Tampere International Center for Signal Processing. TICSP series # 64

José Fonseca, Reija Autio, Andre Ribeiro & Olli Yli-Harja (eds.)

The 11th International Workshop on Computational Systems Biology, WCSB 2014, May 15-16, Lisbon, Portugal

Tampere International Center for Signal Processing
Tampere 2014

ISBN 978-952-15-3297-9 (PDF)
ISSN 1456-2774

PREFACE

The Workshop on Computational Systems Biology (WCSB) has begun more than 10 years ago now (2003). During these years, it was first and most times organized by the Computational Systems Biology Research Group of the Department of Signal Processing of Tampere University of Technology (TUT), Tampere, Finland.

Since its first year, the meeting has increased in size and quality, reflecting the rapid development in Computational Systems Biology. So far, the event has been organized and hosted in Tampere, by TUT but also in the Institute for Medical Informatics, Statistics and Epidemiology (IMISE), Leipzig, Germany, in conjunction with the Bioinformatics Research Centre (BiRC) at Aarhus University (AU), Denmark, in the Luxembourg Centre for Systems Biomedicine (LCSB) at University of Luxembourg, in ETH Zürich, Switzerland and in Ulm University, Germany.

This year WCSB is organized in Caparica, Portugal, for the first time. The scientific program includes nine invited talks by internationally acknowledged experts of their respective fields from Finland and Portugal.

This volume collects together the abstracts submitted to WCSB 2014. We would like to thank the authors and the reviewers for their contributions to the workshop and the proceedings. We are also grateful for the contribution of organizers in Portugal and in Finland for their efforts. Special thanks are also due to Leonardo Martins, João Santinha, André Mora and Virve Larmila in organization of the WCSB 2014. We would also like to thank the Academy of Finland, the Portuguese Foundation for Science and Technology and the Department of Signal Processing of the Tampere University of Technology.

On behalf of the WCSB 2014 Scientific Committee,

Olli Yli-Harja and Jose Manuel Fonseca

WCSB 2014 ORGANIZATION

Organizing institutions

UNINOVA - Instituto de Desenvolvimento de Novas Tecnologias
Tampere University of Technology CSB-group
Tampere International Center for Signal Processing (TICSP)

Chairs of WCSB

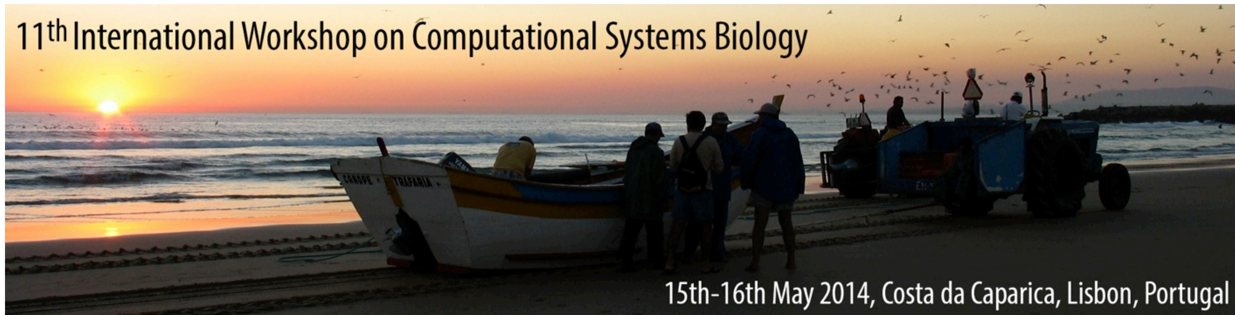
Jose Fonseca (Uninova, Portugal)
Olli Yli-Harja (Tampere University of Technology)

Scientific Committee

André Mora (New University of Lisbon, Portugal)
André Ribeiro (Tampere University of Technology, Finland)
Antti Honkela (University of Helsinki, Finland)
Carsten Wiuf (University of Copenhagen, Denmark)
Christophe Roos (Euformatics, Finland)
Daniel Nicorici (Orion Pharma, Finland)
Harri Lähdesmäki (Aalto University, Finland)
Heinz Koeppl (TU Darmstadt, Germany)
Ilya Shmulevich (Institute for Systems Biology, USA)
Jose Fonseca (Uninova, Portugal)
Juha Kesseli (University of Tampere, Finland)
Korbinian Strimmer (University of Leipzig, Germany)
Matti Nykter (University of Tampere, Finland)
Meenakshisundaram Kandhavelu (Tampere University of Technology, Finland)
Olli Yli-Harja (Tampere University of Technology, Finland)
Ralf Takors (University of Stuttgart, Germany)
Reija Autio (Tampere University of Technology, Finland)
Samuli Niiranen (Mylab Oy, Finland)
Steffen Schober (Ulm University, Finland)
Wei Zhang (University of Texas, MD Anderson Cancer Center, USA)

Organizing Committee

Jose Fonseca (Uninova, Portugal)
André Mora (Uninova, Portugal)
Joao Santinha (Uninova, Portugal)
Leonardo Martins (Uninova, Portugal)
Andre Ribeiro (Tampere University of Technology, Finland)
Olli Yli-Harja (Tampere University of Technology, Finland)
Reija Autio (Tampere University of Technology, Finland)



Scientific Programme

Day 1 – 15th May 2014

09:30 - 10:00 **José Fonseca**
Welcome address

10:00 - 10:30 **Olli Yli-Harja**
Big data, machine learning, and cancer

10.30 - 11.30 **Session A – Chair: André Ribeiro**

Jake Lin and Patrick May
Systemic Integration and Exploration of Genomic Variants

Kristin Blacklock and Gennady Verkhivker
Integrative Studies of Allosteric Regulation in Signal Transduction Networks: a Synergistic Perspective From Computational Systems Biology and Proteomics Approaches

Gianluca Selvaggio, Pedro Coelho and Armindo Salvador
Design Space Analysis Supports Role of Peroxiredoxin/Thioredoxin/Thioredoxin-Reductase System in Integrating Redox Signaling And Anticipatory Blocking

Vincenzo Belcastro
Identification of Signalling Mediators

11:30 - 12:00 **Discussion & Coffee break**

12:00 - 12:30 **Antti Ylipää**
Identifying previously unannotated cancer specific transcripts using RNA-sequencing

12:30 – 13:00 **Meenakshisundaram Kandhavelu**
Advances in understanding noisy signal transmission in neuronal cancer cells - GPR17 as a case study

13:00 – 14:30 **Lunch**

14:30 – 15:00 **Sara Madeira**
The NEUROCLINOMICS project: Understanding NEUROdegenerative Diseases through CLINical and OMICS Data Integration

15:00 – 16:00 **Session B – Chair: Reija Autio**

Christoph Zimmer and Sven Sahle
Parameter inference for stochastic models: A comparison of multiple shooting for stochastic systems to conventional least squares

Eszter Lakatos, Paul Kirk and Michael P. H. Stumpf
Multivariate moment closure techniques for stochastic kinetic models

Charalampos Kyriakopoulos and Verena Wolf
Optimal Observation Time Points in Stochastic Chemical Kinetics

Alexander Andreychenko, Linar Mikeev and Verena Wolf
Distribution reconstruction for stochastic chemical kinetics from conditional moment data

16:00 – 16:30 **Discussion & Coffee break**

16:30 – 17:00 **Isabel Gordo**
Adaptation of bacterial populations approaching a fitness peak

19:30 - 23:00 **Conference Dinner**

Day 2 – 16th May 2014

10:00 - 10:30 **Reija Autio**
Recycling high-throughput data: Combining existing data can lead to new knowledge

10:30 - 11:30 **Session C – Chair: Meenakshisundaram Kandhavelu**

Domokos Meszéna, Eszter Lakatos and Gabor Szederkenyi
Sensitivity analysis and parameter estimation of a human blood glucose regulatory system model

Anton Salykin, Dominique Chu, Petr Fojtik, Miriama Kruta, Katka Schrammova, Petr Skladal, Petr Dvorak and Vladimir Rotrekl
Computational Model of Fgf2 Signalling Network In Human Embryonic Stem Cells Revealed Oxygen Triggered Metabolic Switch

João Santinha, Antti Häkkinen, Abhishekh Gupta, Teppo Annila, Jason Lloyd-Price, Andre S. Ribeiro and José Fonseca
Image Alignment and Lineage Construction Tool to Study Segregation and Partitioning in Division of Unwanted Protein Aggregates in Escherichia Coli

Leonardo Martins, Andre S. Ribeiro and Jose Fonseca
Generator of Synthetic Time-Lapsed, Phase-Contrast Microscopy Images of Growing E. Coli Cell Populations

11.30 - 12-00 **Discussion & Coffee break**

12:00 - 12:30 **Samuli Niiranen**
Convergence in clinical diagnostics

12:30 – 13:00 **Andre Ribeiro**
Effects of stress on the in vivo dynamics of non-stress-responsive genes in Escherichia coli

13:00 – 14:30 **Lunch**

14:30 – 15:00 **Ilda Sanches**
Unraveling the host-specificity within streptococci

