SIBERIAN BRANCH OF THE RUSSIAN ACADEMY OF SCIENCES
INSTITUTE OF COMPUTATIONAL MATHEMATICS AND MATHEMATICAL GEOPHYSICS
FEDERAL RESEARCH CENTER INSTITUTE OF CYTOLOGY AND GENETICS

MATHEMATICAL MODELING AND HIGH PERFORMANCE COMPUTING IN BIOINFORMATICS, BIOMEDICINE AND BIOTECHNOLOGY (MM-HPC-BBB-2016)

Abstracts

MM-HPC-BBB-2016
Novosibirsk, Russia
29 August – 2 September, 2016

Novosibirsk
2016

PROGRAM COMMITTEE

CHAIRS

S.I. Kabanikhin, Prof., RAS corresponding member, Institute of Computational Mathematics and Mathematical Geophysics SB RAS, Russia

N.A. Kolchanov, Prof., full member of the Russian Academy of Sciences, Institute of Cytology and Genetics SB RAS, Russia

VICE-CHAIRS

H.T. Banks, Prof., North Carolina State University, USA

MEMBERS

Yu.E. Anikonov, Prof., Sobolev Institute of Mathematics SB RAS, Russia

Yu.S. Aulchenko, Ph.D., Institute of Cytology and Genetics SB RAS, Russia

M.A. Bektemesov, Prof., Al-Farabi Kazakh National University, Kazakhstan

G.A. Bocharov, Prof., Institute of Computational Mathematics RAS, Russia

M. Chen, Prof., Zhejiang University, China

I.G. Chernykh, Ph.D., Institute of Computational Mathematics and Mathematical Geophysics SB RAS, Russia

A.P. Chupakhin, Prof., Lavrentyev Institute of Hydrodynamics SB RAS

G.V. Demidenko, Prof., Sobolev Institute of Mathematics SB RAS, Russia

G.N. Erohin, Prof., Kant's Baltic Federal University, Russia

S.I. Fadeev, Prof., Sobolev Institute of Mathematics SB RAS, Russia

M.P. Fedoruk, Prof., Novosibirsk State University, Russia

A.M. Fedotov, RAS corresponding member, Novosibirsk State University, Russia

B.M. Glinskiy, Prof., Institute of Computational Mathematics and Mathematical Geophysics SB RAS, Russia

V.P. Golubyatnikov, Prof., Sobolev Institute of Mathematics SB RAS, Russia

S.K. Golushko, Prof., Design Technological Institute of Digital Techniques of SB RAS, Russia

S.S. Goncharov, RAS corresponding member, Prof., Sobolev Institute of Mathematics SB RAS, Russia

R. Hofestaedt, Prof., University of Bielefeld, Germany

A.I. Ilin, Prof., JSC "NTSPP", Kazakhstan

I.V. Koptiug, Prof., International Tomography Center SB RAS, Russia

O. Krebs, Dr., Heidelberg Institute for Theoretical Studies (HITS), Heidelberg, Germany

I.M. Kulikov, Ph.D., Institute of Computational Mathematics and Mathematical Geophysics SB RAS, Russia

G.M. Kuramshina, Ph.D., MSU, Russia

M.M. Lavrentiev, Prof., Novosibirsk State University, Russia

V.A. Likhoshvai, Ph.D., Institute of Cytology and Genetics SB RAS, Russia

V.Yu. Makeev, Ph.D., Vavilov Institute of General Genetics RAS, Russia

A.G. Marchuk, Prof., Institute of Informatics Systems SB RAS, Russia

M.P. Moshkin, Prof., Institute of Cytology and Genetics SB RAS, Russia

D.B. Nurseitov, Ph.D., Al-Farabi Kazakh National University, Kazakhstan

Yu.L. Orlov, Ph.D., Institute of Cytology and Genetics SB RAS, Russia

I.V. Oseledec, Ph.D., Skolkovo Institute of Science and Technology, Russia

N.V. Pertsev, Prof., Omsk branch Institute of Mathematics SB RAS, Russia

A.Y. Rzhetsky, Prof., The University of Chicago, USA

A.M. Samsonov, Ph.D., St.Petersburg State Polytechnical University, Russia

Y.I. Shokin, RAS member, Institute of Computational Technologies SB RAS, Russia

V.N. Snytnikov, Ph.D., Boreskov Institute of Catalysis SB RAS, Russia

Ch. Tsoumpas, Ph.D., University of Leeds, UK

E.E. Vityaev, Prof., Glinsky B.M., Prof., Institute of Computational Mathematics and Mathematical Geophysics SB RAS, Russia

A.G. Yagola, Prof., Lomonosov Moscow State Univerisity, Russia

ISBN 978-5-91291-028-9

ORGANIZING COMMITTEE

CHAIR
S.I. Kabanikhin, ICM&MG SB RAS, Novosibirsk, Russia

VICE CHAIR
M.A. Shishlenin, Sobolev Institute of Mathematics, Novosibirsk, Russia

SCIENTIFIC SECRETARY
D.A. Voronov, ICM&MG SB RAS, Novosibirsk, Russia

MEMBERS
A.Yu. Belonog, NSU, Novosibirsk, Russia
I.G. Chernykh, ICM&MG SB RAS, Novosibirsk, Russia
D.V. Ermolenko, NSU, Novosibirsk, Russia
A.A. Grodz, NSU, Novosibirsk, Russia
E.S. Kabanikhina (designer), ICM&MG SB RAS, Novosibirsk, Russia
V.N. Kashtanova, NSU, Novosibirsk, Russia
D.V. Klyuchinskiy, NSU, Novosibirsk, Russia
O.I. Krivorotko, NSU, Novosibirsk, Russia
I.M. Kulikov, ICM&MG SB RAS, Novosibirsk, Russia
V.A. Latishenko, NSU, Novosibirsk, Russia
V.I. Lifenko, NSU, Novosibirsk, Russia
N.S. Novikov, ICM&MG SB RAS, Novosibirsk, Russia
Yu.L. Orlov, Institute of Cytology and Genetics SB RAS, Russia
N.L. Podkolodnyy, Institute of Cytology and Genetics SB RAS, Russia
E.I. Vostrikova, NSU, Novosibirsk, Russia

CONTACTS

Organizing committee: dmitriy.voronov@sscc.ru
Institute of Computational Mathematics and Mathematical Geophysics
The Siberian Branch of the Russian Academy of Sciences
630090 Novosibirsk, Lavrentyeva, 6
Tel: +7(383) 330-83-53
Fax: +7(383) 333-12-78
URL - ICMMG SB RAS: http://icmmg.nsc.ru/
URL - MM-HPC-BBB-2016: http://conf.bionet.nsc.ru/mm-hpc-bbb-2016/

Organizers

Institute of Computational Mathematics and Mathematical Geophysics,
Siberian Branch of the Russian Academy of Sciences

Federal Research Center Institute of Cytology and Genetics,
Siberian Branch of the Russian Academy of Sciences

Sponsors

Sobolev Institute of Mathematics,
Siberian Branch of the Russian Academy of Sciences

Novosibirsk State University

The Federal Agency for Scientific Organizations

Russian Foundation for Basic Research

ИВМиМГ СО РАН

Директор – член-корреспондент РАН Сергей Игоревич Кабанихин

Институт вычислительной математики и математической геофизики Сибирского отделения РАН (ИВМиМГ СО РАН), ранее Вычислительный Центр СО АН СССР (ВЦ СО АН СССР), создан распоряжением Совета Министров РСФСР № 1693-р от 4 мая 1963 г. и постановлением Президиума Академии наук СССР № 455 от 24 мая 1963 г.

Основные направления научной деятельности ИВМиМГ СО РАН:

- вычислительная математика,
- математическое моделирование и методы прикладной математики,
- параллельные и распределенные вычисления,
- информационные системы.

В состав ИВМиМГ СО РАН входят 19 научных лабораторий. На конец 2015 г. в институте работают 298 сотрудников, в том числе один академик РАН, два члена-корреспондента РАН, 43 доктора наук, 84 кандидата наук.

ИВМиМГ СО РАН является признанным лидером в области разработки численных методов решения прямых и обратных задач математической физики, методов численного статистического моделирования (методов Монте-Карло), геофизики, физики атмосферы, океана и окружающей среды, химии, электрофизики. Разработанные алгоритмы и программы применяются при решении актуальных задач рационального природопользования, поиска месторождений нефти и газа, предсказания природных и техногенных катастроф и оценки их последствий, зондирования Земли из космоса, медицины, наноиндустрии, информационной безопасности, разработки перспективной суперкомпьютерной техники.

В институте работают семь научных семинаров. ИВМиМГ СО РАН является базовым институтом для четырех кафедр НГУ и двух кафедр НГТУ. В институте проводится обучение в аспирантуре по шести образовательным программам (специальностям). В ИВМиМГ СО РАН работают два совета по защите диссертаций на соискание ученой степени кандидата наук, на соискание ученой степени доктора наук.

На базе ИВМиМГ СО РАН созданы Центр коллективного пользования “Сибирский суперкомпьютерный центр СО РАН», Фонд алгоритмов и программ СО РАН.
ИВМиМГ СО РАН издает следующие научные журналы:

- Сибирский журнал вычислительной математики (Numerical analysis and applications)
- Проблемы информатики
- Bulletin of Novosibirsk Computing Center

CONTENTS

INVESTIGATION OF THE INFLUENCE OF GENOTYPE ON THE STRUCTURE OF THE CIRCULATORY SYSTEM LABORATORY MICE

A.E. Akulov[1], S.V. Maltseva[2, 3], M.P. Moshkin[1, 3], A.K. Khe[3, 4], A.A. Cherevko[3, 4], A.P. Chupakhin[3, 4], G.S. Yankova[3]*

[1]*Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2]*Sobolev Institute of mathematics SB RAS, Novosibirsk, Russia*
[3]*Novosibirsk State University, Novosibirsk, Russia*
[4]*Lavrentyev Institute of Hydrodynamics SB RAS, Novosibirsk, Russia*
** Corresponding author: galinayankova2703@gmail.com*

Key words: *variability, MRI imaging, the method of variation of scanning plane, segmentation, statistical analysis, CFD*

Motivation and Aim: Investigation of the structure of the circulatory system is of interest both in scientific terms and in terms of applications. The assumption about the relationship of genotype and vasculature structure is a natural. In this paper we consider models of vessel net for two genetic lines of laboratory mice. The purpose of the work is to compare the vascular network models of two genetic lines of animal.

Methods and Algorithms: MRI scanning was realized using the variation of slope of scanning plane method for the construction of the vascular network configurations [1]. The hemodynamic calculations for the constructed models were carried out using software ANSYS/CFX on the base of supercomputer center of Novosibirsk State University. On the basis of the obtained configurations was carried out statistical analysis of the blood flow parameters.

Results: Construction of models of vascular channel for laboratory animals of two genetic lines using method of variation of slope of scanning plane realized in the paper. On the basis of data obtained as a result of statistical and numerical analysis of two genetic lines of laboratory mice it was shown absence of influence of TNF knock-out on morphological and hydrodynamic characteristic of Willis' circle.

Conclusion: The method of variation of scanning plane shown its efficiency and allowed for construction models of vascular channel suitable for hemodynamic calculations and statistical analysis of vessel architectonics.

Acknowledgements: The study was completed thanks to the support of Russian Science

Foundation (project №14-35-00020, all MRI experimentation studies using Bruker BioSpec 117/16 USR scanner), Russian Foundation for Basic Research (project №14-01-00036, mathematical modelling) at the site of the Centre for Laboratory Animal Genetic Resources, Institute of Cytology and Genetics, SB RAS (RFMEFI61914X0005 and RFMEFI62114X0010). Also, study was supported by grant SS-8146.2016.1.

References:

1. S. V. Maltseva, A. A. Cherevko, A. K. Khe, A. E. Akulov, A. A. Savelov, A. A. Tulupov, E. Yu. Derevtsov, M. P. Moshkin, A. P. Chupakhin. (2016) Reconstruction of Complex Vasculature by Varying the Slope of the Scan Plane in High-Field Magnetic Resonance Imaging. Applied Magnetic Resonance. V 47, Iss 1, pp 23-39.

HLA TYPING PIPELINE FOR AMPLICON SEQUENCING

O.S. Altukhova[1]*, P.I. Borovikov[1], T. Jankevi[2], I.S. Balashov[1]
[1]*Academician V.I. Kulakov Research Center of Obstetrics, Gynecology and Perinatology, Ministry of Health, Russia*
[2]*NRC Institute of Immunology FMBA of Russia, Moscow, Russia*
** Corresponding author: olg333@yandex.ru*

Key words: *HLA, dynamic programming, PAM, semi-global alignment, OpenMP, NVIDIA CUDA, NGS*

Motivation and Aim: Dissimilarity of human leukocyte antigens (HLA) are primary reason of organ transplant rejections. In selecting donors becomes necessary genotyping — identifying specific alleles of HLA genes. One of molecular methods for solving this problem is a high-performance new-generation sequencing (NGS). A feature of HLA genes is their high polymorphism. For example, the database IMGT/HLA [1] contains 14473 alleles I and II classes HLA. Alleles can be different only a single substitution in exon, therefore considering sequencing errors, the correct identification the allele is a specific task. The conception of our work is the development of software for the analysis of high-throughput sequencing data obtained for HLA typing.

Methods and Algorithms: In this work was used pairwise alignment algorithm (semi-global) based on method of dynamic programming and OpenMP, NVIDIA CUDA technologies [2]. For the separation reads into two groups (two alleles) PAM (Partitioning Around Medoids) clustering algorithm was applied. To calculate the distance between reads we have proposed an approach based on identifying the most appropriate nucleotide positions by which a plurality of reads is well divided into two groups (and filtered positions arising from sequencing errors).

Results: As a result of our work, the software has been developed by which identified alleles of genes HLA I and II classes. The result is displayed in the G (HLA alleles that have identical nucleotide sequences across the exons, 2 and 3 for HLA class I and exon 2 only for HLA class II alleles, encoding the peptide binding domains) and P (HLA alleles having nucleotide sequences that encode the same protein sequence for the peptide binding domains) nomenclature. It is possible to trim the primers and configure alignment algorithms.

Conclusion: The use of dynamic programming algorithm for pairwise alignment allows for greater precision than using heuristic algorithms. At the same time, the use of technology for parallel computing significantly reduces the operating time. Further divisions reads group belonging to the same gene/exon helps to distinguish significant positions of nucleotides, which belong to different alleles, and positions that arise due to sequencing errors.

References:

1. Robinson J, Halliwell JA, Hayhurst JH, Flicek P, Parham P, Marsh SGE. (2016) The IPD-IMGT/HLA Database - New developments in reporting HLA, Human Immunology, **77**:233-237
2. Liu, D. Maskell, and B. Schmidt.(2009) CUDASW++: Optimizing Smith-Waterman Sequence Database Searches for CUDA-enabled Graphics Processing Units, BMC Research Notes, **2(1)**:73.

vlincRNA DATABASE: TOOL FOR VERY LONG INTERGENIC NON-CODING RNA FUNCTIONAL ANNOTATION

D. Antonets[1,2,4]*, Y. Vyatkin[2,3], D. Luppov[2,3], P. Kapranov[3,5], M. Ri[2,3], O. Saik[2,3,6], D. Shtokalo[1,2,3]

[1]*A.P.Ershov Institute of informatics systems, Novosibirsk, Russia*
[2]*AcademGene LLC, Novosibirsk, Russia*
[3]*St. Laurent Institute, Woburn, USA*
[4]*State Research Center of Virology and Biotechnology 'Vector', Novosibirsk, Russia*
[5]*Institute of Genomics, School of Biomedical Sciences, Huaqiao University, Xiamen, China*
[6]*Institute of cytology and genetics SB RAS, Novosibirsk, Russia*
** Corresponding author: antonec@yandex.ru*

Key words: *very long intergenic non-coding RNA, systems biology, function annotation, gene ontology, database, web-service*

Motivation and Aim: Vast amount of lncRNA species have been discovered recent years in mammalian genomes and stored on the web. However there is an impressive misbalance between portion of at least 10% of human genome occupied by non-coding RNA and relatively few facts known about its function. Our results suggest that vlincRNAs represent a hitherto hidden layer of regulation involved in critical biological processes and diseases. Many of them are highly expressed in cancers, and some are expressed in stem cells and appear to be regulated by transcription factors involved in stem cell differentiation [1]. Here we present our vlincDB – a web-tool for functional annotation and analysis of human very long intergenic non-coding RNAs.

Methods and Algorithms: vlincDB web-site was created with PHP (v.5) and bootstrap framework (v.3.3.6), all annotation tables are stored in MySQL database (v.14.14).

Results and Conclusion: VlincRNA database, provides both the list of 5151 putative very long intergenic non-coding RNA transcripts and the list of separate 1542 vlinc RNA genes. The integrated analysis of genomic features leveraged by transcription level data measured in 833 tissues and cell lines by FANTOM5 consortium allowed us to predict possible functions of vlincRNA genes [1]. The database provides annotation and a tool of search by gene ontology categories, SNP traits, chromatin modification states, ChIP-seq signals in vlincRNA genes promoter regions, by overlapping known lncRNA, nearby genes and other. Advanced tools implement scenario of overlapping vlincRNA genes with user-defined genomic intervals (provided in BED format), GO terms enrichment analysis and classification into cancer/normal/stem-cell categories according to FANTOM5 gene expression data. All annotation data can be downloaded.

Availability: http://office.nprog.ru:8081/table.php

References:

1. G. St. Laurent et al. (2016) Functional annotation of the vlinc class of non-coding RNAs using systems//*Nucl. Acid Res., doi: 10.1093/nar/gkw162.*

FEATURES OF SOLUTION OF SOME INVERSE PROBLEMS IN PK MODELS WIITH FIRST ODER ABSORPTION

N. Asmanova*, A.I. Ilin
JSC "Scientific center for anti-infection drugs", Almaty, Kazakhstan
** Corresponding author: asmanova@inbox.ru*

Key words: *one- and two-compartment models, first order absorption, direct and inverse problems in Pharmacokinetics*

Motivation and aim: Previously [1] we have shown, that for one- and two-compartment models with first order absorption, the equations of $C-t$ curves of blood have 2 and 3 solutions respectively. When $k_a = k_{21}$, the equation (2) becomes (3) and the number of solutions increases to infinity [2]:

$$C_1 = A\left(e^{-k_{10}t} - e^{-k_a t}\right) \quad (1)$$

$$C_1 = A_1 e^{-\alpha t} + A_2 e^{-\beta t} - A_3 e^{-k_a t} \quad (2)$$

$$C_1 = A_2 (e^{-\beta t} - e^{-\alpha t}) \quad (3)$$

The purpose of this work is to identify how the features of the equations (1-3) are reflected on the solution of inverse problems.

Methods: Inverse problems for models with desired properties were solved by the conjugate gradient method.

Results: It is revealed that the solution of inverse problems depend on the relations between the coefficients k_a, α and β in original system (1-3):

$$k_a \geq \alpha > \beta, 2ev_{(I)} \ (4); \ \alpha > k_a > \beta, \ 2ev_{(II)} \ (5); \ \alpha > k_a = k_{21} > \beta, \ 2ev_{(II)ka=k21} \ (6).$$

PK parameters of direct and inverse problems coincide only for the case (4), original relation (5) is transformed to (4): $2ev_{(I)} \leftrightarrow 2ev_{(II)}$ and any version of the system (6) is irreversibly converted into one-compartment model (1): $2ev_{(II)ka=k21} \rightarrow 1ev_{(1)}$.

The parameters of the inverse problems differ from their actual values an underestimation of peripheral compartment up to its full vanishing (case 6, a well-known phenomenon of the vanishing exponential) and the overestimation of the rate constant of absorption and volume of distribution of drugs. These data are consistent with the results of works [1,2] and experimental data of PK literature.

Conclusion: Discrepancies between the solutions of direct and inverse problems can lead to systematic errors in the evaluation of the properties of drugs. It is likely that the drugs, classified as type (4) in fact are characterized by the relation (5), and one-compartment models are one of the variants of their two-compartment counterparts. Selecting one of the solutions for system (2), or their certain set in the case of (3) can be made only with additional independent information, such as, for example, iv bolus data.

References:

1. Asmanova N., Koloskov G., Ilin A.I. (2013) Coupled solutions of one- and two- compartment pharmacokinetic models with first-order absorption J. Pharmacokinet. Pharmacodyn. 40:229-241
2. Asmanova N., Koloskov G., Ilin A.I. On Issues of Nonuniqueness of Solution to Some Problems in Pharmacokinetics and Appraisal of Excretion dates Capabilities in their Resolving. Abstracts of sectional reports MODERN TECHNOLOGIES FOR DEVELOPMENT OF ANTI-INFECTIOUS DRUG Almaty, 2014, 17-19 September, P.133

FLIP-FLOP PHENOMENON IN TWO-COMPARTMENT MODELS WITH FIRST ORDER ABSORPTION

N. Asmanova*, A.I. Ilin
JSC "Scientific center for anti-infection drugs, Almaty, Kazakhstan
** Corresponding author: asmanova@inbox.ru*

Key words: *two-compartment models, flip-flop phenomenon, inversion line, infinity number of solutions, drug concentration in peripheral compartment*

Motivation and aim: The reduction of PK model dimension at the transition from intra- (iv) to extravascular (ev) drug administration - is a well known fact in practice of the solution of inverse problems. As one of its reasons mistakenly believed the condition: $k_a = k_{21}$. Actually, PK equation (1) for such system combines into itself the properties $1ev$ and $2ev$ models simultaneously. The number of solutions for $1ev$ model remains equal to 2 and in $2ev$ case, as shown from schematic representation (2), increases to infinity [2]:

$C_1 = A_2(e^{-\beta t} - e^{-\alpha t})$ (1) $1ev(I)[\,2ev(II)_1, 2ev(II)_2, \ldots, 2ev(II)_j \ldots.]1ev(II)$ (2)

The aim of this work was to study the patterns of individual solutions $2ev(II)_j$ on the basis of the numerical simulation.

Results: It has been revealed that the totality of solutions for the two-compartment model in the bracketed expression (2), consists of an infinite number of twin solutions differing from one another by the usual ($k_a > k_{10}$) and flip-flop ($k_a < k_{10}$) combinations of the absorption (k_a) and elimination (k_{10}) rate constants:

$$2ev(II)_{(\mathrm{ka=k21})\mathrm{j(I)}} \leftrightarrow 2ev(II)_{(ka=k21)\,j(II)} \quad (3)$$

These solutions coincide on inversion line, where the magnitude of rate constant k_{12} reaches its maximum whose value can be found as follows:

$$k_{12(\max)} = \alpha + \beta - 2\sqrt{\alpha\beta} \quad (4)$$

The drug concentration in compartment 2, as proposed in [2] and also in a number of works, is calculated relative to the volume of central compartment $-V_1$, because it allows comparing the amount of drug in blood and tissues using the same scale. Generally accepted volume of peripheral compartment V_2 such possibility does not give.

Conclusion: An important conclusion in practical terms: due to loss of the initial value k_a in equation (1) inverse problem in this case becomes almost unsolvable [3]. Using $k_{12(\max)}$ one can evaluate the possible contribution of the second compartment and choose one of the groups of solution (3).

References:

1. N. Asmanova, G. Koloskov, A.I. Ilin. (2014) On Issues of Nonuniquenes of Solution to Some Problems in Pharmacokinetics and Appraisal of Excretion dates Capabilities in their Resolving //Abstracts of sectional reports MODERN TECHNOLOGIES FOR DEVELOPMENT OF ANTI-INFECTIOUS DRUG Almaty, 17-19 September, P.133
2. J. Loo, S.A. Riegelman. (1968) New Method for Calculation the Intrinsic Absorption Rate of Drugs, J. Pharm. Sci., 57: 918-928

THE PATTERNS OF CORRELATIONS BETWEEN HALF-LIFE - $t_{1/2}$ AND MEAN RESIDENCE TIME-MRT FOR PK MODELS OF DIFFERENT DIMENSIONS AND MODE OF DRUG ADMINISTRATION

N. Asmanova*, A.I. Ilin
JSC "Scientific center for anti-infection drugs", Almaty, Kazakhstan
** Corresponding author: asmanova@inbox.ru*

Key words: *half-life, mean residence time, one- and two-compartment models, intra- and extravascular drug administration,* $t_{1/2} - MRT$ *correlations*

Motivation and aim: As per physical sense, a half-life - $t_{1/2}$ cannot exceed MRT - mean residence time and their ratio - $t_{1/2} / MRT$ should be equal to ln2. In practice, this rule may be violated and even there are cases, when $t_{1/2} > MRT$. Unlike [1], the purpose of the study is the patterns of change $t_{1/2} / MRT$ values for PK models of different dimensions and take into account the problem of nonuniqueness of solutions of $C - t$ curve equations for them [2].

Methods: Revealed patterns of deviation $t_{1/2} / MRT$ from ln2 as well as conditions of their realization are verified by analysis of mathematical tool, numerical simulation and experimental data from literature.

Results: It is shown that deviations from the rule: $t_{1/2} / MRT = \ln 2$, which are true only for one-compartment model with intravascular administration ($1iv$), are logically related with the number of compartments in a PK model and mode of drug administration. Thus, the deviations $t_{1/2} / MRT$ from ln2 for one-compartment model with extravascular drug administration ($1ev$) are negative only. In two-compartment models with iv administration ($2iv$) they are always positive, but even, it is possible to find cases, when $t_{1/2} > MRT$. The nature of deviation in two-compartment models with ev drug administration ($2ev$) depends on correlation between elimination, redistribution and absorption contribution. Therefore the deviations $t_{1/2} / MRT$ from ln2 can be not only positive or negative, but also close to zero.

Conclusion: According to the results of this study, the deviation of $t_{1/2} / MRT$ values from ln2 is a reflection of unrecorded processes of absorption and redistribution and their interrelations with the various contributions of the drug elimination. MRT and $t_{1/2(\beta)}$ are not unambiguous measure function of eliminating drugs, as is commonly believed.They can be taken as the peculiar characteristics the departure of properties of more complex models from the simplest case -$1iv$ model. Individually they do not demonstrate this property in an explicit form, but their relation $t_{1/2(\beta)} / MRT$ in this regard is more informatively and can be recommended as an additional features of the drug.

References:

1. E. Sobol, M. Bialer. (2004) The relationships between half-life $(t_{1/2})$ and mean residence time (MRT) in two two-compartment open body model. Biofarm. Drug Dispos., 25: 157-162.
2. N.Asmanova, G. Koloskov, A. I.Ilin, (2013) Coupled solutions of one- and two- compartment pharmacokinetic models with first-order absorption J. Pharmacokinet. Pharmacodyn. 40:229-241.

A MODEL OF 3 CELLS INTERACTION IN *D. MELANOGASTER*'S PRONEURAL CLUSTER OF THE WING IMAGINAL DISK

N.B. Ayupova[1,2], V.P. Golubyatnikov[1,2]*
[1]*Sobolev Institute of Mathematics SB RAS, Novosibirsk, Russia*
[2]*Novosibirsk State University, Novosibirsk, Russia*
** Corresponding author: glbtn@math.nsc.ru*

Key words: *gene network, mathematical model, equilibrium points, stability*

Motivation and Aim: We describe phase portrait of 9-dimensional dynamical system which models early stage of morphogenesis of *D.Melanogaster*, namely, appearance of its parental cell in the proneural cluster containing 3 cells K_1, K_2, K_3. We give classification of equilibrium points of this system and show that 3 of them are stable.

Methods and Algorithms: Our description of phase portrait of this gene network model is based on geometrical methods elaborated in [1], [2].

Results: Dynamical system

$$\frac{dx_m}{dt} = f(z_m) - x_m; \frac{dy_m}{dt} = \sigma(x_m) - y_m; \frac{dz_m}{dt} = \varsigma(y_j) + \varsigma(y_k) - z_m; \quad (1)\ j,k,m=1,2,3,$$

describes interaction of 3 cells in the proneural cluster. Indices $j\neq m\neq k\neq j$ correspond to different cells. Positive variables x_j, y_j, z_j denote concentrations of the proteins AS-C, Delta and Notch, respectively, in the cell K_j. Monotonically decreasing function f is positive and describes negative feedback $(\text{Notch})_m \cdots \blacktriangleleft (\text{AS-C})_m$ in K_m; functions σ and ς describe positive feedbacks in the cells $(\text{AS-C})_m \rightarrow (\text{Delta})_m$ and between them $(\text{Delta})_k \rightarrow (\text{Notch})_m \leftarrow (\text{Delta})_j$, respectively, they grow monotonically. We assume that the cells are identical at the early stage of development, thus, for all these 3 cells the equations of 9-dimensional dynamical system (1) for each protein are identical,

$$f(z) = \frac{A}{a+z^2};\ \sigma(x) = \frac{Bx^2}{b+x^2};\ \varsigma(y) = \frac{Cy}{c+y},$$

see [2], where we studied two cells interaction.

We find equilibrium points of the system (1) from the system of 9 algebraic equations $x_m = f(z_m);\ \ z_m = \varsigma(y_k) + \varsigma(y_j)$. For a wide domain of positive parameters A, B, C, a, we show that this system has exactly 7 solutions. One of them corresponds to unstable equilibrium point $x_1=x_2=x_3>0$, $y_1=y_2=y_3>0$, $z_1=z_2=z_3>0$. We show that 3 partially symmetric solutions $0< z_k = z_j < z_m$ etc. also correspond to 3 unstable equilibrium points. The rest partially symmetric solutions $z_k = z_j > z_m > 0$ etc. describe 3 stable equilibrium points; for each of them, corresponding K_m becomes the parental cell in the proneural cluster.

Conclusion: We describe geometry of phase portrait of the model of an early stage of *D.Melanogaster* morphogenesis. All equilibrium points are listed. Biological interpretation is given.

Acknowledgements: Supported by RFBR, grant 15-01-00745.

References:

1. V.A.Likhoshvai, V.P.Golubyatnikov et al. (2008) Theory of gene networks, In: *System computerized biology*, (N.A.Kolchanov, S.S.Goncharov), 397-480, Novosibirsk, SB RAS.
2. A.A.Akinshin, T.A.Bukharina, V.P.Golubyatnikov, D.P.Furman (2014) Mathematical modeling of two cells interaction in the proneural cluster of the wing imaginal disk of *D.Melanogaster*. Vestnik of Novosibirsk state university. **14(4):** 3-10.

PREDICTIVE MODELS OF EARLY-ONSET PREECLAMPSIA BASED ON THE BLOOD PLASMA MICRORNA EXPRESSION LEVEL

I.S. Balashov*, O.S. Altukhova, A.V. Timofeeva, V.A. Gusar, K.N. Prozorovskaya, N.E. Kan, P.I. Borovikov, M.Y. Bobrov
Research Center for Obstetrics, Gynecology and Perinatology Ministry of Healthcare of the Russian Federation, Moscow, Russia
** Corresponding author: ivanov@bionet.nsc.ru*

Key words: *early-onset preeclampsia, microRNA, predictive test, logistic regression*

Motivation and Aim: Preeclampsia, a multisystem pathological condition complicating pregnancy, is a leading cause of maternal and perinatal mortality. The pathogenesis of preeclampsia is associated with differential expression of the protein products of genes involved in the relevant signaling pathways. Epigenetic regulators of these signaling cascades are small non-coding RNAs (miRNAs).

Methods and Algorithms: Preeclampsia, a multisystem pathological condition complicating pregnancy, is a leading cause of maternal and perinatal mortality. The pathogenesis of preeclampsia is associated with differential expression of the protein products of genes involved in the relevant signaling pathways. Epigenetic regulators of these signaling cascades are small non-coding RNAs (miRNAs).

Results: Comparative evaluation of the model quality was based on the maximum likelihood function and the number of predictors. For each model, dichotomy boundary between groups was adapted in accordance with the maximum sensitivity and specificity. Two predictive models based on the analysis of the expression level of miR-423-5p, miR-181a-5p and miR-92b-3p were developed, providing equal values of sensitivity (87%) and specificity (80%), but predicting different values in 23% of cases. Linear combinations of the predicted values were proposed to use in order to compensate the classification errors.

Conclusion: The employment of linear combinations of predictive models for early-onset preeclampsia based on blood plasma miRNA expression level can increase the sensitivity of the test up to 92% and specificity up to 100%.

AN ALGORITHM FOR SELECTING OF ANTIBIOTIC RESISTANCE GENE-PREDICTORS FOR *KLEBSIELLA PNEUMONIAE* HOSPITAL STRAINS

I.S. Balashov*, V.A. Naumov, O.S. Altukhova, P.I. Borovikov, I.S. Mukosey, T.O. Kochetkova, A.B. Gordeev, D.V. Dubodelov, E.S. Shubina, L.A. Lyubasovskaya, T.V. Priputnevich

Academician V.I. Kulakov Research Center of Obstetrics, Gynecology and Perinatology, Ministry of Health, Russia

** Corresponding author: i_balashov@oparina4.ru*

Key words: *K. pneumoniae, antimicrobial resistance, nosocomial infections, NGS*

Motivation and Aim: Klebsiella pneumoniae (*K. pneumoniae*) - one of the causative agents of nosocomial infections. *K. pneumoniae* has the ability to spread rapidly in the hospitals, has the resistance to a wide range of antibiotics, as well as the possibility of prolonged asymptomatic carriage and a high incidence of infectious complications in patients with reduced immunity. Antibiotic resistance is currently based on growth inhibition assay, which is not an optimal method in case of need for early appointment of antimicrobial drugs. The aim of this study is to develop an algorithm for a predicting of antimicrobial drug resistance based on data of the strain genotype.

