and when she finally had a proper position at Tel Aviv University, it was as an Associate Professor and promotion to Full Professor with tenure following closely thereafter.

Ruth's parents (and her husband) followed her success with great pride. She stresses that her father,

who lived through 99% of the 20th century, was there when she started to become a leading figure in the field. Though, she laments, he only saw about half her publications and missed her appointment as Editor-in-Chief of PLOS Computational Biology.

These days, Ruth is increasingly fascinated by the role of structure and interactions in biological mechanisms in disease. She cares about the impact of her work and revels in sustaining fruitful collaborations with scientists of diverse backgrounds. She proceeds to tell me that her passion for research goes hand in hand with her passion for teaching and guiding students (she has had almost thirty Ph.Ds at Tel Aviv University, half of which have been women). She regularly co-advises students from other universities, through her collaborations. As I am writing this all down, she is the midst of co-ordinating a student presentations day.



Ruth and her husband Shmuel.

As it seems we have come to a natural stopping point, she asks me what I have been doing these days. I proceed to tell her about my computational research, and, as I start defining the problem from a computational point of view, she interjects: “What is the biological significance of that?”

ACB BCB 2013 Report

ACM BCB is the flagship conference of the ACM SIGBio (Special Interest Group on Bioinformatics, Computational Biology and Biomedical Informatics). Now in its fourth year, ACM BCB has established itself as a premier annual forum for promoting multidisciplinary research and development in Bioinformatics, Computational Biology and Biomedicine from academia, industry and government.

BCB2013 attracted nearly 280 attendees with world-renowned scientists as keynote speakers, contributed talks at a highly competitive acceptance rate, and the broad participation of the research community serving on the program committee and the organizing committees for workshops, tutorials, and panels. The program featured three highly anticipated keynote lectures by Drs. Kohane, Nussinov, and Salzberg. A total of 43 regular papers and 28 short papers were selected after a rigorous review process from about 150 submissions, covering topics ranging from comparative genomics, protein and RNA structure, to network reconstruction and medical informatics. The program also features 11 contributed highlight papers, 62 posters, 5 demos, 6 tutorials, as well as 7 workshops on special topics such as immunoinformatics, structural bioinformatics, and next-gen sequencing bioinformatics. In addition to Conference Proceedings, special issues of selected papers will be published in the IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB) and the Database journal.

With the merging of the ACM SIGHIT into the SIGBio, this year also marks the new addition of a symposium dedicated to Health Informatics. Another new addition is the Industry Workshop to foster potential collaboration across industry and academia. Three special panels on Funding Agency Roundtable, Women in Bioinformatics and PhD Forum help students and faculty navigate the career and funding maze.

We wish to thank all of the authors who submitted papers and participated in the conference. Our deepest appreciation goes to the 100+ researchers who served on the various committees or helped assess the submitted papers. The conference would not be possible without generous gift of time by many people serving on the organizing committee. The essence of a scientific conference is its technical program. Thanks to Donna Slonim, Srinivas Aluru, and other area chairs, program committee members and reviewers for a diligent and fair review process. Thanks to Dr. Catalyurek for organizing the workshops, Drs. Honavar and Congdon for the tutorials, Drs. Zhu and Yu-Ping Wang for the posters, and Drs. Kann and Payne for putting together the Health Informatics Symposium.

We thank other members of the organizing committee – Drs. Christianson, Liebman, Wang, Vaisman, Sun, Cheng, Ghosh, Gao, and Edwards, for their generous help. Special thanks to Drs. Florea and Shehu for spearheading the local organization, and Kang Li for prompt and excellent management of the website. Last but not the least, we thank the Steering Committee, especially Dr. Aidong Zhang for imparting her institutional memory and wisdom; she was always available to provide feedback and her help was invaluable.

Thanks to the generous funding support from the National Science Foundation, we are able to provide partial financial support to 30 student attendees, many from underrepresented groups. We are also delighted to present the ACM SIGBio Best Paper, Best Student Paper, and Best Poster Awards.

Cathy Wu and Sridhar Hannenhalli

ACM BCB 2013 Paper and Poster awards

The ACM Conference on Bioinformatics, Computational Biology and Biomedical Informatics (ACM BCB 2013) featured best paper and best student paper awards. New this year is the awards are sponsored by the ACM SIGBio, in its role as the primary sponsor of the conference. Top papers from each of the nine areas within the conference have been considered by the award selection committee. The narrowed list of papers are reviewed independently by the committee, in addition to taking the reviews obtained for the conference into account.

The best student paper award is selected from among the papers where the first and primary author is a student.

The following papers received the award:

Best paper award:

Improving discrimination of essential genes by modeling local insertion frequencies in transposon mutagenesis data, by Michael Dejesus and Thomas Joerger



Best paper award

Best student paper award:

MRfY: Remote homology detection for beta-structural proteins using Markov random fields and stochastic search, by Noah Daniels, Andrew Gallant, Normal Ramsey and Lenore Cowen



Best studente paper award

The awards were presented at the conference banquet by program co-chair Srinivas Aluru from the Georgia Institute of Technology.

ACM SIGBio sponsored also a prize for the best poster award. At ACM conference on Bioinformatics, Computational Biology and Biomedical Informatics (ACM BCB 2013), over 60 posters were presented and they were evaluated on research quality, content clarity and presentation/design. The winner is selected from over 30 ballots representing over 70 votes from the program committee. The award was announced on Tuesday Sept 22 and the certificate was presented at the

conference banquet by poster chair Dongxiao Zhu from Wayne State University. Pietro Hiram Guzzi, Mario Cannataro and Pierangelo Veltri from University Magna Graecia of Catanzaro received the prize for the award with the poster titled Modularity and community detection in Semantic Similarity Networks through Spectral Based Transformation and Markov Clustering.



Best Poster Award



SigBio Record - Submission Guidelines

We invite to submit contributions and papers to SIGBio Record, the newsletter of the ACM's Special Interest Group on Bioinformatics, Computational Biology and Biomedical Informatics

The contributions may be any paper considered of interest to SIGBio community, including the application of computer science, Informatics, information technology, and communication technology, and ICT technology.

These can be in any one of the following categories:

- Survey/tutorial articles (short) on important topics
- Topical articles on problems and challenges
- Well-articulated position papers
- Technical articles
- Review articles of technical books, products and methodologies
- Reviews/summaries from conferences, panels and special meetings
- Book reviews and reports on relevant published technical books
- PhD dissertation abstracts
- Calls and announcements for conferences and journals
- Reports on Workshop and Conference in the area of computer science applied to medical systems or biology
- Brief announcements

Submissions should be made via email to editors, following the guidelines report on:

<http://www.sigbioinformatics.org/Main/Newsletter>