15:00 – 16:30 **Session D - Chair: Antti Ylipää**

Tomasz Ignac, Alexander Skupin, Nikita Sakhanenko and David Galas
Detecting Genetic Interactions in Case-Control Data: Information Theory Based Methods

Alessandro Bolli and Armindo Salvador
Microorganism's Growth-Robustness Trade-Off Explained by Simple General Principles

Nadia Goncalves, Ramakanth Neeli Venkata, Antti Martikainen, Joao Santinha, Jose Fonseca and Andre Ribeiro
In Vivo Retention of Protein Aggregates in Escherichia Coli

Samuel M.D. Oliveira, Jerome G. Chandraseelan, Antti Häkkinen, Olli Yli-Harja and Andre S. Ribeiro
Analysis of The Robustness of a Single Copy Repressilator

Jerome G. Chandraseelan, Jason Lloyd-Price, Olli Yli-Harja and Andre S. Ribeiro
Estimating the Duration of Open Complex Formation From in Vivo Measurements

Florbela Pereira, Diogo A. R. S. Latino and Susana P. Gaudêncio
In Silico Screening of Lead-Like Molecules en Route to Antitumor and Antibiotic Drugs from Marine and Microbial Natural Products Using Empirical and Semi-Empirical Quantum-Chemical Descriptors

16:30 – 17:00 **Olli Yli-Harja**
Closing speech

TABLE OF CONTENTS

INVITED TALKS 9

Big data, machine learning, and cancer

Olli Yli-Harja, *Laboratory of Biosystem Dynamics, Computational Systems Biology Research Group, Department of Signal Processing, Tampere University of Technology, Tampere, Finland and Institute for Systems Biology, Seattle, USA* 10

Identifying previously unannotated cancer specific transcripts using RNA-sequencing

Antti Ylipää*, Department of Signal Processing, Tampere University of Technology and Institute of Biomedical Technology, University of Tampere, Tampere, Finland* 11

Advances in understanding noisy signal transmission in neuronal cancer cells - GPR17 as a case study

Meenakshisundaram Kandhavelu, *Molecular Signaling Lab, Computational Systems Biology Research Group, Signal Processing Department, Tampere University of Technology, Tampere, Finland* 12

The NEUROCLINOMICS project: Understanding NEUROdegenerative Diseases through CLINical and OMICS Data Integration

Sara Madeira, *Instituto Superior Técnico, Universidade de Lisboa and INESC-ID, Knowledge Discovery and Bioinformatics Group, Lisbon, Portugal* 13

Adaptation of bacterial populations approaching a fitness peak

Isabel Gordo, *Instituto Gulbenkian de Ciência, Oeiras, Portugal* 14

Recycling high-throughput data: combining existing data can lead to new knowledge

Reija Autio, *Department of Signal Processing, Tampere University of Technology, Tampere, Finland* 15

Convergence in clinical diagnostics

Samuli Niiranen*, Department of Signal Processing, Tampere University of Technology and Mylab Oy, Tampere, Finland* 16

Effects of stress on the in vivo dynamics of non-stress-responsive genes in Escherichia coli

Andre Ribeiro, *Laboratory of Biosystem Dynamics, Department of Signal Processing, Tampere University of Technology, Tampere, Finland* 17

Unraveling the host-specificity within streptococci

Ilda Sanches, Centro de Recursos Microbiológicos, Departamento de Ciências da Vida, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Caparica, Portugal 18

ABSTRACTS 19

Systemic Integration and Exploration of Genomic Variants

Jake Lin and Patrick May 20

Integrative Studies of Allosteric Regulation in Signal Transduction Networks: a Synergistic Perspective From Computational Systems Biology and Proteomics Approaches

Kristin Blacklock and Gennady Verkhivker 21

Design Space Analysis Supports Role of Peroxiredoxin/Thioredoxin/Thioredoxin-Reductase System in Integrating Redox Signaling And Anticipatory Blocking

Gianluca Selvaggio, Pedro Coelho and Armindo Salvador 22

Identification of Signalling Mediators
Vincenzo Belcastro 23

Parameter inference for stochastic models: A comparison of multiple shooting for stochastic systems to conventional least squares
Christoph Zimmer and Sven Sahle 24

Multivariate moment closure techniques for stochastic kinetic models
Eszter Lakatos, Paul Kirk and Michael P. H. Stumpf 25

Optimal Observation Time Points in Stochastic Chemical Kinetics
Charalampos Kyriakopoulos and Verena Wolf 26

Distribution reconstruction for stochastic chemical kinetics from conditional moment data
Alexander Andreychenko, Linar Mikeev and Verena Wolf 27

Sensitivity analysis and parameter estimation of a human blood glucose regulatory system model
Domokos Meszéna, Eszter Lakatos and Gabor Szederkenyi 28

Computational Model of Fgf2 Signalling Network In Human Embryonic Stem Cells Revealed Oxygen Triggered Metabolic Switch
Anton Salykin, Dominique Chu, Petr Fojtik, Miriama Kruta, Katka Schrammova, Petr Skladal, Petr Dvorak and Vladimir Rotrekl 29

Image Alignment and Lineage Construction Tool to Study Segregation and Partitioning in Division of Unwanted Protein Aggregates in Escherichia Coli
João Santinha, Antti Häkkinen, Abhishekh Gupta, Teppo Annila, Jason Lloyd-Price, Andre S. Ribeiro and José Fonseca 30

Generator of Synthetic Time-Lapsed, Phase-Contrast Microscopy Images of Growing E. Coli Cell Populations
Leonardo Martins, Andre S. Ribeiro and Jose Fonseca 31

Detecting Genetic Interactions in Case-Control Data: Information Theory Based Methods
Tomasz Ignac, Alexander Skupin, Nikita Sakhanenko and David Galas 32

Microorganism's Growth-Robustness Trade-Off Explained by Simple General Principles
Alessandro Bolli and Armindo Salvador 33

In Vivo Retention of Protein Aggregates in Escherichia Coli
Nadia Goncalves, Ramakanth Neeli Venkata, Antti Martikainen, Joao Santinha, Jose Fonseca and Andre Ribeiro 34

Analysis of The Robustness of a Single Copy Repressilator
Samuel M.D. Oliveira, Jerome G. Chandraseelan, Antti Häkkinen, Olli Yli-Harja and Andre S. Ribeiro 35

Estimating the Duration of Open Complex Formation From in Vivo Measurements
Jerome G. Chandraseelan, Jason Lloyd-Price, Olli Yli-Harja and Andre S. Ribeiro 36

In Silico Screening of Lead-Like Molecules en Route to Antitumor and Antibiotic Drugs from Marine and Microbial Natural Products Using Empirical and Semi-Empirical Quantum-Chemical Descriptors
Florbela Pereira, Diogo A. R. S. Latino and Susana P. Gaudêncio 37

INVITED TALKS

BIG DATA, MACHINE LEARNING, AND CANCER

Olli Yli-Harja[1,2]

[1]Laboratory of Biosystem Dynamics, Computational Systems Biology Research Group, Department of Signal Processing, Tampere University of Technology, FI-33101 Tampere, Finland
[2]Institute for Systems Biology, 1441N 34th Street, Seattle, WA 98103-8904, USA
yliharja@cs.tut.fi

ABSTRACT

As biology is becoming an information science, Big data holds a promise of revealing secrets of complex diseases. But computers just compute and machine learning is just a tool. In order to utilize it, an intelligent user is needed to define suitable data structures that match purpose and context. We apply machine learning in the context of prediction of tumour phenotype. Such prediction from genomic data is notoriously difficult. Therefore we utilize the known association of genes with pathways, adding the layer of pathways to the internal representation of a machine learning system. Association of genomic data to pathways facilitates prediction of tumour phenotype. We conclude that integration of relevant background knowledge can lead to feasible solutions of complex problems in Big data.

IDENTIFYING PREVIOUSLY UNANNOTATED CANCER SPECIFIC TRANSCRIPTS USING RNA-SEQUENCING

Antti Ylipää[1,2] and Matti Nykter[2]

[1] Department of Signal Processing, Tampere University of Technology, P.O. Box 553, FI-33101 Tampere, Finland
[2] Institute of Biomedical Technology, University of Tampere, FI-33014 Tampere, Finland
antti.ylipaa@tut.fi

ABSTRACT

Cancer is a leading cause of death worldwide accounting for 13% of all deaths in 2008, and the percentage is projected to continue rising. Cancer is a generic term for a group of genetic diseases that arise due to accumulation of advantageous mutations to a cell's DNA. Genomic alterations bring about changes in the way the genome is transcribed which enable the cancer cells to proliferate uncontrollably. Although different tumors often share a number of mutations, their genomes and transcriptomes are essentially unique in many respects which makes it difficult to treat them effectively. Recently, genome-wide RNA-sequencing studies have identified several hundreds of previously unannotated long non-coding RNA molecules (lncRNAs) that may only be expressed in certain subtypes of cancers. Despite lacking protein coding potential, some of these rare transcripts have been shown to play a key role in oncogenesis. For example, Prostate Cancer Associated Transcript 1 (PCAT1) has been shown to regulate an important tumor suppressor gene BRCA2, and to control homologous recombination in prostate cancer. We have performed RNA-sequencing experiments from 41 prostate tumors to screen for novel lncRNAs. By conducting transcriptome assembly, we identified 145 previously unannotated intergenic prostate cancer associated transcripts or isoforms, including an ERG regulated transcript PCAT5 whose knockdown resulted in a dramatic effect on cell growth, migration, invasion and apoptosis specifically in PCAT5-positive prostate cancer cell lines. Genome-wide expression analysis indicated that PCAT5 works as a regulator of proliferation pathways. We have also developed alternative approaches to computationally intensive transcriptome assembly for discovering novel loci of expression from very large sample cohorts, such as those provided by the Cancer Genome Atlas (TCGA) consortium. Application of these methods to TCGAs RNA-sequencing data has resulted in identification of novel transcripts in a brain cancer that have subsequently been experimentally validated.