Methods: Klebsiella strains were analyzed by next-generating sequencing. The reads were aligned with the nucleotide sequences from the database of antibiotic resistance genes ResFinder. Selecting of the most promising gene-predictors was performed by estimating of concordance of the phenotypic resistance and the presence of the gene in samples using the Jaccard similarity measure with the normalized proportion of the covered nucleotides as a weighting factor.

Predictors were clustered according to similarity measure based on editorial distance.

Results: The clusters of antimicrobial resistance genes-predictors for *K. pneumoniae* hospital strains have been identified. The median of the accuracy of the prediction antimicrobial resistance by identifying one or more genes from identified clusters - 0.79.

Conclusion: The proposed algorithm could be used in the development of test systems for rapid diagnosis and epidemiological monitoring of antibiotic resistance.

Acknowledgements: The study is supported by agreement № 14.607.21.0019 Ministry of Education of the Russian Federation "Development of molecular-genetic test systems to evaluate the pathogenicity and resistance of pathogens nosocomial and opportunistic infections in the mother and the newborn"(2014-14-579-0001-065).

CYCLES OF DISCRETE DYNAMICAL SYSTEMS OF A CIRCULANT TYPE WITH A THRESHOLD FUNCTION IN THE VERTICES OF THE NETWORK

Ts.Ch.-D. Batueva
Sobolev Institute of Mathematics SB RAS, Novosibirsk, Russia
Corresponding author: batueva@math.nsc.ru, tsyn.batueva@gmail.com

Key words: *discrete dynamical system, circulant, gene network, functional graph, cycle*

Motivation and Aim: This study originates in biology and genetics. We consider discrete dynamical systems that can serve as a model of a regulatory circuit of a gene network (for more detail see [2, 4]). The vertices of the carrier graph correspond to various chemicals in a cell. The vertex labels indicate the concentrations of substances, and the closed paths describe periodic processes.

Methods and Algorithms: A *circulant* is the oriented graph whose adjacency matrix is a circulant matrix.

General model of the discrete dynamical system of a circulant type is introduced in the article [3]. All circulants vertices are marked by elements of Z_2^n. Set of such labels is called the *state of the system*. At each step, state is recalculated by mapping $A_{f,2}$: $Z_2^n \to Z_2^n$, each vertex receives a value of a Boolean threshold function *f* depending on values labels previous $k \leq 3$ vertices. A threshold function is defined by the hyperplane.

A functional graph $G_{f,n}$ is an oriented graph with vertices from Z_2^n such that there is an edge between *v* and *u* if and only if $A_{f,2}(v) = u$. Each component of these graphs consists of finitely many trees that are directed toward the roots, and the roots form a closed path. The number of components of the functional graph is the number of its cycles. In the [1], we have proved theorem describing some cycles of the functional graph formed by cyclic shifts.

Results: This theorem describes almost all cycles of functional graphs of the systems with Boolean threshold functions of at most three variables in the network vertices. The states of the other cycles go in two steps of the system in its circular shift.

Conclusion: Obtained result describes periodic states and makes it possible to estimate the number of components of the functional graph.

References:

1. Ts.Ch.-D. Batueva. (2016) Circulant Discrete Dynamical Systems with Threshold Functions of at Most Three Variables, *Diskretn. Anal. Issled. Oper.*, **23**, No. **1**: 17-32.
2. G.V. Demidenko, N.A. Kolchanov, V.A. Likhoshvai, Yu.G. Matushkin, S.I. Fadeev. (2004) Mathematical Modeling of Regular Contours of Gene Networks, *Zh. Vychisl.Mat. Mat. Fiz.*, **44**, No. **12**: 2276–2295.
3. A.A. Evdokimov, A.L. Perezhogin. (2011) Discrete dynamical systems of a circulant type with linear functions at the vertices of a network, *Diskretn. Anal. Issled. Oper.*, **18**, No. **3**, 39-48.
4. V.A. Likhoshvai, V.P. Golubyatnikov, G.V. Demidenko, A.A. Evdokimov, I.I. Matveeva, S.I. Fadeev. (2008) Theory of Gene Networks, In: *Computational Systems Biology*, 397–482 (Sibirsk. Otdel. Ross. Akad. Nauk, Novosibirsk).

HIGH-PERFORMANCE INTELLIGENT ANALYSIS OF BIOMECHANICAL PROCESSES CONTROL AND MANAGEMENT OF BLOOD PRESSURE IN HUMAN KIDNEY

A.A. Bedelbayev
Kazakh National University them. al-Farabi, Almaty, Kazakhstan
Corresponding author: agyn08@gmail.com

Key words: *kidneys, blood pressure, control, data mining, HPC, SSAS, MPI, models*

Motivation and goal. Study of monitoring and control mechanism of blood pressure in the human body, which involved delicate mechanisms of the nervous, endocrine system, implemented on a cellular level, it is the most important task for the understanding and protection of human health. Such a study is carried out within the computer simulation of biochemical and physical processes in the human kidney vital activity conducted under a scientific research topic of the author together with the medical staff of research centers in Almaty. The issues concern the biomechanical processes regulating the osmotic pressure occurring in the distal convoluted tubule of the nephron human kidney. The aim of this study is to identify ways to prevent disturbances in the mechanisms for monitoring and controlling the operation of the human circulatory system, which involved many of his organs, significant of which is a complete work of human kidneys.

Methods and algorithms. Computer modeling of the studied processes carried out on Microsoft Windows HPC Server 2012 R2 platform using WPF4 technology by the C# language. The studied processes are divided in their place of origin, which are located in different parts of the human, but the main ones being the human kidney nephrons, whose number in each kidney reaches one million pieces. The number of such processes is extremely large, and they are different in their physical and biochemical characteristics. For efficient modeling of these processes used cluster technology and MPI technology with the accumulation of data in a distributed database system with SQL Server 2012 R2. The main center of information collection and analytical analysis is performed on the base cluster by the SSAS 2012 with helps of MDX language.

Results. Our investigations are in the initial path and mostly are illustrative and debug. Building a computer system architecture and the use of traditional technology of intellectual analysis of the economic information for the study of the biochemical processes in human organs and their subsequent use to monitor and focused management of these processes is the main result of our work at this stage. Subsequently, our research will be related to verification of the adequacy of the resulting model with real performance and results of the impact of these processes on the blood pressure of a person and the possibility of targeting him.

Conclusion and availability. The aim of the report is to present the problem that arises in a natural way and is extremely important for human health.

THE ORGANIZATION OF HIGH-PERFORMANCE COMPUTING PROCESSES MONITORING AND CONTROL OF BLOOD SALINE COMPOSITION IN THE HUMAN KIDNEY NEPHRONS

A.A. Bedelbayev
Kazakh National University them. al-Farabi, Almaty, Kazakhstan
Corresponding author: agyn08@gmail.com

Key words: *nephron, kidney, saline, blood, HPC, MPI, model, clusters*

Motivation and Aim: Complex biochemical processes salt composition of human blood monitoring and management occurring in the distal convoluted tubules of the kidney nephron that involve many organs of the endocrine system of humans, are essential to his health and fulfilling his life. The purpose of research is to build an adequate model of human kidney functioning under natural conditions for the development of effective methods of influence on the place in which biochemical processes.

Methods and Algorithms: Modelling of biochemical processes monitoring and control salt composition of human blood by aldosterone, which is the process of developing a chain of interdependent processes in different parts of the human body, is carried on the platform of Microsoft HPC Server 2012 R2 by using the MPI technology. Visualization of the processes carried out on the basis of technology WPF 4 by C # language. In the process of monitoring and decision-making on the blood cleaning from various salts involves many hormones that are produced by various glands of the endocrine system, and ensure the immutability of the salt composition of the blood. The process of regulating the blood is carried out in the human kidney nephrons, the number of which is extremely large, and each of them is controlled independently and collectively organize such treatment. Therefore, for the organization of all processes involved emerging cluster technology based on MPI. Collecting and storing the received information is carried out on a distributed server Databases SQL Server 2012 R2. This database allows us to build different scenarios for the development of the processes and to explore the consequences of such a course of events.

Results: A cluster computing system based on MPI technology with a distributed database on SQL Server 2012 R2 on the Microsoft HPC Server 2012 R2 platform. Established structural and logical connections to the main system modules that mimic the functioning of the major organs, involved in the test process of blood purification from excess salt. Features a computer system for modeling and monitoring of working process visualization and cleaning the blood from unwanted salts give a way to research into the causes of various diseases of the abundance of unwanted salts.

Conclusion and availability: Research on modeling of human kidney functioning in conditions of natural its activity at an early stage. The results are more experimental than the final.

ALGORITMS COMPARISON OF INVERSE PROBLEM SOLUTION FOR PHARMACOKINETIC MODELS

A.Yu. Belonog[1]*, D.A. Voronov[1,2]
[1]*Novosibirsk State University, Novosibirsk, Russia*
[2]*Institute of Computational Mathematics and Mathematical Geophysics SB RAS, Novosibirsk, Russia*
** Corresponding author: anatolybelonog@gmail.com*

Key words: *pharmacokinetics, inverse problem, mathematical modeling, numerical solution, optimization algorithms*

Motivation and Aim: The aim of work is researching and developing new algorithms for pharmacokinetic models parameters estimation

Methods and Algorithms: This problem reduces to an optimization problem of data fitting which is solved by minimization of selected functional. Minimization is done by optimization methods such as direct, gradient or stochastic methods.

Consider a direct problem:

$$\begin{cases} \dot{x} = F(q, x, t) \\ x(0) = x_0, \end{cases}$$

where x is a model functions vector, q is a model parameters vector. Inverse problem consist in estimation of parameters q using additional data of direct problem solution.

Methods of solving inverse problem differ at this stage. It can be reduced to the optimization form where problem is solved by fitting the additional data through minimizing some selected functional that depends on parameters and data. To do so optimization algorithms are used such as gradient, direct or stochastic methods.

Another approach is to reduce problem to the following form:

$$Aq = f,$$

where A is an operator that image parameters vector q to additional data vector f. That goal can be reached through linearization and discretization of model.

Results: Either examples of pharmacokinetic models with synthetic and real data is given. The results of numerical solution using different approaches are presented including inverse problem with noisy data. Approaches comparison is showed. A matter of choice of measurement time points is considered.

Conclusion: The domains of applicability of each proposed algorithm were explored.

Acknowledgements: The reported study was supported by Russian Foundation for Basic Research, research project No. 16-31-00382.

References:

1. Ilyin A.I., Kabanikhin S.I., Voronov D.A., Krivorotko O.I., Belonog A.Yu. (2015) Numerical solution of inverse problem for secretion and kinetics of c-peptide model, *The proceedings of the 2015 International Conference on Biomedical Engineering and Computational Technologies (SIBIRCON).*
2. R. Eaton, R. Allen, D. Schade, K. Erickson, J. Standefer "Prehepatic Insulin Production in Man: Kinetic Analysis Using Peripheral Connecting Peptide Behavior", *Journal of Clinical Endocrinology and Metabolism, Vol.51*
3. N. Yamamura, Y. Magata, F. Yamashita, M. Hashida, H. Saji (1999) Pharmacokinetic analysis of ^{123}I -labeled medium chain fatty acid as a radiopharmaceutical for hepatic function based on beta-oxidation, *Annals of Nuclear Medicine, Vol. 13, No.4.*

SYMMETRICAL GENETIC CODE AND GENETIC MUTATIONS

B.O. Biletskyy*, A.M. Gupal
V.M. Glushkov Institute of Cybernetics NAS Ukraine
** Corresponding author: borys.biletskyy@gmail.com*

Key words: *genetic codes, genetic mutations, mutation-associated diseases*

Motivation and Aim: Genetic code maps a triplet of nucleotides to one of 20 amino acids or to one of 3 STOP commands terminating the synthase of a protein. In current research, we study noise immunity of arbitrary genetic codes with respect to polarity of resulting amino acids.

Methods and Algorithms: We consider a noise immunity coefficient (NIC) of an arbitrary genetic code, that is a ratio of a single nucleotide polymorphisms (SNP) of a triplet preserving the polarity of a resulting amino acid residue under given genetic code to all possible SNP's of a triplet. Genetic algorithms with NIC as a scoring function were used to construct various highly mutation immune genetic codes. Pairwise permutations were used as mutation operator to modify genetic codes preserving important qualities of the standard genetic code. Mutation immunity properties we also compared on a disease-associated mutations in human genome. Genetic mutations were obtained from the NCBI server.

Results: An artificial symmetrical genetic code with NIC equals 77.86% has been constructed using pairwise permutations preserving the properties of the standard genetic code. NIC of the new symmetrical genetic code has been compared to NIC coefficient using 300 disease associated SNP' in the human genome. In 80% of cases, when the mutation in condone led to a mutation of a resulting amino acid residue polarity, the symmetrical genetic code was able to preserve the original amino acid polarity.

Conclusion: Mutation immunity characteristics of the standard genetic code are not optimal, since there are genetic codes with higher mutations immunity. Computations revealed that both standard and constructed symmetric genetic codes have mutation immunity much higher than arbitrary genetic codes.

Availability: The source code of the R script used for extracting SNP-related data from NCBI server is available at https://github.com/bbiletskyy/snp-retrieval.

References:

1. И. В. Сергиенко, А.М. Гупал.(2013) Помехоустойчивость генетических кодов. Оптимальные коды, Доклады Академии Наук, 5:492-495.
2. И. В. Сергиенко, А.М. Гупал, А.В. Островский.(2014) Устойчивость генетического кода к точечным мутациям, Кибернетика и системный анализ, 5:17-24
3. Б.А. Белецкий, А.М. Гупал, Н.А. Гупал.(2015) Симметричный код и генетические мутации, Компьютерная математика, 2:63-69

HOW NEW SCIENCE EMERGES: A CASE STUDY OF MICRORNA RESEARCH

A.A. Blinov[1], A.B. Firsov[1], S.I. Demurin[1], M.V. Pankova[1], I.I. Titov[1,2]*
[1]*Novosibirsk State University, Novosibirsk, Russia*
[2]*Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
** Corresponding author: titov@bionet.nsc.ru*

Key words: *co-authorship network, miRNA, SVM, SIR model*

Motivation and Aim: Scientific and technological advances are critical for modern society development. The recent accumulation of digital information made it possible to reveal how new research areas emerge and grow.

Results: From the PubMed database we have extracted almost 50 000 papers that contain the word miRNA and its synonyms. Then, using SVM we have identified the unique institutions and authors relating to these papers and have built two time-evolving collaboration networks: the co-authorship network and the network of institution interactions. Next we have analyzed these networks growth in terms of information spreading and have fitted their evolution to the epidemiological SIR model.

CENSORING OF NOISY OBJECTS AND ATTRIBUTES WITH FUNCTION OF RIVAL SIMILARITY IN MEDICAL AND BIOLOGICAL TASKS

I.A. Borisova, O.A. Kutnenko*
Sobolev Institute of Mathematics SB RAS, Novosibirsk, Russia
** Corresponding author: olga@math.nsc.ru*

Key words: *Censoring, outliers detection, FRiS-function, classes separability function*

Censoring or outliers detection task is one of the important problems in Data Mining, particularly in case of biomedical datasets analysis. Nonrelevant attributes, misclassified objects, nonrelevant week correspondences can cause major distortion in collected data, mask real dependences and make biomedical data processing a non-trivial task.

In the presented work the most complex case was considered, when there are noisy attributes and noisy objects in a dataset at the same time. We distinguish between two types of noise objects sources. In the first case noise objects in a training dataset appear because of some errors or failures during expert classification process. In the second case noise objects are originated by some unaccounted weak correspondences which inclusion in general model of a dataset results in classification accuracy increasing. Attributes which irrelevant to the solving task and uncorrelated with the class attribute are considered as noisy ones.

To solve censoring problem we offered two major approaches. First one restores the structure of the dataset by selecting the set of the most typical objects of the dataset, called stolps. After that noisy objects are detected among objects, which are dissimilar to any stolp of their class. Second approach uses local characteristics of objects neighborhood and detect noisy objects based on their dissimilarity to the closest objects of their classes. Both approaches uses function of rival similarity (FRiS-function) [1] to estimate similarity of an object z to class *A* in competition with rival class *B*. For noisy attributes filtering we propose some special classes separability function, based on FRiS-function. It is calculated after noisy objects filtering to estimate relevance of this or that subset of attributes to the solving task.

Both approaches were tested on a wide range of model tasks and on a few real biomedical problems. During this testing we estimate improving of classification accuracy after censoring procedure. Also sensitivity and specificity of the approaches for noise detection were calculated for model datasets with given structure of artificially added noise. The presented algorithms were compared with the Edited Nearest Neighbor (ENN) [2] Rule and the Repeated Edited Nearest Neighbor Rule (RENN) [2], which are well known approaches to outliers detection in case of classes with complex structure. Our experiments have shown high efficiency of censoring procedure both objects and attributes, based on function of rival similarity.

Given work is executed at financial support of the Russian Fund of Basic Researches, the Grant №14-01-00039.

References:

1. Zagoruiko N.G., Borisova I.A., Dyubanov V.V., Kutnenko O.A. (2011) A quantitative measure of compactness and similarity in a competitive space //Journal of Applied and Industrial Mathematics. V. 5, №1. P. 144–154.
2. Wilson, D. R., & Martinez, T. R. (2000). Reduction techniques for instance-based learning algorithms. Machine learning, 38(3), 257–286.

STOCHASTIC PATTERN FORMATION INDUCED BY CELL-TO-CELL COMMUNICATIONS IN ELASTIC EPITHELIAL TISSUE

D.A. Bratsun[1,2]*, I.V. Krasnyakov[2]
[1]*Perm National Research Polytechnic University, Perm, Russia*
[2]*Perm State Humanitarian Pedagogical University, Perm, Russia*
** Corresponding author: bratsun@pspu.ru, krasnyakov_ivan@mail.ru*

Key words: *signaling, stochastic modeling, gene regulation, time-delay, systems biology*

Motivation and Aim: It is known that the noise during gene expression comes about in two ways. The inherent stochasticity of biochemical processes such as transcription and translation generates "intrinsic" noise. "Extrinsic" noise refers to variation in identically-regulated quantities between different cells. The small number of reactant molecules involved in gene regulation can lead to significant fluctuations in intracellular mRNA and protein concentrations, and there have been numerous recent studies devoted to the consequences of such noise at the regulatory level.

Methods and Algorithms: To study the spatial effects of intrinsic and extrinsic noises on gene regulation within the cellular system we have applied a multiscale chemo-mechanical model proposed recently in [1] for tumor development. The epithelium is represented by an elastic 2D array of polygonal cells with its own gene regulation dynamics. The model allows for the simulation of evolution of multiple cells interacting via the chemical signaling or mechanically induced strain. The algorithm includes also the division and intercalation of cells, as well as the transformation of normal cells into a cancerous state triggered by a local failure of spatial synchronization of the cellular rhythms driven by transcription/translation processes. To model the delay-induced stochastic chemical signaling we have used a generalization of the Gillespie algorithm that accounts for delay suggested in [2]. The possibility of the stochastic pattern formation produced by time delay and noise was demonstrated for simple regulatory scheme in [3].

Results and Conclusion: In this work, we study the pattern formation excited by stochastic dynamics in cells and cell-to-cell communications. The influence of stochasticity on the emergence of tumor is discussed. Both the intrinsic and extrinsic contributions to stochastic pattern formation have been explored.

Availability: Software is not yet publicly available.

Acknowledgements: The research has been supported by the Perm Ministry of Education and Science (MIG's project C-26/004.4) and RFBR's grant #14-01-96022r_ural_a.

References:

1. D.A.Bratsun, D.V.Merkuriev, A.P.Zakharov, L.M.Pismen. (2016) Multiscale modeling of tumor growth induced by circadian rhythm disruption in epithelial tissue, *J. Biol. Phys.*, **42**: 107-132.
2. D.Bratsun, D.Volfson, J.Hasty, L.S.Tsimring. (2005) Delay-induced stochastic oscillations in gene regulation, *Proc. Natl. Acad. Sci. U.S.A.*, **102**: 14593-14598.
3. D.A.Bratsun, A.P.Zakharov. (2016) Spatial Effects of Delay-Induced Stochastic Oscillations in a Multi-scale Cellular System, In: *Springer Proceedings in Complexity*, S. Battiston et al. (Eds.), 93-103 (Springer).

UGENE: A TOOLKIT FOR TEACHING STUDENTS

I.V. Bykova[1]*, O.I. Golosova[1], A.Y. Bakulina[2,3], D.A. Afonnikov[2,4], D.Y. Kandrov[1], A.Y. Palyanov[2,5], G.A. Grekhov[1], Y.E. Danilova[1]

[1]Unipro Center of Information Technologies, Novosibirsk, Russia
[2]Novosibirsk State University, Novosibirsk, Russia
[3]State Research Center of Virology and Biotechnology VECTOR, Koltsovo, Novosibirsk region, Russia
[4]Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
[5]Institute of Informatics Systems SB RAS, Novosibirsk, Russia
** Corresponding author: ibykova@unipro.ru*

Motivation and Aim: Modern biological experiments in many cases require bioinformatics methods for planning and subsequent data analysis with application of complex computing algorithms.

Though there are serious commercial bioinformatics packages for these purposes, they sometimes are not available for the students. On the other hand free useful program tools and algorithms are uncoordinated in many cases and young biologists have to be experienced in programming to work with them successfully. Thus, young scientists encounter big obstacles in mastering bioinformatics hands-on.

Methods and Algorithms: UGENE is developed as an open-source free software aimed to assist a molecular biologist. It comprises a lot of analysis tools, including both the experiments design and data processing without any programming skills for a user.

UGENE provides an easy way to work with DNA, RNA and protein sequences. The functionality list is very wide: sequence annotation with access to remote databases, multiple alignment and phylogenetic trees, 3D protein structures, processing of Sanger and NGS sequencing data (genome assembling and variations, processing of RNA-Seq and ChIP-Seq data), etc. UGENE can be run on MS Windows, Linux and Mac OS X platforms.

Due to its accessibility and wide functionality UGENE can be used as an excellent tool to teach different biological methods.

Results and Conclusion: Currently UGENE is used in tutorials of several universities, for example, Roskilde University (Denmark). "Practical Bioinformatics", interdisciplinary course utilizing UGENE, starts in September 2016 at NSU. Growth of educational programs using UGENE promotes its popularization within the international community of molecular biologists.

Availability: http://ugene.net/download.html

SOFTWARE MODULE FOR INTEGRATION OF SBML-WRITTEN MATHEMATICAL MODELS OF MOLECULAR GENETIC SYSTEMS FOR THE HAPLOID EVOLUTIONARY CONSTRUCTOR 3D SOFTWARE PACKAGE

A.D. Chekantsev*, S.A. Lashin
Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
Novosibirsk State University, Novosibirsk, Russia
** Corresponding author: chekantsev@bionet.nsc.ru*

Motivation and Aim: Microbes forming large communities in nature regularly exchange genes via horizontal transfer. It gives microbial cells the ability to acquire novel metabolic functions [1] and consequently may lead to ecological changes in community as a whole. Nowadays, there are a lot of resources for warehousing mathematical models of metabolic [2-4] and gene regulatory [4] systems. It is a great challenge to integrate these data in a complex hierarchical model of evolving microbial community. The Haploid Evolutionary Constructor 3D (HEC 3D) framework allows constructing and simulating such communities consisting of cells of various strains/species/populations living in spatially heterogeneous habitats. Cells consume, utilize, synthesize and secrete metabolites according to genetic programs written modeled as gene networks [5]. Some of cells consume one metabolites and synthesize another, which respectively may be consumed by third cells, i.e. they form trophic cycles of exchanging metabolites.

The aim of this study is development and implementation of a software module for HEC 3D framework in order to import mathematical models of molecular-genetic systems from the existing databases to the HEC.

Methods and Algorithms: We used models written in SBML format. The integration is provided via libSBML and SOSlib libraries. To resolve issues with different synonyms of the same metabolites, we used the REST API for ChEBI database.

Results and discussion: The module designed allows us to extract the parameters and formulas reactions from the model loaded from the repositories such as BioModels or SABIO-RK and to replace existing HEC 3D generalized synthesis strategies with real SBML models. ChEBI database of chemical names was integrated which solved the problem of metabolites wrong usage. Thus, the novel module allows users to use the "real-world" models in the HEC 3D and to investigate the behavior and the evolution of complex microbial communities.

Acknowledgement: The study has been partially supported by the RFBR grant №150703879 and Budget Project 0324-2015-0003.

References:

1. T.Y. Pang, S.S. Maslov (2011) A toolbox model of evolution of metabolic pathways on networks of arbitrary topology, *PLOS Comp. Biol.*, **7**: e1001137.
2. U. Witting et al. (2012) SABIO-RK – database for biochemical reaction kinetics, *Nucleic Acid Res.*, **40**: D790–D796.
3. S.A. Coskun et al. (2013) An online model composition tool for system biology models, *BMC Syst. Biol.*, **7**: 88.
4. F.V. Kazantsev et al. (2016) MAMMOTh: a new database for curated MAthematical Models of bioMOlecular sysTems, *bioRxiv*, 053652.
5. S.A. Lashin et al. (2014) HEC 2.0: improved simulation of the evolution of prokaryotic communities, *In Silico. Biol.* **11**: 125-135.

MODELING AND OPTIMIZATION THE PROCESS OF EMBOLIZATION ARTERIOVENOUS MALFORMATION ON THE BAZIS OF TWO-PHASE FILTRATION MODEL

A.A. Cherevko[1, 2], T.S. Gologush*[2], V.V. Ostapenko[1, 2], I.A. Petrenko[3], A.P. Chupakhin[1, 2]

[1]*Lavrentyev Institute of Hydrodynamics SB RAS, Novosibirsk, Russia*
[2]*Novosibirsk State University, Novosibirsk, Russia*
[3]*Vladimir State University, Vladimir, Russia*
** Corresponding author: tatiana_06.08@mail.ru*

Key words: *arteriovenous malformation (AVM), embolization, two-phase filtration, Buckley-Leverett equation, CABARET scheme*

Motivation and Aim: Cerebral AVM is a difficult, dangerous, and most frequently encountered vascular failure of development. Endovascular embolization of AVM is effective treatment of such pathologies. However, the danger of intraoperative rupture during embolization still exists. The purpose is to model this process and build an optimization algorithm for AVM embolization.

Methods and Algorithms: The process of embolization was described by Buckley-Leverett equation, which was solved numerically by using a new modification of CABARET scheme.

Results: To study the different embolization variants, the initial-boundary value problems, describing the process of embolization, were solved numerically by using a new modification of CABARET scheme. The essential moments of embolization process were modeled in our numerical experiments.

Conclusion: This approach well reproduces the essential features of discontinuous two-phase flows, arising in the embolization problems. It can be used for further study on the process of AVM embolization.

References:

1. A. A. Cherevko, T. S. Gologush, V. V. Ostapenko, I. A. Petrenko, A. P. Chupakhin (2016) Modeling process of embolization arteriovenous malformation on the basis of two-phase filtration model, JCPS, in press.

NYQUIST DIAGRAMS FOR THE GENERALIZED VAN DER POL – DUFFING EQUATION DESCRIBING LOCAL CEREBRAL HEMODYNAMIC

A.A. Cherevko[1, 2], E.E. Bord[2]*, A.K. Khe[1, 2], V.A. Panarin[3], K.Yu. Orlov[3], A.P. Chupakhin[1,2]

[1]*Larentyev Institute of Hydrodynamics SB RAS, Novosibirsk, Russia*
[2]*Novosibirsk State University, Novosibirsk, Russia*
[3]*Academician E.N. Meshalkin Research Institute of Circulation Pathology, Novosibirsk, Russia*
** Corresponding author: bord.elizaveta@yandex.ru*

Key words: *Nyquist diagram, generalized Van der Pol – Duffing equation, aneurysm, hemodynamic, solution bifurcation*

Motivation and Aim: Pathologies of cerebrovascular system, such as aneurysms [1] and AVM [2] lead to violation of blood circulation and cause a threat to the patient. From the medical point of view, it is important to know the vascular response to the changes in blood flow. The aim of this study is to obtain information about typical behaviour of blood flow parameters in the vicinity of pathologies.

Methods and Algorithms: The blood velocity and pressure are measured during the operations [3]. The model of generalized Van der Pol - Duffing equation (VPD) is used to identify the characteristic behaviour of hemodynamic parameters near pathologies [4]. We use non-linear generalization of the Nyquist diagrams (ND) to describe the solution behaviour [5].

Results: To predict the vascular response to changes in bloodstream we analyse the solutions of VPD built for different patients. We replace the right part with a harmonic function and then build ND to analyse the solution. As a result we have analysed about 300 ND.

Conclusion: There is study of VPD built on the clinical data of concrete patients in this paper. It is shown that ND are divided into several classes. The diagrams move from one class to another in the definite order with the increasing of the right side amplitude. The dependence of the characteristic solution behavior on the form of ND is found.

Acknowledgements: The study was completed thanks to the support of Russian Foundation for Basic Research (project № 14-01-00036). Also, study was supported by grant SS-8146.2016.1.

References:

1. A. B. Klimov, V. E. Ryabuhin, K. Y. Orlov and V. V. Krilov (2012) Endovascular treatment of cerebral arteriovenous malformations. *Surgery of the Cerebral Aneurysms*, **3:** 231-268.
2. V. V. Krilov (ed) (2011) *Surgery of the cerebral aneurysms*, **1** (Moscow: T. A. Alexeeva).
3. A. P. Chupakhin, A. A. Cherevko, A. K. Khe, N. J. Telegina, A. L. KrivoshapkinAL, K. J. Orlov, V. A. Panarin, V. I. Baranov (2012) Measurement and analysis of cerebral hemodynamic parameters in the presence of brain vascular anomalies. *Circulation Pathology and Cardiac Surgery*, **16**, **44:** 27-31.
4. A. K. Khe, A. A. Cherevko, A. P. Chupakhin, A. L. Krivoshapkin, K. J. Orlov (2016) Endovascular blood flow measurement system. *Journal of Physics: Conference Series.* 5. H. Nyquist (1932). Regeneration theory. *Bell System Technical Journal*, **11:** 126 – 147.

VAN DER POL – DUFFING'S EQUATION AS A RELAXATION OSCILLATION MODEL OF HEMODYNAMIC PARAMETERS IN DIFFERENT CEREBRAL VESSELS

A.A. Cherevko[1,2], I.V. Ufimtseva[1]*, A.P. Chupakhin[1,2], A.L. Krivoshapkin[3], K.Yu. Orlov[3]

[1]*Novosibirsk State University, Novosibirsk, Russia*
[2]*Lavrentyev Institute of Hydrodynamics SB RAS, Novosibirsk, Russia*
[3]*Academician E.N. Meshalkin Research Institute of Circulation Pathology, Novosibirsk, Russia*
** Corresponding author: ufimtsevaiv@gmail.com*

Key words: *Van der Pol – Duffing equation, oscillator, blood flow, cerebral vessels, pressure, velocity*

Motivation and Aim: Simulation of a blood flow under normality as well as under pathology is extremely complex problem of great current interest both from the point of view of fundamental hydrodynamics, and for medical applications. This paper proposes a model of Van der Pol – Duffing nonlinear oscillator equation describing relaxation oscillations of a blood flow in the cerebral vessels.

Methods and Algorithms: The model is based on the patient-specific clinical experimental data flow obtained during the neurosurgical operations in Meshalkin Novosibirsk Research Institute of Circulation Pathology. Experimental clinical data due to measuring system Combo Map device specifics describes the blood rate and pressure connection.

Results: The simplest model for the description of these processes is the equation of a nonlinear oscillator of Van der Pol - Duffing

$$\varepsilon q'' + P_2(q)q' + Q_3(q) = ku(t),$$

where one of the values – pressure $q = q(t)$ or velocity $u = u(t)$ – acts as the control function, i.e, a right-hand side of the equation. It is important that equation has a stable periodic solution. At the same time, the equation of this form has a large and complex set of solutions, the structure of which is determined by the values of its coefficients.

Conclusion: The stability of the model is demonstrated through the variations of initial data and coefficients. It is universal and describes pressure and velocity fluctuations in different cerebral vessels (arteries, veins, sinuses), as well as in a laboratory model of carotid bifurcation. The model was tested on data sets of 50 neurosurgical operations and laboratory experiments with elastic tee. This model allows to describe the characteristics of blood flow in the presence of anomalies.

Acknowledgements: The theoretical research of this paper was supported by the Russian Foundation for Basic Research (Project No. 14-01-00036). The experiments with using Combo Flow were supported by the Russian Science Foundation (Project No. 14-35-00020).

ALGORITHMS AND TOOLS DEVELOPED BY NOVEL COMPUTING SYSTEMS IN BIOLOGY LLC

E. Cheryomushkin*, S. Nikitin, T. Valeev, T. Konovalova, A. Ryabova, K. Golosov, I. Mikerova, N. Gorokhov, D. Babiy
Novel Computing Systems in Biology LLC, Novosibirsk, Russia
** Corresponding author: evgeny.cheryomushkin@gmail.com*

Key words: *Transcription Factor Binding Sites prediction algorithms*

Motivation and Aim: Since 2000 we're working with Biobase GmbH (two years ago it was acquired by Qiagen). We developed many algorithms and tools since that time. All of them cover transcription regulation analysis using Biobase databases, mostly Transfac. The tools are supported by graphical interfaces and integrated into Biobase products.

Methods and Algorithms: Philogenetic footprinting[1] is algorithm to reveal TFBS (Transcription Factor Binding Sites) filtering evolutionary conserved sites. Idea of our method was to realign alignment fragments that underlie predicted TFBS with question – maybe alignment is wrong and site is conserved? Optimal alignment may not be conserved and suboptimal alignment may be conserved.