ADVANCES IN UNDERSTANDING NOISY SIGNAL TRANSMISSION IN NEURONAL CANCER CELLS - GPR17 AS A CASE STUDY

Gnanavel Mutharasu[1] and Meenakshisundaram Kandhavelu[1]

[1]Molecular Signaling Lab, Computational Systems Biology Research Group, Signal Processing Department, Tampere University of Technology, P.O. Box 553, FI-33101, Tampere, Finland
meenakshisundaram.kandhavelu@tut.fi

ABSTRACT

Study of signaling networks in neuronal cells provide the unique opportunity to uncover the basis of many diseases. It helps in the development of novel therapeutic strategies to treat neurological disorders, neuronal tumors and cancers. Many of the genetic and epigenetic aberrations leads to cancer states are due to inappropriate activation/inactivation of intracellular signaling pathways. Such signaling cascades usually proceed from the cell surface, where extra cellular factors interact with their specific receptors i.e. G-Protein Coupled Receptors (GPRs). A report shows [1] the activation or suppression of GPR17 in diseased neuronal cells has potential impact in altering the tumor conditions. Many neuronal cancer [1, 2] cells showed differing expression levels of GPR17. The role of GPR17 receptor and their interaction with various signaling pathways and its importance in neuronal signal transduction are currently not well understood. Recent developments in sequencing methods provide informations on wide varieties of mutated states of GPR17, which makes it a difficult pharmacological target. A report shows critical mutation in this receptor evidently affects the binding of known ligands [3]. From the caner sequencing and 1000 genome projects many forms of mutations reported. Implications of these mutations in signal fidelity, transduction accuracy and consequences in pathways yet to be studied. Present study focus on such novel signals in the complex network of pathways in neuronal cancer cells. About 500 unique cases of GPR17 with mutations from 4 different types of neuronal cancers were analyzed with major focus on Glioma. From our study we found that many of the mutated GPR17 are not able recognize and bind with previously reported ligands or binds with different orientation and interaction energy. It implies the necessity of more specified and customized designing of ligand molecules with the consideration of mutations present, towards the development of personalized medicine.

References

[1] Lecca D, Trincavelli ML, Gelosa P, Sironi L, Ciana P, Fumagalli M, Villa G, Verderio C, Grumelli C, Guerrini U, Tremoli E, Rosa P, Cuboni S, Martini C, Buffo A, Cimino M, Abbracchio MP (2008) The recently identified P2Y-like receptor GPR17 is a sensor of brain damage and a new target for brain repair. *PLoS One*. 3(10): e3579.
[2] Boda E, Viganò F, Rosa P, Fumagalli M, Labat-Gest V, Tempia F, Abbracchio MP, Dimou L, Buffo A. The GPR17 receptor in NG2 expressing cells: focus on in vivo cell maturation and participation in acute trauma and chronic damage. (2011) *Glia*. 59(12): 1958-1973.
[3] Parravicini C, Abbracchio MP, Fantucci P, Ranghino G (2010) Forced unbinding of GPR17 ligands from wild type and R255I mutant receptor models through a computational approach. *BMC Struct Biol*. 10(8): 1-18.

THE *NEUROCLINOMICS* PROJECT: UNDERSTANDING *NEURO*DEGENERATIVE DISEASES THROUGH *CLINI*CAL AND *OMICS* DATA INTEGRATION

Sara Madeira[1,2]

[1]Instituto Superior Técnico, Universidade de Lisboa,
Av. Rovisco Pais, 1049-001 Lisbon, Portugal
[2]INESC-ID, Knowledge Discovery and Bioinformatics Group,
R. Alves Redol, 9, 1000-029 Lisbon, Portugal
smadeira@kdbio.inesc-id.pt

ABSTRACT

The need for integrative approaches to provide a broader understanding of brain related pathologies, in general, and neurodegenerative diseases, in particular, has been largely recognised. These approaches should infer relationships between omics, clinical, and personal data. In NEUROCLINOMICS, we are interested in the development of innovative approaches to understanding neurodegenerative diseases through heterogeneous data integration. We work on the development of a sophisticated knowledge discovery system to integrate powerful data mining algorithms to unravel potentially relevant links between omics and clinical data. Disease diagnostic and prognostic markers, disease progression rates, and patient profiles, are tackled. Together with the challenging task of studying complex diseases we also embrace the challenging topic of developing efficient and effective mining algorithms for biomedical data integration. We now use Amyotrophic Lateral Sclerosis and Alzheimer's disease as case studies.

ADAPTATION OF BACTERIAL POPULATIONS APPROACHING A FITNESS PEAK

Isabel Gordo[1]

[1]Instituto Gulbenkian de Ciência, Rua da Quinta Grande, PT-2780-156 Oeiras, Portugal
igordo@igc.gulbenkian.pt

ABSTRACT

Populations adapt to novel environments through the successive accumulation of adaptive substitutions. This adaptive process is dynamic and expected to depend both on the fitness landscape characteristic of a given environment and on the genetic background on which new mutations arise. One of the oldest models of adaptive evolution was proposed by R.A. Fisher and it's known as Fisher's geometrical model (FGM). FGM assumes that organisms climb a fitness landscape characterized by a single peak by performing an adaptive walk towards that phenotypic peak. Here we model bacterial adaptation under a range of parameter values which have reasonable biological support. We find that FGM extended to allow for small random environmental variations is able to explain several observations made recently in experimentally evolved populations. Consistent with data on populations evolving under laboratory controlled conditions, the model predicts that: mean population fitness increases rapidly when populations face novel environments and then achieves a dynamic plateau; the rate of molecular evolution is remarkably constant over long periods of evolutionary time; bacteria with a mutator phenotype are expected to frequently invade such evolving populations; and patterns of epistasis vary along the adaptive walk. Negative epistasis is expected in the initial steps of adaptation, as observed in currently available data, but not at later steps, a prediction that remains to be tested. Furthermore, populations are expected to exhibit high levels of phenotypic diversity at all times during their evolution. This implies that populations are possibly able to adapt rapidly to novel abiotic environments.

RECYCLING HIGH-THROUGHPUT DATA: COMBINING EXISTING DATA CAN LEAD TO NEW KNOWLEDGE

Reija Autio[1]

[1]Department of Signal Processing, Tampere University of Technology,
P.O. Box 553, FI-33101 Tampere, Finland
reija.autio@tut.fi

ABSTRACT

High-throughput studies providing large amounts of data are in routine use in biomedical research. In addition to the importance of the original findings, this data may have value in wider context when combined with other previously acquired and published results. These published genome-wide data can be exploited also to answer questions other than those addressed in the original studies. However, the means to perform such re-analyses of existing data are currently limited. Also the information of the samples is often available only as free text in publications or public databases, which in turn is making identification of appropriate samples difficult. Further, the measurement data are often available in heterogeneous formats, and the lack of systematic preprocessing hampers the comparisons between the data from different experiments. In addition to selecting suitable data for the combination studies, it is relevant to identify best methods for the data analysis. We have created data resources for these data analysis, specifically for stem cell and nanoparticle related data and research questions. The resources ESTOOLS Data@Hand [1] and NanoMiner [2] are freely available for the research community. The resources include powerful tools for analyzing and illustrating the data. The users may combine various data sets and results, and can gain deeper knowledge on the data. With the computational analysis and illustration tools within these resources, the users can create new hypothesis as well as test their existing ones utilizing the recycled high-throughput data.

References

[1] Kong L, Aho K-L, Granberg K, Lund R, Järvenpää L, Seppälä J, Gokhale P, Leinonen K, Hahne L, Mäkelä J, Laurila K, Pukkila H, Närvä E, Yli-Harja O, Andrews PW, Nykter M, Lahesmaa R, Roos C and Autio R (2013) ESTOOLS Data@Hand: human stem cell gene expression resource. *Nat Meth*. 10(9): 814–815.
[2] Kong L, Tuomela S, Hahne L, Ahlfors H, Yli-Harja O, Fadeel B, Lahesmaa R and Autio R (2013) NanoMiner - Integrative Human Transcriptomics Data Resource for Nanoparticle Research. *PloS One*. 8(7): e68414.

CONVERGENCE IN CLINICAL DIAGNOSTICS

Samuli Niiranen[1,2]

[1]Department of Signal Processing, Tampere University of Technology,
P.O. Box 553, FI-33101, Tampere, Finland
[2]Mylab Oy, Tampere, Finland
samuli.niiranen@mylab.fi

ABSTRACT

There is an increasing functional convergence both within the traditional specialties of the clinical laboratory (clinical chemistry and hematology, anatomic pathology, clinical microbiology and clinical genetics) and also between clinical imaging diagnostics and laboratory diagnostics. This convergence is due to a number of new and emerging technological and operational innovations including, but not limited to, molecular diagnostics, digital pathology and the LEAN production system. These innovations and their implications for informatics solutions in clinical diagnostics are discussed.

EFFECTS OF STRESS ON THE *IN VIVO* DYNAMICS OF NON-STRESS-RESPONSIVE GENES IN *ESCHERICHIA COLI*

Andre Ribeiro[1]

[1]Laboratory of Biosystem Dynamics, Department of Signal Processing,
Tampere University of Technology, P.O. Box 553, FI-33101 Tampere, Finland
andre.ribeiro@tut.fi

ABSTRACT

Under stress, *Escherichia coli* alter the transcriptomic profile. We study the *in vivo* transcription dynamics of genes uninvolved in stress-response pathways in cells under sub-lethal acidic shift and oxidative stress, by following with single-molecule sensitivity the activity of a probe gene encoding RNA target for MS2d-GFP. From the measurements of cells, when under stress and when in optimal conditions, we address the following questions. Does stress affect the kinetics of activation of the probe gene? Does stress affect the kinetics of transcript production following activation? Are the effects immediate or gradual? Do the effects of the two stress conditions differ? What are the changes occurring at various intermediate stages of transcription responsible for the changes in the kinetics of RNA production from the probe gene? We find that stress slows the activation of the probe gene, likely due to changes in the intake kinetics of inducers. From distributions of intervals between consecutive transcription events in individual cells, we show that, unlike in optimal conditions, stressed cells reduce, gradually, the probe gene's mean transcription rate. This change, confirmed by qPCR, is faster and wider under oxidative stress. Meanwhile, the transcriptional noise is unaffected by acidic shift but decreases under oxidative stress. Next, we show that under acidic shift, transcription has two rate-limiting steps, as in optimal conditions, although increased in duration by the same relative amount, explaining how the mean transcription rate decreases while noise is unaltered. Under oxidative stress, a third step becomes rate-limiting, explaining how noise and mean rate decrease. Our results suggest that cells' responses to sub-lethal stress include a gradual regulation of the speed and number of rate-limiting steps in transcription of non-stress related genes. The changes are stress-dependent. In the end, we discuss not only the effects of stress on the kinetics of transcription of active genes but also how to make use of single-molecule detection techniques to study how cells regulate the intake of molecules from the environment when under stress conditions.

UNRAVELING THE HOST-SPECIFICITY WITHIN STREPTOCOCCI

Ilda Sanches[1]

[1]Centro de Recursos Microbiológicos, Departamento de Ciências da Vida,
Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa,
Campus de Caparica, 2829-516 Caparica, Portugal
isanches@fct.unl.pt

ABSTRACT

Animal disease-associated streptococcal species including Streptococcus agalactiae, Streptococcus dysgalactiae subsp. dysgalactiae and Streptococcus uberis are relevant mastitis pathogens (a highly prevalent and costly disease in dairy industry). In spite of their veterinary importance, knowledge of virulence factors of most animal streptococcal species is limited or almost inexistent. Virulence genes of the human pathogen Streptococcus pyogenes phages encoding superantigens, DNase, and/or streptodornase were detected in bovine S. dysgalactiae. Phylogenetic analysis of superantigen gene sequences revealed a high level of identity among genes of bovine S. dysgalactiae, of the horse pathogen Streptococcus equi subsp. equi, and of the human pathogen S. pyogenes. These findings indicate that S. pyogenes phages may play a role in animal streptococci evolution, and that S. dysgalactiae subsp. dysgalactiae (considered a strictly-animal pathogen) should not be disregarded as a possible human pathogen. Currently, bovine S. dysgalactiae subsp. dysgalactiae and human related streptococcal subspecies - S. dysgalactiae subsp. equisimilis and S. pyogenes are being investigated for comparison of virulence associated traits such as biofilm formation, to uncover the genomic structure of the phages and to assess their involvement in mobile genetic transfer. Human and bovine cell lines and Zebrafish are used for in vitro and in vivo testing of bovine strains infection potential. With these studies we aim to establish a rational basis for the development of a clinical assessment panel for novel biomarkers of virulence in livestock streptococcal species and substantially contribute to assessing the infection potential of a microorganism frequently detected in clinical and subclinical mastitis.

ABSTRACTS

SYSTEMIC INTEGRATION AND EXPLORATION OF GENOMIC VARIANTS

Jake Lin[1,2] *and Patrick May*[1]

[1]Bioinformatics Core, Luxembourg Centre for Systems Biomedicine, 7, Avenue des Hauts-Fourneaux, L-4362 Esch-Sur-Alzette, Luxembourg
[2]Department of Signal Processing, Tampere University of Technology, P.O. Box 553, FI-33101 Tampere, Finland
jake.lin@tut.fi and patrick.may@uni.lu

ABSTRACT

Contrasting personal genomes to the human reference reveals many variants. Variant data from whole genome sequencing is complex due to scale and calls are sensitive to frequency and familial context. We are establishing a computational paradigm for annotation, integration and web interface to systemically explore different scales of genomic variants. Taking advantage of advances in sequencing platforms and their reduced costs, studies conducted by Luxembourg Centre for Systems Biomedicine have focus on sequencing personal genomes consisting of 45 families and 263 individuals for various diseases using Complete Genomics whole genome platform. Most of the families studied have a history of epilepsy and focus on Parkinson disease families has begun. By applying published and novel analytic tools along with established variant references, many millions of variants have been annotated and further classified. High quality variant calls are stored and existing variants further classified upon integration of new personal genomes. Data exploration and visualization tools are designed and built with modern Internet and database technologies. Researchers are able to access and investigate variant findings from multiple studies using modern web browsers. The intuitive interface, designed for identification of important variants, encodes the data and interactively shows multiple scales and perspectives. Variant interpretation is aided by integration of 1000 Genomes and Personal Genome Project. Furthermore, genome wide network of variants can be dynamically plotted. All developed methods and programs are open-sourced and freely available to the research community. Personal genomics together in a familial context have the power to reveal disease variants and form the basis of personalized medicine. We believe that adoption of scalable tools combined with interactive multi-scaled visualizations can lead to greater insights, allowed for improved systematic genomic comparison between diseases and ultimately assist with biomarker discovery.

INTEGRATIVE STUDIES OF ALLOSTERIC REGULATION IN SIGNAL TRANSDUCTION NETWORKS: A SYNERGISTIC PERSPECTIVE FROM COMPUTATIONAL SYSTEMS BIOLOGY AND PROTEOMICS APPROACHES

Kristin Blacklock[1] and Gennady Verkhivker[2]

[1]Department of Chemistry and Chemical Biology, Center for Integrative Proteomics Research, Rutgers University, 174 Frelinghuysen Road, Piscataway, NJ 08854-8076, USA
[2]Department of Computational Biosciences, Crean School of Health and Life Sciences, Chapman University, One University Drive, Orange, CA 92866, USA
verkhivk@chapman.edu

ABSTRACT

The allosteric interactions and regulation of molecular chaperones and protein kinases allow for molecular communication and event coupling in signal transduction networks. The overarching goal of understanding molecular principles underlying differentiation of protein kinase clients and chaperone-based modulation of kinase activity is fundamental to understanding activity of many tumor-inducing signaling proteins. We report the results of integrative systems biology and proteomics studies of the Hsp90 chaperone and protein kinases with an atomic level analysis of the communication pathways regulating conformational equilibrium of theses protein systems in signaling networks. By integrating structural bioinformatics, computational systems biology approaches and experimental proteomics studies, we have identified and characterized evolutionary and functionally conserved elements of the Hsp90 chaperones and protein kinases that may serve as key regulators of collective motions and hubs of long-range communication networks. The results of computational systems biology analysis and proteomics experiments have been integrated into a graph-based network model of allosteric regulation. The evolution of protein structure networks in molecular chaperones and protein kinases during allosteric activation has revealed the increased cooperativity reflecting a preferential attachment of allosterically interacting functional sites. Among our primary findings is the emerging evidence that a small number of functional motifs may be utilized by the chaperone and protein kinases to act collectively as central regulators of the intermolecular communications, ATP hydrolysis, and protein client binding in signaling networks. The diversity of allosteric communication mechanisms could ensure a proper balance of the network efficiency and functional redundancy required to maintain resilience against random attacks in the fluctuating protein environment. The additional layers of protection in regulatory mechanisms can be provided through recruitment of cochaperones and posttranslational modifications. The proposed model suggests that this phenomenon may be a common functional requirement encoded across allosterically regulated molecular systems.