Composite Module Analyst[2] was revealing overrepresented TFBS combinations and predicts composite modules. It uses genetic algorithm, because complete enumeration fails in such big space. It was wrapped into ExPlain system, that was also contains many different tools. ExPlain was developed by our company.

We were participating in COGANGS Grant of FP7 and re-implemented Match algorithm in java. Using various optimizations we reached very fast speed. Now this implementation is used in Biobase tools.

Last 3 years support of Transfac web interface was done mostly by our team. In this interface we implemented match, fmatch (revealing statistically over or underrepresented sites), comatch (reveal a factor that works cooperatively with given factor), functional analysis and network analysis (from TransPath). Code backed by perl containing 696 modules with 334 560 lines of code. It uses MySQL or Oracle database. We support matrix generation pipeline and recently updated it with method for finding matrices that have two cores.

Availability: www.biobase-international.com All rights on the tools and interfaces belong to Biobase/Qiagen.

Acknowledgements Work was done by request of Biobase/Qiagen and supervised by Alexander Kel (ExPlain), Jennifer Hogan, Volker Matys, Sanjeev Pillai (Transfac/Proteome).

References:

1. Cheremushkin E, Kel A. (2003) Whole genome human/mouse phylogenetic footprinting of potential transcription regulatory signals. *Pac Symp Biocomput.* **2003**: 291-302.
2. Waleev T,, et. al. (2006) Composite Module Analyst: identification of transcription factor binding site combinations using genetic algorithm. *Nucleic Acids Res.* **2006 Jul 1**;34(Web Server issue):W541-5.

DISCRETE CONTROL OF IMMUNE REACTION IN CONDITIONS WITH INCOMPLETE INFORMATION

M.V. Chirkov*, S.V. Rusakov
Perm State National Research University, Perm, Russia
** Corresponding author: ozs-50@mail.ru*

Key words: *discrete control, parameter identification, mathematical model of infectious disease, immunotherapy*

Motivation and Aim: Mathematical models infectious diseases usually nonlinear systems ordinary differential equations and contains large number of parameters. Parameter estimation equation system on basis clinical data allows forecasting course and outcame disease by individual simulate treatment program. In practice, this is possible, as general rule, measuring values of some state variables to certain points of time. In this regard, it is necessary developing effective parameter identification model methods that allows build control function and in same time to identifing parameters. The aim of this research is solving numerically problem of discrete control by immune response in conditions with incomplete information. Process of immune response based on the basic mathematical model of infectious disease that proposed by G.I. Marchuk [1]. The conditions with incomplete information implies that parameter values are unknown and parameter estimation has adjusted new experimental values becomed available.

Methods and Algorithms: In order to solve this task algorithm allowing construct a control function and at same time to identifying parameters with mathematical model of the immune response [2]. Algorithm on basis of Monte-Carlo method. The goal of control is reaching "ideal" immune response corresponds to high stimulation of the immune system and lack of delay at reaction to infection.

Results: Control for main forms of disease: acute, chronic and lethal outcome. Treatment programs has based on immunotherapy which consists in introduction of preparing immunoglobulins or donor antibodies. Immunotherapy allows reducing chronic and lethal form in acute form with recovery. In the acute form of disease decreases maximum concentration of antigens and recovery occur more rapidly.

Conclusion: From the analysis for results that proposed algorithm has able estimate the model parameters with one sufficient accuracy and relatively lowly computational cost as well as the construct treatment program with simultaneously identification parameters.

References:

1. Marchuk G.I. (2013) *Mathematical Modeling of Immune Response in Infectious Diseases*. Dordrecht: Springer Science & Business Media.
2. S.V. Rusakov, M.V. Chirkov. (2014) Parameter identification and control in mathematical models of the immune response, *Russian Journal of Biomechanics*, 18 (2): 226 – 235.

ON A METHOD OF APPROXIMATION OF SOLUTIONS TO DELAY DIFFERENTIAL EQUATIONS

G.V. Demidenko
Sobolev Institute of Mathematics SB RAS, Novosibirsk, Russia
Novosibirsk State University, Novosibirsk, Russia
Corresponding author: demidenk@math.nsc.ru

Key words: *delay differential equations, ordinary differential equations of large dimension, approximation of solution*

Motivation and Aim: We consider delay differential equations and suggest a method of approximation of their solutions.

Methods and Algorithms: Our method of approximation of solutions to delay differential equations is based on the wavelet analysis technique and properties of solutions to some classes of systems of ordinary differential equations of large dimension.

Results: At present there are some methods of approximation of solutions to delay differential equations

$$\frac{d}{dt}y(t) = f(t, y(t), y(t-\tau)), \qquad t > \tau, \tag{1}$$

by solutions to special classes of systems of differential equations of large dimension

$$\frac{dx}{dt} = Ax + F(t, x) \tag{2}$$

(for examples, see [1-6]). In our paper we describe a wide class of systems of the form (2) whose solutions are closely connected with solutions to equations of the form (1). We study properties of solutions to these systems. Using the wavelet analysis technique, we propose a new method of approximation of solutions to delay differential equations (1). This paper continues our investigations [6-8].

Conclusion: This new method allows us to study properties of solutions to (1).

Acknowledgements: The research was supported by the Russian Foundation for Basic Research (project no. 16-01-00592).

References:

1. Yu.M. Repin. (1965) On the approximate replacement of systems with lag by ordinary dynamical systems, *J. Appl. Math. Mech.*, 29(2): 254-264.
2. H.T. Banks, J.A. Burns. (1978) Hereditary control problems: numerical methods based on averaging approximations, *SIAM J. Control Optim.*, 16(2): 169-208.
3. I. Györi. (1988) Two approximation techniques for functional-differential equations, *Comput. Math. Appl.*, 16(3): 195-214.
4. A. Ponosov, A. Shindiapin, J.J. Miguel. (2002) The W-transform links delay and ordinary differential equations, *Funct. Differ. Eq*., 9(3-4): 437-469.
5. B. Krasznai et al. (2010) The modified chain method for a class of delay differential equations arising in neural networks. *Math. Comput. Modelling*, 51(5-6): 452-460.
6. G.V. Demidenko, I.A. Mel'nik. (2010) On a method of approximation of solutions to delay differential equations, *Sib. Math. J.*, 51(3): 419-434.
7. G.V. Demidenko, V.A. Likhoshvai. (2005) On differential equations with retarded argument, *Sib. Math. J.*, 46 (3): 417-430.
8. G.V. Demidenko. (2012) Systems of differential equations of higher dimension and delay equations, *Sib. Math. J.*, 53(6): 1021-1028.

EXPERIMENTAL RESEARCH OF THE VISCOUS FLUID FLOW IN THE ELASTIC MODELWITH THE APPLICATION IN HEMODYNAMICS

N.S. Denisenko[1,3]*, A.A. Cherevko[1,3], V.M. Kulik[2], A.P. Chupakhin[1,3]
[1]*Lavrentyev Institute of Hydrodynamics SB RAS, Novosibirsk, Russia*
[2]*Kutateladze Institute of Thermophysics SB RAS, Novosibirsk, Russia*
[3]*Novosibirsk State University, Novosibirsk, Russia*
** Corresponding author: raselkroy@gmail.com*

Key words: *nonlinear oscillator equation, fluid flow in elastic tubes*

Motivation and Aim: The research of blood flow rate in vessels is a fundamental and important problem. One way of studying this problem is a direct building mathematical model, based on the neurosurgical data. Such models include an influence of external environment and fluid properties. To perform more detailed analysis of this model and its property it proposes to move on from neurosurgical measurements to experimental studies. This work dedicated to an experimental research of the viscous fluid flow in the elastic model.

Methods and Algorithms: The object of study is the fluid flow in the elastic model. The fluid flow is generated by programmable pump CompuFlow 1000 MR. The liquid used in the experiment has the same mechanical properties as a blood. The measurements of velocity and pressure run by the intravascular ComboWire guidewire. Laboratory models represent cylindrical tubes with length 200 mm, the wall thickness 123 mm and diameter of bore 75 mm. The models are made of the CKTH-A material with various admixture of PMS oil to vary the mechanical properties of a material. Basing on measuring data of velocity and pressure, it builds a nonlinear oscillator equation with a right part (generalized equation of Van-der-Pol-Duffing with load): $\varepsilon q'' + f(q)q' + g(q) = ku(t),$ where functions $f(q) = a_0 + a_1 q + a_2 q^2$ and $g(q) = b_1 q + b_2 q^2 + b_3 q^3$ determine dissipative and elastic properties of the system. The quantities $q = q(t)$ and $u = u(t)$ represent normalized values for pressure and flow rate $(|q| \le 1, |u| \le 1)$. The coefficients $k, a_i, b_j (i = 0,1,2,\ j = 1,2,3)$ calculate with the measured data and define individual characteristic of environment.

Results, Conclusion and Availability: For each experimental models, we had researched an interrelation between a mechanical properties and a character of flow, the differential equation are built and studied. An obtained information plays an important role in analysis of the real neurosurgical operations.

Acknowledgements: ComboWire measurements were supported by Russian Foundation for Basic Research (project No. 14-01-00036). Works with the pump CompuFlow 1000 MR were supported by Russian Science Foundation (project No. 14-35-00020). Design and fabrication of elastic models are supported by the the Program of state support of leading scientific schools of RF (project No. NSh-8146.2016.1).

MR AND DOPPLER ULTRASOUND VELOCIMETRY MEASUREMENTS OF VISCOUS FLUID FLOW IN THE MODEL OF THE COMMON CAROTID ARTERY BIFURCATION

N.S. Denisenko*[1,6], A.P. Chupakhin[1,6], A.K. Khe[1,6], A.A. Cherevko[1,6], A.A. Yanchenko[1,6], A.A. Tulupov[2,6], A.V. Boiko[3], A.L. Krivoshapkin[4], K.Yu. Orlov [4], M.P. Moshkin[5], A.E. Akulov [5]

[1]*Lavrentyev Institute of Hydrodynamics, Novosibirsk SB RAS, Russia*
[2]*International Tomography Center, Novosibirsk SB RAS, Russia*
[3]*Khristianovich Institute of Theoretical and Applied Mechanics SB RAS, Novosibirsk, Russia*
[4]*Academician E.N. Meshalkin Research Institute of Circulation Pathology, Novosibirsk, Russia*
[5]*Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[6]*Novosibirsk State University, Novosibirsk, Russia*
** Corresponding author: raselkroy@gmail.com*

Key words: *MRI flow measurements, Comparative study, Carotid bifurcation phantom*

Motivation and Aim: This work is dedicated to an experimental research of the viscous fluid flow in the silicone model of the common carotid artery bifurcation.

Methods and Algorithms: The object of study is fluid flow in the silicone model of the common carotid artery. Fluid flow is generated by programmable pump CompuFlow 1000 MR. The liquid used in the experiment has the same mechanical properties as blood. The measurements are performed on three facilities: 1) the intravascular ComboWire guide, 2) the tomograph Philips Achieva (1.5 T), 3) The tomograph Bruker BioSpec 117/16 USR (11.7 T). A variety of flow regimes are presented: with constant rate, periodic, measured experimentally during neurosurgical operation, etc.

Results, Conclusion and Availability: About 300 measurements on the facilities (Philips, Bruker, ComboWire) were performed during the experimental sessions The result of the experimental studies is qualitative and quantitative information about the dynamics of fluid flow inside the model of the carotid artery for different flow regimes. The measurement equipment consisting of the intravascular ComboWire guide wire, CompuFlow 1000 MR pump and two Philips and Bruker MR scanners can be used to obtain reliable and consistent data in the investigation of fluid flow in the vessel model.

Acknowledgements: ComboWire measurements were supported by the Russian Foundation for Basic Research (project No. 14-01-00036). MRI measurements were supported by the Russian Science Foundation (project No. 14-35-00020). Experimental research is supported by the the Program of state support of leading scientific schools of RF (project No. NSh-8146.2016.1).

MATHEMATICAL MODELLING OF ARTIFICIAL HEART VALVE PERFORMANCE

D.A. Dolgov, Y.N. Zakharov
Kemerovo State University, Kemerovo, Russia
Corresponding author: 9erthalion6@gmail.com

Key words: *viscous inhomogeneous fluid, artificial heart valve, immersed boundary method*

Motivation and Aim: In recent years interest in mathematical modeling of blood flow in vessels and artificial human heart valves significantly increased because of development of new methods of cardiovascular system diseases treatment. An artificial heart valve is an extremely complex system, which must meet a number of requirements, and mathematical modeling can simplify valve development and optimization process. In this paper we propose the mathematical model and its numerical implementation to describe three dimensional blood flow dynamics in artificial heart valve and its numerical implementation.

Methods and Algorithms: The mathematical model, proposed to solve a nonstationary problem of blood flow in valve, allows taking into consideration main features of heart valve performance: the inhomogeneous blood structure, the valve leaflets flexibility and complex geometry. Blood is described as a viscous incompressible inhomogeneous fluid and consists of two components (e.g. plasma and formed elements). The fluid motion can be defined by Navier-Stokes nonstationary system of differential equations with variable density and viscosity [1]. We model a valve leaflet as a flexible impenetrable surface which is deformed under the fluid pressure. The leaflet deformation and interaction with the fluid can be described by the immersed boundary method [2]. The valve leaflets influence on the fluid is described by body forces in the equation of fluid motion.

Results and conclusions: The mathematical model and its numerical implementation were applied to the problem of blood flow inside aortic valve with different forms and admixture distribution to define the dynamics of described biological system, including the flow rate, leafters geometry, stress distribution.

Acknowledgements: The research is performed as part of the government contract 1.630.1.2014/K.

References:

1. Milosevic H., Gaydarov N. A., Zakharov Y. N. (2013) Model of incompressible viscous fluid flow driven by pressure difference in a given channel, *International J. Heat and MassTransfer*. Vol. 62, P. 267–274
2. Boyce E. G. (2011) Immersed boundary model of aortic heart valve dynamics with physiological driving and loading conditions, *International J. Numerical Methods in Biomedical Engineering.* Vol. 29, No 5, P. 317–345.

EUCLIDEAN ANALOGUES OF GENETIC DISTANCES BETWEEN NUCLEOTIDE SEQUENCES

V.M. Efimov[1–4]*, K.V. Efimov[5], V.Y. Kovaleva[2]
[1]*Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*
[2]*Institute of Systematics and Ecology of Animals, SB RAS, Novosibirsk, Russia*
[3]*Novosibirsk State University, Novosibirsk, Russia*
[4]*Tomsk State University, Tomsk, Russia*
[5]*Moscow Institute of Physics and Technology (State University), Moscow, Russia*
** Corresponding author: vmefimov@ngs.ru*

Key words: *nucleotide sequences, evolution tree, Jaccard index*

Motivation and Aim: It was even noticed by Cavalli-Sforza&Edwards that the gene data are only available, in general, for presently living populations, so we may only be able to reconstruct a projection of the evolution tree onto the "now" plane. Reconstruction of the tree in space and time will be possible, however, if we are ready to make hypotheses about the mode and speed of evolution. The Euclidean representation of such space-time is the simplest possible and suffice for this goal [1]. Just because their famous genetic distance based on allelic frequencies is Euclidean. But current genetic distances between nucleotide sequences are almost all not Euclidean [2]. Most of them are built by common pattern. It is suggested that own probability of substitutions exist for each nucleotide pair in each site. Distance between sequences estimates as divergence time·2 from common ancestor but any metric or non-metric features are supposed, unlike Cavalli-Sforza&Edwards idea. Different models differ by number of independent parameters only, from one to six in symmetrical case. The difference can be accounted for in the synonymy and/or the position in the codon. One exception is a p-distance which estimates a number of mismatched nucleotides only and isn't time-oriented. P-distance completely coincides Cavalli-Sforza&Edwards conception. It is metric distance between sequences in "now" space. More over square root from p-distance is Euclidean and it is an analogue of Jukes–Cantor distance. Previously we built an Euclidean analogue for two-parameter Kimura distance [2]. Now our aim is a construction the Euclidean analogues for other time-oriented distances.

Methods and Algorithms: We consider the problem in the general case. Various nucleotide pair frequencies can be either set or estimated for the sample. Let us denote by F_X, F_Y nucleotide frequencies and by F_{XY} – nucleotide pair frequency for each nucleotide pair X,Y. Define $d_{XY}^2=1-F_{XY}/(F_X+F_Y-F_{XY})$. It is one's complement to well-known Jaccard index, d^2 is metric distance, and $d=(d^2)^{1/2}$ is Euclidean one [3]. Summing d^2 by all sites we receive D^2 – metric distance between nucleotide sequences, and $D=(D^2)^{1/2}$ – Euclidean one.

Results: An example from MEGA6 is considered [4]. The results are quite reasonable.

Conclusion: The Euclidean analogues for certain set of time-oriented evolution distances are built.

Acknowledgements: This work has been supported by Budget Project (grant #0324-2015-0003).

References:

1. Cavalli-Sforza LL, Edwards A.W. (1967). Phylogenetic analysis. Models and estimation procedures. American journal of human genetics. 19(3 Pt 1): 233.
2. Efimov VM, Melchakova MA, Kovaleva VY. (2013). Geometric properties of evolutionary distances. Vavilov Journal of Genetics and Breeding. 17 (4/1): 202–211. (in Russian).
3. Gower JC, Legendre P. (1986). Metric and Euclidean properties of dissimilarity coefficients. Journal of classification. 3(1): 5–48.
4. Tamura K, et al. (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. Molecular biology and evolution. Mst197.

GENOME GENESIS. APPEARANCE OF MULTICELLULAR ORGANISMS

I.L. Erokhin
National Biotechnological Company LLC, Moscow, Russia
Corresponding author: i.erokhin@inbox.ru

Key words: *genome genesis, multicellular organisms, junk DNA*

All the genes may be divided into two groups: regulatory genes that encode the transcription factors and the structural genes that encode enzymes and structural proteins, ribosomal and soluble RNAs. Improvement of DNA-dependent RNA polymerases turned the mobile elements integrated into the genome of prokaryotes and their mutated endogenous copies into non-coding regulatory genes, which made it possible to create complicated cell programs and resulted in the appearance of unicellular eukaryotes. Further increase in the number of regulatory genes on the basis of mobile elements made it possible to create a supracellular structure responsible for ontogenesis and morphogenesis of multicellular organisms.

The organism becomes multicellular if its genome acquires a more complicated structure in several directions. First of all, the programs of its cells should ensure formation of intercellular structures which prevent the daughter cells from separating from each other and determine their interaction. Secondly, several cell programs should accumulate in the genome as a result of amplification, and these programs will differ from each other due to mutations of regulatory genes. Third, the structural genes, household genes and their gene networks had to be drawn beyond the cell programs. Now, sufficiently extended structural genes and their total may be used by different cell programs with different intensity. Fourth, supracellular structures should be formed which use cell programs as subroutines. It allows to considerably reduce the genome size due to multiple use of the same programs by different (but identical by functions) cells and their generations.

As a rough approximation, the multicellular organism genome is structured (in terms of the graph theory) in the form of an oriented binary tree, with tops corresponding to homogeneous logical elements "iteration step of the genome tree" (GT step), and arcs corresponding to transfer of management between the "steps". Each cell of a multicellular organism corresponds to its "step". It receives management from the "step" of the mother cell, initiates realization of the "cell program" that determines its development and division, returns management upon termination of the cell program realization and transfers management to the "steps" of daughter cells The “cell program” is a network where the regulatory, as well as structural genes, are engaged in certain order.

This genome model is based on the hypothesis in accordance with which the management between the “steps” is transferred with the help of short RNAs. Being the product of expression of non-coding interspersed repeats, these short RNAs which make up a part of nucleoproteins formed with general factors of polymerase transcription Pol III, can initiate transcription of other repeats. Thus, non-coding interspersed repeats should be the main, the most numerous regulatory genes in the genome. The cell program iteration step may be represented as a pair of regulatory genes with identical complex promotors, and GT step – with three regulatory genes. GT and cell programs represent a multitude of dispersed repeats which are transcribed very rarely and to a small degree and look like junk DNA. Multiplication and mutation of repeats create new and new variants of GT and cell programs. Now is the time for natural selection which leaves the best variants of multicellular organisms. The model and the supporting hypotheses belong to the author.

GENOME GENESIS. ORIGIN OF EUKARYOTES

I.L. Erokhin
National Biotechnological Company LLC, Moscow, Russia
Corresponding author: i.erokhin@inbox.ru

Key words: *genome genesis, eukaryote, junk DNA*

The model of transition from prokaryote to eukaryote is suggested. The model is based on the hypothesis in accordance with which mobile elements, a considerable part of junk DNA, are the main regulatory genes of the genome. All the genes may be divided into two groups: regulatory genes that encode the transcription factors and the structural genes that encode enzymes and structural proteins, ribosomal and soluble RNAs. The fundamental difference of eukaryotes from prokaryotes consists in DNA-dependent RNA polymerases. Eukaryotes have three of them, with POL III transcribing dispersed repeats, which cannot be done by bacterial polymerase. Unnecessary mobile elements that are periodically integrated into the genome of prokaryotes are gradually ousted by natural selection. The mechanism using the products of mobile elements expression as the key transcription factors to initiate the transcription of other mobile elements, transforms these latter into basic regulatory genes of eukaryotes. They help to build complex programs of cell development, especially on the condition that the key transcription factors considerably increase affinity of POL II preinitiation complex to promoters of protein-coding genes. It should be noted that there was a case of sites complementary to short RNA sequences being present in the double-stranded DNA; this is the CRISPR/CAS system.

It is insufficient for the cell to have relevant genes in the genome and produce a certain product to form complex structures. It is necessary for this product to be found in the certain area of the cell in the time needed along with a set of other products. One of the most important trends of cell programs complication is the formation of the targeted transport organs. First of all, these are various membrane formations: nuclear membrane which prevents the transcription products from "roaming" all over the cell, endoplasmic reticulum, Golgi's apparatus and transport vesicles. Targeted delivery of vesicles to the necessary part of the cell is done with the help of motor proteins and microtubules. The product itself and its precursors should contain signal sequences of amino acids and nucleotides, enabling the sorting complex to orient them in the right direction. Intron-exon structure of protein-coding genes and the alternative splicing mechanism are the preferred methods of forming signal sequences for delivered products. Upon completion of their function, signal sequences are removed in the process of maturation. Acquisition of mitochondria and plastids completes appearance of eukaryotes.

Now the well-powered cell, possessing the targeted transport mechanism, can have a large size and an extended genome. The genome increases considerably, mainly by acquiring new and reproduction of endogenous mobile elements. They mutate and thus create new and new variants of the cell programs. Now is the time for the natural selection which secures the most appropriate variants of cells programs whose regulatory genes are the mobile elements. The model and the supporting hypotheses belong to the author.

STRUCTURE OF UNICELLULAR EUKARYOTES ONTOGENESIS PROGRAMS

I.L. Erokhin
National Biotechnological Company LLC, Moscow, Russia
Corresponding author: i.erokhin@inbox.ru

Key words: *ontogenesis, cell program, junk DNA*

Construction and development of such a complex object as the eukaryote cell is impossible without a relevant program. The gene surely forms the material basis of the cell ontogenesis program. All the genes may be divided into regulator and structural ones. Examples of structural genes include the genes coding structural proteins, enzymes, ribosomal and transfer RNAs. Regulator genes code transcription factors. We may single out general transcription factors which render transcription of any gene impossible, as well as unique key transcription factors (KTF) which, jointly with general transcription factors, are responsible for initiating of transcription of particular genes.

If one gene codes the KTF initiating another regulator gene, the latter initiates a third one, etc., there appears a structure called a chain (in graph theory). Let's supplement each gene of the chain with one more regulator gene, and let both genes forming a pair have the same promoters. Both genes will be initiated by one KTF (its copies). The first gene of this pair ensures the chain structure, and the second one initiates some process, for instance, a structural gene or other chain of regulator genes. A considerable defect of such a construction is simultaneous performance of chain elements and the processes initiated by them. It would be better if the next chain element started execution only upon termination of the process initiated by the previous element.

This requires introducing a notion of a complex promoter. Unlike a simple promoter for which one KTF is enough, transcription of a gene with a complex promoter may only be initiated in presence of two different KTFs. Let (i) all the regulator genes that form pairs of the chain have complex promoters, (ii) the second KTF be identical for all these complex promoters, (iii) all the processes initiated by chain elements have a gene at the end that codes the second KTF identical for all the pairs of the chain.

Such a construction makes a loop called Dijkstra's loop. A pair of genes with the same complex promoters forms a step of the loop. The process initiated by the loop step is a subroutine. If all the loop steps refer to the same subroutine, it is executed in accordance with the number of steps contained in the loop. If the subroutine itself represents a loop, it becomes an embedded loop. Embedding may have multiple levels. The advantage of Dijkstra's loop is that each step may refer to a unique subroutine that differs from the others. Several loops may refer to the same subroutine. For this purpose, it should contain several genes at the end that code the second KTF specific for its loop. The step of the loop may contain several genes with the same complex promoters, which turns the chain into a network. In total, such a multi-level embedded loop forms the program of the cell whose lower level consists of structural genes and enhancers and silencers that serve them.

The existence of such a structure is only possible within the hypothesis under which KTFs are represented by short RNAs. Being the product of mobile elements expression, these short RNAs that form a part of nucleoproteins formed with general transcription factors of polymerase Pol III can initiate the transcription of other mobile elements. Thus, mobile elements should be the main and the most numerous regulator genes in the genome of eukaryotes. The model and the supporting hypotheses belong to the author.

A NEW ALGORITHM TO THE RECONSTRUCTION OF A SET OF POINTS FROM THE MULTISET OF N^2 PAIRWISE DISTANCES IN N^2 STEPS FOR THE *DE NOVO* SEQUENCING PROBLEM

E.S. Fomin
Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
Corresponding author: fomin@bionet.nsc.ru

Key words: *cyclic peptides, sequencing, polynomial-time*

Motivation and Aim: The problem of amino acid sequence reconstruction in cyclic peptides is particularly important, because such peptides are of special significance in a cell. They include antibiotics, antitumor agents, toxins, immunomodulatory agents, and a variety of peptides with still unknown properties. The sequences of many cyclic peptides cannot be reconstructed from the DNA sequence, because they are not encoded by genes and their synthesis occurs in a nonribosomal way [1].

Methods and Algorithms: De novo sequencing, i.e., the direct reconstruction of the primary sequences of peptide chains from mass spectra, has been in use since the early 2000s [2]. Its major advantage is that it can be applied when no genomic information is available. The problem of sequencing linear and cyclic peptides is reduced to the known turnpike and beltway problems [3], the latter of which having no polynomial-time algorithm in the general case.

Results: Many algorithms were proposed for de novo sequencing, see the review of [4]. Despite the enormous efforts made in recent years, the turnpike and beltway problems are still considered open. A new simple approach to reducing overhead costs for the problems, based on the sequential removal of redundant information from inputs, is proposed. For error-free inputs that simulate de novo sequencing spectra with high accuracy, up to 10^{-3} Da, the size of inputs decreases drastically, from n^2 to O(n), which permits one to eliminate exhaustive search from the algorithm almost completely and reconstruct sequences in numbers of steps in direct ratio to the input size, n^2. Computational experiments show high efficiency of the algorithm for both the turnpike and beltway cases, with the reconstruction time for sequences of lengths up to several thousand elements being within one second on a modern PC.

References:

1. M.A.Marahiel et al. (1993) Regulation of peptide antibiotic production in Bacillus, Mol Microbiol, 7(5): 631-636.
2. R.S.Johnson, J.A.Taylor. (2002) Searching sequence databases via de novo peptide sequencing by tandem mass spectrometry, Mol Biotechnol, 22: 301-315.
3. P.Lemke et al. (2003) Reconstructing Sets From Interpoint Distances, Discrete and Computational Geometry Algorithms and Combinatorics, 25: 597-631.
4. J.Allmer. (2011) Algorithms for the de novo sequencing of peptides from tandem mass spectra, Expert Review of Proteomics, 8(5): 645-657.

A FENOMENON OF MULTISTABILITY IN A SIMPLE ECOLOGICAL EVOLUTIONARY POPULATION MODEL

E.Ya. Frisman[1], O.L. Zhdanova[2]*
[1]*Institute of Complex Analysis of Regional Systems FEB RAS, Birobidzhan, Russia*
[2]*Institute of Automation and Control Processes FEB RAS, Vladivostok, Russia*
** Corresponding author: axanka@iacp.dvo.ru*

Key words: *population dynamics, evolution, natural selection, adoptive locus, mathematical modeling, multistability*

Motivation and Aim: A new phenomenon discovered recently is multistability [1, 2, etc.], it appears as a simultaneous coexistence of different dynamic regimes in the system; at that initial values of the system's characteristics define resulted type of its dynamic regime. The phenomenon of multistability attracted attention of investigators because it allows an explaining some cases of changes in population dynamics regimes observed in real biological populations (for example, [3, 4]). Ecological-genetic models are naturally multistable, because they usually have several genetically different fixed points and switch in those is determined by the initial conditions. This effect is of great evolutionary significance as direction of evolution of the population in that case is determined by a set of initial conditions, or in the other words by random factors. Therefore, investigation of multistability areas in ecological-genetic models seems useful.

Methods and Algorithms: Mathematical and computer modeling is conducted to reveal a variety of possible bifurcations, dynamical regimes and to investigate multistability regions in the model of uniform population, which dynamics is described by Ricker's model and values of its intrapopulation parameters are defined on genetic level under influence of natural selection [5].

Results & Conclusion: It has been showed that this simple ecological-genetic model has strong sensitivity to the initial conditions. So a coexistence of several different asymptotic dynamic regimes (with own attraction basins) is possible in numerous enough parametric regions which are meaningful biologically. It means that results of natural selection in biological population depend from both initial conditions and intrapopulation parameters.

Acknowledgements: The present work was partially supported by the Russian Foundation for Basic Research (project 15-29-02658).

References:

1. A.P. Kuznetsov, A.V. Savin, Yu.V. Sedova, L.V. Tyuryukina. (2012) *Bifurcations of mappings*. (Nauka, Saratov).
2. M.P. Kulakov, G.P. Neverova, E.Ya. Frisman. (2014) Multistability in models of dynamics of migrationally-connected populations, *Nelineynaya dinamika*, 10: 407-425.
3. K.L. Kausrud, A. Mysterud, H. Steen, et al. (2008) Linking climate change to lemming cycles, *Nature*, 456: 93-97.
4. T.C.R. White. (2011). What has stopped the cycles of sub-Arctic animal populations? Predators or food? *Basic Appl. Ecol*., 12: 481-487.
5. O.L. Zhdanova, E.Ya. Frisman. (2005) Dynamic regimes in a model of single-locus density-dependent selection, *Rus. J. Genetics*, 41(11): 1302–1310.

SyGraph – WEB SYSTEM FOR VISUALIZATION OF SYNTENY ALIGNMENTS AND COMPARISON OF ASSEMBLY CONTIGS

M.A. Genaev[1], D.A. Afonnikov[1,2]*
[1]*Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2]*Novosibirsk National Research State University, Novosibirsk, Russia*
** Corresponding author: mag@bionet.nsc.ru*

Motivation and Aim: Analysis of synteny between genomic sequences is an important step for quantifying evolutionary relatedness between genomes in terms of chromosomal rearrangements [1]. Comparison of large genomic segments (contigs) can also be useful in comparison of the genome assemblies to estimate their quality and consistency. An important step of such analysis is the visualization of syntenic regions between genomes or assemblies, in particular, to visualize relationships between several genome segments (chromosomes, contigs) simultaneously. A number of tools were developed to visualize the relationships between several contigs (IGV[2], Circos [3]). However, they are not sufficient enough when relationships between hundreds or thousands contigs need to be visualized.

Methods and Algorithms: To resolve the problem of visualization of synteny between large set of genomic fragments we suggests a method based on reconstruction and visualization of graphs representing the relationships between genomic fragments. The method is based on building directed, weighted, bipartite graph, which nodes are contigs and the weight of edges are the coverage of contig pairwise alignments.

Results: The SyGraph web application use the information about genomic alignments in coords format as input. It reconstruct the graph of sequence similarity and generates graphic representing this graph carry out its statistical analyses.

Conclusion: The SyGraph output provide useful information about the similarity between genomic sequences and the phylogenetic relations between the compared organisms according to the graph topology.

Availability: SyGraph system is available at http://pixie.bionet.nsc.ru/SyGraph

Acknowledgements: This work supported by budget project № 0324-2015-0003.

References:

1. Kurtz S. et al. Versatile and open software for comparing large genomes //Genome biology. – 2004. – Т. 5. – №. 2. – С. R12.
2. Thorvaldsdóttir H., Robinson J. T., Mesirov J. P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration //Briefings in bioinformatics. – 2013. – Т. 14. – №. 2. – С. 178-192.
3. Krzywinski M. et al. Circos: an information aesthetic for comparative genomics //Genome research. – 2009. – Т. 19. – №. 9. – С. 1639-1645.

MUTATIONAL LANDSCAPE OF PROSTATE TUMORS BASED ON WHOLE EXOME SEQUENCING

I.R. Gilyazova[1,2*], M.A. Yankina[1], G.B. Kunsbaeva[2], A.A. Izmaylov[3], A.T. Mustafin[3], V.N. Pavlov[3], E.K. Khusnutdinova[1,2]

[1]*Institute of biochemistry and genetics Ufa Scientific Centre RAS, Ufa, Russia*

[2]*Bashkir State University, Ufa, Russia*

[3]*Bashkir State Medical University, Ufa, Russia*

** Corresponding author: gilyasova_irina@mail.ru*

Key words: *prostate cancer, candidate genes, tumor tissue, variants, functional significance, rare alleles*

Motivation and Aim: Prostate cancer (PC) is one of the most common malignancies of the male population worldwide. Extremely rapid increase in the incidence of prostate cancer and the high prevalence of the disease in the world testify to the need to study the mechanisms underlying the development of prostate cancer.