DESIGN SPACE ANALYSIS SUPPORTS ROLE OF PEROXIREDOXIN/THIOREDOXIN/THIOREDOXIN-REDUCTASE SYSTEM IN INTEGRATING REDOX SIGNALING AND ANTICIPATORY BLOCKING

Gianluca Selvaggio[1] , *Pedro M. B. M. Coelho*[1] *and Armindo Salvador*[1,2]

[1]Computational & Systems Biology Group, Center for Neuroscience and Cell Biology, University of Coimbra
[2]Chemistry Department, University of Coimbra, 3004-535 Coimbra, Portugal
salvador@cnc.uc.pt

ABSTRACT

Cells are occasionally exposed to high H_2O_2 concentrations, often preceding exposure to other electrophylic compounds. Both H_2O_2 and these compounds can irreversibly modify protein thiols, causing deleterious loss of function and protein aggregation. The induction of enzymatic defenses against those agents is too slow to avoid significant damage between the onset of stress and the expression of the defenses. Cells may solve this conundrum by “blocking” the thiols through reversible covalent modification once H_2O_2 concentrations begin to increase. We term this mechanism “anticipatory blocking” because it acts in anticipation of irreversible damage upon detection of early signs of stress. It can simultaneously accomplish H_2O_2-dependent signaling. We hypothesize that the ubiquitous Peroxiredoin/Thioredoxin/Thioredoxin-Reductase/Protein-Dithiol System (PTTRDS) drives anticipatory blocking of protein dithiols as disulfides. We applied the design space approach [1, 2] to examine the design requirements for the PTTRDS system to effectively integrate H_2O_2 signaling and anticipatory blocking, and we compared these requirements to the actual design in human erythrocytes. To that effect, we developed a minimal model of the PTTRDS and defined a set of quantitative performance criteria that embody the requirements for (a) efficient scavenging capacity, (b) low NADPH consumption, (c) effective signal propagation, and (d) effective anticipatory blocking control. We then sought the design principles (relationships among rate constants and species concentrations) that warrant satisfaction of all the criteria. These were as follows, for human erythrocytes: (i) the equilibrium constant for thiol-disulfide exchange between thioredoxin and the protein dithiol $[T(SH)_2 + PSS \rightarrow TSS + P(SH)_2]$ should be in the range $0.01<K<20$ to allow the protein to fully accumulate in the oxidized form as soon as the H_2O_2 concentration increases; (ii) the maximum flux of thioredoxin reduction must be lower than the maximum flux of peroxiredoxin disulfide reduction and formation. Additionally, we identified a trade-off between the robustness of signal propagation and the NADPH expenditure in the process. Human erythrocytes have a limited capacity for NADPH regeneration, and should thus sacrifice the former performance criteria to some extent in order to save NADPH. A comparison of experimental data to the theoretical predictions above indicates that the design of the PTTRDS in human erythrocytes accomplishes effective integration between anticipatory blocking, antioxidant protection and redox signaling. A more general analysis suggests that these principles hold in a wide variety of cell types and organisms.

References

[1] Coelho PMBM, Salvador A, et al. (2009) Quantifying global tolerance of biochemical systems: design implications for moiety-transfer cycles. *PLoS Comp Biol*. 5(3): e1000319.

[2] Savageau MA, Coelho PMBM, Fasani RA, Tolla DA, and Salvador A (2009) Phenotypes and tolerances in the design space of biochemical systems. *Proc Natl Acad Sci U S A*. 106(16):6435–6440.

Acknowledgments

We acknowledge fellowship SFRH/BPD/90065/2012 and grants PEst-C/SAU/LA0001/2013-2014, PEst-OE/QUI/UI0612/2013, FCOMP-01-0124-FEDER-020978 financed by FEDER through the “Programa Operacional Factores de Competitividade, COMPETE” and by national funds through “FCT, Fundação para a Ciência e a Tecnologia” (project PTDC/QUI-BIQ/119657/2010).

IDENTIFICATION OF SIGNALLING MEDIATORS

Vincenzo Belcastro[1]

[1]Telethon Institute of Genetics and Medicine, Naples, Italy
belcastro@tigem.it

ABSTRACT

Different are the use of information encoded in functional and causal gene networks, from the discovery of novel interactions to the identification of new transcription factor targets. Similarly, protein-protein interaction networks have been interpreted to associating new functions to known proteins. On the other hand, drug induced genome-wide transcriptional responses are also widely adopted as source of information for drug repositioning. However, no attempt is pro-vided that combines functional and physical gene and protein networks and drug-induced transcriptional responses. Here we describe a procedure to identifying signaling mediators, i.e. transcription factors (TFs) with key roles in regulating signaling cascades induced by drug treatments, and other types of cellular perturbations. In the specific, we explored the regulatory mechanisms induced by ouabain and low temperature (lowT) on a cystic fibrosis cellular model (F508del CFTR). Cystic fibrosis is an autosomal recessive genetic disorder that mainly affects lungs. Cystic fibrosis is caused by mutations to the CFTR gene, via misfolding of the translated CFTR protein. Misfolded CFTR is then unable to reach the plasma membrane where it acts as a transmembrane conductance regulator. The procedure combines a functional gene network, a protein-protein interaction network, and genome-wide responses induced by treatment with ouabain and lowT, shown to provide F508del CFTR phenotype rescue. Three parts compose the approach: first we identify candidate TFs; i.e., TFs that are functionally related to changes induced by ouabain and lowT; then we apply a filtering step to retain signal mediator TFs; TFs that maximize the similarity in response between ouabain and lowT. Lastly, mediators TFs are clustered to identify transcription regulator complexes. Mediators TFs acting as hubs, and with a detectable lung expression, are then experimentally tested via small interference RNA. The experiments consist in silencing the mRNA associated to the selected mediator TFs. Experimental results shown a phenotype rescue on cystic fibrosis cell lines for a significant subset of the tested TFs. This approach pro-vides a step-forward to identifying drug-induced signaling mediators; i.e., key signaling components important to de-signing new and less invasive treatments to diseases.

PARAMETER INFERENCE FOR STOCHASTIC MODELS: A COMPARISON OF MULTIPLE SHOOTING FOR STOCHASTIC SYSTEMS TO CONVENTIONAL LEAST SQUARES

Christoph Zimmer[1] and Sven Sahle[2]

[1]Bioms/IWR, University of Heidelberg
[2]BioQuant, University of Heidelberg, INF 267, 69120 Heidelberg, Germany
christoph.zimmer@bioquant.uni-heidelberg.de

ABSTRACT

Due to progress in experimental techniques producing single cell measurements there is more and more need for parameter inference for stochastic models. This poster introduces the multiple shooting for stochastic systems method to estimate parameters in stochastic models and compares it to a least squares functional which is widely used for parameter estimation. The comparison is performed on a stochastic model of a genetic toggle switch which shows a switching behavior between to steady states which cannot be shown in deterministic modeling. The comparison shows that only the MSS method is able to estimate the parameters with high accuracy. When using resulting estimates of both methods as input for new stochastic simulations, only the MSS estimates are able to reproduce the switching behavior. This show the need for the use of estimation methods suited to stochastic models such as the MSS method when dealing with intrinsic stochastic models.

MULTIVARIATE MOMENT CLOSURE TECHNIQUES FOR STOCHASTIC KINETIC MODELS

Eszter Lakatos[1], Paul Kirk[1] and Michael P.H. Stumpf [1]

[1]Division of Molecular Biosciences, Theoretical Systems Biology Group,
Imperial College London,
London SW7 2AZ, United Kingdom
e.lakatos13@imperial.ac.uk

ABSTRACT

The behaviour of eukaryotic cells is determined by the intricate interplay between signalling and regulatory processes, which can be described as chemical reaction systems between different molecular species. As often the key species have low copy numbers, stochastic modelling is necessary in order to capture the essential kinetics of the system. One possible technique to approximate stochastic dynamics is describing the evolution of the system's probability distribution in terms of its moments. Most studies concentrate on the first two moments of the variables, but in many cases more moments are necessary to describe the dynamics of the system; this is particularly true for oscillating, non-linear and multi-stable systems. Ale et al. [1] derived a general Moment Expansion Algorithm (MEA) to generate equations of any order central moments automatically. In systems involving non-linear connections, the set of differential equations generated by the method is in principle infinite and therefore unsolvable. One can overcome this by applying a moment closure (mc) approximation, to close the equations by expressing moments of order $n+1$ with lower order ones based on the properties of an assumed underlying distribution. Normal and Poisson approximations have been successfully applied to 1D systems, but the advantages of considering various multivariate distributions are still to be explored. In our work we extend the approaches of Ale et al. and Gillespie [2] by developing multivariate type mc methods based on multivariate normal, lognormal and gamma distributions, the latter taken from the definition in [3]. In each case the entire distribution and therefore all higher order moments are completely determined by parameters which can be obtained using up to second order moments. We use this property to automatically generate expressions for all moments of an arbitrary order and number of variables and embed them in the general MEA framework. Instead of applying zerotruncation (setting all higher order terms to zero) as in [1], for an *nth*-order closure approximation $n+1$ order moments are calculated in terms of means, variances and covariances and substituted into the ODE system describing the time evolution of central moments. We demonstrate the method on the oscillating p53 system, for which in the zero-truncation case at least 6 moments are needed to approximate the mean behaviour observed in an ensemble of stochastically evolving cells. The performance of each mc is measured by their squared approximation error compared to zero-truncation (100%). As expected, normal closure causes the smallest change in the quality of estimation (and only for odd-order cases), but the error is still reduced to half. Lognormal and gamma closures both perform very well (decreasing the approximation error for 3rd-order closure to 4 and 17%, respectively), and are able to capture the characteristic dampening in oscillations from as few as 3 moments. These results are further substantiated by comparing the distribution of p53 molecules obtained from exact stochastic simulations with the ones based on the respective moment closure assumptions. As the underlying distribution is asymmetric, lognormal and gamma distributions offer a better approximation than normal. MEA with mc can be used to efficiently characterize the dynamics of large non-linear biochemical systems with low molecular numbers, when deterministic approaches or the linear noise approximation would fail. The use of mc methods improves the approximation 10-fold and can capture the same behaviour with a 5-times smaller equation set. mc-MEA also provides a natural framework for sensitivity analysis and parameter inference, where the reduction in system size proves crucial. mc-MEA is available as a Python package containing the above applications.

References

[1] Ale A, Kirk P, and Stumpf MPH (2013) A general moment expansion method for stochastic kinetic models. *J Chem Phys*, 138(17):174101.

[2] Gillespie CS (2009) Moment-closure approximations for mass-action models. *IET Syst Biol*, 3(1): 52–58.

[3] Furman E (2008) On a multivariate gamma distribution. *Statistics & Probability Letters*, 78(15): 2353–2360.

OPTIMAL OBSERVATION TIME POINTS IN STOCHASTIC CHEMICAL KINETICS

Charalampos Kyriakopoulos[1] *and Verena Wolf* [1]

[1]Department of Computer Science, University of Saarland,
P.O. Box 66123, Saarbruecken, Germany
charalampos.kyriakopoulos@uni-saarland.de; wolf@cs.uni-saarland.de

ABSTRACT

The estimation of the reaction rate parameters in stochastic biochemical networks is usually based on taking observations of the real system at discrete time points and computing then the value of the parameters that maximizes the likelihood of the observations. Nevertheless, this choice of the observation time points can be far from optimal in the sense of the information it provides for the true value of the parameters. Using optimal design criteria based on Fisher Information Matrix we set up best informative experiments for estimating the system's rate parameters. We apply our design in three biochemical models considering the general case of n possible observation time points and k unknown parameters.

DISTRIBUTION RECONSTRUCTION FOR STOCHASTIC CHEMICAL KINETICS FROM CONDITIONAL MOMENT DATA

Alexander Andreychenko[1], Linar Mikeev[1] and Verena Wolf [1]

[1]Department of Computer Science, University of Saarland,
P.O. Box 66123, Saarbruecken, Germany
makedon@mosi.cs.uni-sb.de

ABSTRACT

The dynamics of stochastic chemical kinetics can be accurately described by the Chemical Master Equation (CME) that governs the flow of the associated probability distribution. An analytic solution for the CME can be obtained only in a limited number of cases therefore computationally extensive numerical methods have to be applied. Another way to describe the stochastic evolution of a biochemical reaction network is to approximate the moments of the associated random process. However for the species that are present only in small quantities a moment description may be not adequate. To overcome this difficulty we integrate conditional moments such that the distribution of the species that are present in small quantities is governed by a "small" CME. The distribution of those species present in large quantities is represented by a finite number of moments and the evolution is given by the corresponding moment's equations. In addition to the conditional moments computed at a certain time point of interest and probabilities of the species that are present in small quantities, we want to reconstruct the joint probability distribution of the stochastic process to be able to compute likelihoods or the probabilities of certain events. We propose a reconstruction based on the maximum entropy approach combined with a fast integration of the conditional moments and compare our results both to the direct numerical solution of the CME and a reconstruction based on the unconditional moments. For our comparison we consider two case studies: the exclusive switch and a multi-attractor model.

SENSITIVITY ANALYSIS AND PARAMETER ESTIMATION OF A HUMAN BLOOD GLUCOSE REGULATORY SYSTEM MODEL

Domokos Meszéna[1], Eszter Lakatos[1,2] and Gábor Szederkényi[1]

[1]Pázmány Péter Catholic University, Faculty of Information Technology and Bionics,
Práter u. 50/a, H-1083 Budapest, Hungary
[2]Division of Molecular Biosciences, Theoretical Systems Biology Group,
Imperial College London, London SW7 2AZ, United Kingdom
szederkenyi@itk.ppke.hu

ABSTRACT

Quantitative dynamical mathematical models are certainly useful in the thorough understanding and possible manipulation (control) of biological processes. However, complex nonlinear models typically contain large numbers of parameters whose values are not precisely known. Determining these parameters from experimental data compromised by various types of disturbances and noise is often a challenging task. Sensitivity analysis can provide valuable insights about the robustness of biological responses and how parameters contribute to model output variations. As this impact of model parameters changes over time, time-dependent sensitivity analysis is needed in models where the underlying dynamics are likely to be composed of different time-scales. In our work we further investigate a model of human blood glucose control for which structural identifiability was previously shown [2], now focusing on separated estimations of the parameters based on the their sensitivities. We analyse the mathematical model published in [1]. The state variables tracked in the model are: plasma and cellular hormones, hormone-bound receptor concentrations for glucagon and insulin, together with blood glycogen and blood glucose levels, the latter being the measured output. The feedback is incorporated in the glucose-dependent hormone infusion rates (see the detailed explanation of the model in [1]). In this study we investigate which parameters might be neglected at some time intervals to avoid problems arising from lack of identifiability. Our aim is to quantify the importance of parameters at different stages of the regulatory process, and thus to achieve improvement in model fit and obtain a technique that is more suitable to examine real experiments and later design artificial control for the accurate tracking of insulin level. Time-dependent parameter sensitivities are analysed with the CVODES solvers in the SUNDIALS library [3]. Based on the obtained time-dependent sensitivity coefficients we divide the parameters into three subsets and estimate them separately while considering the values of other parameters fixed. The estimation procedure is set to minimize the standard normed quadratic error between the experimental data taken from literature and the simulated output. Least squares method and the pattern search optimization algorithm are applied to the parameter sets with linear and nonlinear dependencies, respectively. During sensitivity analysis we find that the time-scale on which different parameters affect the output naturally divides them into three groups: parameters with transient, middle-range and long-term impact. This is consistent with the assumption that the overall response consists of subprocesses where different parameters have temporary significance. Strictly glucagon connected parameters have nonzero sensitivities only up to 10 minutes, corresponding to an initial glucagon release, which is quickly shut down by the increased blood glucose level. On the other hand, parameters related to insulin-dependent glucose uptake have negligible impact on this time-scale but exert influence on long-term dynamics and steady-state concentrations. We demonstrate that the refined estimation technique can improve model fitting and is able to reduce the objective function value by about 4% compared to [1]. Our analysis reveals the time-scales of the key mechanisms behind blood glucose control. These results can be applied to study and compare abnormal and healthy glucose regulatory systems, as well as to design an advanced nonlinear control method, which acts directly on the crucial processes.

References

[1] Liu W and Tang F (2008) Modeling a simplified regulatory system of blood glucose at molecular levels. *J Theor Biol*. 252(4): 608–620.
[2] Lakatos E, Meszéna D, Szederkényi G (2013) Identifiability analysis and improved parameter estimation of a human blood glucose control system model. *Lecture Notes in Computer Science*, A. Gupta and T.A. Henzinger (Eds.): CMSB 2013, LNBI 8130 Springer, pp. 248–249.
[3] Hindmarsh AC, Brown PN, Grant KE, et al. (2005) SUNDIALS: suite of nonlinear and differential/algebraic equation solvers. *ACM Trans Math Softw*. 31(3): 363–396.

COMPUTATIONAL MODEL OF FGF2 SIGNALLING NETWORK IN HUMAN EMBRYONIC STEM CELLS REVEALED OXYGEN TRIGGERED METABOLIC SWITCH

Anton Salykin[1], Dominique Chu[2], Petr Fojtik[1], Miriama Kruta[1], Katka Schrammova[1], Petr Skladal[4,5], Petr Dvorak[1,3] and Vladimir Rotrekl[1,3]

[1]Department of Biology, Faculty of Medicine, Masaryk University, Brno, Czech Republic
[2]School of Computing, University of Kent, UK
[3]International Clinical Research Centre St.Anne's University Hospital, Brno, Czech Republic
[4]Department of Biochemistry, Faculty of Science, Brno, Czech Republic
[5]CEITEC, Masaryk University, Brno, Czech Republic
P.O. Department of Biology, Kamenice 5, 625 00 Brno, Czech Republic
asalykin@med.muni.cz

ABSTRACT

Human embryonic stem cells (hESC) have the ability to differentiate into any cell type of human body and as such they may provide for cornerstone of future regenerative medicine. Contrary to differentiated cells, pluripotent stem cells meet their energy demands mostly via glycolysis rather than by oxidative phosphorylation. Such metabolic switch seems to be important for stem cell's identity, self-renewal and early differentiation. Hence, understanding the regulation of the choice of metabolic pathway is fundamental to *in vitro* hESC production. Basic fibroblast growth factor (FGF2) is a key factor responsible for stemness maintenance in pluripotant stem cells *in vivo* as well as upon *in vitro* cultivation. Contrary to self-renewal and antiapoptotic regulation the FGF2 effect on the metabolism has not been so far described. Our data show, that the key metabolic targets regulated by FGF2 include hexokinase (HK1), importing glucose into the cell and pyruvate dehydrogenase (PDH), converting pyruvate to acetyl-CoA and thus allowing for the Krebs cycle and coupled oxidative phosphorylation in mitochondria. Analysis of the response of mitogen activated kinase and AKT pathways to FGF2 and their coupling with PDH deactivation revealed differential response of hESC metabolism to FGF2 stimulus in normoxic and hypoxic conditions. Computer model based on the experimental data a) showed that the whole model can be explained by only limited number of interactions setting base for future dynamic modeling and b) allowed for the simulation of the hESC metabolic behavior upon FGF2 stimuli in varying oxygen concentration. The model revealed elaborate metabolic switch responding to FGF2 and oxygen concentration, enabling the hESC to mine maximum energy under actual oxygen condition without compromising the cell's stability.