Methods and Algorithms: New possibilities in molecular genetics investigations using NGS technology can expand the possibilities of tumor heterogeneity investigation, which is especially important for the genetic heterogeneity of pathologies with a large number of candidate genes. To find new genes involved in the pathogenesis of prostate cancer, we conducted a whole exome sequencing in samples of normal and tumor tissue for 8 patients with prostate cancer. The variants revealed were annotated using ANNOVAR software tool [1]. To identify the events associated with the tumor, the variants with frequencies more than 0.03 in any of the databases 1000 Genomes Project (European, East Asian, All), ExAC, ESP were excluded. The functional significance of the observed changes was carried out using following databases: SIFT, PolyPhen (using annotations for rare alleles), MutationTaster, MutationAssessor [2], FATHMM, CADD.

Results: We detected 41542 variants in normal tissue sample, 45948 - in tumor, in average. All the samples contained mutations in the *ATM*, and *TP53* genes. After all stages of bioinformatic analysis 35 candidate genes, involved in cell cycle control, apoptosis signaling, androgen processes of cell growth and differentiation, transcription repair were selected - *MUC16*, *MUC6*, *MTCH2*, *ZNF844*, *PRSS3*, *SSTR1*, *PDE11A*, *L2HGDH*, etc.

Conclusion: The study revealed a number of genes which role in prostate cancer has not been described previously. The analysis is to be continued to determine the involvement of the identified genes in the pathogenesis of prostate cancer.

References:

1. H.Yang, K.Wang. (2015) Genomic variant annotation and prioritization with ANNOVAR and wANNOVAR, Nature Protocols 10,1556–1566.
2. F.Gnad, A.Baucom, K.Mukhyala, G.Manning, Z.Zhang. (2013) Assessment of computational methods for predicting the effects of missense mutations in human cancers, Gnad et al. BMC Genomics, 14 (Suppl 3):S7.

SIBERIAN SUPERCOMPUTER CENTER AS A SERVICE FOR BIOINFORMATICS RESEARCH

B. Glinskiy*, I. Chernykh, N. Kuchin
Institute of Computational Mathematics and Mathematical Geophysics SB RAS, Novosibirsk, Russia
** Corresponding author: chernykh@ssd.sscc.ru*

Key words: *high performance computing, GPU, bioinformatics and life sciences*

Introduction: Sequencing and protein docking are very compute-intensive tasks that see a large performance benefit by using a CUDA-enabled GPU. There is quite a bit of ongoing work on using GPUs for a range of Bioinformatics and life sciences codes [1]. The main part of Siberian Supercomputer Center (SSCC) is HPC cluster with total peak performance 115 TFlops. This system is well designed for bioinformatics researches due to the flexible computational resource management.

Siberian Supercomputer Center resources: SSCC offers computer resources for bioinformatics researches to its users. We have supercomputer with hybrid architecture and consists of NKS-30T (platform BL2h220c hp) system with 576 Intel Xeon processors E5450/E5540/X5670 (2688 cores) and hybrid cluster that based on 40 servers HP SL390s G7 (80x CPU X5670 - 480 cores) with 3x NVidia Tesla M2090 GPU on each node. All cluster nodes are connected via Infiniband QDR network interface. Cluster file system IBRIX (4 servers, 32 TB of available disk space) is also connected by Infiniband interface. In addition, we provide access to HP DL380 G8 node with 2x Intel Xeon E5-2650 CPU (8 cores each) and NVIDIA K40M GPU [2]. This architecture is well suited for open source packages like MUMmerGPU: High-through DNA sequence alignment using GPUs [3], Parallel-META: a GPU- and multi-core-CPU-based open-source pipeline for metagenomic data analysis, which enabled the efficient and parallel analysis of multiple metagenomic datasets [4], and Molecular Dynamics packages like GROMACS [5], LAMMPS [6]. We also support The MGSmodeller [7] software package. The MGSmodeller includes modules for the mathematical model reconstruction and specification routine simulation as well as for inverse problems. The model reconstruction and simulation experiment's specification are describing in terms of model specification language SiBML. The complex mathematical models in the SiBML[8] depicted by the set of MGS elementary subsystems. MGS elementary subsystem defines by set of biological entities of interest and the law of the interaction between them. The model specification language allows to present full information about biological entities such as name, type (gene, RNA, protein etc.), localization in the compartment structure (organelles, cells, tissue etc.) and/or other object's properties. Thus, the language allows to take into consideration a hierarchical structure and complex organization of modeling objects, on basis of that we can model such organizational properties of the MGS as mutual positions of genes, promoters, terminals and others genetics elements on DNA, anisotropy of spatial compartments and transport processes. All this resources are available to all organizations that are operated by Federal Agency of Scientific Organizations Russia.

References:

1. ttp://www.nvidia.com/object/bio_info_life_sciences.html
2. http://www2.sscc.ru/ENG/Resources.htm
3. https://sourceforge.net/projects/mummergpu/
4. Xiaoquan Su, J. Xu and Kang Ning, "Parallel-META: A high-performance computational pipeline for metagenomic data analysis," 2011 IEEE International Conference on Systems Biology (ISB), Zhuhai, pp. 173-178, 2011.
5. http://www.gromacs.org/
6. http://lammps.sandia.gov/
7. Kzantsev F, et al, Proc. of the 6th International Conference on BGRS, p.113, 2008. Likhosvai V, et al, Generalized chemokinetic method for gene network simulation, Mol Biol, 35, pp. 1072-1079, 2001.

GEOMETRY OF PHASE PORTRAIT OF ONE GENE NETWORK MODEL WITH VARIABLE FEEDBACKS

V.P. Golubyatnikov[1,2]*, M.V. Kazantsev [3]
[1]*Sobolev Institute of Mathematics SB RAS, Novosibirsk, Russia*
[2]*Novosibirsk State University, Novosibirsk, Russia*
[3]*Polzunov Altai State Technical University, Barnaul, Russia*
** Corresponding author: glbtn@math.nsc.ru*

Key words: *Positive and negative feedbacks, non-linear dynamical systems, integral manifolds, cycles*

Motivation and Aim: We study one simple piece-wise linear (**PL**) dynamical system and show existence of a stable cycle in its phase portrait. Dynamical systems of this type appear often in gene networks modeling, see [1].

Methods and Algorithms: The approaches to modeling of similar gene networks functioning and description of corresponding phase portraits are presented in [2] and [3].

Results: For wide domain of parameters A and b, we consider 3D-dynamical system

$$\frac{dx}{dt} = L(z) - x;\ \ \frac{dy}{dt} = \Gamma(x) - y;\ \ \frac{dz}{dt} = \Pi(y) - z. \qquad (1)$$

All variables are non-negative, and L, Γ are monotonic step-functions which correspond to negative and, respectively, positive feedbacks: $L(0,1)=A_1>2$, $L(1,\infty)=0$; $\Gamma(0,1)=0$, $\Gamma(1,\infty)=A_2>2$. The function Π describes a variable feedback in this gene network model: $\Pi(y) = A_3>2$ for $1<y<b<A_2$; $\Pi(y)=0$ for $y<1$, and for $b<y<A_2$. Actually, in our calculations the case $A_1=A_2=A_3=A=3$, $b=1.8$ was considered. We decompose the invariant domain $Q=[0,A]\times[0,A]\times[0,A]$ to 12 blocks and construct discrete scheme (diagram) of the phase portrait of the system (1). One of the remaining blocks is its attractor. Existence of a stable **PL**-periodic trajectory ***C*** of (1) which passes through union U_8 of 8 of these blocks is shown. An invariant 2-dimensional **PL** surface $P^2 \subset U_8 \subset Q$ containing the cycle ***C*** is constructed, it has infinitely many faces which tends very rapidly to ***C*** in a spiral way, and so do trajectories of all points of P^2. They pass eventually through all blocks of U_8 as well.

In order to find ***C*** and to analyze other trajectories behavior, we have elaborated software written in **R** programming language, which uses the Cauchy problem solver *lsoda* from the package **deSolve** (see https://cran.r-project.org/web/packages/deSolve/index.html) in construction of trajectories of the dynamical system (1).

Conclusion: The cycle ***C*** is stable, and in its small neighborhood trajectories of the system (1) wind around ***C*** in a spiral way tending to it very rapidly.

Acknowledgements: RFBR, grant 15-01-00745.

References:

1. V.A.Likhoshvai, V.P. Golubyatnikov et al. (2008) Theory of gene networks, In: *System computerized biology*, 397-480, Novosibirsk, SB RAS.
2. N.B.Ayupova, V.P. Golubyatnikov. (2015) On two classes of nonlinear dynamical systems: the 4-dimensional case, *Sib. Math. Journ.*, **56(2)**: 231-236.
3. M.V.Kazantsev On some properties of the domain graphs of dynamical systems. *Sib. Journ. Industrial Math*. (Russian), 2015, **18(4)**: 42-48.

IDENTIFIABILITY OF MATHEMATICAL MODELS OF PHISIOLOGY

A.A. Grodz[2]*, S.I. Kabanikhin[1,2], D.A. Voronov[1,2], O.I. Krivorotko[1,2]
[1]*Institute of Computational Mathematics and Mathematical Geophysics SB RAS, Novosibirsk, Russia*
[2]*Novosibirsk State University, Novosibirsk, Russia*
** Corresponding author: nastya-grodz@yandex.ru*

Key words: *systems of ordinary differential equations, identifiability, inverse problem, software*

Motivation and Aim: Identifiability is an a priori property of a mathematical model, which guarantees an unambiguous determination of the unknown parameters through the available experimental data (measurements). A priori identifiability is a key step in the formulation of a structural model whose parameters are estimated from a set of data. The question that a priori identifiability addresses is: do the data contain enough information to estimate all of the unknown parameters of the postulated model structure? If the answer to this question is positive, it is possible to carry out a series of measurements to obtain the experimental data, without possibility of the fact that these data may not be useful, and on the other hand – expensive.

Results: The problem of the identifiability of dynamic systems is considered in this work. The review of methods of analysis of identifiability of dynamic models is given. In this review article following approaches are considered: transfer function method; the Taylor series expansion method; method based on differential algebra theory and its modification covered; method based on graph theory; method based on definition of Lie derivatives; method based on likelihood function. The necessity of making an a priory identifiability analysis before estimating parameters characterizing any process is demonstrated on several examples. The examples of identifiability analysis of mathematical models of medical biology are presented.

References:

1. N. Meshkat, et al. (2014) Finding and Using Identifiable Parameter Combinations in Nonlinear Dynamic Systems Biology Models and COMBOS: A Novel Web Implementation, Plos One, 19(11): 163–175.
2. A. Raue, et al. (2009) Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood, Bioinformatics, 25(15): 1923-1929.

THE APPLICATION OF OPTIMAL PARTITIONING BASED APPROACHES FOR ESTIMATION OF THE ADVERSE OUTCOME RISK IN PATIENTS DISCHARGED AFTER ACUTE CORONARY SYNDROME

R.R. Guliev[1]*, O.V. Senko[2], D.A. Zateyshchikov[3], V.V. Nosikov[1], A.V. Kuznetsova[1], M.A. Evdokimova[3], V.A. Brazhnik[3], I.N. Kurochkin[1,4]
[1]*Emanuel Institute of Biochemical Physics of Russian Academy of Science, Moscow, Russia*
[2]*Computer Center of Russian Academy of Science, Moscow, Russia*
[3]*Central State Medical Academy of Department of Presidential Affairs, Moscow, Russia*
[4]*Lomonosov Moscow State University, Moscow, Russia*
** Corresponding author: glvrst@gmail.com*

Key words: *data mining, collective decision making, optimal partitioning, prognosis acute coronary disease, ischemic heart disease*

Motivation and Aim: Cardiovascular diseases are the most common cause of death in the world. There are certain tasks which could help people to improve a quality of the disease treatment. One of them is an estimation of the adverse outcome risk in patients discharged after acute coronary syndrome (ACS). In order to develop a good predictive model a wide range of parameters (genetic, biochemical, physiological, etc.) should be taken into account. Relatively new approaches based on the optimal partitioning seem very suitable for such complex clinical data. Thus, the aim of this study is to create a risk estimating model by means of optimal partitioning based approaches and to evaluate it possibilities in comparison with the described approaches and methods.

Methods and Algorithms: For that aim, a dataset collected from 16 centers of 7 cities of Russia during 7 years was used. It contains a wide range of clinical, biochemical and genetic characteristics. In order to create a predictive model, approaches based on the optimal partitioning, such as the optimal valid partitioning (OVD) [1, 2] and the multimodel statistically weighted syndromes method (MSWS) [3], were used. Created model was compared with mostly used methods like logistic regression, decision trees, DFA, SVM, neural networks.

Results: A predictive model, allowing to estimate 6 months cardiovascular risk in patients discharged from hospital after ACS, has been created. Accuracy of the result model is quite well and estimated as AUC=0,72. The model has shown the better predictive ability compared to mostly used methods.

Conclusion: Optimal partitioning based approaches are a powerful tool for analysis of complex clinical data. And it opens up additional opportunities for the creation of predictive models on clinical trials data, containing a variety of medical, biochemical, genetic, and other information.

References:

1. O.V. Senko, A.V. Kuznetsova. (2006) The Optimal Valid Partitioning Procedures, InterStat. April. № 2.
2. A.V. Kuznetsova, I.V. Kostomarova, O.V. Senko. (2014) Modification of the method of optimal valid partitioning for comparison of patterns related to the occurrence of ischemic stroke in two groups of patients, Pattern Recognition and Image Analysis. 24:114–123.
3. O.V. Senko, A.V. Kuznetsova. (2010) A recognition method based on collective decision making using systems of regularities of various types, Pattern Recognition and Image Analysis. 20:152–162.

PHYSIOLOGICAL MEASUREMENTS WITH MICROENCAPSULATED FLUORESCENT SENSORS AND DONNAN'S EFFECT

A.N. Gurkov*, E.V. Borvinskaya, I.A. Belousova, E.P. Shchapova, D.S. Bedulina, M.A. Timofeyev
Institute of Biology, Irkutsk State University, Irkutsk, Russia
** Corresponding author: a.n.gurkov@gmail.com*

Key words: *Donnan's effect, fluorescent sensors, microcapsules*

Motivation and Aim: The microencapsulation of fluorescent molecular sensors of a variety of chemical and physical characteristics of solutions is a promising technology for *in vivo* measurements of physiological parameters in human and animal tissues [1]. However, it has been shown that due to Donnan's effect, the concentration of charged molecules of interest may be shifted inside the microcapsule compared to bulk media, which can potentially lead to distorted signals from the sensor [2]. This shift is caused by charged polymers and, importantly for physiological applications, can be produced by proteins in the fluids of organisms. In the present study, we perform modelling of possible errors of the microencapsulated fluorescent sensors of small charged molecules caused by Donnan's effect in the blood of rat (*Rattus norvegicus*) and zebrafish (*Danio rerio*).

Methods and Algorithms: The value of shift in concentrations of small charged molecules due to Donnan's effect was calculated according to [2]. Serum albumin was treated as the only rat plasma protein in our model for simplification. Protein composition of blood plasma of zebrafish was identified based on LC-MS data [3], 28 most abundant proteins were included in the model. With LC-MS data on Universal Proteomics Standard-2 (UPS2, Sigma) [4] precision of label-free LC-MS was evaluated to take into account possible ranges of relative proteins' concentrations in zebrafish plasma. Mature amino acid sequences of rat albumin and zebrafish plasma proteins were obtained from UniProt to model these proteins with I-TASSER server. The obtained 3D-models were transferred to PROPKA 3.1 algorithm to precisely predict their charge in physiological pH range and finally calculate the value of Donnan's effect.

Results and Conclusion: We conclude that the differences in pH and concentrations of charged molecules such as ions and metabolites between blood and microcapsule hollows may not be identifiable by means of fluorescent sensors, which justifies the use of the microencapsulated sensors *in vivo* without additional *in loco* calibration.

Acknowledgements: This work was supported by Russian Science Foundation (#15-14-10008).

References:

1. A. Sadovoy, C. Teh, V. Korzh et al. (2012) *Laser Phys Lett.*, **9:** 542-546.
2. D. Haložan, U. Riebentanz et al.. (2009) *Col Surf Phys Eng Asp.*, **342:** 115–121.
3. F. Babaei, R. Ramalingam et al. (2013) *J Proteome Res.*, **12:** 1580–1590.
4. L. Dicker, X. Lin, A.R. Ivanov. (2010) *Mol Cell Proteomics.*, **9:** 2704–18.

ANDSYSTEM: AN INTERNET-ACCESSIBLE TOOL FOR AUTOMATED LITERATURE MINING IN THE AREA OF BIOLOGY

V.A. Ivanisenko*, O.V. Saik, E.S. Tiys, T.V. Ivanisenko, P.S. Demenkov
Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
PBSoft LLC, Novosibirsk, Russia
** Corresponding author: salix@bionet.nsc.ru*

Key words: *text mining, data mining, ANDSystem, ANDVisio, ANDCell, gene networks, pathways*

Motivation and Aim: The number of publications in the areas of biology, medicine, and biotechnology increases dramatically, which makes imperative computer-based analysis. To date, over 20 million abstracts highly relevant to biology and medicine are stored in the PubMed database and this number keeps growing. To address the confounding problem of extraction of information on molecular-genetic objects from texts, approaches based on algorithms such as search for co-occurrence of the biological object names in texts, linguistic-semantic analysis of texts, and machine learning have been suggested. Previously we have developed the ANDSystem package that incorporates utilities for automated extraction of knowledge from Pubmed published scientific texts and databases (Demenkov et al, 2011, Ivanisenko et al, 2015). The goal of this work was a development of a new version of ANDSystem supplied with extended utilities for automated extraction of knowledge and associative gene networks reconstruction related to the studied phenotypic trait or biological process.

Results: ANDSystem was developed for the purpose of scanning literature for extracting relationships between diseases, pathways, proteins, genes, microRNAs and metabolites. The ANDSystem tool consists of ANDCell database, ANDVisio tool and utilities for automated extraction of knowledge from Pubmed published scientific texts and analysis of factographic databases. The ANDCell database contains information on molecular-genetic events retrieved from texts and external databases. ANDVisio is a user's interface to the ANDCell, it provides graphic visualization, editing and search features as well as possibilities to save an associative gene networks in different formats resulting from user's request. Such networks describe semantic relationships between molecular-genetic objects (proteins, genes, metabolites and others), biological processes, and diseases. ANDVisio is provided with various tools supporting automated reconstruction of associative networks with taking into account the specific relation of objects in the network to the studied phenotypic trait or biological process limiting uncontrolled network expansion. The ANDSystem can assist in the interpretation of complex multifactorial experimental data.

References:

1. Demenkov P.S. et al. (2011) ANDVisio: a new tool for graphic visualization and analysis of literature mined associative gene networks in the ANDSystem, In silico biology, 11:149-61.
2. Ivanisenko V.A. et al, (2015) ANDSystem: an Associative Network Discovery System for automated literature mining in the field of biology. BMC Syst Biol. 9 Suppl 2:S2.

A NUMERICAL ALGORITHM OF PARAMETER IDENTIFICATION IN MATHEMATICAL MODEL OF TUBERCULOSIS TRANSMISSION WITH CONTROL PROGRAMS

V.N. Kashtanova[2]*, S.I. Kabanikhin[1,2], O.I. Krivorotko[1,2], D.A. Voronov[1,2]
[1]*Institute of Computational Mathematics and Mathematical Geophysics SB RAS, Novosibirsk, Russia*
[2]*Novosibirsk State University, Novosibirsk, Russia*
** Corresponding author: vikakashtanova@ya.ru*

Key words: *system of ordinary differential equations, tuberculosis transmission, control programs, parameter identification, inverse problem, optimization approach, fast simulate annealing, gradient descent method*

Motivation and Aim: The development of an individual mathematical model describing the process of the propagation of Tuberculosis (TB) infection in the population is one of the most effective methods for prediction of the epidemic spread in a particular region. Such models are described by systems of nonlinear ordinary differential equations (ODE). The coefficients of these systems characterize the features of population and disease spread. Consequently, it is necessary to qualitatively evaluate parameters of model (or their combinations) [1] for specification model for special population.

Methods and Algorithms: The purpose of this work is the construction and investigation of the numerical algorithm for determining the coefficients of nonlinear ODE system which describes TB transmission processes with treatment and drug resistance [2] using additional information about a special population according to statistical data for the previous few years (namely, the number of healthy, latently infected and infectious diseases individuals). The numerical algorithm is based on combination of very fast annealing and gradient approaches for minimization of least squares objective function [3].

Results and Conclusion: The results of numerical calculations show that above approach determines the set of more sensitive parameters to a particular region that differs significantly from its widely used standard values. The numerical results are analyzed and discussed.

Acknowledgments: This work is supported by the Ministry of Education and Science of the Russian Federation and the Russian Foundation of Basic Research No. 16-31-00189.

References:

1. S.I. Kabanikhin. (2011) Inverse and Ill-Posed Problems: Theory and Applications (Berlin: de Gruyter).
2. J. M. Trauer, J. T. Denholm, E. S. McBryde. (2014) Construction of a mathematical model for tuberculosis transmission in highly endemic regions of the Asia-pacific, Journal of Theoretical Biology, **358**: 74–84.
3. H. T. Banks, Sh. Hu, W. C. Thompson. (2014) Modeling and Inverse Problems in the Presence of Uncertainty (Chapman and Hall/CRC press).

IDENTIFICATION OF MASTER-REGULATORS FOR PROGRAMMING OF SPERMATOGONAL STEM CELLS PLURIPOTENCY BY THE USE OF THE GENEXPLAIN/BIOUML PLATFORM

A.E. Kel[1,2]*, D.E. Stelmashenko[2]
[1]*Institute of Chemical Biology and Fundamental Medicine SB RAS, Novosibirsk, Russia*
[2]*BIOSOFT.RU, Ltd, Novosibirsk, Russia*
** Corresponding author: alexander.kel@biosoft.ru*

Key words: *search for master-regulators, BioUML/geneXplain platform, upstream analysis, stem cells programming, unipotency to pluripotency stem cell transformation*

Motivation and aim: It has been shown [1] that a transformation from unipotent to pluripotent spermatogonal stem cells can be carried out by cell culturing and does not require transgenic intervention into the cells which opens a "clean" way of generating iPS-like cells. However the molecular mechanism of such transition and signaling pathways activated during such transformation are still not known. Understanding of the signaling pathways involved is extremely important as it allows identification of master molecules, responsible for such transformations. Such master regulators appear to be the key nodes of signaling pathways and their affection can eventually lead to targeted cell reprogramming into the pluripotential state.

Methods and Algorithms: We analyzed data from the experiment [1] with the use of the geneXplain/BioUML platform. The upstream analysis approach [2] was used and the intracellular signaling pathways, which emerge during the cell transformation into the pluripotent state, were reconstructed. We identified several transcription factors involved and master-regulatory molecules, which could be clearly seen as the key nodes, responsible for all the variety of changes which eventually happen during the cell potency transformation process. Among found transcription factors there were Oct4 and NANOG – well-known pluripotency genes, but there also were other, not yet well-studied transcription factors which can now be proposed to the scientific community for further research. Upstream from the revealed transcription factors we identified few potential master-regulators that can be used as novel targets for iPS development. An article [3] was recently published on the analysis of the same data set [1]. The authors experimentally confirmed the transcription factors that were independently found in our study, but were not able to go beyond that and identify master regulators, which becomes possible with the use of geneXplain/BioUML platform.

Results and conclusion: We constructed an intracellular signaling network guiding transformation of unipotent spermatogonal cells to a pluripotent state. Novel key-node master regulators were revealed. Namely: Gli2, TGFbeta4, BMP4, E-cadherin, FGF-4, SHIP, Tie2.

References:

1. Query DataSets for GSE74151. Expression data from three types of spermatogonial stem cells. https://www.ncbi.nlm.nih.gov/geo/query/ acc.cgi?acc=GSE74151
2. Koschmann, J., Bhar, A., Stegmaier, P., Kel, A. E., & Wingender, E. (2015). "Upstream Analysis": An Integrated Promoter-Pathway Analysis Approach to Causal Interpretation of Microarray Data. *Microarrays*, *4*(2), 270-286.
3. Jo, J., Hwang, S., Kim, H. J., Hong, S., Lee, J. E., Lee, S. G., ... & Lee, D. R. (2016). An integrated systems biology approach identifies positive cofactor 4 as a factor that increases reprogramming efficiency. *Nucleic acids research*, *44*(3), 1203-1215.

A SOFTWARE TOOL FOR VISUALIZATION AND CONTROL OF BIOLOGICAL NEURAL NETWORKS ACTIVITY BASED ON THE NEURON SIMULATION ENVIRONMENT

S.S. Khayrulin[1,2]*, N.A. Serdtseva[2], A.Yu. Palyanov[1,2]
[1]*Institute of Informatics Systems SB RAS, Novosibirsk, Russia*
[2]*Novosibirsk State University, Novosibirsk, Russia*
** Corresponding author: skhayrulin@iis.nsk.su*

Key words: *The NEURON, visualization, computer simulation, 3D graphics*

Motivation and Aim: Construction and investigation of biological neural networks models is one of the most important instruments in computational neurobiology. Biologically reasonable computer simulation of a single neuron or neural network of a living organism needs to be built upon the morphological and electrophysiological data, including gap junctions and synaptic contacts, neuromediators, ion channels and ion currents, and other neurophysiologic parameters. The NEURON [1] is one of the best simulation environments, well-suited to problems that are closely linked to experimental data, especially those that involve cells with complex anatomical and biophysical properties. It is used in classrooms and laboratories around the world. Unfortunately, its system for visualization of neural networks lacks many important features which could make an interaction with it significantly more efficient and convenient. Modern high-quality 3D graphics for scientific and educational purposes also seems beneficial.

Methods and Algorithms: Being a part of OpenWorm international project [2] aimed at simulation of the *C. elegans* nematode, including its nervous system, we are interested in development of a two-way data exchange interface between NEURON and Sibernetic [3], our software for simulation of C. elegans body and environment, as well as in ability to use custom 3D visualization and control for the NEURON-based simulation of the *C. elegans* neurons. NEURON provides well-documented Python interface. We've chosen PyOpenGL for 3D visualization, PyQt for GUI support and MathPlotLib for 2D graphs drawing.

Results: A cross-platform software tool for custom 3D visualization and control for the NEURON-based neural networks simulations had been developed.

Conclusion: Currently available version can be already used as an alternative 3D viewer for NEURON-based simulations, with elements of control and visualization of signals propagation over the network. Integration with Sibernetic is expected to be developed soon.

Availability: The source code is here:
https://github.com/openworm/sibernetic_NEURON

Acknowledgements. The work of A.P., S.K. and N.S. was supported by Russian Federation President grant MK-5714.2015.9.

References:

1. N.T. Carnevale and M.L. Hines (2016) *The NEURON Book*, Cambridge, UK: Cambridge University Press.
2. B. Szigeti, P. Gleeson, M. Vella, S. Khayrulin, A. Palyanov, J. Hokanson, M. Currie, M. Cantarelli, G. Idili and S. Larson (2014) OpenWorm: an open-science approach to modeling Caenorhabditis elegans. *Front. Comput. Neurosci.,* **8**:137.
3. A. Palyanov, S. Khayrulin, S.D. Larson (2016) Application of smoothed particle hydrodynamics to modeling mechanisms of biological tissue. *Advances in Engineering Software*, **98**: 1-11.

IMPROVED SBGN (ML) SUPPORT IN BIOUML

I.N. Kiselev[1,2]*, S. Kinsht[3], F.A. Kolpakov[1,2]
[1]*Design Technological Institute of Digital Techniques, SB RAS, Novosibirsk, Russia*
[2]*Institute of Systems Biology, Ltd, Novosibirsk, Russia*
[3]*Novosibirsk State University, Novosibirsk, Russia*
** Corresponding author: axec@developmentontheedge.com*

Key words: *visual modeling, BioUML, Systems Biology Graphic Notation, markup language*

Motivation and aim: In BioUML mathematical model is represented as visual diagram. Each element of the diagram may be associated with element (variable, equation) of mathematical model. BioUML provides several visual notations including Systems Biology Graphic Notation (SBGN PD) [1]. Diagram should be semantically correct according to both SBML and SBGN rules e.g. each reaction should have at least one reactant or product, each arc should be connected to appropriate glyph, etc. This is controlled automatically as user creates model. This process implies conversion process between SBML and SBGN and vice-versa and was implemented in BioUML earlier. However visual diagram created in BioUML could be easily passed to other tools only as SBML with layout and rendering information, moreover not every SBGN element was supported properly and several improvements were needed.

Results: Recently we have significantly improved support of SBGN and SBGN-ML in BioUML.

1. Logical operator and phenotype are supported.
2. Ability to set brush, pen and title font for each separate diagram element.
3. Creation of styles for similar customization of groups of elements.
4. Corrected visualization of process glyph.
5. BioUML-specific elements “subdiagram” and “port” replaced by SBGN “submap” and “tag”.
6. Automatic redraws of the diagram as user drags mouse.
7. SBGN-ML export and import including.
8. SBML objects are automatically parsed to correspondent SBGN classes (complex, macromolecule, simple chemical, association etc.) based on SBO annotation.

For automatic nightly testing we use reimport with automatic models comparison. Another nightly test imports example SBGN-ML documents from www.sbgn.org and generates report with comparison between diagrams from BioUML, other tools (VANTED [2], SBML Layout [3]) and figure from specification.

Availability: described software is freely available as a part of open source platform BioUML at www.biouml.org.

References:

1. Moodie S. et al. (2011) Systems Biology Graphical Notation: Process Description language Level 1. Available from Nature Precedings <http://dx.doi.org/10.1038/npre.2011.3721.4>.
2. Junker B. H. et al. (2006) VANTED: a system for advanced data analysis and visualization in the context of biological networks. *BMC Bioinformatics.* 2006; **7**(1): 109.
3. SBML Layout: http://sysbioapps.dyndns.org.

VALIDATION OF THE HUMAN ARTERIAL TREE MODEL

I.N. Kiselev[1,2]*, E.A. Biberdorf[3,4], V.I. Baranov[5], T.G. Komlyagina[5],
I.Y. Suvorova[5], V.N. Melnikov[5], S.G. Krivoshchekov[5], F.A. Kolpakov[1,2]
[1]*Institute of Systems Biology Ltd. Russia, Novosibirsk*
[2]*Design Technological Institute of Digital Techniques SB RAS, Novosibirsk, Russia*
[3]*Novosibirsk State University, Novosibirsk, Russia*
[4]*Sobolev Institute of Mathematics SB RAS, Novosibirsk, Russia*
[5]*State Scientific-Research Institute of Physiology & Basic Medicine, Novosibirsk, Russia*
** Corresponding author: axec@developmentontheedge.com*

Key words: *arterial system, mathematical modeling, experimental data, model validation*

Motivation and aim: This work is concerned with the problem of validation of a one-dimensional model of the human cardiovascular system including the arterial tree of 59 vessels [1]. This kind of models is widely known and used in many studies. However to our best knowledge, there are very few works on validation of models prediction against significant amount of real clinical data. Usually validation is performed on single individual or small group of volunteers.

Methods: In this work we use results of a clinical study [2] comprising more than 1500 subjects. We restricted ourselves to predicting of systolic (maximal) and diastolic (minimal) pressures. Validation goes as follows: individual subject data is used to tune (personalize) model parameters, numerical simulation is performed, then obtained pressures are compared to those from database. Several simulations were conducted using different combinations of parameters to personalize. Subject characteristics from database are: height, weight, heart rate (HR), diastolic pressure, systolic pressure and parameters of 4 arteries in limbs: effective radius, arterial elastic resistance, characteristic impedance. Total peripheral resistance (TPR) and stroke volume (SV) are estimated on the base of these characteristics. Cross-sectional area and elasticity for 59 arteries are extrapolated on the base of information about 4 particular arteries.

Results: The simulation results show that the single parameter – TPR is able to explain up to 70% of systolic ($r\approx0{,}838$) and diastolic ($r\approx0{,}81$) pressures dispersion while other parameters cannot account for more than 10% of dispersion. The best result is obtained when personalization of the HR and SV is added to TPR: $r\approx0{,}897$ for systolic and $r\approx0{,}925$ for diastolic respectively. Thus, the structure of the arterial tree and arterials cross-section area and elasticity are unlikely to play a significant role in prediction of systolic and diastolic pressures. Meanwhile none of the used personalization parameters could explain the variability of pulse pressure.

Availability: The model described in this paper is available as a part of the free open-source platform BioUML at www.biouml.org.

Acknowledgement: The reported study was supported by RFBR research project No. 16-01-00779 A.

References

1. Biberdorf E. A. et al. (2012) Global modeling of the human arterial system. Circulatory System and Arterial Hypertension: Experimental Investigation, Mathematical and Computer Simulation / Eds. L. N. Ivanova A. L. et al. *Nova Science Publishers, Inc*. PP. 115-142.
2. Melnikov V. N. et al. (2012) Elasticity of the arteries and hemodynamics in healthy and sick people: the database is registered in the Registry database of the Federal Service for Intellectual Property at June 9, 2012, number 2012620540.