Acknowledgments
The work was supported by the Grant Agency of the Czech Republic No. GA13-19910S and GBP302/12/G157.

IMAGE ALIGNMENT AND LINEAGE CONSTRUCTION TOOL TO STUDY SEGREGATION AND PARTITIONING IN DIVISION OF UNWANTED PROTEIN AGGREGATES IN ESCHERICHIA COLI

João Santinha[1], Antti Häkkinen[2], Abhishekh Gupta[2], Teppo Annila[2], Jason Lloyd-Price[2], Andre S. Ribeiro[2] and José Fonseca[3]

[1]CA3 – UNINOVA - Center of Technology and Systems, FCT Campus, Universidade Nova de Lisboa, 2829-516, Caparica, Portugal
[2]Laboratory of Biosystem Dynamics, Department of Signal Processing, Tampere University of Technology. P.O. Box 553, FI-33101 Tampere, Finland
[3]Department of Electrotechnical Engineering, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, 2829-516, Caparica, Portugal
jsantinha@uninova.pt

ABSTRACT

Escherichia coli is a model organism of several biological processes, including cellular aging. Regarding the aging-related processes in *E. coli*, time-lapse images of cells are needed, so that the processes of segregation and partitioning in division of unwanted aggregates can be studied. Further, two different image acquisition techniques are needed, namely, phase-contrast and confocal, in order to obtain morphological and functional information. Since the two images are acquired with different resolution and non-coincident fields of view their alignment is mandatory. The large number of images hampers the use of manual analysis. To assist in studies of cellular aging, based on the software "CellAging", we developed additional means for image alignment, automatic or computer-aided manual cell segmentation and tracking, division detection and lineage construction. Additionally to these functions, the tool can also perform the detection of intracellular fluorescent spots. As output, this tool provides, for each cell, information on the cell morphology (location, size and, orientation angle), on cell lineages (division time and parent-daughter relationships) and on intracellular fluorescent spots (location, size and, intensity). The tool should be of assistance to studies of the kinetics and spatial distribution of fluorescently tagged protein aggregates in bacteria. Also, it can aid in studying the growth kinetics of populations in various conditions.

Acknowledgments
Work partially supported by the Portuguese funding institution FCT - Fundação para a Ciência e a Tecnologia – project grant PTDC/BBB-MET/1084/2012.

GENERATOR OF SYNTHETIC TIME-LAPSED, PHASE-CONTRAST MICROSCOPY IMAGES OF GROWING *E. COLI* CELL POPULATIONS

Leonardo Martins[1], Andre S. Ribeiro[2] and José Fonseca[1]

[1]Department of Electrotechnical Engineering, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, 2829-516, Caparica, Portugal
[2]Department of Signal Processing, Tampere University of Technology, P.O. Box 553, FI-33101 Tampere, Finland
lmartins@uninova.pt

ABSTRACT

Escherichia coli is a model organism in studies of biomolecular and cellular processes, including gene expression and aging. Nowadays, these studies rely on temporal single-cell imaging. The analysis of the images is, however, complex and cumbersome and, thus, requires automated image analysis methods of cell segmentation and tracking, which require validation. The best means to validate such methods is to know the “ground truth”, which is only possible provided synthetic images. Further, such synthetic images should have similar statistical distributions and noise levels as Phase-Contrast Microscopy images. We present a tool, currently under development, for generating artificial time series of Phase-Contrast microscopy images of *E. coli* cells with realistic morphological features (e.g. shape, curvature, and size). Further, the tool models cell growth and division. When generating synthetic images, it is possible to define image size, initial number of cells and degree of noise of the images. This tool, once complete, will be validated by comparing the morphological properties of synthetic cells along with their kinetics of cell growth and division, with the same properties as extracted from microscopy measurements. Finally, once validated, this tool should also be of use in generating null-models that will aid in the quantification of cellular processes such as the degree of asymmetry in polar segregation of aggregates or in the positioning of the nucleoid in the cells.

Acknowledgments
Work supported by the Portuguese funding institution FCT - Fundação para a Ciência e a Tecnologia – scholarship SFRH/BD/88987/2012.

DETECTING GENETIC INTERACTIONS IN CASE-CONTROL DATA: INFORMATION THEORY BASED METHODS

Tomasz Ignac[1,2], Alexander Skupin[1,2], Nikita A. Sakhanenko[2], and David J. Galas[1,2]

[1]Luxembourg Centre for Systems Biomedicine, 7, Avenue des Haute Forneaux,
L-4362 Esch-sur-Alzette, Luxembourg
[2]Pacific Northwest Diabetes Research Institute, 720 Broadway, Seattle, Washington, 98122 USA
tomasz.ignac@uni.lu

ABSTRACT

Phenotypic variations, including those which underlay health and disease in humans, result from multiple interactions among both genetic variations and environmental factors. While diseases caused by single gene variants can be identified by established association methods and family-based approaches, complex phenotypic traits resulting from multigene interactions remain very difficult to characterize. Recently, new information theory based methods that are aimed specifically at detecting complex, non-additive interactions have been proposed. Such approaches have the advantage of being intrinsically model-free and parsimonious and thus offer an unbiased and potentially statistically powerful approach to detection of gene interactions. Moreover, even in situations when the sample size is not large enough for making statistically confident assessments, these methods can often be used to filter candidate gene interactions and to generate useful hypotheses. Here we present a set of tools based on information theory. We start from discussing naïve straightforward applications of conditional entropy and mutual information, and demonstrate that they become inefficient when applied to large data sets. Then, we present the Interaction Distance (ID) – a measure of information among three variables. It combines two concepts: the interaction information, which is a generalization of the mutual information to three variables, and the normalized information distance which measures informational sharing between two variables. The specific design of ID allows for a computationally efficient detection of pairwise effects on phenotype of genetic markers that have negligible effect while considered alone. We demonstrate how ID improves on previous approaches to identify genetic interactions in data sets emerging from case-control studies. While ID aims at detecting pairwise interactions, it is now hypothesised that complex diseases are caused by complex and non-linear multi-gene interactions that determine phenotypic traits. Based on our work, we argue that information theory related methods could be extended to these cases.

MICROORGANISM'S GROWTH-ROBUSTNESS TRADE-OFF EXPLAINED BY SIMPLE GENERAL PRINCIPLES

Alessandro Bolli[1] and Armindo Salvador[1,2]

[1]Computational Systems Biology Group, Center for Neuroscience and Cell Biology, University of Coimbra
[2]Chemistry Department, University of Coimbra, Portugal Departamento de Química, University of Coimbra, 3004-535 Coimbra, Portugal
aless.bolli@gmail.com

ABSTRACT

Microorganisms show some conserved relations between physiological state and environmental conditions. Growth Robustness Reciprocity (GRR) is an intriguing example: genetic and environmental factors that impair microorganism's vegetative performance (growth rate) enhance their ability to resist abiotic stresses, and *vice-versa*. Mechanistically, this relationship may be explained by regulatory interactions that determine higher expression of protection mechanisms in response to low growth rates. However, these mechanisms are not conserved. Therefore, the observed GRR must result from convergent evolution. Why does natural selection favor such an outcome? Here, we used mathematical models of an idealized cellular self-replicating system to identify the general evolutionary and physiological principles that may explain why GRR is widespread. These models account for the cell protein components involved in: (i) substrate uptake from environment, (ii) metabolic transformation, (iii) biosynthesis (e.g. ribosomes and lipid synthesizing enzymes), (iv) protein inactivation, representing stress, and (v) stress defenses. A non-linear optimization algorithm has been used to generate allocation patterns of biosynthetic resources (ribosomes devoted to the synthesis of each cell component) which maximize the growth rate. This method reproduces the ability of real cells to regulate gene expression and thus cell physiology following environmental changes. We found that optimal resource allocation determines that at high substrate availabilities and low stress intensities stress defenses are not expressed. Under these conditions, stress tolerance ensues from growth-related damage dilution: the higher the substrate availability, the highest the growth rate, the fastest the dilution of damaged proteins by newly synthesized proteins, the highest the stress that can be tolerated until the inactive pool drains all the cell resources necessary for growth. In turn, under low substrate availability or high stress lower growth rates can be attained, and thus growth-related damage dilution is less effective. Under these conditions, growth is maximized when stress defenses are expressed, and their expression is higher the lower the nutrient concentrations and biosynthetic efficiency. Our optimality models reproduce the negative correlation between the expression of stress protection and the expression of growth-promoting genes that is observed in microorganisms. As a consequence of this phenomenon, slow-growing but not fast-growing cells are pre-adapted to withstand acute stresses. Overall, these results show that GRR can be explained by the interplay among three general principles. Namely (a) damage and biosynthesis errors that inactivate cellular components are inevitable, (b) faster-growing cells are often selectively favored, (c) growth-associated synthesis of new components dilutes damage, with the consequence that cells in more favorable environments grow fastest when expressing defenses the least.