HAPLOID EVOLUTIONARY CONSTRUCTOR 3D: A FRAMEWORK FOR MULTILAYER MODELING OF SPATIALLY DISTRIBUTED MICROBIAL COMMUNITIES

A.I. Klimenko, Yu.G. Matushkin, Z.S. Mustafin, A.D. Chekantsev, R.K. Zudin, S.A. Lashin*

Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
Novosibirsk State University, Novosibirsk, Russia
** Corresponding author: lashin@bionet.nsc.ru*

Motivation and Aim: Next-generation ecological modeling approaches are based on the principles of emergence, predictability and structural realism [1]. Nowadays it becomes clear that to make reasonable ecological models it is necessary to take into account not only species interrelations but also their microevolution and organisms' response to environment heterogeneity and its dynamic nature. It also tends to use functional types rather than taxonomical units in ecological models. We find microbial communities to be promising objects to manifest the power of such next-generation approaches. On the one hand, a community is a strictly localized spatially distributed structure where spatial localization of cells of different functional types determines the variety its metabolic functions. On the other hand, microevolution of microbes runs relatively fast and may be observed in both natural and experimental conditions [2].

Methods and Algorithms: In this study, we present a software package Haploid Evolutionary Constructor 3D (HEC 3D) [3] designed for modeling and simulating spatially distributed multispecies microbial communities. The HEC 3D specifies a model on several layers of biological organization, namely, on genetic, metabolic, cellular, population, and ecological ones. Using HEC 3D one may combine various mathematical modeling approaches (agent-based modeling, differential equations, automata etc.) in one model to simulate such processes as cellular chemotaxis, substrates flow and diffusion, mutations, horizontal gene transfer, gene regulation, reproduction and metabolism. We provide the import of SBML models into the HEC 3D model (metabolism layer) as well as the graphical user interface and high-performance calculations on various platforms.

Results: Several already published HEC 3D models (for example, in [3]) show the close interrelations between evolutionary and ecological processes occurred in microbial communities, where spatial organization of a community may predetermine the evolutionary scenarios manifesting during its whole lifetime.

Conclusion: HEC 3D allows building comprehensive models of microbial communities taking into account both high-level and low-level processes. We hope that this tool will contribute to the development of general theory in this field and understanding the underlying mechanisms.

Availability: http://evol-constructor.bionet.nsc.ru

Acknowledgements: The study has been partially funded by the RFBR grant 150703879 and Budget Project 0324-2015-0003.

References:

1. V.Grimm, U.Berger. (2016) Structural realism, emergence, and predictions in next-generation ecological modelling: Synthesis from a special issue, Ecological modelling, 326: 177-187.
2. R.Kassen. (2014) Experimental evolution and the nature of biodiversity (Roberts).
3. A.I. Klimenko et al. (2016) Bacteriophages affect evolution of bacterial communities in spatially distributed habitats: a simulation study, BMC Microbiology, 16: 31.

CHAOS AND HYPERCHAOS IN THE ALTERNATIVE SPLICING MODEL

V.V. Kogai[1,2]*, V.A. Likhoshvai[2,3], S.I. Fadeev[1,2], T.M. Khlebodarova[3]
[1]*Sobolev Institute of Mathematics SB RAS, Novosibirsk, Russia*
[2]*Novosibirsk State University, Novosibirsk, Russia*
[3]*Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
** Corresponding author: kogai@math.nsc.ru*

Key words: *modeling, gene networks, alternative splicing, deterministic chaos, hyperchaos, Feigenbaum's, intermittency and quasiperiodicity route to chaos*

Motivation and Aim: Establishing the relationship between the structural and functional organization of genetic networks and its dynamic properties is a fundamental problem of mathematical biology of gene. The aim is to study the methods of mathematical modeling of dynamic properties of the genetic system of alternative splicing, which is one of the basic formation mechanisms of protein diversity in eukaryotes.

Methods and Algorithms: We use standard techniques as Lyapunov exponents, Poincare maps and instability with respect to initial data.

Results: Developed a mathematical model of a genetic system constituting a single transcription factor-encoding gene self-regulated by a feedback loop that involves protein isoforms. Alternative splicing results in the synthesis of protein isoforms providing opposite regulatory outcomes — activation or repression.

$$\frac{dp(t)}{dt} = k_s \frac{\delta_0 + \delta_1 (\frac{p(t-\tau_1)}{K})^h}{(1 + (\frac{p(t-\tau_1)}{K})^h + (\frac{p(t-\tau_2)}{K})^h)^2} - k_d\, p(t),$$

$p(t-\tau_1)$ and $p(t-\tau_2)$ are the concentration of the transcription factor isoforms — autoactivator and autorepressor.

We investigated the conditions of transition to chaos in this model of alternative splicing consisting of one differential equation with delayed parameters. It was found that model exhibits hyperchaos with two positive Lyapunov exponents and all basic scenarios of transition to chaos: through a cascade of period-doubling bifurcations, through quasiperiodicity and through type-I, type-II and type-III intermittency. It should be noted that the values of the parameters of transition to chaos in all scenarios investigated are within the physiological limits of live systems functioning.

Conclusion: It has been theoretically proved the possibility of transition to chaotic dynamics in the genetic system of alternative splicing based on real biological prototype. Our model for the time being is the only example of biological system demonstrating all classical routes to chaotic dynamics as well as the possibility of transition from chaos to hyperchaos.

Acknowledgements: The work was partially supported by the RFBR (No 16-01-00237a) and Project Fundamental Research of SB RAS (No 0324-2015-0003).

A MULTIDIMENSIONAL APPROACH TO PERRSONALITY TRAITS ASSESSMENT FOR PSYCHOMETRIC EXAMINATIONS

N.Yu. Kolomenskii
Republican Psychiatric Hospital, Republic of Karelia, Russia
Corresponding author: n.c@nm.ru

Key words: *multidimensional approach, multiaxial approach, psychodiagnostics, psychometry test, MMPI, personality traits, personality disorders*

Motivation and Aim: The Diagnostic and Statistical Manual of Mental Disorders (DSM) uses multiaxial or multidimensional approach to psychiatric diagnostics. However, personality disorders historically are viewed in a categorical classification. The most widely used phychodiagnostic test, MMPI (Minnesota Multiphasic Personality Inventory), is based on clinical scales in accordance to a "single-dimension" model. In the recent, fifth edition of the DSM, an alternative hybrid categorical-dimensional model for personality disorders is included to stimulate further research. On practice, personality traits assessment plays a significant role in psychiatry diagnostics, but present methods lack exactness. For example, in forensic psychiatry patient's fate may depend on the result of such method. This work proposes a new approach to process data, received from psychometric examination. I.e., the aim is to improve existing methods of psychometrics and to elaborate new ones.

Methods and Algorithms: In this work human personality is presented as a figure in a multidimensional space, one dimension representing one personality trait. Each subtest of psychodiagnostic examination estimates the expression of some trait. In MMPI test human personality is figured as a broken line in 2-dimensional space, and accentuation is diagnosed if the line goes beyond some level of "normality". For certain reasons this approach lacks accuracy. In multidimensional space the level of "normality" becomes a multidimensional sphere, and certain personality disorders are pictured as areas in the space beyond this sphere. The multidimensional figure that depicts patient's personality may situate within one or more of those areas, which represents accentuation of some traits. Meanwhile the surface area of that figure is the combined index representing total "harmonicity" of patient's psyche. This representation solves some theoretical difficulties concerning advantages and disadvantages of use of clinical-based scales in the MMPI-2 vs use of "atheoretical" scales in the old MMPI.

Results: The work proposes a new approach to assess personality traits which resolves some theoretical problems and may make the diagnostics of personality disorders more precise. The reliability of this method is being implemented and tested on virtual models.

IT ANALYSIS OF CORNEA ENDOTHELIUM TRANSPORT ABILITY IN CORNEAL TRANSPLANTS AFTER HYPOTHERMIC CONSERVATION

A.A. Konev[1], I.G. Palchikova[1], I.A. Iskakov[2], L.E. Katkova[3], G.S. Baturina[3], E.I. Solenov[3]*

[1]*Technological Design Institute of Scientific Instrument Engineering SB RAS*
[2]*Multidisciplinary Science and Technology Complex "Eye Microsurgery" named after S.N. Fyodorov Federal State Institution, Novosibirsk Branch, Novosibirsk, Russia*
[3]*Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
** Corresponding author: eugsol@bionet.nsc.ru*

Key words: *analitic software; cornea endothelium transport; hypothermic conservation*

Motivation and Aim: Despite advances in surgery, corneal edema is a usual complication after cornea transplantation. Removing of extra fluid from the swelled cornea tissue is function of corneal endothelium. Little is known about the influence of hypothermic conservation on ion and water transport in corneal endothelial cells. Cell transport mechanisms that provide cell osmotic balance are determined, in a great extent, by Na,K-ATPase pumping capacity. The intracellular concentration of sodium ($[Na^+]_i$) is result of balance between sodium entering the cell and outward current determined by Na,K-ATPase. The main purpose of hypothermic conservation is to save transport ability of endothelial cells in corneal transplants. The methods of transplants conservation are subject of intense studies nowadays. The aim of the present work is development of effective method for analysis of corneal endothelial cells transport function by creating and using experimental protocol for measuring of $[Na^+]_i$ and analytical software for automatic computer analysis of experimental data.

Methods and Algorithms: Intracellular sodium concentration was measured using as specific intracellular fluorescent probe for sodium Sodium Green (MolProbes, USA) by the method developed in the group of cell molecular physiology (Institute of Cytology and Genetics SB RAS)(1). Microscopic fluorescent images were captured with AxioCam HSm (Zeiss, Germany) and stored on PC. The intensity of cell fluorescence was evaluated by analysis of microscopic images of cells using software developed in TDISIE, SB RAS.: program "CytoDynamics" (state registration number 2016612766).

Conclusions: The kinetics of intracellular sodium concentration was measured in endothelial cells from isolated eyes of rat, pig and human. Experimental protocol was developed by the group of cell molecular physiology (Institute of Cytology and Genetics SB RAS). The software and user friendly interface was developed for automatic outline fluorescent images of cells as AOI and analysis of fluorescence intensity. The software was registered as "CytoDynamics" (state register number 2016612766). The program analyses consequences of microscopic images of living cells that undergone experimental protocol and calculates the values of $[Na^+]_i$. The results of calculations reflect the dynamics of $[Na^+]_i$ in individual cells. Program uses libraries OpenCV / JavaCV, it accepts colors or monochrome images and presents the results as Excel tables and plots.

Complex approach includes experimental protocol for experimental measurement of $[Na^+]_i$ in endothelial cells and analytical software enables analyze large scale experimental data sets. The program "CytoDynamics" could be useful as analytical tool for transplantology and drug discovery.

References:

1. Solenov EI. Cell volume and sodium content in rat kidney collecting duct principal cells during hypotonic shock. J Biophys. 2008;2008:420963. doi: 10.1155/2008/420963.

AN INVERSE PROBLEM FOR A SYSTEM WITH A SMALL PARAMETER IN KINETICS MODELS

L.I. Kononenko* [1,2]
[1]*Sobolev Institute of Mathematics SB RAS, Novosibirsk, Russia*
[2]*Novosibirsk State University, Novosibirsk, Russia*
** Corresponding author: volok@math.nsc.ru*

Key words: *mathematical modeling, singularly perturbed system, integral manifold, slow surface, inverse problem*

Motivation and Aim: Inverse problems are very important for mathematical modeling of biochemical kinetics and for its biological applications.

Methods and Algorithms: Our approach is based on the method of integral manifolds and its modification to the applied problems [1, 2].

Results: A system of ordinary differential equations modeling some catalytic reactions involves slow and fast variables, and it can be regarded as a singularly perturbed system of the form

$x' = f(x, y, t,\varepsilon)$, $\varepsilon y' = g(x, y, t,\varepsilon)$, (1)

for $x \in R^m$ and $y \in R^n$ with the time derivatives x' and y', an infinitely small positive number ε, and some sufficiently smooth functions f and g.

We study an inverse problem for the system (1) under the following conditions:

1) the case of degenerate system $\varepsilon=0$ is considered;

2) the right-hand side f of the slow sub-system is a linear function $f = \sum_{i+j\le p} b_{ij} x^i y^j$, $p=1$;

3) the system (1) has one slow and one fast variable $m=n=1$;

4) the function $g(x, y, t,\varepsilon)$ is known and satisfies the implicit function theorem in the domain, in particular, $\frac{\partial}{\partial y} g(x, y, t, 0) \neq 0$, and thus the slow variable surface $g(x, y, t, 0) = 0$ is known;

5) the slow surface has only one sheet.

Under these conditions consider the following dynamical system:

$x' = p_0 + p_1 x + p_2 y;\ g(x, y, t, 0) = 0;$ (2)

$x \in R,\ y \in R,\ t \in R,\ p_i \in R\ i=0,1,2,\ x(t_i) = \alpha_i,\ x'(t_i) = \beta_i\ \ i=1,2,3,$

where the function g satisfies conditions 4), 5). Given t_i, α_i, β_i we find unknown coefficients p_0, p_1, p_2 which satisfy the system (2).

Proposition. If the conditions 1) - 5) are satisfied for t_i, α_i, β_i $i=1,2,3$, and determinant of the system (2) is non-zero, then the inverse problem has a unique solution.

Conclusion: Existence and uniqueness conditions can be obtained for the coefficients of polynomial in the right-hand part of the slow well-conditioned subsystem when some data are given on the slow surface.

Acknowledgements: Supported by RFBR, grant 15-01-00745.

References:

1. V. M. Gol'dshtein and V. A. Sobolev, (1988) *Qualitative Analysis of Singularly Perturbed Systems*. Sobolev Institute of Mathematics, Novosibirsk. (Russian)
2. L.I.Kononenko. (2016) An identification problem for singular systems with a small parameter in chemical kinetics. *Siberian Electron. Math. Reports*, **13**, 175-180. (Russian)

PERSONALIZED SIMULATION BASED ON THE MODIFIED ANALYTICAL MODEL OF THE LEFT VENTRICLE OF THE HUMAN HEART

A.A. Koshelev[1,2]*, A.E. Bazhutina[1], K.S. Ushenin[1,3]
[1]*Ural Federal University, Ekaterinburg, Russia*
[2]*Institute of Mathematics and Mechanics UB RAS, Ekaterinburg, Russia*
[3]*Institute of Immunology and Physiology UB RAS, Ekaterinburg, Russia*
** Corresponding author: Anton.Koshelev@urfu.ru*

Key words: *Ventricular myocardium, myofibres direction field, mathematical anatomy, analytical cardiac model, personalized modeling, heart simulation*

Motivation and Aim: Cardiovascular diseases are ones of the leading causes of worldwide mortality. Mathematical models of the heart and corresponding computer tools are becoming widely used to elucidate mechanisms of the diseases. Progress in computer hardware made it possible to perform simulations of the physiological functions of whole organs, including the heart [1, 2]. The LV is the largest and most powerful of the four chambers of a mammalian heart including that of humans. It pumps arterial blood with the large pressure into the aorta where the blood comes to all organs and tissues of the body.

Methods and Algorithms: To take into account all individual features we use the personalized or individual map approaches to construct the shape of the left ventricle and to define myofibres direction field. They are essential for dynamics of the electrical waves (e.g. spiral waves, see [3]) and following LV contraction. For these we use spline modification of the analytical model of the left ventricle of the mammal heart [4, 5].

Results: For the model verification, we used DT-MRI data of a human heart [6]. From these data we generate likeness of echocardiography two projections of the left ventricle. For these two projections we construct a shape and myofibres direction field. The average of vector fields deviation is 30.5 °, mean square deviation is 19.7 °.

Conclusion: The modified model approximates the left ventricle shape and the myofibres direction field with sufficient accuracy in the case of a limited data or bad quality.

Acknowledgements: supported by the Russian Science Foundation (grant no.14-35-00005).

References:

1. R.L.Winslow, N.Trayanova, D.Geman, M.I.Miller (2012) Computational medicine: translating models to clinical care, *Sci. transl. medicine*, 4(158): 158rv11-158rv11.
2. P.J.Hunter, W.W.Li, A. D.McCulloch, and D.Noble (2006) Multiscale modeling: Physiome project standards, tools, and databases, *Computer*, (11): 48-54.
3. S.F.Pravdin, H.Dierckx, L.B.Katsnelson, O.Solovyova, V.S.Markhasin, A.V.Panfilov, (2014) Electrical wave propagation in an anisotropic model of the left ventricle based on analytical description of cardiac architecture, *PLOS one*, 9(5): e93617.
4. S.F.Pravdin, V.I.Berdyshev, A.V.Panfilov, L.B.Katsnelson, O.Solovyova, V.S.Markhasin, (2013) Mathematical model of the anatomy and fibre orientation field of the left ventricle of the heart, *Biomedical engineering online*, 54(12): 21.
5. S.Pravdin, (2013) Nonaxisymmetric mathematical model of the cardiac left ventricle anatomy, *Russian Journal of Biomechanics*, 17(4): 75-94.
6. Diffusion tensor magnetic resonance imaging dataset of a human heart http://gforge.icm.jhu.edu/gf/project/dtmri_data_sets/docman/?subdir=93

INVERSE PROBLEMS OF POPULATION DYNAMICS

A.I. Kozhanov*, Yu.A. Kosheleva
Sobolev Institute of Mathematics SB RAS, Novosibirsk, Russia
Sakhalin State University, Yuzhno-Sakhalinsk, Russia
** Corresponding author: kozhanov@math.nsc.ru*

Key words: *ultra parabolic equations, inverse problems, problems of population dynamics*

Motivations and Aims: Under study are the problems of population dynamics whose primarily feature is that the life activity of a particular specie is determined by the astronomic time t and the age of the specie. This means mathematically that the population dynamics is described by ultraparabolic equations. If the mathematical model has as unknown not only density but also the coefficient at a solution (mortality) or some external impact, then the problem under study is an inverse problem of population dynamics.

Results: This talk presents a series of results on solvability of inverse problems of population dynamics or, in other words, a series of solvability results on inverse problems for ultraparabolic equations.

TWO MODELS OF THE DROSOPHILA GAP GENE NETWORK WITH VARIATION OF MATHERNAL INPUT

K.N. Kozlov[1]*, A.V. Svichkarev[1], V.V. Gursky[1,2], I.V. Kulakovskiy[3], S.Y. Surkova[1], M.G. Samsonova[1]
[1]*Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia*
[2]*Ioffe Institute, St. Petersburg, Russia*
[3]*Engelhardt Institute of Molecular Biology, RAS, Moscow, Russia*
** Corresponding author: kozlov_kn@spbstu.ru*

Key words: *morphogen, transcription, thermodynamics, reaction-diffusion, drosophila*

Motivation and Aim: Bicoid (Bcd) as a classic morphogen plays a key role in the expression of segmentation genes in Drosophila embryo. The segmentation gene network controls the determination of the fruit fly segments during the blastoderm stage, just before the onset of gastrulation. We applied the systems-level approach to understand the spatial dynamics of gap gene expression domains under different Bicoid dosages.

Methods and Algorithms: We considered two modeling approaches. The first phenomenological model [1] uses the reaction-diffusion differential equations and the matrix of regulatory coefficients characterizing the action of regulators on their targets. The second sequence-based model [2, 3] uses thermodynamic approach to model target gene expression at the RNA level and two sets of reaction-diffusion differential equations for mRNA and protein concentrations to describe the dynamics of the system. We model the expression of 4 gap genes – *hb*, *Kr*, *gt*, and *kni* – under control of 11 transcription factors (TF) – the products of *hb*, *Kr*, *gt*, *kni*, *bcd*, *tll*, *cad*, *hkb*, *cic*, *slp*, and *run* genes. We predicted TF binding sites in the potential regulatory region from 12Kbp upstream to 6Kbp downstream of transcription start site for each gene using enhanced dinucleotide positional weight matrices. The unknown model parameters were obtained with the DEEP method by fitting the model solutions to both expression patterns of gap genes and data on the *hb* anterior domain shifts in embryos with varying Bcd concentration.

Results: Both models successfully reproduce the characteristic features of experimental data. The sequence-based model reproduces the spatial dynamics of the *hb* anterior expression domain more precisely.

Availability: The software is available from authors upon request.

Acknowledgments: The study was supported by the RFBR grant 14-04-01522 and the "5-100-2020" Program of the Ministry of Education and Science of the Russian Federation.

References:

1. Andreev, S. A., Samsonova, M. G., and Gursky, V. V. (2015). Modeling of the Drosophila gap-gene network with the variation of the Bcd morphogen. Biophysics, 60(2), 173–180.
2. K. Kozlov, V. Gursky, I. Kulakovskiy, A. Dymova and M. Samsonova, Analysis of functional importance of binding sites in the Drosophila gap gene network model. BMC Genomics 2015, 16(Suppl 13):S7
3. K. Kozlov, V. Gursky, I. Kulakovskiy, M. Samsonova (2014). Sequence-Based Model of Gap Gene Regulatory Network. BMC Genomics, 15(Suppl 12):S6

GENELO – PROGRAM FOR STATISTICAL ANALYSIS OF GENES LOCATION RELATIVE TO CHROMOSOME CONTACTS REVEALED BY CHIA-PET AND HI-C

E.V. Kulakova*, A.M. Spitsina
[1]*Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2]*Novosibirsk State University, Novosibirsk, Russia*
** Corresponding author: kylakovaekaterina@gmail.com*

Key words: *sequencing, ChIA-PET, Hi-C, chromosome contacts, genome, CTCF sites*

Motivation and Aim: Several technologies based on chromatin immunoprecipitation (ChIP) have been developed to study the binding of transcription factors (TF) to genomic DNA including microarray (ChIP-chip), ChIP-PET and ChIP-Seq [1]. The challenge is to define whether such distal binding sites are functional, i.e. physically proximal to target gene promoters via chromosome loops attracting RNA polymerase II complex for gene transcription [2]. Chromatin Interaction Analysis with Paired-End-Tag sequencing (ChIA-PET) method fits these demands still requiring development of specialized high-throughput software for data integration [2]. The aim of the work was to develop a computer program for statistical data analysis and test it on CTCF (CCCTC-binding factor) binding sites, genes and spatial topological domains.

Methods and Algorithms: We used data on the location of CTCF binding sites clusters obtained by ChIA-PET as well as obtained experimentally by methods Hi-C, ChIA-PET [2]. Gene annotation was obtained from UCSC Genome Browser (http://genome.ucsc.edu).

Results: In result has been developed computer software for statistical analysis and visualization of results for experimental data obtained by ChIA-PET and Hi-C. The program was developed in Java language that calls modules based on R and Matlab environment using library such as Rserve and MatlabControl. The program has graphical user interface. This tool has function such as identification gene location near to domains boundary; near to binding sites of transcription factor; visualization of heatmap based on pairs of CTCF binding sites; distributions of human genes relative CTCF binding sites and a randomly generated list of such sites.

Conclusion: We considered a model the location of genes relative chromosome loops and binding sites. Genes of RefSeq are located inside the loop between the sites accounted for half of the total. It was revealed that most of the genes in the chromosomal loops are arranged individually

Availability: Software is available from the author upon request

Acknowledgements: The work is supported by ICG budget project 0324-2015-0003.

References:

1. E.V. Ignatieva et al. (2015) Regulatory Genomics: Combined Experimental and Computational Approaches, *Russian J. Genetics*, V. 51, 4: 334–352.
2. G. Li et al. (2012) Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation, *Cell*, V.148. 1-2: 84-98.

REGULARIZATION METHODS IN DETERMINATION OF BIOLOGICAL MOLECULE FORCE FIELDS

G.M. Kuramshina*, A.Ya. Korneichuk, S.A. Sharapova
Faculty of Chemistry, Department of Physical Chemistry, Moscow State University, Moscow, Russia
** Corresponding author: kuramshi@phys.chem.msu.ru*

Key words: *regularizing algorithms, inverse vibrational problem, molecular force field model*

Motivation and Aim: Fast growing computational resources leads to the appearance of new methods of quantum chemistry for solving many problems of structural chemistry in application to the large biological systems. But there are also existed obvious severe limitations of using purely *ab initio* methods for the analysis of molecular systems consisting of a few hundred atoms. In cases where such systems are organized from separate smaller size units the most successful approaches for their analysis are based on the joint use of theoretical results with some empirical approach and in many cases result in good descriptions of investigated systems.

Methods and Algorithms: Inverse vibrational problem is a problem of finding the molecular force field parameters on a base of experimental data. This problem can be written in a form of nonlinear operator equation

$$A_h F = \Lambda_\delta \,, \qquad (1)$$

where $F \in Z \in R^{n(n+1)/2}$ (Z is a set of possible solutions) is the unknown force constant matrix (real and symmetrical), $\Lambda \in R^m$ represents the set of available experimental data (vibrational frequencies, etc.) determined within δ errorlevel: $||\Lambda - \Lambda_\delta|| \le \delta$. A is a nonlinear operator which maps matrix F on the Λ, h is an error of operator A: $||A - A_h|| \le h$. This problem relates to the class of nonlinear ill-posed problems (it does not satisfy any of three the well-posedness conditions by Hadamard).

Results: For overcoming the ill-posedness of inverse vibrational problem the quantum mechanical data are served as an additional information and as a stabilizer within the regularization procedure while looking for the Regularized Quantum Mechanical Force Field (RQMFF) [1]. Quantum mechanical data also provide a priori information on molecular force field model.

Conclusion: As the result we obtain the regularized quantum mechanical force field (RSQMFF) in different set of generalized coordinates including the scaled Cartesian ones [2] which are convenient in applications to biological molecules. Examples of application of this approach to bulky systems are presented.

Acknowledgements: This work was partly supported by the RFBR grant No 14-03-00929a.

References:

1. G.M.Kuramshina, F.A.Weinhold, I.V.Kochikov, Yu.A.Pentin, A.G.Yagola (1994) Joint treatment of ab initio and experimental data in molecular force field calculations with Tikhonov's method of regularization, J. Chem. Phys., 100(2): 1414-1424.
2. A.G.Yagola, I.V.Kochikov, G.M.Kuramshina, Yu.A.Pentin (1999) *Inverse problems of Vibrational Spectroscopy*. VSP. Zeist, The Netherlands.
3. I.V.Kochikov,G.M.Kuramshina,A.V.Stepanova (2009) New approach for the correction of Ab initio molecular force fields in Cartesian coordinates, Int. J. Quant.Chem., 109 : 28-33.

POPULATION-BASED MATHEMATICAL MODELING OF HUMAN IMMUNOGLOBULIN G *N*-GLYCOSYLATION

E. Kutumova[1,2]*, I. Yevshin[1,2], E. Basmanova[1,2,3], N. Mandrik[1,2,4], R. Sharipov[1,2,3], F. Kolpakov[1,2]

[1]*Institute of Systems Biology Ltd., Novosibirsk, Russia*
[2]*Design Technological Institute of Digital Techniques, Novosibirsk, Russia*
[3]*Novosibirsk State University, Novosibirsk, Russia*
[4]*Sobolev Institute of Mathematics SB RAS, Novosibirsk, Russia*
** Corresponding author: elena.kutumova@biouml.org*

Key words: *population-based modeling, immunoglobulin G, N-glycosylation, parameter fitting, genome-wide association study*

Motivation and Aim: *N*-linked glycosylation of immunoglobulin G (IgG) is the important process that plays a fundamental role in a range of human pathologies including autoimmune and inflammatory diseases, as well as rheumatoid arthritis. The formal description of *N*-glycosylation was significantly enriched for the last years and a range of mathematical models considering consequent formation of sugar molecule oligosaccharides known as glycans was created. Relative values of IgG glycan levels in individual human plasma proteins could be detected by ultra performance liquid chromatography (UPLC), whereas genome-wide association study (GWAS) of these values in human populations could be used to identify new genetic loci possibly implicated in the glycosylation process. Thus, the problem of GWAS performed for relative concentrations of enzymes that appear in the mathematical model of the IgG *N*-glycosylation and catalyze reactions of the glycan formation became actual.

Results: In this work, we have progressed from experimental studies of IgG *N*-glycosylation to the population-based mathematical modeling and *in silico* studies of this process. We developed a mathematical model of IgG *N*-glycosylation on the base of data from the KEGG database and models constructed by Krambeck *et al*. [1]. The model includes 45 glycans from the human IgG glycome, 7 enzymes and 106 biochemical reactions occurring in the Golgi apparatus represented by four compartments: *cis*-, *medial*- and *trans*-cisterns, and *trans*-network. To train the model parameters we used quantitative UPLC data of IgG *N*-glycosylation obtained for 1723 individuals from two European populations [2]. The data represent relative values of glycan levels in 22 chromatographic peaks generated for each individual. To approximate these data, we defined 28 model parameters reflecting concentrations of seven main enzymes catalyzing formation of glycans in four Golgi compartments. We performed consecutive reduction of these parameters to decrease the risk of the model overfitting and received a set containing 13 parameters adjusted for each individual. Then we searched for genomic associations with enzyme concentrations using previously published genome-wide single-nucleotide polymorphism (SNP) data and fitted model parameters. As a result, we detected four known loci (*FUT8*, *MGAT3*, *ST6GAL1*, and *SMARCB1-DERL3*) provided by only few traits, while GWAS performed for raw values of glycan peaks did not give clear associations with the traits.

References:

1. F.J. Krambeck, et al. (2009) A mathematical model to derive N-glycan structures and cellular enzyme activities from mass spectrometric data, Glycobiology, 19(11): 1163–1175.
2. G. Lauc, et al. (2013) Loci Associated with *N*-Glycosylation of Human Immunoglobulin G Show Pleiotropy with Autoimmune Diseases and Haematological Cancers. PLoS Genet 9(1): e1003225. doi:10.1371/journal.pgen.1003225.

STOCHASTIC AND GRADIENT APPROACHES FOR SOLVING OF THE INVERSE PROBLEM FOR BASIC MATHEMATICAL MODEL OF INFECTIOUS DISEASE WITH DELAY

V. Latyshenko[1]*, O. Krivorotko[1,2], S. Kabanikhin[1,2]
[1]*Novosibirsk State University, Novosibirsk, Russia*
[2]*Institute of Computational Mathematics and Mathematical Geophysics SB RAS, Novosibirsk, Russia*
** Corresponding author: Latushenko_varia@mail.ru*

Key words: *basic mathematical model of infectious disease, inverse problem, parameter identification, genetic algorithm, gradient method, optimization approach, regularization*

Motivation and Aim: Immune response characteristics are individual for specific pathogen. In mathematical models these characteristics describe coefficients in systems of nonlinear ordinary differential equations (ODE). It is necessary to identify the specific parameters so called inverse problem for constructing an individual treatment plan. The aim of work is development and validation a numerical algorithm for solving an inverse problem for ODE using additional measurements of system states at fixed times.

Methods and Algorithms: The paper focuses on numerical investigation of the inverse problem for the basic mathematical model of infectious disease developed by G.I. Marchuk [1]. This model describes dynamic of concentrations of pathogenic antigens, antibodies, plasma cells and relative characteristic of affected organ. The direct problem is solved by Runge-Kutta-Felberg method. Inverse problem consists in determining of parameter vector $q \in R^{10}$ using additional measurements at fixed time points t_k, $k = 1,2,...,K$, that could be defined by blood test, urinalysis, MRI, *etc*. Inverse problem reduces to the minimization of the misfit function [2]. The solution of minimization problem is identified by the combination of two methods, *i.e*. genetic algorithm and gradient method of minimal errors [3]. Firstly, genetic algorithm finds the global minima domain. Then the gradient approach determines the inverse problem solution *q* using the regularization technique. The evident formula of the gradient of the misfit function based on an adjoint problem solution is obtained.

Results: A combination of stochastic and deterministic methods restores the parameters of the model quite well. It helps one to predict the course of disease.

Conclusion: A combined numerical algorithm for parameter identification in basic mathematical model of infectious disease using measurements of system states in fixed times is constructed and validated.

Acknowledgments: This work is supported by the Ministry of Education and Science of the Russian Federation and the Russian Foundation of Basic Research No. 16-31-00189.

References:

1. G.I. Marchuk. Mathematical Modeling of Immune Response in Infectious Diseases. Dordrecht, Springer, 1997.
2. S.I. Kabanikhin. Inverse and Ill-Posed Problems: Theory and Applications. Berlin, de Gruyter, 2011.
3. A.I. Il'in, S.I. Kabanikhin, O.I. Krivorotko. On the determination of the model parameters of systems of nonlinear differential equations, Siberian Electronic Mathematical Reports. – 2014. - V. 11. – P. C62-C76.

ON STAPP'S APPROACH TO THE MIND-MATTER PROBLEM: AN ATTEMPT TO INCORPORATE IT INTO THE DLF-MODEL

A.V. Levichev[1]*, A.Yu. Palyanov[2]
[1]*Sobolev Institute of Mathematics SB RAS, Novosibirsk, Russia,*
[2]*A.P. Ershov Institute of Informatics Systems SB RAS, Novosibirsk, Russia*
** Corresponding author: alevichev@gmail.com*

Key words: *psychophysical events and quantum mechanics, Penrose-Hameroff model of consciousness, DLF-theory, parallelizations of space-times bundles*

Motivation and Aim: In [1], on the basis of the D-component of DLF-theory (see [2]), we suggested a modification of the theoretical basis of the Penrose-Hameroff model of consciousness. Detailed computations of [3] indicate that our modification turned out to be quite a working one. As a follow-up, we now try to incorporate the fundamental notion of a *psychophysical* (rather than purely *physical*) *event* into a rigorous model (which necessarily involves quantum mechanics, QM). Here are the basics of Stapp's approach (see [4]) to the mind-matter problem. These dual-aspect psychophysical events are now the basic entities. The psychological description and the physical description specify two aspects (sides) of a single event-type entity. The conscious self is a stream of conscious events. These events are the psychologically described aspects of a sequence of psychophysical events whose physical aspects are a sequence of physical events in a single brain. Mental process is to be understood in terms of this richer dualistic ontological base, rather than the impoverished purely physical part that survives contraction to the classical approximation. The first author believes (see [5]) that it is the presence of the (space-time, ST) component F in the DLF-triad which may allow to adequately describe psychological part of person's stream of conscious experiences (as shown by Segal [6], the D-component models the physical part). It is worth reminding that the main idea of the DLF-approach to model particles and interactions is that there are three Hamiltonians (the "Russian Troika", see [2]) to drive evolution of a physical system. The aim of our current study is to (somehow) circumvent singularities (they are unavoidable in the F-component whereas they are absent in the D-component) and to develop methods of the F-based analysis of ST-bundles. Here is a purely mathematical reason for such a research: of all four-dimensional Lie algebras (there are infinitely many such algebras) only u(2) and u(1,1) are the reductive ones.

Methods and Algorithms: The QM-approach to consciousness necessarily involves the notion of a parallelization of ST-bundles (elements of which are wave functions). Segal's theory (see [6, 7]) exploits D, the Lie group U(2), whereas F=U(1,1). Matrix Lie groups G_D, G_F are certain representations of SU(2,2). G_D acts globally on D by linear-fractional transformations, whereas there are singularities of the linear-fractional G_F-action on F. The main new constructs describe properties of finite and of infinitesimal G_F-actions on F. The data/results are presented similarly to how it has been done in [7] for the D-case. In particular, we keep the tables' numeration (there are ten tables in [7] which present important vector fields, their actions, and other parallelization-related data): for each D-table its F-counterpart has to be determined.

Results: So far, we have determined Tables I, III и IV. Overall, Table IV (having in mind part of the Table I initial data) can be viewed as the starting point in order to determine all other necessary information about the infinitesimal G_F-actions on F. Our lines of reasoning are based on the linear-fractional action. Hence, they can be applied to both D- and F-cases. Notice that [6] does not provide much details on how the data has been determined.

Conclusion: The obtained results (together with some earlier findings) provide mathematical ground which allows to start the analysis of ST-bundles on the basis of the F=U(1,1) parallelization group. Such an analysis and its correlation with the analysis of ST-bundles on the basis of the D=U(2) parallelization group are of utmost interest. In particular, we hope to incorporate Stapp's approach to the mind-matter problem into the DLF-model.

References:

1. A. Levichev, A. Palyanov (2014) On a modification of the theoretical basis of the Penrose-Hameroff model of consciousness. International Conference MM-HPC-BBB-2014, Abstracts, p.49.
2. A.V. Levichev (2011) Pseudo-Hermitian realization of the Minkowski world through DLF theory, *Physica Scripta,* **83(1)**: 1-9.
3. A. Levichev and A. Palyanov (2015) On Separation between Metric Observers in Segal's Compact Cosmos. Journal of Modern Physics, **6**: 2040-2049.
4. H.P. Stapp (2008) Philosophy of Mind and the Problem of Free Will in the Light of Quantum Mechanics. http://arxiv.org/ftp/arxiv/papers/0805/0805.0116.pdf
5. A. V. Levichev (2015) New trends in space-time problematic: the DLF-model and its possible applications. *Proceedings of all-Russia conference with the international participation "Knowledge-Ontology-Theories"*, 201-206.
6. Segal Archive, MIT, http://math.mit.edu/segal-archive/Segalrefs3-9-08.pdf
7. S.M. Paneitz, I.E. Segal (1982). Analysis in space-time bundles I: General considerations and the scalar bundle. *Journal of Functional Analysis*, **47**: 78-142.

RULE-BASED MODELING IN BIOUML

N. Mandrik[1,2,3]*, E. Kutumova[1,2], F. Kolpakov[1,2]
[1]*Institute of Systems Biology Ltd., Novosibirsk, Russia*
[2]*Design Technological Institute of Digital Techniques SB RAS, Novosibirsk, Russia*
[3]*Sobolev Institute of Mathematics SB RAS Novosibirsk, Russia*
** Corresponding author: manikitos@biouml.org*

Key words: *rule-based modeling*

Motivation and Aim: The traditional approach to mathematical modeling of biological systems involves usage of nonlinear systems of ordinary differential equations (ODEs) with given initial conditions. Talking about the modeling, we emphasize the fact that we consider an abstraction and study the mathematical description of some qualitative and quantitative characteristics of biological processes. The level of detail is dependent on the problem and based on the knowledge of the researcher. On the one hand, many meaningful models consist of few non-linear equations. On the other hand, a detailed study of the biochemical networks leads to development of large-scale models consisting of hundreds of variables and, therefore, equations. Moreover, if we incorporate to the model site-specific details of protein-protein interactions, the number of protein modifications increases dramatically, and complexity of the model becomes combinatorial. For example, a protein comprising *n* amino acids can be potentially found in 2^n distinct phosphorylation states.

Investigation of such models using formalism of differential equations is difficult in view of the fact that we need to analyze thousands of variables whose values are often small. Visualization of the models (graphical representation of the reaction network as diagram) using one of the conventional standards (e.g., SBGN or KEGG) does not simplify the problem, although the diagram is easier to interpret than the corresponding system of equations, and readability of the diagram can be improved.

Methods and Algorithms: The main idea to deal with such models is based on representations of protein-protein interactions using rules serving as generators of species and biochemical reactions (or discrete events). This approach is known as «rule-based» modeling. Each rule describes a class of reactions with a common kinetic law and establishes the correspondence between reactant and product patterns defining a set of species with similar chemical compositions and properties.

Conclusion: The principles for creation of the «rule-based» models were implemented in several software resources including KaSim (http://dev.executableknowledge.org/) and BioNetGen (www.bionetgen.org). BioUML supports the BioNetGen language (BNGL) and a special graphical notation created on the basis of SBGN and uses it to visualize the «rule-based» models.

Availability: BioUML is available to download on www.biouml.org (free).

COMPUTATIONAL TOOLS FOR DATA PROCESSING OF MEDICAL IMAGING

An.G. Marchuk[1], F.P. Kapsargin[2], L. Cadena[3], M.A. Kurako*[3], K.V. Simonov[4]
[1]*Institute of Computational Mathematics and Mathematical Geophysics SB RAS, Novosibirsk, Russia*
[2]*Krasnoyarsk State Medical University, Krasnoyarsk, Russia*
[3]*Siberian Federal University, Krasnoyarsk, Russia*
[4]*Institute of Computational Modelling SB RAS, Krasnoyarsk, Russia*
** Corresponding author: mkurako@gmail.com*

Key words: *approximation algorithms, spectral decomposition, image processing, medical imaging*

Motivation and Aim: The paper offers the theoretical bases and algorithmic support for processing and analysis of medical imaging data (urology), signal and image processing.

The first group of tools is Fourier series based algorithms for data approximation. Combination of these algorithms allows us to construct experimental functions for any medical imaging signals and solve the data restoration and data extrapolation problems.

Other group of tools is algorithms for spectral decomposition and spatial clustering of spatial data with respect to an anisotropic features and geometric singularities of studied structures. These tools helps to extract fine-grained features of noisy data.

The last tool is a technique of detecting patterns in data based on nonlinear multivariable regression. Technique allows to get an adequate estimates of parameters of detected patterns with respect to "quality" of studied data.

Results: As a result, combination of existing medical imaging data processing algorithms and proposed algorithms enables us to increase the quality and effectiveness of medical imaging data modelling.

ON PROPERTIES OF SOLUTIONS TO SOME NONLINEAR SYSTEMS WITH PARAMETERS

I.I. Matveeva
Sobolev Institute of Mathematics SB RAS, Novosibirsk, Russia
Novosibirsk State University, Novosibirsk, Russia
** Corresponding author: matveeva@math.nsc.ru*

Key words: *systems of ordinary differential equations, Cauchy problem, large coefficients, estimates for solutions, limit theorems.*

Motivation and Aim: We consider a class of systems of nonlinear ordinary differential equations with parameters. We study properties of solutions to the systems and propose a method for approximate solving the systems in the case of very large coefficients.

Results: Consider the following system of nonlinear ordinary differential equations

$$\frac{dx_1}{dt} = -\frac{n-1}{\tau}\frac{x_1}{1+\rho x_1^{\gamma}} + g(t, x_n), \quad \frac{dx_j}{dt} = \frac{n-1}{\tau}\frac{x_{j-1}}{1+\rho x_{j-1}^{\gamma}} - \frac{n-1}{\tau}\frac{x_j}{1+\rho x_j^{\gamma}}, \quad j = 2,\ldots,n-1,$$
$$\frac{dx_n}{dt} = \frac{n-1}{\tau}\frac{x_{n-1}}{1+\rho x_{n-1}^{\gamma}} - \theta x_n, \qquad \tau,\ \rho,\ \gamma > 0,\ \theta \geq 0,\ g(t,u) \in C\left(R^2\right). \tag{1}$$

In the case of $\rho = 0$ the system is often termed the `Goodwin' model. Ordinary differential equations of such kind and more complicated equations containing Hill's type functions arise when modeling gene networks. In the case of, $n \gg 1$ in order to study effectively solutions to this system [1], one can use methods proposed by G.V. Demidenko for various systems of nonlinear ordinary differential equations of large dimension (for example, see [2, 3]). In the present paper we study properties of solutions to (1) for $\tau \ll 1$ when n is fixed. As result, in order to find approximately the last component $x_n(t,\tau)$ of the solutions, it is sufficient to solve one ordinary differential equation. This result gives us an effective method for approximate calculating $x_n(t,\tau)$. Moreover, the less τ, the more exact the method; i.e., the larger the coefficients of (1), the more exact the approximate solution. Thereby this method allows us to avoid difficulties arising inevitably when solving systems of nonlinear differential equations with very large coefficients.

Acknowledgements: The research was supported by the Russian Foundation for Basic Research (project no. 16-01-00592).

References:

1. I.I. Matveeva, I.A. Mel'nik. (2012) On the properties of solutions to a class of nonlinear systems of differential equations of large dimension, *Sib. Math. J.*, **53(2)**: 248-258.
2. G.V. Demidenko et al. (2006) On one class of systems of differential equations and on retarded equations, *Sib. Math. J.*, **47(1)**: 45-54.
3. G.V. Demidenko. (2012) Systems of differential equations of higher dimension and delay equations, *Sib. Math. J.*, **53(6)**: 1021-1028.

DEVELOPMENT OF DNA BARCODING WEB-SERVICE FOR ROBUST AND QUICK YERSINIA IDENTIFICATION IN LARGE-SCALE EPIDEMIOLOGICAL STUDY

A.U. Medvedev[1]*, D.A. Romashko[1], A.M. Stenkova[1,2], E.P. Bystritskaya[1,2], M.P. Isaeva[1,2]

[1]*Far Eastern Federal University, Vladivostok, Russia*

[2]*Pacific Institute of Bioorganic Chemistry, Vladivostok, Russia*

** Corresponding author: rf_alexmedvedev@mail.ru*

Key words: *gyrB, barcoding, Yersinia*

Motivation and Aim: The barcoding of Yersinia genus is important due to some Yersinia species are human pathogens. A full biochemical identification of Yersinia species and its biotyping is time-consuming (about 20-25 tests). Recently, the gyrB gene sequencing was proposed as an identification tool instead of 16S rDNA one, which cannot resolve several phylogenetic relationships. The gyrB region of 430 bp was chosen to develop Yersinia barcoding.

Methods and Algorithms: Our Database contains the 344 verified gyrB sequences of Yersinia species. Genetic distances between each sequence were estimated using Kimura's two-parameter model (K2P). For each sequence was calculated the average K2P distance to sequences inside its cluster and the minimal K2P distance to outside sequences. The data were used in SVM binary classification to find the best barcoding threshold value which separates distances inside a cluster from other distances. For classification of an unknown Yersinia sample our algorithm calculates its average K2P distance to all clusters and picks two nearest ($d1 < d2$). If $d1 <$ threshold $< d2$ then the sequence belongs to the first cluster. If $d1 <$ threshold and $d2 <$ threshold then the sequence falls between the selected clusters and we cannot classify this sample. If threshold $< d1$ and threshold $< d2$ then the sequence forms its own cluster. The cross-validation of our algorithm was performed with the leave-one-out method.

Results: The best threshold value for separating intra- and interspecific distances was 2.14%. This value provided classification accuracy of 99.4%, and cross-validation score of 98.2%.

Conclusions and Availability: Based on these results, we developed the public web-service which allows identifying any Yersinia species. This service is suitable for routine, large-scale Yersinia screening.

Acknowledgements: This work supported by the Far Eastern Federal University, project № 14-08-06-04_м

ADAPTER FILTERING IN NGS DATA USING APACHE SPARK FRAMEWORK

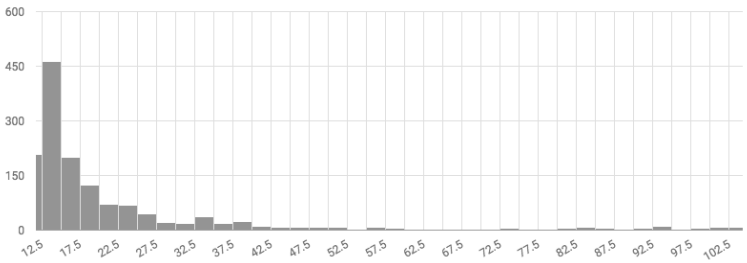
D.S. Musatov[1]*, Y.V. Vyatkin[2], E.S. Cheryomushkin[3]
[1]*Novosibirsk State University, Novosibirsk, Russia*
[2]*AcademGene LLC, Novosibirsk, Russia*
[3]*Ershov Institute of Informatics Systems SB RAS, Novosibirsk, Russia*
** Corresponding author: d.s.musatov@gmail.com*

Key words: *adapter sequence, DNA sequencing, Illumina, apache spark*

Motivation and Aim: In DNA sequencing technology, adapters are not trimmed completely or even left as is, which results foreign sequences in data[1]. We solved task of predicting adapter sequences without having original adapter sequence. The purpose if this work is to devise a program using Apace Spark framework that would discover adapter sequence remained in NGS data.

Methods and Algorithms: Algorithm comprises enumerating all k-mers (overlapping subsequences of length k) in original data and search for overrepresented k-mers. After that overrepresented k-mers are glued together to form adapter sequence, considering k-1 overlap.

Results: We implemented a program that searches for overrepresented sequences in NGS data using framework for distributed computation Apache Spark. A sequence joining algorithm, which assembles an original adapter sequence, was devised. Software was tested on an experimental data of one person produced by Illumina X Ten sequencer. We archived linear computation-time speedup on the number of computing nodes.



Histogram of values of frequencies of top 1500 k-mers (k=10) on test analysis of fastq file of patienet NA12878 (137G). Axis X – ratio of number of found k-mers to average value. Axis Y – frequency of first 1500 k – mers (sorted descending of frequency).

Conclusion: We developed an algorithm that was used to search for foreign sequences in NGS data. It was implemented in Scala language. The algorithm runs on an Apache Spark cluster. It was used to find an adapter sequence in raw Illumina HiSeq® sequencing data.

Availability: Software is available on request

References:

1. P.A. Pevzner. Computational Molecular Biology. 2001 The MIT Press, Cambridge, Massachusetts, USA

A CONGESTION GAME MODEL FOR VIRTUAL DRUG SCREENING IN A DESKTOP GRID

N.N. Nikitina*, E.E. Ivashko
Institute of Applied Mathematical Research, Karelian Research Center RAS, Petrozavodsk, Russia
** Corresponding author: nikitina@krc.karelia.ru*

Key words: *virtual drug screening, protein-ligand docking, desktop grid, BOINC, congestion game*

Motivation and Aim: Virtual drug screening is a significant part of drug development process as it allows reducing the chemical space of the order of 10^{60} potentially synthesizable compounds down to a manageable set for laboratory testing. However, pre- and post-filtering of the chemical space consume a lot of effort and time which makes structure-based virtual drug screening over very large databases unfeasible. Small focused libraries are not always readily available for new drug targets. In this work we address a game-theoretic model for filtering the explored compounds space on the fly when performing structure-based virtual drug screening over very large databases.

Methods and Algorithms: Being a computational technique to process many independent fine-grained tasks, virtual drug screening essentially involves a set of computational nodes that may be seen as independent agents, each of them willing to perform as much useful work as possible. We model virtual drug screening process as a congestion game between computational nodes who compete for a shared pool of resources, namely subsets of computational tasks. The utility of each player depends not only on the value of the chosen resource, but also on the number of other players choosing it (the "congestion level"). According to drug development principles, two primary characteristics of the resulting set of compounds to be laboratory tested are their estimated efficiency of interaction with the disease-relevant target and their structural diversity. Ranking compounds by the former characteristic is performed by molecular docking software. We propose to attain the structural diversity of the results by employing the competition between computational nodes that tend to select the computational task subsets that are in less demand by other nodes.

Results: We propose a congestion game model for virtual drug screening. The game has at least one Nash equilibrium in pure strategies; best- and better-response dynamics are guaranteed to converge to equilibrium in polynomial time. The social utility function expresses efficiency and diversity of the resulting set of compounds. The developed algorithms for taskflow management are being implemented and tested within the BOINC-based Enterprise Desktop Grid for virtual drug screening [1].

Acknowledgements: This work is supported by the Russian Fund for Basic Research (projects 16-07-00622 A and 15-29-07974 ofi_m).

References:

1. N. Nikitina, E. Ivashko, S. Möller. (2014) BOINC-based Desktop Grid Infrastructure for Virtual Drug Screening, Abstracts of the International Conference on Mathematical Modeling and High Performance Computing in Bioinformatics, Biomedicine and Biotechnology, MM-HPC-BBB-2014: 62.

INVERSE AND ILL-POSED PROBLEMS IN TOMOGRAPHY, BASED ON THE PROPAGATION OF THE ACOUSTIC WAVES

N.S. Novikov[2,3], M.A. Shishlenin[1,2,3]*
[1]*Sobolev Institute of Mathematics SB RAS, Novosibirsk, Russia*
[2]*Institute of Computational Mathematics and Mathematical Geophysics SB RAS, Novosibirsk, Russia*
[3]*Novosibirsk State University, Novosibirsk, Russia*
** Corresponding author: mshishlenin@ngs.ru*

Key words: *inverse and ill-posed problems, regularization methods, numerical methods, continuation problem, coefficient inverse problem*

Motivation and Aim: We consider the problem of detecting the inclusions in the medium by using acoustic sounding. These problems can be considered in medicine tomography, based on the propagation of ultrasound waves. We consider different approaches to the problem: continuation of the acoustic field from the boundary inside the domain and the inverse problem of determining coefficients of the system of acoustic equations, which describes the parameters of the medium.

Methods and Algorithms: We reduce the ill-posed problem to the operator equation [1,2]. For numerical solution of the continuation problem we apply singular value decomposition method and gradient method. We also consider the gradient-based optimization approach for solving coefficient inverse problem [3].

Results: Theory and numerical methods are developed for the continuation problem of the acoustic field. As the result, the formulae to calculate the singular values of the continuation problem operator and have been obtained.

Acknowledgements. The work was supported RFBR (projects No. 16-01-00755 and 14-01-00208) and the Ministry of Education and Science of the Russian Federation.

References:

1. S.I. Kabanikhin et al. (2013) Inverse Problems for the Ground Penetrating Radar, *Journal of Inverse and Ill-Posed Problems*, 21 (6).
2. S.I. Kabanikhin et al. (2013) Regularization of the continuation problem for elliptic equations, *Journal of Inverse and Ill-Posed Problems*, 21 (6).
3. I.M. Kulikov, N.S. Novikov, M.A. Shishlenin. (2015) Mathematical modeling of the propagation of the ultrasound waves in 2D medium: direct and inverse problem. *Siberian electronic mathematical reports*, 12(1), Proceedings of the VI International scientific school-conference for young scientists "Theory and numerical methods for solving inverse and ill-posed problem" (in Russian).

REALISTIC 3D SIMULATION OF *C. ELEGANS* SWIMMING AND CRAWLING WITH SIBERNETIC ENVIRONMENT

A.Yu. Palyanov[1,2,3]*, S.S. Khayrulin[1,2,3]
[1]*Institute of Informatics Systems SB RAS, Novosibirsk, Russia*
[2]*Novosibirsk State University, Novosibirsk, Russia*
[3]*OpenWorm Project*
** Corresponding author: palyanov@iis.nsk.su*

Key words: *C. elegans, swimming, crawling, computer simulation, mathematical modeling, PCI SPH, biomechanics, high-performance parallel computing*

Motivation and Aim: Building *in silico C. elegans* is an attractive idea because creation of the first simulation which correctly reproduces neuroinformational processes of a whole real nervous system will assure that it is highly probable to do it for more complex organisms as well. Realistic body is also required, either robotic (which is hardly realizable for a tiny worm) or simulated on a computer, since the nervous system needs to receive signals from sensors and send them to drive muscles. A feedback, for example, between a locomotory command and the subsequent feeling of sliding on a substrate may also be quite important for correct functioning of the nervous system. This thesis describes our new highly realistic simulation of *C. elegans* body capable of crawling and swimming. The real one is spindle-shaped, 0.8..1.0 mm long and 60-80 μm in diameter. It can be represented as an elastic shell filled with a pressurized liquid provided with body wall muscles attached to the shell's inner surface.

Methods and Algorithms: Sibernetic is a cross-platform simulation environment [1] designed for 3D simulation of physical (liquid, elastic and solid matter) and biomechanical objects (muscles, membranes, tissues) dynamics, which we have developed to simulate *C. elegans* body. It is based on parallel OpenCL/C++ high-performance realization of prediction-correction incompressible smoothed particle hydrodynamics algorithm used for simulation of liquid, which is supplemented with a number of original features providing ability to construct and deal with biomechanical objects. The data for *C. elegans* body form and muscular cells geometry and layout were extracted from the microphotographs.

Results: *C. elegans* body was constructed and parameters describing it and surrounding environment (viscosity, elasticity, surface tension, muscle contraction force etc) were tuned to provide realistic swimming and crawling (without cross-slipping) when muscles were activated with appropriate sinusoidal signals. Simulated movement is comparable with real worm locomotion properties - wavelength, amplitude, velocity etc.

Conclusion: An important step towards building a virtual *C. elegans* was taken - realistic 3D simulation of swimming and crawling in an explicit environment was performed.

Availability: Sibernetic source code is available at https://github.com/openworm/sibernetic

Swimming and crawling video are here: https://www.youtube.com/user/CyberElegansTeam

Acknowledgements: The work was supported by Russian Federation President grant MK-5714.2015.9

References:

1. A. Palyanov, S. Khayrulin, S.D. Larson (2016) Application of smoothed particle hydrodynamics to modeling mechanisms of biological tissue. *Advances in Engineering Software*, 98: 1-11.

FUNCTIONAL GRAPHS OF DISCRETE DYNAMICAL SYSTEMS OF ALMOST CIRCULANT TYPE

A.S. Parfinenko
Sobolev Institute of Mathematics SB RAS, Novosibirsk, Russia
Corresponding author: anastasia.s.parfinenko@gmail.com

Key words: *discrete dynamical system, circulant, gene network, functional graph*

Motivation and Aim: Gene networks serve as the basis for modelling of processes going on in a cell. Circuits with positive and negative feedback control gene network. Some features of the structure of regulatory circuits can be stated in terms of digraphs; the process of the substance concentration redistribution can be described as the discrete model. This research considers the discrete dynamical system which is one of such regulatory circuit models.

Methods and Algorithms: A circulant is the digraph whose adjacency matrix is a circulant matrix. General model of the discrete dynamical system of a circulant type is introduced in the article [1]. Consider the following case.

Let $G_{n,q}$ be a complete digraph with n vertices. At each time the vertices are labelled with the elements $x_0, x_1, \ldots, x_{n-1}$ of a finite field F of order q. Let us call the tuple $v = (x_0, x_1, \ldots, x_{n-1})$ the state of the system. At the next moment the state of the system changes, and dynamics of its change is defined by the mapping $Add: F^n \to F^n$; each vertex gets new label which is equal to the sum of old labels of those vertices which has an outgoing edge incoming to the considered vertex.

A functional graph is the digraph whose vertices are the elements of F^n and there is an edge from w to v if and only if $Add(w) = v$.

The research is dedicated to the study of functional graphs $FG_{n,q}$ in case of some "errors", i.e. removing of one or two edges from circulant $G_{n,q}$. The influence of these errors to the structure of functional graph is also a subject to study.

Results: The functional graphs $FG_{n,q}$ are fully described in case of $q = 2$ and $n > 3$: any functional graph $FG_{n,2}$ can be divided into subgraphs which are isomorphic to $FG_{s,2}$ where $s = n \bmod 2 + 4$.

Conclusion: Obtained result adds new information to descriptions of the functional graph of a discrete dynamical system of gene networks.

References:

1. A.A. Evdokimov, A.L. Perezhogin. (2011) Discrete dynamical systems of a circulant type with linear functions at the vertices of a network, *Diskretn. Anal. Issled. Oper.*, **18**, No. **3**, 39-48.

MATHEMATICAL MODEL OF CEREBRAL HAEMODYNAMICS IN PRESENCE OF ANEURYSM

D.V. Parshin*, I.V. Ufimtseva, A.A. Cherevko, A.K. Khe, K.Yu. Orlov, A.L. Krivoshapkin, A.P. Chupakhin
Lavrentyev Institute of hydrodynamics SB RAS, Novosibirsk, Russia
Novosibirsk State University, Novosibirsk, Russia
Academician E.N. Meshalkin Research Institute of Circulation Pathology, Novosibirsk, Russia
** Corresponding author: danilo.skiman@gmail.com*

Key words: *cerebral aneurysm, oscillating model, Van der Pol equation, cerebral haemodynamics*

Motivation and Aim: The present paper discusses the method of identification arterial aneurysms of human cerebral vessels by using of mathematical model. Human cerebral circulation as a single tuned circuit, which consists of blood flow, elastic vessels and elastic brain gel tissue is under consideration and non linear Van der Pol − Duffing equation is assumed as mathematical model of cerebrovascular circulation.

Methods and Algorithms: The current study deals with analysis of data processed. Clinical data was obtained at Meshalkin institute of circulation pathology, Novosibirsk, Russian Federation. This data was obtained during realtime cerebrovascular measurements for neurosurgery operations. The pressure and velocity were measured with ComboWire sensor and ComboMap station (Volcano Corp.)[1]. Mathematical model which is under consideration is empirical and described by Van der Pol − Duffing nonlinear ordinary differential equation:

$$\varepsilon q'' + (a1 + a2 \cdot q + a3 \cdot q^2) q' + (b1 \cdot q + b2 \cdot q^2 + b3 \cdot q^3) = ku, \quad (1)$$

here q- pressure, u-velocity, ε is a small parameter. The method of calculating coefficients *a1,a2,a3,b1,b2,b3,k* were discussed in [2]. In present study this equation considered as slow-fast dynamic system.

Results: The classification of all hyperbolic singularities of slow subsystem for eq. 1 was completely obtained. Let us talk that measurement position has *index n* if this slow system has n different singularities. It was noted that before the operation the measurement near arterial aneurysm has index 2 or 4 in terms discussed above.

Conclusion: Comparison of the current analysis results to the clinical visualization data allows us to formulate hypothesis.

Hypothesis(Test). The presence of arterial aneurysm in blood vessel could be detected by index value of the measurement for this position

Acknowledgements: The study was completed thanks to the support of Russian Foundation for Basic Research (project No 14-01-00036).

References:

1. A Ferns S P, Schneiders J J, Siebes M, van den Berg R, van Bavel E T and Majoie C B 2010 AJNR. 31 324-26.
2. Mikhailova A V, Cherevko A A, Chupakhin A P, Krivoshapkin A L, Orlov K Yu and Panarin V A 2013 Modeling of the brain haemodynamic based on experimental clinical data Proceedings of International conference on Mathematical Control Theory and Mechanics (Suzdal, Vladimir state university) 166-67 (in Russian)/

INVERSE MODELING OF DIFFUSION PROCESSES IN BIOLOGICAL TISSUES

A.V. Penenko[1,2,3]*, S.I. Baiborodin[2], A.V. Romaschenko[2], S.V. Nikolaev[2]*
[1]*Institute of Computational Mathematics and Mathematical Geophysics SB RAS, Novosibirsk, Russia*
[2]*Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[3]*Novosibirsk State University, Novosibirsk, Russia*
** Corresponding authors: a.penenko@yandex.ru; nikolaev@bionet.nsc.ru*

Key words: *coefficient inverse problem, diffusion model, FRAP, sensitivity operator, adjoint problem*

Motivation and Aim: Study of diffusion transport processes in materials is important for tissue engineering applications. The tissues may be either optically transparent or not. In the work we study material transport in collagen and gelatin thick films, as well as in mineral matrix of bone tissues.

Methods and Algorithms: The collagen and gelatin films of about 100 μm thick were kindly provided by I. Zakharova, and the bone specimens were kindly provided by I. Kirilova. We used fluorescent antibody in FRAP experiments with films, and MnO ions and nanoparticles in MRT experiments with bone specimens. The process of substance particles penetration into the specimen can be described by a diffusion-reaction model. The coefficients of the model describe the specimen transport properties. In order to study coefficient inverse problems we use sensitivity operator that is constructed with adjoint functions [1]. This operator allows analyzing the inverse problem with singular value decomposition. Multidimensional diffusion-reaction problems are dealt with splitting approach. This require direct and adjoint numerical schemes to be consistent in order to produce correct numerical algorithms. To reconstruct coefficients we use gradient based optimization methods for discrepancy functional [2] and Newton-type methods based on the sensitivity operators.

Results: Protocols of experiments were designed, and preliminary experimental data were obtained, and processed. Algorithm for diffusion coefficient reconstruction and inverse problem analysis based on sensitivity operators for image-based measurement data was developed, and diffusion coefficients were evaluated, and compared with published ones.

Conclusion: The developed algorithm is demonstrated to be applicable to study diffusion transport processes in biologically significant materials.

Availability: Algorithms are available on personal inquiry to authors.

Acknowledgements: The work was supported by RFBR grant 15-29-04875.

References:

1. A.V. Penenko. (2010) On solution of inverse heat conduction problem with gradient projection method, *SEMR*, 7: 178-198. (In Russian)
2. A.V. Penenko. (2012) Discrete-analytic schemes for solving an inverse coefficient heat conduction problem in a layered medium with gradient methods, *Numerical Analysis and Applications*, **5 (12):** 326-341.

COMPUTATIONAL MODEL FOR MAMMALIAN CIRCADIAN OSCILLATOR INTERACTING WITH NAD+ / SIRT1 PATHWAY

O.A. Podkolodnaya[1], N.N. Tverdokhleb[1,3], N.L. Podkolodnyy[1,2,3]*
[1]*Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2]*Institute of Computational Mathematics and Mathematical Geophysics SB RAS, Novosibirsk*
[3]*Novosibirsk State University, Novosibirsk*
** Corresponding author: pnl@bionet.nsc.ru*

Key words: *computer modeling, mammalian circadian oscillator, SIRT1 pathway*

Motivation and Aim: The mammalian circadian timing system is a finely tuned hierarchical system which regulates a wide range of processes in the body (molecular genetic, physiological and behavioral) with a period close to 24 hours, allowing the body to optimally adapt to the cyclical changes of environment [Reppert, Weaver, 2001]. Almost every cell in an organism contains autonomous molecular genetic circadian oscillator (CO). The structure of this oscillator can be described by a complex gene network with the feedback process mediated transcription, post-translational modification of proteins, protein-protein interactions, chromatin modification, and others. The basis of the circadian oscillator structure constitute by the two interlocked negative feedback loops generating the circadian rhythm. Additional feedbacks of this gene network provide stability of the oscillator functioning and its relationship with the other molecular-genetic systems and pathways of the organism. One of them is formed with the participation of NAD-dependent deacetylase SIRT1, which couples the deacetylation of a number of transcription factors and co-factors to the cleavage of NAD+. Therefore, SIRT1 play a vital role in metabolism, inflammation, apoptosis, stress resistance, energy responses to restriction and high calorie intake, development, and reproduction, which will ultimately affect the processes of aging and disease.

Methods and Algorithms: The circadian rhythm gene network was reconstructed using the GeneNet system (Ananko et al., 2005). Computer model for mammalian circadian oscillator was implemented in MATLAB (Mathworks) as a system of 187 ordinary differential equations.

Results and conclusion: We have modified and extended the most detailed circadian clock mathematical model, developed by Kim and Forger in 2012 [J.K. Kim and D.B. Forger, 2012]. In particular, the subsystem comprising genes / proteins NAMPT and SIRT1, as well as NAD + and NAM was added into the model. The clock gene expression data, kinetic constants and characteristics of the dynamics of the mammalian circadian processes for the wild type genotype and different mutations in the clock genes were collected and used to verify the extended mathematical model. A numerical study have demonstrated that the dynamic characteristics of the model, including the period, the amplitude and phase changes of concentrations agrees well with the experimentally observed values.

Acknowledgements: The work was supported by the RSF (the project № 14-24-00123).

References:

1. Reppert S.M., Weaver D.R. Molecular analysis of mammalian circadian rhythms. Annu Rev Physiol. 2001;63:647-76.
2. Chang H.C., Guarente L. SIRT1 mediates central circadian control in the SCN by a mechanism that decays with aging. Cell. 2013 Jun 20;153(7):1448-60.
3. Ananko EA, Podkolodnyy NL, Stepanenko IL, Podkolodnaya OA, et al. GeneNet in 2005 // Nucleic Acids Res.,2005, vol. 33, (Database issue), pp. D425--D427.
4. Kim J.K., Forger D.B. A mechanism for robust circadian timekeeping via stoichiometric balance // Mol Syst Biol. 2012;8:630.