Acknowledgments
We acknowledge fellowship SFRH/BPD/90065/2012 and grants PEst-C/SAU/LA0001/2013-2014, PEst-OE/QUI/UI0612/2013, FCOMP-01-0124-FEDER-020978 financed by FEDER through the "Programa Operacional Factores de Competitividade, COMPETE" and by national funds through "FCT, Fundação para a Ciência e a Tecnologia" (project PTDC/QUI-BIQ/119657/2010).

IN VIVO RETENTION OF PROTEIN AGGREGATES IN ESCHERICHIA COLI

Nadia Goncalves[1], Ramakanth N. Venkata[1], Antti Martikainen[1], Joao A. Santinha[2], Jose M. Fonseca[2], and Andre S. Ribeiro[1,]*

[1]Department of Signal Processing, Tampere University of Technology, 33101 Tampere, Finland
[2]Department of Electrotechnical Engineering, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Lisboa, Portugal
*andre.ribeiro@tut.fi

ABSTRACT

Recently, it has been established that *Escherichia coli* cells accumulate protein aggregates at their polar regions. Using fluorescent tagging of IbpA proteins, which are known to co-localize with protein aggregates, it has been further shown by following the aggregates along cell lineages that, due to their retention at the poles, they become asymmetrically distributed by the cells of the lineage. However, there is no evidence for a transport mechanism responsible for the accumulation at the poles. Instead, it has been hypothesized that this process is associated with the presence of the nucleoid at midcell. Here, we describe ongoing as well as planned measurements which aim to establish and characterize the role of the nucleoid in the processes of segregation and subsequent retention of protein complexes at the cellular poles. In particular, in these experiments, we will change the size of the nucleoid by subjecting cells to different media and chemical stress conditions and observe the consequences on the spatial distribution of the complexes. For this, we are using *E. coli* MGAY strain expressing the fluorescent IbpA-YFP proteins and DAPI (4', 6-diamidino-2-phenylindole) staining of the nucleoid. Finally, we will present preliminary results that suggest that the segregation and retention of aggregates at the cell poles, while it a robust process, it is not immune to perturbations of the nucleoid structure, which provides supporting evidence that the nucleoid plays a role in the segregation and subsequent retention of these unwanted protein aggregates.

ANALYSIS OF THE ROBUSTNESS OF A SINGLE COPY REPRESSILATOR

Samuel M.D. Oliveira[1], Jerome G. Chandraseelan[1], Antti Häkkinen[1]
Olli Yli-Harja[1,2] and Andre S. Ribeiro[1]

[1]Laboratory of Biosystem Dynamics, Department of Signal Processing, Tampere University of Technology, P.O. Box 553, FI-33101 Tampere, Finland
[2]Institute for Systems Biology, 1441N 34th Street, Seattle, WA 98103-8904, USA
samuel.oliveira@tut.fi; jerome.chandraseelan@tut.fi; antti.hakkinen@tut.fi;
olli.yli-harja@tut.fi; andre.ribeiro@tut.fi

ABSTRACT

An effective way to understand the functioning of natural biological systems is to create synthetic networks, so that all components are known *a priori* and probes are set to track their behavior inside cells. One of the most studied of such synthetic circuits is the repressilator, a system of three genes connected in a negative feedback loop that, as it oscillates, tracks time throughout cell generations. Together with the repressilator, there is a reporter system that reproduces the oscillations of the system in the temporal levels of a green fluorescent protein (GFP), under the control of a promoter that is also present in the repressilator network. One variable that is able to affect protein numbers in this system is temperature. Another is the number of copies of the repressilator within a cell. Here, we study the effects of temperature in two systems. The first is the low-copy Elowitz et al repressilator (LCR), which contains 3-5 copies of the genetic circuit in each cell. The second we engineered, and it is a single-copy repressilator (SCR). We measure in multiple cells, for 10 hours, the mean intensity of GFP signal at the single cell level, in each of the systems. The results of the data analysis show that the robustness of the periods, as measured from the reporter signal in both LCR and SCR, is identical. This is unexpected, given the differences in copy numbers. Thus, we hypothesize that the LCR system is, somehow, being “favored” by the reporter system. To explain this observation and test our hypothesis, we make use of a stochastic model to compute the possible real robustness of the two repressilator systems. We will present our results from ongoing studies that may shed a light on this surprising observation.

ESTIMATING THE DURATION OF OPEN COMPLEX FORMATIONFROM *IN VIVO* MEASUREMENTS

Jerome G. Chandraseelan[1], Jason Lloyd-Price[1], Olli Yli-Harja[1,2] and Andre S. Ribeiro[1]

[1]Department of Signal Processing, Tampere University of Technology,
P.O. Box 553, FI-33101 Tampere, Finland
[2]Institute for Systems Biology, 1441N 34th Street, Seattle, WA 98103-8904, USA
jerome.chandraseelan@tut.fi; jason.lloyd-price@tut.fi; olli.yli-harja@tut.fi; andre.ribeiro@tut.fi

ABSTRACT

In *Escherichia coli*, gene expression regulation occurs mostly during transcription initiation. This multi-stepped process begins when an RNA polymerase (RNAp) binds to a promoter leading to the formation of a closed complex (CC), which, is subsequently followed by a series of intermediate isomerization steps leading to the formation of an open complex (OC), where the strands of the double-stranded DNA are separate and becomes catalytically active, thus enabling the RNAp to start the synthesis of RNA. A number of previous studies, using *in vitro* techniques, have been done to determine the number and duration of the steps involved in the OC formation, on various promoters. However, little is known about the kinetics of this process *in vivo*. Our group, using inference methods and live, single RNA detection techniques, has established a protocol to determine the number and duration of the steps in transcription initiation in vivo. One major limitation of the present method is the inability to obtain temporal relationships between the steps inferred. To address this issue, here we present a novel method that can be used to directly estimate the time taken for the formation of open complex *in vivo*.

IN SILICO SCREENING OF LEAD-LIKE MOLECULES EN ROUTE TO ANTITUMOR AND ANTIBIOTIC DRUGS FROM MARINE AND MICROBIAL NATURAL PRODUCTS USING EMPIRICAL AND SEMI-EMPIRICAL QUANTUM-CHEMICAL DESCRIPTORS

Florbela Pereira[1], Diogo A. R. S. Latino[1,2] and Susana P. Gaudêncio[1]

[1]REQUIMTE/CQFB, Departamento de Química, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, 2829-516 Monte de Caparica, Portugal
[2]CCMM, Departamento de Química e Bioquímica, Faculdade de Ciências, Universidade de Lisboa, Campo Grande, Lisboa 1749-016, Portugal
florbela.pereira@fct.unl.pt

ABSTRACT

Natural products (NPs), or synthetic products inspired by NPs, have been the single most productive source of leads for the development of drugs. In fact, more than half of the approved drugs from 1981 to 2010 were based on NPs [1]. A Quantitative Structure−Activity Relationship approach was used for the classification and prediction of active/inactive compounds relatively to overall biological activity, antitumor and antibiotic activities using a data set of 1746 compounds from PubChem. Models were built with Rfs, SVMs and decision trees using empirical CDK descriptors and semiempirical quantum-chemical descriptors. The best classification models for antibiotic and antitumor activities were used to screen a data set of marine and microbial natural products from the AntiMarin database. The screen originates 25 and 4 possible lead-like compounds for antibiotic and antitumor drug design, respectively. The present work corroborate by one side the results of our previous work [2] and enable the presentation of a new set of possible lead like bioactive compounds. In other side it is shown the usefulness of quantum-chemical descriptors in the discrimination of biological active and inactive compounds. The use of the ε_{HOMO} quantum-chemical descriptor in the discrimination of large scale data sets of lead-like or drug-like compounds has never been reported. This approach results in the reduction, in great extent, of the number of compounds used in real screens, and it reinforce the results of our previous work.

References

[1] Cragg GM, Newman DJ (2013) Natural products: A continuing source of novel drug leads. *Biochimica et Biophysica Acta*. 1830(6): 3670-3695.
[2] Pereira F, Latino DARS, Gaudêncio SP (2014) A chemoinformatics approach to the discovery of lead-like molecules from marine and microbial sources en route to antitumor and antibiotic drugs. *Mar Drugs*. 12(2): 757–778.

Acknowledgments

Financial support from FCT and FEDER (PTDC/QUI-QUI/119116/2010: Ocean Treasures) and European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement n° PCOFUND-GA-2009-246542. We thank ChemAxon Ltd. for access to JChem and Marvin, Molecular Networks GmbH for access to CORINA. The authors thank Professor W. Fenical for access to AntiMarin707 database, while S.P.G. was a postdoctoral researcher at Scripps Institution of Oceanography, San Diego, CA, USA.