COMPUTER ANALYSIS OF BIOLOGICAL NETWORKS OF MAMMALIAN CIRCADIAN OSCILLATOR

N.L. Podkolodnyy[1,2,3]*, N.N. Tverdokhleb[1,3]*, E.O. Sambilova[3], S.A. Lobynya[3], Z.D. Yakubova[3], O.A. Podkolodnaya[1]*

[1]*Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*

[2]*Institute of Computational Mathematics and Mathematical Geophysics SB RAS, Novosibirsk, Russia*

[3]*Novosibirsk State University, Novosibirsk, Russia*

** Corresponding authors: opodkol@bionet.nsc.ru, nata@bionet.nsc.ru, pnl@bionet.nsc.ru*

Key words: *computer network analysis, gene network, interatomic network, mammalian circadian oscillator*

Motivation and Aim: The methods of network analysis and searching patterns of structural organization of biological networks including gene networks, interactomics networks, gene co-expression networks, the diseases networks, etc. currently used for solving practical problems of bioinformatics and systems computational biology. The paper presents the methods for analysis of biological networks, including methods for analysis the local and global topological properties of networks, methods for identifying the subsystems on the network, the methods for comparing of network structures, etc., and the results of analysis of gene network and protein-protein interaction network which was applied to research mammalian circadian oscillator.

Methods and Algorithms: A method for constructing structural models of biological networks (null model) as random graphs with structural patterns similar to the patterns found in the analyzed biological network are developed. Null structural models of biological networks are important for building statistical hypotheses for solving various types of applications of bioinformatics and computational systems biology is proposed. The original method for description of the integrated structural characteristic of the biological network was proposed. The integrated structural characteristic of the network corresponds to a principal component, built on the basis of the structural characteristics of local units - graphlets frequencies (small related isomorphic induced subgraphs), which include the vertex. On this basis, we developed a method of comparing the network and a statistical criterion for testing the hypothesis under consideration the network of its structural model in the form of random graphs. The circadian rhythm gene network was reconstructed using the GeneNet system (Ananko, 2005). A network of protein-protein interactions (PPI) in the liver at different times of day was reconstructed using experimental data on protein-protein interactions, gene/protein expression data in liver tissue.

Results and conclusion: A PPI network in the liver at different times of day and an expanded version of the gene network of the mammalian circadian oscillator have been reconstructed. The structure of PPI network changed during circadian rhythm that was used to search for structural network biomarkers of the circadian oscillator disorders. A computer analysis of the gene regulatory network of the circadian oscillator and biological interpretation of the identified structural features, including central peaks gene network (hubs), structural patterns of regulation and non-random structural motifs, strongly connected components, regulatory circuits and structural-functional units (clusters) were done. As a result, we identify the central component of the circadian oscillator, which includes basic regulatory circuits passing through the key element of the circadian clock the protein Clock/Bmal1. The reconstructed structural model, which includes both the central component and functional subsystems interacting with it, became the basis for building an extended mathematical model of the dynamics of the gene network regulating the circadian oscillator.

Acknowledgements: The work was supported by the RSF (the project № 14-24-00123).

ABOUT CLASSIFICATION OF ECG SIGNALS BASED ON HIGH-FREQUENCY WAVELET COMPONENTS

P.N. Podkur, N.K. Smolentsev*
Kemerovo Institute (branch) of Plekhanov Russian University of Economics, Russia
Kemerovo State University, Kemerovo, Russia
** Corresponding author: smolennk@mail.com*

Key words: *wavelet analysis, electrocardiogram, high-frequency ECG components, statistical pattern recognition*

Motivation and Aim: In studying the electrocardiogram (ECG) is usually analyzed form of waves, length of waves and complexes of the segment, variability of the lengths of the various intervals of the ECG. Frequencies above 100 Hz are ignored. At present, however, for ECG are used the high-resolution electrocardiographs with a sampling rate of 20 kHz. Therefore, the signal recorded on such cardiographs, also contains high frequency components. The aim of this work is to show that high-frequency components of the ECG are a significant diagnostic information regarding ECG.

Methods and Algorithms: The multilevel wavelet decomposition of the ECG signal and the extraction the high frequency components. Then calculate the number of signs these high frequency components, including the energy, entropy and frequency characteristics of the wavelet decomposition component. The next step is the reduction dimensionality of the feature space using the theory of statistical pattern recognition [1] and the construction of linear classifiers that are able to distinguish between groups of ECG signals.

Results: Developed automated classification system that divides the ECG of patients and healthy patients using only the high frequency components of the ECG. To build a classifying system was used 120 ECG records of patients who have had a recent myocardial infarction (acute period) and 96 ECG records of group healthy patients.

Conclusion: To test the classification system used the ECG records of 96 patients with normal cardiogram and 120 ECG records of patients with myocardial infarction. It turned out that the probability of recording the ECG of a healthy patient to enter the group of patients the ECG is 0.049, and the probability of a patient ECG record patient classified as healthy ECG is 0.152. The results of the test for good classification accuracy. Therefore, high frequency components of the ECG contain significant diagnostic information. This system can be used as a supplement to the classification systems based on the analysis of complex waves and segments of ECG for more accurate separation of patient groups.

References:

1. Fukunaga K. Introduction to Statistical Pattern Recognition. Acad. Press, Boston, 1990, 592 p.

HETERODIMERIZATION OF SEROTONIN RECEPTORS 5-HT1A AND 5-HT7 DIFFERENTIALLY REGULATES RECEPTOR SIGNALING AND TRAFFICKING

E. Ponimaskin
Cellular Neurophysiology, Center of Physiology, Hannover Medical School, Hannover, Germany
Corresponding author: Ponimaskin.evgeni@mh-hannover.de

Serotonin receptors 5-HT_{1A} and 5-HT_7 are highly co-expressed in brain regions implicated in depression. However, their functional interaction has not been established. We have recently demonstrated that 5-HT_{1A} and 5-HT_7 receptors form heterodimers both *in vitro* and *in vivo*. Functionally, heterodimerization is critically involved in initiation of the serotonin-mediated 5-HT_{1A} receptor internalization and also enhances the ability of the 5-HT_{1A} receptor to activate the mitogen-activated protein kinases. Moreover, heterodimerization markedly decreases the ability of the 5-HT_{1A} receptor to activate G-protein gated inwardly rectifying potassium channels in a heterologous system and in hippocampal neurons, demonstrating a physiological relevance of heteromerization *in vivo*. In addition, heterodimerization is critically involved in initiation of the serotonin-mediated 5-HT1A receptor internalization and enhances the ability of the 5-HT_{1A} receptor to activate the mitogen-activated protein kinases. Finally, we found that production of 5-HT_7 receptors in hippocampus continuously decreases during postnatal development, indicating that the relative concentration of 5-HT_{1A}-5-HT_7 heterodimers and, consequently, their functional importance undergoes pronounced developmental changes.

Generally, our data suggest that the regulated and balanced ratio of homo- and heterodimerization on pre- and postsynaptic neurons may be critically involved in the onset of psychiatric diseases (e.g. depression and anxiety) and addiction.

THERSHOLD FUNCTIONS RECOVERY ALGORITHMS IN DISCRETE DYNAMIC SYSTEMS

N.V. Prytkov[1]*, A.L. Perezhogin[2]
[1]*Novosibirsk State University, Novosibirsk, Russia*
[2]*Sobolev Institute of Mathematics SB RAS, Novosibirsk, Russia*
** Corresponding author: n.v.prytkov@gmail.com*

Key words: *Discrete dynamic system, gene network, threshold function, testing, system recovery*

Motivation and Aim: Discrete dynamic systems are a powerful instrument for modeling various complex structures. Specifically, they are important in describing architecture and behavior of gene networks describing real processes in the cells of living organisms. Besides analyzing the functioning of such systems, discrete dynamic systems recovery from experimental data is of great interest [1,2].

Methods and Algorithms: Discrete dynamic system is a directed graph $G_n\ (V, D)$ with the set of vertices $V=\{v_0, \dots, v_{n-1}\}$. Value x_i from Z_p is attributed to each vertex v_i. The set of values of all the vertices determines a state s of the dynamics system. Change of value in the vertex v_i is determined by a threshold function $f_i(x_0,\dots,x_{n-1})$ with weights $W=\{w_0, \dots, w_{n-1}\}$ from Z_q. Significant variables of these functions are determined by the graph G_n. The system functioning means changing each vertex value in the next moment – transition from the current state to the next one. 2 consistent system states are called a test. Discrete dynamic system recovery requires computation of the test set, uniquely defining this system.

Results: Discrete dynamic system recovery with integer threshold functions was investigated in this study; arbitrary discrete dynamic system graph structure recovery algorithm with $n^2(p-1)-n+1$ test complexity was designed. Full system recovery algorithms with pseudopolynomial complexity were discovered. Upper bounds estimations of testing complexity were also found.

Conclusion: The presented algorithms recover discrete dynamic systems from experimental data of their functioning, which allows predicting their further behavior and influence them in order to achieve the needed system state.

Acknowledgements: This work was supported by the RFBR grant 14-01-00507.

References:

1. A.A. Evdokimov, et al. (2005) Reconstruction the structure of discrete model of gene networks, Herald of Tomsk State University, **14**: 213-217.
2. A.V. Komarov,et al. (2006) On the reconstruction of gene networks discrete model, Herald of Tomsk State University, **18**: 66-77.

LOCALISATION OF CENTERS OF NEURON-VESSEL INTERCONNECTIONS FOR NEUROBIOFEEDBACK

P.D. Rudych*, V.S. Rudnev, L.I. Kozlova, A.A. Savelov
Novosibirsk State University, Novosibirsk, Russia
Institute of Molecular Biology and Biophysics, Novosibirsk, Russia
International Tomography Center, Novosibirsk, Russia
** Corresponding author: rudych@ngs.ru*

Key words: *functional magneto-resonant imaging, electroencephalography, neuromapping, selfregularisation, neurobiofeedback*

Motivation and Aim: Dynamical registration and simultaneous fMRI-EEG mapping of neuronetwork formations is a one of the promising way to study the neurobiofeedback fenomena.

Methods and Algorithms: The group of respondents were trained to improve self regularization state using neurobiofeedback and studied with simultaneous fMRI and EEG measurements during their trainings. To reduce the experiment dimensions the fMRI brain images were divided in small cube zones. The statistical distribution features were extracted from the fMRI-cube zone and EEG. Extracted features were correlated.

Results: The zones with high correlation and anti correlation effect are highly involved in neurobiofeedback training and leads to neuron-vessel interconnection for the training. This zones can be used as targets for simultaneous EEG-fMRI neurobiobeedback to get the high efficient biofeedback level and high efficient training.

Acknowledgements: The work is supported by Russian Fond of Fundamental Research, grant №14-04-00480

References:

1. M. B. Shtark, K. Mazhirina, M. Rezakova, A. Savelov, M. Pokrovskiy, O. Jafarova. (2014) Neuroimaging phenomenology of the central self-regulation mechanisms, *Journal of Behavioral and Brain Science*, **1**.
2. L.I. Kozlova, A.A.Savelov et al. (2015) Synergetic fMRI-EEG Brain Mapping in Alpha-Rhythm Voluntary Control Mode. *Bulletin of Experimental Biology and Medicine*, **3**.

NEW IMAGE ANALYSIS AND BASE CALLING ALGORITHM FOR SEQLL SEQUENCING MACHINE ACHIEVED BETTER SENSITIVITY ON SYNTHETIC OLYGONUCLEOTIDES SET

N.E. Russkikh[1]*, D.V. Antonets[2,3,4]
[1]*Novel Software Systems LLC, Novosibirsk, Russia*
[2]*AcademGene LLC, Novosibirsk, Russia*
[3]*A.P. Ershov Institute of informatics systems, Novosibirsk, Russia*
[4]*State Research Center of Virology and Biotechnology 'Vector', Novosibirsk, Russia*
** Corresponding author: russkikh@nprog.ru*

Key words: *image analysis, new generation sequencing, single molecule sequencing, nucleotide sequence, DNA, RNA*

Motivation and Aim: Currently next generation sequencing is rapidly shifting from an expensive and dedicated basic research tool into a powerful technology capable of revolutionizing the medical field. The single molecule sequencing strategy used by SeqLL (USA) simplifies the DNA sample preparation process, avoids PCR-induced bias and errors, simplifies data analysis and tolerates degraded samples. Here we present a new algorithm for analyzing serial images from SeqLL sequencing machines and for basecalling – decoding the primary structures of sequenced oligonucleotide reads.

Methods and Algorithms: Initial image processing such as denoising and signal amplification were performed with high pass filtering by applying custom deconvolution matrix. Registration of images translation (relative to background) was performed by cross-correlation maximization. To highlight the areas of interest and subsequent extraction of oligonucleotide sequences we developed a method based on Jaccard similarity coefficients of binary vectors corresponding to columns of pixels in a series of aligned images. The algorithm was implemented with Python programming language (v.2.7.10) using cv2 computer vision library (v.3.0.0) and NumPy package for scientific computing (v. 1.9.2). Alignments of extracted oligonucleotide sequences with reference synthetic control oligonucleotides were performed using either BLAST software or the SeqLL package IndexDP.

Results and Conclusion: The developed program demonstrates consistent and reproducible results that are in good agreement with both overall oligonucleotides concentration and recovered oligonucleotide counts between the lanes (correlation coefficients exceeded 0.9). In comparison with original image analysis/basecalling software, the new program gives higher yield of sequenced oligonucleotides – the sensitivity is about 30% higher. In terms of specific types of errors the new program calls fewer insertions and substitutions compared to the previous algorithm, but it calls more deletions. The average number of errors is 3% (3 per 100 bp).

FIRST PASSAGE RANDOM WALK MESHFREE METHODS FOR BIOLOGICAL REACTION-DIFFUSION FLUCTUATION INDUCED SYSTEMS

K.K. Sabelfeld
Institute of Computational Mathematics and Mathematical Geophysics SB RAS, Novosibirsk, Russia
** Corresponding author: karl@osmf.sscc.ru*

Key words: *first passage distribution, survival probabilities, bimolecular reactions, random search, meshfree methods*

At cellular scales, different types of molecules interactions in spatially complex systems, and the fluctuation induced reactions require a fast and multi-scale process simulation. In this talk we present a meshfree Monte Carlo algorithm which is based on the first passage technique and simple event probability evaluations [1]. In many conventional stochastic methods based on molecule trajectory modeling on meshes, the molecules diffusing in solutions can not react with each other, nor can surface molecules diffuse and react with themselves [2]. We introduce new type of stochastic methods where the particles are not tracking on a mesh, but the desired events of reactions are explicitly evaluated in a continuous space. We were able to describe in a general unified scheme a volume and surface diffusion simulation. We stress the role of boundary conditions in the correct stochastic simulation of mixed volume-surface reactions. This is related to mixed Dirichlet-Neumann-Robin boundary conditions. Explicit transitions of the Markov chain states are found for the case of semi-cylinders which enable to simulate linear and non-linear diffusion in narrow channels. The method can be considered as a generalization of the classical Random Walk on Spheres algorithm in two directions: (1) the random walk lives on semi-cylinders, and (2) all possible boundary conditions are permitted. In nonlinear diffusion problems, in addition to reactants, the systems include trappings and obstacles. Some results of computer simulation which include enzyme reactions and segregation phenomenon are presented.

Acknowledgements: Support of Russian Science Foundation under Grant N 14-11-0083 is kindly acknowledged.

References:

1. K.K. Sabelfeld, (1991) Monte Carlo methods in boundary value problems. Springer Verlag. New York - Heidelberg – Berlin.
2. R. A. Kerr et al., (2008) Fast Monte Carlo simulation methods for biological reaction-diffusion systems in solution and on surfaces. SIAM J. Sci. Comput. v. 30:6, 3126–3149.

AIMEDICA - INTELLIGENT SYSTEM FOR DISEASE DIAGNOSTICS BASED ON TEXT-MINING ANALYSIS OF SCIENTIFIC PUBLICATIONS AND DIFFERENT MEDICAL DATA SOURCES

O.V. Saik[1, 2]*, P.S. Demenkov[1, 2], A.V. Starkov[3,4], T.V. Ivanisenko[1, 2], E.V. Gaisler[3,4], V.A. Ivanisenko[1, 2]

[1]*Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2]*PB-soft LLC, Novosibirsk, Russia*
[3]*Managing company "Lomonosov Capital" LLC, Novosibirsk, Russia*
[4]*Intelmed Ltd, Novosibirsk, Russia*
** Corresponding author: saik@bionet.nsc.ru*

Key words: *disease diagnostics, symptom checker, laboratory analyses, text-mining*

Motivation and Aim: Physicians in their medicine practice have to make multiple decisions related for a given patient. Computer systems can provide direct or indirect assistance in making these decisions (Degoulet, Fieschi, 1997). Nowadays, there is a need to implement medical decision support systems based on knowledge bases in the daily practice of health workers. The aim of this work was to develop an intelligent system that makes probability diagnostics by communicating with a user, on the basis of described symptoms, risk factors and/or laboratory tests.

Methods and Algorithms: The automatic extraction of data from sources in the world's major specialized medical databases and the Internet.

Results: We have developed an intelligent system AIMedica for probability diagnostics of diseases by communicating with a user, on the basis of described symptoms, risk factors and/or laboratory tests based on the data from medical databases, scientific publications and the Internet sources. AIMedica system database contains 1715 diseases and 13492 symptoms. AIMedica system has user friendly interface. It provides user with reference data about medications, diets, medical procedures, physiotherapy, and health supplements. Smart polling system used for probability diagnostics asks the minimal essential and sufficient number of questions about the current user's health condition.

AIMedica is intended for use by medical institutions of various specialties, and could be adapted for special medical requirements. Also the AIMedica database could be expanded by hospitals internal data.

A statistical analysis of medical data available via Internet used for AIMedica database creation showed that near 200 diseases cover 80% of all medical documents mentioning diseases. This result was in consistent with the WHO World Health Statistics (http://www.who.int/gho/en/).

Conclusion: An intelligent system AIMedica for disease diagnostics based on text-mining analysis of scientific publications and different medical data sources was developed.

Availability: http://aimedica.com/, http://aimedica.ru/

References:

1. P. Degoulet, M. Fieschi. (1997) Medical decision support systems, Introduction to Clinical Informatics, Springer New York. 153-167.

MATHEMATICAL MODELING OF ACTIVE SUBSTANSES AND FACTORS INFLUENCE ON FUNCTIONING OF PLANT ROOT MERISTEM

M.S. Savina[1,2,*], F.V. Kazantsev[1,2], V.V. Mironova[1,2]
[1]*Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2]*Novosibirsk State University, Novosibirsk, Russia*
** Corresponding author: captains-daughter@yandex.ru*

Key words: *mathematical model, auxin, salicylic acid, Arabidopsis thaliana*

Motivation and Aim: Plant hormone auxin is the most important regulator of plant growth and development. Auxin maintains of meristems, involves in gravity response and organogenesis. The influence of external factors and active substances on plant root development is usually mediated through variations in auxin synthesis or transportation, which subsequently affects auxin distribution pattern. With auxin regulating division, growth and differentiation of cells in a dose-dependent manner, the variations in auxin distribution would lead to morphological changes in plants. The problem is that auxin distribution is still elusive as no direct methods exist to study cellular concentrations of the low-molecule substance in a tissue. Mathematical modeling of the auxin transportation and synthesis allows to predict auxin distribution in a tissue and the range of auxin-dependent morphological changes. In order to study the influence of various environmental factors on plant root development, we proposed an approach based on the mathematical modeling of auxin distribution in the tissue.

Methods and Algorithms: Here we extended the mathematical model [1] to describe in more details auxin synthesis and transportation. Rectangular cell layout MxN corresponded to the root tip of *Arabidopsis thaliana*. Concentration changes of auxin and four of its transporters (PIN1, PIN2, PIN3, PIN7) in every cell of cell layout were described by ordinary differential equations. The model took into account synthesis and degradation of all substances, auxin diffusion and active transport. Experimental data on the changes of PINs expression in control and after salicylic acid (SA) treatments were used to fit the model, as an example of the approach application (Pasternak et al., unpublished).

Results: The model was extended to describe the effects of SA treatments on synthesis rates of PIN transporters. At the first step, the model parameters for control were adjusted. At the second step, the model parameters for PINs expression were adjusted to the experimental data. Auxin and PINs distribution were analysed in the steady state solution. The model predicted auxin accumulation in the root tip tissues after SA treatments. The sites of auxin accumulations coincide with the phenotypic changes induced by the treatments.

Conclusion: Developed approach can be applied for studying the effects of various environmental factors on the plant root growth.

Acknowledgements: The work was supported by RSF 14-14-00734. We thank Nadya Omelyanchuk for valuable advice and Taras Pasternak for experimental data.

References:

1. Mironova V.V. et al. (2012) Combined in silico/in vivo analysis of mechanisms providing for root apical meristem self-organization and maintenance, *Annals of botany*, 110(2): 349-360.

INVERSE PROBLEMS FOR NONLINEAR PDE: APPLICATIONS TO BIOLOGY AND MEDICINE

M.A. Shishlenin
Sobolev Institute of Mathematics SB RAS, Novosibirsk, Russia
Institute of Computational Mathematics and Mathematical Geophysics SB RAS, Novosibirsk, Russia
Novosibirsk State University, Novosibirsk, Russia
Corresponding author: mshishlenin@ngs.ru

Key words: *inverse and ill-posed problem, nonlinear PDE, Burgers' equation, regularization methods, numerical methods*

Motivation and Aim: The practical application of theoretical mathematical models helps us to unravel the underlying mechanisms involved in processes from mathematical physics and biosciences. We consider nonlinear partial differential equations (either stationary or evolutionary) that are applied to understand some important applications related to phenomena such as: population dynamics, epidemic spread, the interaction of cells, dead core phenomena, etc. We consider the several statements of inverse problems for nonlinear PDE which arise in biology. The ill-posedness is investigated for the reverse time problem for Burgers' equation.

Methods and Algorithms: Inverse and ill-posed problems are reduced to the operator equation. Due to the nonlinearity we apply gradient type methods for numerical solution.

Results: Numerical methods are proposed for the inverse problem for nonlinear PDE. The numerical results are presented.

Conclusion: The investigation of existing and the development of new models described by nonlinear PDE and methods of their numerical solutions are a very important in mathematical modeling of biology.

Acknowledgements: The work was supported RFBR (projects No. 16-01-00755 and 14-01-00208) and the Ministry of Education and Science of the Russian Federation.

BIOINFORMATIC EXPERT SYSTEM OF ANALYSIS AND INTERPRETATION OF OMICS SEQUENCE OF THE HUMAN GENOME

A.G. Shlikht*, N.V. Kramorenko
Far-Eastern Federal University, Vladivostok, Russia
** Corresponding author: schliht@mail.ru*

Key words: *genomics, proteomics, metabolomics, motif, expert systems*

Motivation and Aim: The most important task in genomics, transcriptomics, the proteomics is to automate the processes of analysis and interpretation of primary sequence of nucleotides, of amino acids related to the search for genes, exons, introns, motifs, consensus sequences. It is necessary to create a qualitatively new representation omics data [1].

Methods and Algorithms: To increase the level of automation of processes of analysis and interpretation of data, need to move from interactive methods of working with data on the Internet portals to automated methods based on local back-end databases and knowledge bases. To do this, semi-structured omics data Internet portals were loaded and converted into a relational data format, additionally coded, structured and indexed. Thus, we went from a poorly structured line to highly structured data set, while ensuring end-to-end forward and reverse flow of information from DNA nucleotides to the RNA transcripts, and on to amino acid sequences, protein, enzymes, reactions, metabolites, mutations, diseases.

Results: Developed bioinformatic expert system (ES) allowing to move from traditional string representations omics data to indexed relational data arrays based on database technology and knowledge. ES allows to fully automatically interpret many omics sequences for multiple motifs, consensus sequences, promoters etc. ES through a coded system of relations and links is a single information space integrated with world portals that provides an update.

Conclusion: Bioinformatic expert system is the basis for the creation of diagnostic systems and interpretation of omics data.

Availability: ES runs on a standalone server in the local network on the basis of workstations in the terminal mode, with access if necessary to the global network.

References:

1. A.G.Shlikht, N.V.Kramorenko. (2014) Integrated genome-oriented information system for monitoring and control of biological systems. Tne Ninth International Conference on Bioinformatics of Genome Regulation and Structure\Systems Biology, ISHG-2014, June 24-25 2014, Novosibirsk, Russia, p.151.

DEVELOPMENT OF A METHOD OF BASIC TRAJECTORIES OF G.I. MARCHUK FOR PARAMETRICAL IDENTIFICATION OF THE NONLINEAR DIFFERENTIAL EQUATIONS

B.M. Shumilov
Tomsk State University of Architecture and Building, Tomsk, Russia
Corresponding author: sbm@tsuab.ru

Key words: *splines, wavelets, measurements, data processing, homogeneous group of objects of exponential type, modeling*

Motivation and Aim: The aim of work is creation of a new method for approximation of linearly-expanded spatial objects and justification of its application for the solution of problems of processing of results of measurements on individual trajectories of group.

Methods and Algorithms: We study recovery problems for nonlinear differential equations which describe a complicated dynamical system by means of experimental states of the system. Results of observations are described as approximation splines and wavelets. Linearized equations of a system are used for a group of homogeneous exponential objects. Sufficient conditions are given implying that this description is correct. The right-hand sides of the differential equations under study are found as linear-fractional functions which approximate derivatives of the above splines.

Results: One example is given to illustrate the estimation of the state, of the organism of a one-year old child by counting *T*- and *B*-lymphocytes in blood. The other - for the solution of problems of processing of results of laser scanning in road construction: restoration of mathematical model of a surface of a roadbed and detection of cracks and damages to the places demanding repair.

Conclusion: Generalization of a method of middle trajectories on the basis of parametrical identification of the nonlinear differential equations of G. I. Marchuk to a case of several basic trajectories is for the first time considered. In relation to the highway it is admissible to use well distinguishable on the scanned image of a brow, borders and other, linearly extended elements of the highway.

Availability: Still not.

Acknowledgements: Research is executed with financial support of the RFBR and the Tomsk region within the scientific project 16-41-700400 “r_a”.

References:

1. G.I. Marchuk. (1984) *Mathematical Models in Immunology*, (New York, Optimization Software)
2. L.I. Konstantinova, V.A. Kochegurov, B.M. Shumilov. (1997) Nonlinear differential equations: Parametric identification by exact polynomial spline schemes, *Autom. Remote Control*, **58:** 756-764.
3. E.A. Esharov, B.M. Shumilov. (2015) Processing of materials of laser scanning of roads on the basis of recursive cubic splines, *Advanced Materials in Construction and Engineering (15–17 October 2014, Tomsk, Russia) / IOP Conf. Series: Materials Science and Engineering*, **71:** 012046.
4. B.M. Shumilov. (2015) Matrix sweep algorithm for computing multiwavelets of odd order orthogonal to polynomials, *Optoelectronics, Instrumentation and Data Processing*, **51:2:** 1-9.
5. B.M. Shumilov, N.A. Baigulov. (2015) A study on modeling of road pavements based on laser scanned data and a novel type of approximating hermite wavelets, *WSEAS Transactions on signal processing*, **11:** 150-156.

ESTIMATES OF SOLUTIONS TO A SYSTEM DESCRIBING THE SPREAD OF AVIAN INFLUENZA

M.A. Skvortsova
Sobolev Institute of Mathematics SB RAS, Novosibirsk, Russia
Novosibirsk State University, Novosibirsk, Russia
Corresponding author: sm-18-nsu@yandex.ru

Key words: *bird's migration, avian influenza, delay differential equations, Lyapunov differential equation, asymptotic stability, estimates of solutions*

Motivation and Aim: We consider a system of delay differential equations [1]

$$\begin{cases} S_w'(t) = -\left[\mu_w^s + m_{wb}(t)\right]S_w(t) + \alpha_{bw}^s m_{bw}(t-\tau_{bw})S_b(t-\tau_{bw}) - \dfrac{\beta_w S_w(t) I_w(t)}{S_w(t)+I_w(t)}, \\ S_b'(t) = -\left[\mu_b^s + m_{bw}(t)\right]S_b(t) + \alpha_{wb}^s m_{wb}(t-\tau_{wb})S_w(t-\tau_{wb}) - \dfrac{\beta_b S_b(t) I_b(t)}{S_b(t)+I_b(t)} + b(t)S_b(t)\left(1-\dfrac{S_b(t)}{K}\right), \\ I_w'(t) = -\left[\mu_w^i + m_{wb}(t)\right]I_w(t) + \alpha_{bw}^i m_{bw}(t-\tau_{bw})I_b(t-\tau_{bw}) + \dfrac{\beta_w S_w(t) I_w(t)}{S_w(t)+I_w(t)}, \\ I_b'(t) = -\left[\mu_b^i + m_{bw}(t)\right]I_b(t) + \alpha_{wb}^i m_{wb}(t-\tau_{wb})I_w(t-\tau_{wb}) + \dfrac{\beta_b S_b(t) I_b(t)}{S_b(t)+I_b(t)}. \end{cases}$$

This system describes the spread of avian influenza between birds migrating between two territories. Here $S_w(t)$ and $S_b(t)$ are the numbers of healthy birds in summer and winter territories, $I_w(t)$ and $I_b(t)$ are the numbers of infected birds in summer and winter territories respectively. It is assumed that all the coefficients of the system are T - periodic and nonnegative. In [1] it was obtained conditions on the coefficients of the system under which there exists asymptotically stable periodic solution $(S_w(t), S_b(t), I_w(t), I_b(t)) = = (S_w^*(t), S_b^*(t), 0, 0)$ corresponding to the healthy population. Our aim is to obtain estimates of the rate of convergence of solutions to this periodic solution.

Methods and Algorithms: We obtain estimates using a solution $H(t)$ to the special boundary value problem for Lyapunov differential equation [2] $H' + HA(t) + A^*(t)H = -I$, $t \in [0,T]$, $H(0) = H(T) > 0$. It is important to note that this problem is well-conditioned and it does not require finding the spectrum of the monodromy matrix [2].

Results: Our main results are estimates of solutions characterizing the rate of convergence of solutions to the periodic solution.

Acknowledgements: The author is grateful to Professor G.V. Demidenko for the attention to the research. The work was supported by RFBR, grant 15-01-00745.

References:

1. X.-S. Wang, J. Wu (2012) Periodic systems of delay differential equations and avian influenza dynamics, Sovrem. Mat. Fund. Naprav., 45: 32-42 (Russian); transl. in: J. Math. Sci., 2014, 201(5): 693-704.
2. G.V. Demidenko, I.I. Matveeva (2001) On stability of solutions to linear systems with periodic coefficients, Sib. Mat. Zh., 42(2): 332-348 (Russian); transl. in: Sib. Math. J., 2001, **42(2):** 282-296.

PROCESSING AND ANALYSIS OF GENE EXPRESSION DATA BY EXPGENE SOFTWARE

A.M. Spitsina*, Y.L. Orlov
Novosibirsk State University, Novosibirsk, Russia
Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
** Corresponding author: aspitsina@bionet.nsc.ru*

Key words: *DNA microarray, RNA-Seq, bioinformatics software*

Motivation and Aim: Analysis of gene expression molecular mechanisms has a great fundamental importance in various fields of science, particularly in medicine and statistics. Currently, there is a rapid development of genomic and biological technologies, which are leads to the accumulation of large experimental gene expression data in publicly available databases (most popular and free is BioGPS [1], GEO NCBI). Processing of such data requires development of new computer analysis methods, what will allow solve problems of gene expression regulation integration transcription factors binding and expression data [2]. The aim of this work is the creation of software for the analysis and visualization of transcriptomic and microarray data, which will be easy to use and multifunctional.

Methods and Algorithms: The Affymetrix GeneChip data on the human genome and genomes of model organisms (mice and rats) were used as test data. We also used RNA-seq data on gene expression levels obtained at ICG SB RAS. The program is written in *Python* language using *JSON* modules, and also popular libraries for processing and visualization text and numerical data (*pandas*, *numpy*, *scipy*, *matplotlib*).

Results: Software package ExpGene has been developed. It includes a set of options to work with a large array of microarray data - preprocessing, statistical analysis of gene expression correlations and visualization. This tool is also designed to work with gene ontologies. It is versatile for any type of text databases (allows the user to pre-select processed data). The program has a user-interaction interface (menu) and is easy to handle even by an inexperienced user.

Conclusion: Using this program we performed a comparative analysis of different samples of genes [1], such as genes from gene networks annotated in the ICG SB RAS regulating cholesterol levels and circadian rhythm. Gene expression correlation matrices for gene lists were reconstructed as basis for qualitative analysis of the gene network studied.

Availability: Software is available from the author upon request.

References:

1. Spitsina A.M. et al. (2015) Supercomputer analysis of genomics and transcriptomics data revealed by high-throughput DNA sequencing, Program systems: theory and applications, 6:1(23): 157–174. (In Russian)
2. Orlov Y. et al. (2012) Computer and Statistical Analysis of Transcription Factor Binding and Chromatin Modifications by ChIP-seq data in Embryonic Stem Cell // J Integr Bioinform. 9(2):211.

NUMERICAL MODEL OF DROSOPHILA SENSORY ORGAN PRECURSOR CELL DETERMINATION

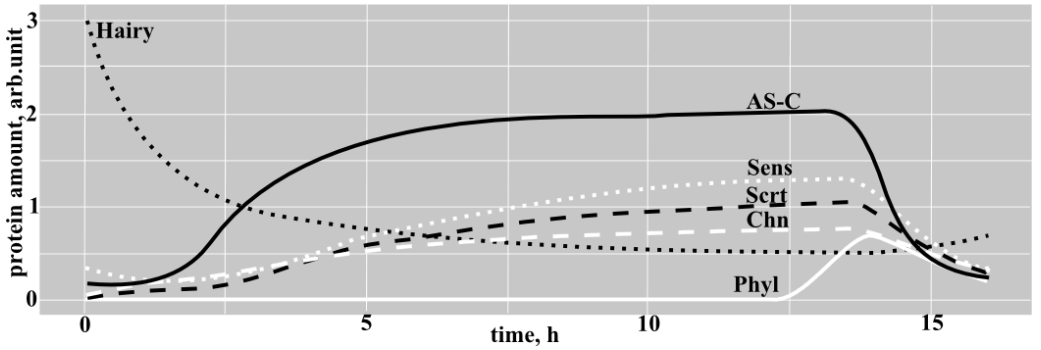
T.A. Bukharina[1], D.P. Furman[2], V.P. Golubyatnikov[3]*, M.V. Kazantsev[4]
[1]*Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2]*Novosibirsk State University, Novosibirsk, Russia*
[3]*Sobolev Institute of Mathematics SB RAS, Novosibirsk, Russia*
[4]*Polzunov Altai State Technical University, Barnaul, Russia*
** Corresponding author: glbtn@math.nsc.ru*

Key words: *Sensory Organ Precursor Cell (SOPC), Central Regulatory Circuit (CRC), dynamical system, stability*

Motivation and Aim: SOPC determination is the main event in development of *D. melanogaster* bristles. We describe one mathematical model of SOPC formation under the control of CRC in order to perform numerical simulations of this process.

Methods and Algorithms: Modeling of CRC functioning and analysis of the numerical results follows mathematical constructions presented in [1].

Results: We study phase portrait of 6-dimensional nonlinear dynamical system which simulates two stages formation of SOPC. This process takes up to 16 hours [2]. In our mathematical model this interval was split into two periods with different dynamics. The first period takes 10 hours, and it is characterized by absence of the Phyllopod protein in the CRC dynamics. Here, the AS-C proteins concentration grows with decreasing speed due to feedbacks structure in the system. By the end of the first period, the system approaches to its equilibrium state with maximal AS-C concentration. Appearance of Phyllopod in CRC during the second period (10-16 hours) induces decreasing of AS-C amount to almost zero values by the end of this period.



Dynamics of the amount of proteins during two phases of SOPC determination.

Conclusion: The two-phases model is the fullest description of SOPC determination. Results of our numerical experiments correspond to available biological data, see [2].

Acknowledgements: Budget project 0324-2015-0003 and RFBR, grant 15-01-00745.

References:

1. V.P. Golubyatnikov et al. (2015) A model study of the morphogenesis of *D. melano-gaster* mechanoreceptors: the central regulatory circuit, *JBCB*, 13(1): 1540006.
2. P.J. Chang et al. (2008) Negative-feedback regulation of proneural proteins controls the timing of neural precursor division, *Development*, 135(18): 3021-3030.

GIANT GLOBAL ALIGNMENTS PERFORMED ON A NOVEL GRID-SYSTEM BASED ON THE CLIENT-SIDE SCRIPTING ONLY

R.K. Tetuev*, A.N. Pankratov, M.I. Pyatkov
Institute of Mathematical Problems of Biology, Keldysh Institute of Applied Mathematics, Russian Academy of Sciences, Moscow, Russia
** Corresponding author: ruslan.tetuev@gmail.com*

Key words: *Rich Internet Applications, DNA alignments, grid systems, cloud computing*

There is a vast variety of computational tools have been recently developed for comparative genomics, including ones seeking for giant inexact DNA repeats. Earlier the authors found huge tandem repeats [1, 2] using our own tools (SpectralRevisor, SBARS), however they were semi-automated since the final verification needs a third-side tool to globally align the potential sites, which is still a big problem for such a huge fragments (>50K). Later one of the authors worked awhile on the project [3], it's a tool seeking the potential off-target sites for a pair of user-defined TALENs, and there formulates a new conception for computing: What if the server-side parallel processes bring a whole bunch of potential sites directly to the user browsers, because the client-side scripting today is strong enough to do all the rest – verifying sites and sorting out the complete list of the *in-silico* prediction. So, finally, we decided to develop a new type of grid systems when the server is the central node of a star network and its role is nothing but to dispatch code/data and to do 'messaging' between the peripheral nodes (any user navigated to the server's main page and voluntarily stays *on-line* while some task computing on its internal JS-engine). As a formal model to test a new type of grid computations, we choose the mentioned problem of our current interest – a pairwise global alignment for giant DNA sequences. It allows us to use all the advantages of 'divide and conquer' design paradigm for Needleman-Wunsch algorithm, as was indicated in [4] on a typical grid system. In our work, besides the technical methods, it is shown that the global alignment memory complexity could be reduced from quadratic to linear space. We explore the possibilities of using HTML5 technologies to run parallel computations entirely on a set of browsers Now it means potential usage of any computing device in the Internet, or even any CPU (when we're considering background web workers). Whether it is a standard PC, or a laptop, or even a cellphone, they could be used all in one for the same goal since all of them have modern browsers. Interesting, the core idea of this like computations previously was coined to as a 'reverse cloud computing' [5], but actually this idea has a great advantage over both, clouds and grids, because the same code can be run everywhere with no need for additional software portability or installation. The desktop version of the tool is available at http://mpyatkov.github.io/sbars, and all the on-line versions located at http://spectral.psn.ru. This research was supported by RFBR grants 14-07-00924, 15-29-07063.

References:

1. R.Tetuev, F.Dedus, N.Nazipova, *IMPB_02* (2010) Repbase Reports, V.10, Is.8, P. 1185.
2. M.Pyatkov, V.Filippov, A.Pankratov, *MSU1* (2012) Repbase Reports, V.12, Is.3, P. 256.
3. R.Tetuev, *TALEN Design In-Silico*, (2014), http://tandis.sabanciuniv.edu/index.php.
4. T.Naveed, I.Siddiqui, S.Ahmed, (2005) Parallel Needleman-Wunsch Algorithm for Grid, *Proceedings of the HONET 2005,* Dec 19 - 21.5.

HOW SOLVING THE INVERSE PROBLEM HELPS TO DESIGN A GENE NETWORK AND TO REVEAL ITS PATHWAYS

K.Y. Tkachev* [1], P.G. Emelyanov[1, 2], I.I. Titov[1, 3]
[1]*Novosibirsk State University, Novosibirsk, Russia*
[2]*A.P. Ershov Institute of Informatics Systems SB RAS, Novosibirsk, Russia*
[3]*Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
** Corresponding author: ktkachev@mail.ru*

Key words: *synchronous Boolean networks, gene regulatory networks*

Motivation and Aim: Synchronous Boolean networks is a convenient model of gene networks, control networks, and so on. Due to time-consuming step of logic rules reconstruction only small-size networks have been considered yet.

Results: We present a software tool for building a synchronous Boolean model of gene networks and for investigating its dynamical properties. The intelligent assistant built-in in the software and based on inverse problem solving helps to find the Boolean networks which have a given fixed point. It minimizes the undefined number of logic table elements thus saving human efforts of network design. Simulating a perturbed network relaxation allows to reveals the network pathways. We demonstrate our software features on a number of stem cell differentiation networks and show how it can be used for experiment design.

ARGO_CUDA: A FULL-EXHAUSTIVE GPU BASED APPROACH FOR A MOTIF DISCOVERY IN THE LARGE DNA DATASETS

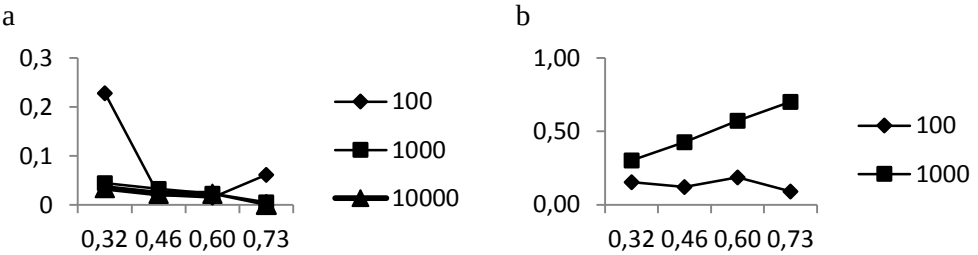
O.V. Vishnevsky[1,2]*, A.V. Bocharnikov[2], N.A. Kolchanov[1,2]
[1]*Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2]*Novosibirsk State University, Novosibirsk, Russia*
** Corresponding author: oleg@bionet.nsc.ru*

Key words: *degenerated oligonucleotide motif, transcription regulation, CUDA, GPU*

Motivation and Aim: A motif discovery in ChIP-Seq datasets remains a challenging issue. A low effectiveness of classic heuristic motif discovery approaches on a whole ChIP-Seq datasets forces the researchers to take into analysis only a fraction of top "peak" segments.

Methods and Algorithms: Argo_CUDA web service is designed to process the massive DNA data. This program for detection of degenerate oligonucleotide motifs of fixed length is based on the full-exhaustive approach and uses high-performance GPU technologies.

Results: We compared an effectiveness of Argo_CUDA and Info-gibbs [1]. Info-gibbs is a Gibbs sampling algorithm that compares well with existing heuristic methods like MEME, BioProspector, Gibbs or GAME on both synthetic and biological data sets. The sets of random sequences of 128bp in length and of 100, 1000, and 10000 sequences in size were generated. The sample motifs of different degeneracy level were placed in a 60 percent of the sequences. The similarity between a motifs obtained by the programs and the sample motifs were measured with the average Kullback-Leiber distance (KLD).



The dependence of KLD (Y axis) and information content (IC) of the sample motifs (X axis) for Argo_CUDA (a) and Info-gibbs (b). The picture demonstrates that the effectiveness of Argo_CUDA is acceptable for all examined IC levels and sample sizes. It was shown (pic 1b) that Info-gibbs is effective only for sets of few hundred sequences in size. The same results were obtained for the sets with 30% and 10% presence of the sample motifs (data not shown). The sets of ChIP-Seq segments corresponding to the mouse TFBS were analyzed with Argo_CUDA. The significant motifs were revealed and classified.

Conclusion: An effective web service for motif discovery in ChIP-Seq datasets is developed. It is not as fast as a classic heuristic approaches, but it considerably reduces the restrictions on the size of the sample under analysis.

Availability: wwwmgs.bionet.nsc.ru/mgs/programs/Argo_CUDA.

Acknowledgements: The work was supported by the budget project 0324-2015-0003.

References:

1. Defrance M, van Helden J. (2009) Info-gibbs: a motif discovery algorithm that directly optimizes information content during sampling. *Bioinformatics* 25(20):2715-22.

IPE PACK SOFTWARE FOR MODELING DYNAMIC PROCESSES

D.A. Voronov[1,2]*, A.Yu. Belonog[1,2], S.I. Kabanikhin[1,2]
[1]*Institute of Computational Mathematics and Mathematical Geophysics SB RAS, Novosibirsk, Russia*
[2]*Novosibirsk State University, Novosibirsk, Russia*
** Corresponding author: dmitriy.voronov.89@gmail.com*

Key words: *pharmacokinetics, inverse problem, IPE pack, identifiability, epidemiology*

Motivation and Aim: There are a lot of different software platforms for managing, analyzing and reporting PK/PD data nowadays at the market. There are a lot of areas of applications of such development. Most of them are intended for basic and clinical research scientists and is designed to facilitate the discovery, exploration and application of the underlying pharmacokinetic and pharmacodynamic properties of drugs. Some applications are even developed for mobile devices. The IPE Pack software is developed for modeling pharmacokinetic, immunology and epidemiology processes.

Methods and Algorithms: IPE Pack provide the possibility for: solving direct and inverse problems for SODE, modeling PK processes, analyze the identifiability properties of the model, modeling TB spread. IPE Pack is developed in C# platform. The problem of the identifiability of mathematical models of physiology, pharmacokinetics and epidemiology is considered in this report. The methods covered in this talk could reduce the costs of conducting experiments on obtaining an experimental data. The necessity of making a priory identifiability analysis before estimating parameters characterizing any process is demonstrated on several examples. Different approaches of estimating physiological parameters are covered in this talk. Numerical results of inverse problem solving are presented in this report including the experiments with noisy data.

Results: Special software for modeling pharmacokinetic, epidemiology and immunology processes is presented in this report. PK model building process for specific drug is shown in this talk.

Conclusion: IPE Pack software package could be useful for large amount of specialists in different areas: mathematics, pharmacokinetics, epidemiology and pharmacy. IPE Pack could be used for teaching basics of pharmacokinetic modeling techniques and had already been used in the MA course in Novosibirsk State University.

Availability: upon request.

Acknowledgements: The reported study was supported by Russian Foundation for Basic Research, research project No. 16-31-00382.

References:

1. Kabanikhin S.I., Voronov D.A., Grodz A.A., Krivorotko O.I. (2015) Identifiability of mathematical models in medical biology, *Vavilov J. Genetics and Breeding*. V. 19, No. 6. P. 738-744. DOI: 10.18699/VJ15.097.
2. D.A. Voronov, A.Yu. Belonog, S.I. Kabanikhin, O.I. Krivorotko. Certificate of program registration №PR16001 - «Numerical algorithm for solving inverse problem for system of differential equations», 2016-03-23.

MATHEMATICAL MODELING AND PARAMETERS ESTIMATION FOR PK EXPERIMENTAL DATA

E.A. Vostrikova[1]*, A.Yu. Belonog[1], D.A. Voronov[1,2]
[1]*Novosibirsk State University, Novosibirsk, Russia*
[2]*Institute of Computational Mathematics and Mathematical Geophysics SB RAS, Novosibirsk, Russia*
** Corresponding author: vostrikova-liza@inbox.ru*

Key words: *pharmacokinetics, inverse problem, modeling, numeric solution, system of ordinary differential equations*

Motivation and Aim: The aim of work is pharmacokinetic modeling and further data fitting for real data provided by Scientific Center for Anti-Infectious Drugs.

Methods and Algorithms: Considering a pharmacokinetic model described by a system of ordinary differential equations. Its form is determined by the compartment modeling method and given data. Statements of direct and inverse problems were considered.

The inverse problem is to estimate model parameters using additional data. This problem reduces to an optimization problem of data fitting which is solved by minimization of selected functional.

Results: The models that allow real data fitting are built. The results of numerical solution using different approaches are presented. A matter of choice of initial parameter approximation is considered.

Conclusion: Parameters estimation lets us know drug absorption rate and movement rate between organs in each organism.

Acknowledgements: The reported study was supported by Russian Foundation for Basic Research, research project No. 16-31-00382.

References:

1. E. Carson, C. Cobelli (2001) Modelling Methodology for Physiology and Medicine, *Academic Press.*
2. H.T. Banks, Shuhua Hu, W. Clayton Thompson (2014) Modeling and Inverse Problems in the Presence of Uncertainty, *Chapman and Hall/CRC*.
3. A.I. Ilyin, S.I. Kabanikhin, D.B. Nurseitov, A.T. Nurseitova, N.A. Asmanova, D.A.Voronov, D. Bakytov (2012) Analysis of Ill-Posedness and Numerical Methods of Solving a Nonlinear Inverse Problem in Pharmacokinetics for the Two-Compartmental Model with Extravascular Drug Administration, *Journal of Inverse and Ill-Posed Problems*, vol. 20, p. 39-64
4. D.Z. D'Argenio, A. Schumitzky, X. Wang (2009) ADAPT 5 user's guide: pharmacokinetic/pharmacodynamic systems analysis software, *Biomedical Simulations Resource, Los Angeles*.
5. J.A. Nelder, R. Mead (1965) A simplex method for function minimization, *The computer journal*. T. 7. – №. 4. – C. 308-313.

STREPTOMYCES SP. IB 2014 011-1 STRAIN, ISOLATED FROM *TRICHOPTERA* SP. LARVAE OF LAKE BAIKAL: DRAFT GENOME SEQUENCE

I.V. Voytsekhovskaya[1]*, D.V. Axenov-Gribanov[1], B.T. Tokovenko[2], Y.V. Rebets[2], E.S. Protasov[1], A.N. Luzhetskyy[2], M.A. Timofeyev[1]

[1]*Irkutsk State University, Institute of Biology, Irkutsk, Russia*

[2]*Helmholtz Center for Infectious Research, Helmholtz Institute for Pharmaceutical Research Saarland, Saarbrücken, Germany*

** Corresponding author: irina.voytsekhovskaya@gmail.com*

Key words: *Actinobacteria, Baikal Lake, biodiversity, Streptomyces sp. IB2014 011-1*

Motivation and Aim: Unique ecosystems with specific environmental conditions is a promising source for isolation of new actinobacteria strains. Lake Baikal is one of the greatest examples of ecosystem with high species biodiversity and endemicity caused by the long isolated evolution. The main aim is estimation of the *Streptomyces sp*. IB2014 011-1 strain to produce natural compounds through draft genome sequence.

Methods and Algorithms: The *Streptomyces sp.* IB2014 011-1 strain was isolated from *Trichoptera sp*. larvae collected from the bottom of Lake Baikal close to the Listvyanka settlement. The raw sequencing data was obtained using Illumina HiSeq 2500 technology. High molecular mass DNA was extracted from *Streptomyces* sp. IB2014 011-1. Manufacturer-recommended standard protocol was used to prepare two paired-end libraries. After quality control, only the 2nd library was used for genome assembly using SPAdes v3.7.A total of 3985 contigs were assembled, of them 84 longer than 1kbp. Scaffolding was performed by SSPACE 2.1 Premium using both libraries, and resulted in 49 scaffolds. Genome annotation was performed using prokka and antiSMASH v.3, followed by manual GenBank pre-submission curation. 16S rRNA delineation was performed using both ARB-SILVA database and NCBI's non-redundant database blast search. 16S rDNA sequences were multiple-aligned using MAFFT v7.222 (algorithm: auto, scoring matrix: 200PAM / k=2, gap open penalty 1.53, offset value 0.123). The phylogenetic consensus tree was built and formatted using Geneious 9.0.4 (Tamura-Nei model, NJ tree build method, *S. avermitilis* as an outgroup, 1000 bootstrap replicates).

Results: The genome of *Streptomyces* sp. IB2014 011-1 has a total length of about 8.1 Mbp, including a possible 100 kbp plasmid (scaffold STIB_19). The GC content, the number of protein coding genes, tRNA and rRNA genes are in accordance with other streptomycetes.

Conclusion: The chromosome of *Streptomyces sp.* IB2014 011-1 contains 31 putative gene clusters involved in the biosynthesis of secondary metabolites (or 84 gene clusters, if we also include ClusterFinder predictions)

Acknowledgements: This study was supported by the Ministry of education and science of Russian Federation as a part of Goszadanie projects (№6.382.2014/K, 6.734.2016 DAAD, 6.696.2016 DAAD), Russian science foundation (project N 14-14-00400), Russian foundation for basic research (projects N 14-04-00501), Grants of Irkutsk State University for researchers and Deutscher Akademischer Austauschdienst.

A VARIATION APPROACH FOR SOLVING OF A PARAMETER IDENTIFICATION PROBLEM FOR THE MATHEMATICAL MODEL OF HIV DYNAMICS

D.V. Yermolenko[1,2]*, O.I. Krivorotko[1,2], S.I. Kabanikhin[1,2]
[1]*Novosibirsk State University, Novosibirsk, Russia*
[2]*Institute of Computational Mathematics and Mathematical Geophysics SB RAS, Novosibirsk, Russia*
** Corresponding author: ermolenko.dasha@mail.ru*

Key words: *mathematical model of HIV dynamics, ordinary differential equations, parameter identification, optimal treatment control, inverse problem, genetic algorithm*

Motivation and Aim: Mathematical models in immunology are described by systems of nonlinear ordinary differential equations. It is important to determine parameters of these systems that characterize features of immunity and disease for constructing an individual treatment plan. The purpose of this work is the construction and investigation of the numerical algorithm for determining of coefficients in nonlinear system that describes HIV dynamics with treatment [2, 3] using additional information about given concentrations at fixed times.

Methods and Algorithms: The parameter identification problem (inverse problem) for mathematical model of HIV dynamics (*i.e.* dynamics of infected and uninfected CD4 T-lymphocytes, infected and uninfected macrophages, free virus, immune effectors (CD8 - cells)) with treatment using additional measurements of some concentrations in fixed times is numerically investigated. In this paper two inverse problems are considered [1]: firstly, the inverse problem for mathematical model without treatment and then optimal control problem for identify effective treatment function. The first inverse problem is reduced to minimization problem of least square function that describes the deviation between model and measured data. Then the problem of optimal treatment control is solved by minimizing another misfit function that characterizes combination of viral load and treatment costs. The numerical algorithm for solving inverse problems is based on stochastic approach (genetic algorithm).

Results: Individual parameters that characterize the human immune system have been identified. The optimal treatment control for an individual patient has been received.

Conclusion: It is shown that optimal treatment plan is better determined if previously individual parameters for patient are identified well. The results of the numerical calculations are presented and discussed.

Availability: Using the results of this paper one can make an individual patient's treatment plan. It will extend the duration of the patient's life.

Acknowledgments: This work is supported by the Ministry of Education and Science of the Russian Federation and the Russian Foundation of Basic Research No. 16-31-00189.

References:

1. S.I. Kabanikhin (2011) Inverse and Ill-Posed Problems: Theory and Applications (Berlin: de Gruyter).
2. G. Bocharov, A. Kim, A. Krasovskii, V. Chreshnev, V. Glushenkova, A. Ivanov. (2015) An extremal shift method for control of HIV infection dynamics, *Russ. J. Numer. Anal. Math. Modelling*, 30(1): 11–25.
3. B.M. Adams, H. T. Banks, M. Davidian, Hee-Dae Kwon, H.T. Tran, S.N. Wynne, E.S. Rosenberg. (2005) HIV dynamics: Modeling, data analysis, and optimal treatment protocols, *Journal of Computational and Applied Mathematics*, 184: 10-49.

ESTIMATING THE SURVIVAL RATES OF NORTHERN FUR SEALS (*CALLORHINUS URSINUS*, TYULENIY HERD) AND MODELING THE POPULATION NUMBER DYNAMICS

O.L. Zhdanova[1]*, A.E. Kuzin[2], E.Ya. Frisman[3]
[1]*Institute of Automation and Control Processes FEB RAS, Vladivostok, Russia*
[2]*Pacific Research Fisheries Center, Vladivostok, Russia*
[3]*Institute of Complex Analysis of Regional Systems FEB RAS, Birobidzhan, Russia*
** Corresponding author: axanka@iacp.dvo.ru*

Key words: *population dynamics, mathematical modeling, estimating survival rate, fur seal*

Motivation and Aim: In the middle of the last century the northern fur seal became the attractive object for population investigations. During 30-years period of existence of the Interim Convention of Conservation of North Pacific Fur Seals a unique set of data on the population dynamics of this species has been accumulated and it becames a good base for estimating population parameters and developing various mathematical models of population dynamics. In particular, the detailed model of fur seals dynamics [1] and techniques of calculating its parameters were developed [1-3]. Significantly increased data series by now allow us to verify the suitability of constructed model and the techniques of calculating the intrapopulation parameters, taking into account possible changes in intrapopulation and harvesting processes.

Methods and Algorithms: Mathematical modeling, numerical simulations and multivariate statistical technique are used to estimate intrapopulation parameters and to model population dynamics. Following data set is utilized: number of pups and bulls at the rookery (for 1958 – 2013 years), the age composition of animals caught in coastal harvest, as well as the age structure and physiological state of adult females form sea samples (1958–1988).

Results & Conclusion: Harvest process in fur seal population changed, at that the previously used method [2] for estimating the juvenile survival rate of males does not properly work now. We developed a new method for estimating the juvenile survival rate of males without any requirements to harvest process. Satisfactory estimates for all characteristics of bulls' lifecycle have been gained. It was revealed that structural change in survivability of harem bulls occurred at the end of 80s. The model dynamics for number of harem bulls consisting enough with the data (mean error of approximation 3.2%) has been obtained. A set of numerical simulations shows, that juvenile survival rate of females had to increase too to be able to give an adequate dynamics of adult females' number.

Acknowledgements: The research is partially supported by RFBR (project 15-29-02658).

References:

1. E.Ya. Frisman, E.I. Skaletskaya, A.E. Kuzyn. (1982) A mathematical model of the population dynamics of a local northern fur seal with seal herd, *Ecol. Mod.*, 16: 151-172.
2. R.H. Lander. (1975) Method of determining natural mortality in the northern fur seal (*Callorhinus ursinus*) from known pups and kill by age and sex, *J. Fish. Res. Board Can.*, 32(12): 2447-2452.
3. A.W. Trites. (1989) Estimating the juvenile survival rate of male northern fur seals (*Callorhinus ursinus*), *Can. J. Fish. Aquat. Sci.*, (46): 1428-1436.

WEB-BASED APPLICATION FOR FLOW CYTOMETRY DATA ANALYSIS

Y. Zhou[1], W. Ni[2], M. Chen[1]*
[1]*Department of Bioinformatics, College of Life Sciences, Zhejiang University, Hangzhou, China*
[2]*Molecular Diagnosis Centre, Zhejiang Provincial People's Hospital, Hangzhou, China*
** Corresponding author: mchen@zju.edu.cn*

Key words: *Flow cytometry, Data analysis, Web-based, Bioinformatics, AutoFCM*

Motivation and Aim: Flow cytometry is a widespread single-cell measurement technology with a multitude of clinical and research applications. In the past few years, a number of methods and tools have emerged for processing flow cytometry high dimensional data. However, these tools are coded in different programming languages and work only on specific platforms. Users need certain skills to deal with tool installation and numerous parameters. We are motivated to develop an easy-to-use web-based application, named autoFCM, with friendly UI and multi-user system.

Methods and Algorithms: Different methods such as k-means and t-SNE are integrated into this tool to cover the shortage of each other. Users can make the optimal analysis as need.

Results: After uploading flow cytometry data, users can simply click a few buttons to achieve a result. Based on constant data accumulation and data training, the system can automatically provide more accurate groups by supervised learning kit.

Availability: The tool is available at http://bis.zju.edu.cn/autoFCM.

HIGH-PERFORMANCE COMPUTATIONS SUPPORT FOR THE HAPLOOID EVOLUTIONARY CONSTRUCTOR 3D SOFTWARE PACKAGE

R.K. Zudin[1,2]*, S.A. Lashin[1,2]
[1]*Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2]*Novosibirsk State University, Novosibirsk, Russia*
** Corresponding author: rockalocular@gmail.com*

*Motivation and Aim:*Spatial factors are one of the main sources of biological evolution. Spatial redistribution of species and populations is taking one of the first places when we talk about evolution of microbial communities, because evolution cannot happen without communication of species between each other and concurrency for resources that are important for living.

There are a lot of software packages for computer modeling of microbial communities that allow calculating models with spatial redistribution. Among them, the HEC 3D [1] is the package allowing modeling communities of one of the most complicated spatial and trophic structure. At the same time HEC 3D provides one of the largest number of biological processes that are taken into account during modeling (substrates consumption and secretion, metabolism, mutations, horizontal transfer and loss of genes, phage infection etc. [1]). All these factors (such as large amount of cells and many complex inner processes to be calculated) make HEC 3D calculations too heavy (in comparison with other software packages). It leads to the fact that some models with complex structure and sophisticated spatial hierarchy are calculated for several days.

In this study, we have developed several high-performance versions for HEC. Parallelization efficiency was obtained up to 180% while acceleration was obtained up to 60 times.

Methods and Algorithms: Parallelization was implemented using MPI for computer clusters and SMP-machine usage and QTConcurrent technologies for desktop usage.

Results:

HEC was optimized using MPI and QTConcurrent technologies.

The speed of the original program was increased up to 2.8 times using QtConcurrent on desktop machine with the following specification: AMD PhenomTM II X6 1055T (6 cores), 2.8 Ghz, 12Gb RAM.

The speed of the original program was increased up to 60 times with 140% of parallel efficiency (testing simple models) and up to 40 times with 80% efficient (testing complex models) using MPI version.

The time of calculations was decreased from several days to several hours.

Analysis of parallelization efficiency showed that in case of 1D model calculations it was reached up to 180% (complex model) and up to 140% (simple model); in case of 2D model calculations it was reached up to 145% (complex model) and up to 60% (simple model); in case of 3D model calculations it was reached up to 50% (simple model) and up to 120% (complex model).

Analysis of data match was held and it showed that the output data of sequential and parallel versions are equal.

Conclusion: The optimized versions of the HEC 3D allow users to reduce the time of calculations in comparison with the original program.

References:

1. S.A. Lashin et al. (2014) HEC 2.0: improved simulation of the evolution of prokaryotic communities, *In Silico. Biol.* 11: 125-135.

ALTORFEV: A NOVEL TOOL FOR PREDCITION OF ALTERNATIVE ORFS BASED ON THE LINEAR SCANNING MODEL

B.S. Zuraev[1]*, A.V. Kochetov[1,2], A.I. Klimenko, S.A. Lashin[1,2]
[1]*Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2]*Novosibirsk State University, Novosibirsk, Russia*
** Corresponding author: bulatzuraev@gmail.com*

Key words: *Cytoscape plugin; ortholog; paralog; metabolic pathway; gene regulatory network; evolution; phylostratigraphy; evolution*

Motivation and Aim: The ribo-seq and proteomics techniques have revealed a large number of alternative ORFs (altORFs) within eukaryotic mRNAs. Some bioinformatics resources were developed to explore the available ribo-seq data to locate altORFs within mRNAs of interest (e.g., Ribotools, RiboGalaxy, GWIPS-viz). Indeed, knowledge on the full set of polypeptides encoded by a eukaryotic gene under study is essential for detailed investigation of its functions. However, published ribo-seq data are still very limited and conventional nucleotide sequence databanks do not provide information on the altORFs. In addition, the individual genetic variants may cause changes in mRNA coding potential: if a nucleotide sequence of mRNA under study is non-identical to the available ribo-seq-checked reference sequence, the positions of altORF(s) and their relative translation rates may differ. Thus, development of new tools for altORFs prediction remains quite actual. However, an accurate prediction of altORFs is very complicated because of a large number of various parameters influencing their recognition and translation efficiency.

Methods and Algorithms: The altORFev is based on the linear scanning model of translation [1]. It also considers the leaky scanning and reinitiation mechanisms. In brief, 40S ribosomal subunits bind to 5'-end of mRNA and move linearly along mRNA until start AUG codon is found. The probability of AUG recognition depends on its nucleotide context: start codon in the optimal context is recognized by the majority of 40S ribosomal subunits. Thus, if AUG codon is located in the optimal context and its ORF is larger than 30 codons, this ORF is defined as "terminal" since the majority of incoming 40S ribosomal subunits can't move beyond it. If AUG codon is located in a suboptimal context, some 40S ribosomal subunits will recognize it and initiate translation, whereas others skip it and may initiate translation downstream (leaky scanning). Finally, if AUG is located in the optimal context but the ORF size is small (lesser than 30 codons), the reinitiation is possible: in this case, some 40S ribosomal subunits after termination of translation of small ORF remain connected to mRNA and may continue movement in 3'-direction. During scanning they restore their initiation competence by acquiring the lacked eIFs and met-tRNAi and may initiate translation further downstream.

Results: We have implemented two versions of the altORFev: (1) web application (Java 1.8, Vaadin); (2) desktop application (Java 1.8, Swing).

Conclusion: The altORFev may be used to get additional information on eukaryotic genes taking into consideration alternative coding abilities of their mRNAs.

Availability: web-version: http://www-bionet.sscc.ru:7780/AUGWeb/, desktop version: upon the requests to the authors.

Acknowledgements: The study is supported by the RSF 14-24-00123 grant.

References:

1. Kozak,M. (2005) Regulation of translation via mRNA structure in prokaryotes and eukaryotes. *Gene*, 361:13–37.

Author Index

Научное издание

MATHEMATICAL MODELING AND HIGH PERFORMANCE COMPUTING IN BIOINFORMATICS, BIOMEDICINE AND BIOTECHNOLOGY (MM-HPC-BBB-2016)

Abstracts
на английском

Printed without editing

«МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ И ВЫСОКОПРОИЗВОДИТЕЛЬНЫЕ ВЫЧИСЛЕНИЯ В БИОИНФОРМАТИКЕ, БИОМЕДИЦИНЕ И БИОТЕХНОЛОГИИ»

Тезисы докладов

Составители: С.И. Кабанихин, М.А. Шишленин, Д.А. Воронов, А.Ю. Белоног

Выпуск подготовлен информационно-издательским отделом ИЦиГ СО РАН.
Начальник отдела Т.Ф. Чалкова
Дизайн и верстка: А.В. Харкевич
Дизайн обложки: Е.С. Кабанихина

Подписано к печати 17.08.2016. Формат 70 × 108 $^{1}/_{16}$. Усл. печ. л. 9,63.
Тираж 100 экз. Заказ № 207

Федеральный исследовательский центр
«Институт цитологии и генетики Сибирского отделения Российской академии наук»
630090, г. Новосибирск, проспект Академика Лаврентьева, 10

Отпечатано в типографии ФГУП «Издательство СО РАН»
630090, г. Новосибирск, Морской проспект, 2