RUSSIAN ACADEMY OF SCIENCES
SIBERIAN BRANCH

INSTITUTE OF CHEMICAL BIOLOGY
AND FUNDAMENTAL MEDICINE

INTERNATIONAL CONFERENCE ON POSTGENOMIC TECHNOLOGY FOR BIOMEDICINE

Abstracts

PTB'12
Novosibirsk, Russia
June 25–29, 2012

Novosibirsk
2012

PROGRAM COMMITTEE

Valentin Vlassov
Institute of Chemical Biology and Fundamental Medicine SB RAS, Novosibirsk, Russia
(Co-Chairman of the Conference)

Renad Sagdeev
International Tomographic Center SB RAS, Novosibirsk, Russia
(Co-Chairman of the Conference)

Vadim Govorun
Research Institute for Physical-Chemical Medicine, Moscow, Russia
(Co-Chairman of the Conference)

Alex Borovkov
Synbuild LLC, USA

Maxim Filipenko
Institute of Chemical Biology and Fundamental Medicine SB RAS, Novosibirsk, Russia

Aleksander Gabibov
Institute of Bioorganic Chemistry RAS, Moscow, Russia

Marsel Kabilov
Institute of Chemical Biology and Fundamental Medicine SB RAS, Novosibirsk, Russia
(Academic Secretary)

Valery Puzyrev
Research Institute of Medical Genetics SB RAMS, Tomsk, Russia

Dmitry Pyshnyi
Institute of Chemical Biology and Fundamental Medicine SB RAS, Novosibirsk, Russia

Vladimir Richter
Institute of Chemical Biology and Fundamental Medicine SB RAS, Novosibirsk, Russia

Alexander Sinyakov
Institute of Chemical Biology and Fundamental Medicine SB RAS, Novosibirsk, Russia

Dmitry Stetsenko
University of Reading, Reading, UK

LOCAL ORGANIZING COMMITTEE

Marsel Kabilov, ICBFM SB RAS (Chairperson)

Alexander Chernonosov, ICBFM SB RAS

Irina Hasanova, ICBFM SB RAS

Irina Zhilovskaya, ICBFM SB RAS

ISBN 978-5-91291-012-8

ORGANIZERS

Institute of Chemical Biology and
Fundamental Medicine SB RAS

Siberian Branch of the Russian Academy
of Sciences

"METABOLOMICS"
Limited Liability Company

WITH SUPPORT

Russian Foundation for Basic Research

Program of SB RAS
"Genomics, Proteomics, Bioinformatics"

SPONSORS

GOLD SPONSOR

ДИА•М
современная лаборатория

www.dia-m.ru
заказ on-line

Roche

ХИМЭКСПЕРТ

SILVER SPONSORS

БГА
БиоГен-Аналитика
ваши задачи - наши решения
www.bga.su

ИнтерЛабСервис
www.interlabservice.ru

life technologies™

OLYMPUS
Your Vision, Our Future

helicon

Научное оборудование

ГРУППА КОМПАНИЙ
ВСМ
БИОХИММАК

BRONZE SPONSORS

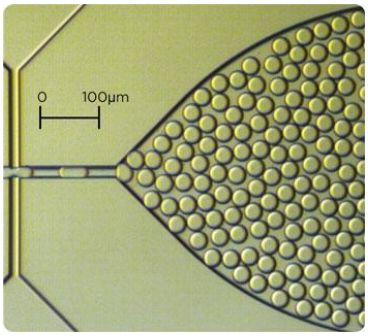
ОПТЭК
Объединяя решения

BIO-RAD

Микрофлюидика: система приготовления капель и эмульсий

Dolomite

Микрофлюидика - развивающаяся технология, революционно изменяющая облик лабораторных приборов и использующая концепцию построения "лаборатории-на-одном-чипе".



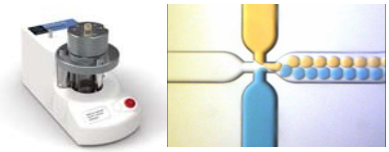
Возможности:

- Точные манипуляции с микрообразцами жидкостей, например:
  - воспроизводимое формирование микрокапель микро- и пиколитрового объема;
  - получение "капель в капле".
- Возможность работы с объектами на микроуровне: с микрочастицами, микрокаплями и с живыми клетками.
- Возможность изготовления конфигурации чипа "на заказ".

Преимущества:

- уменьшение объема образцов и количества используемых реагентов;
- точный контроль над процессами в жидкостях, например, нагревом / охлаждением, скоростью потока, давлением жидкости, скоростью перемешивания, протеканием химических реакций;
- широкий выбор геометрии микрфлюидики, разнообразные интерфейсы и коннекторы;
- высокий массоперенос из-за высокого отношения площади поверхности к объему жидкости;
- превосходная воспроизводимость;
- интеграция отдельных стадий процессов, например, реакций, отделения продукта, детекции.

dolomite



Диаэм с 1988 года поставляет оборудование, расходные материалы, наборы и реактивы для нужд молекулярно-биологических, генетических и криминалистических лабораторий, решения для Single Cell PCR, для персональной медицины, клеточной биологии, роботы для отбора отдельных клеток и т.п.

Диаэм в Москве:
ул. Космонавта Волкова, д. 10
тел./факс: (495) 745-0508
info@dia-m.ru

Диаэм в Новосибирске:
Академгородок, пр. Ак. Лаврентьева, 6 / 1
тел / факс: (383)328-00-48
nsk@dia-m.ru

Диаэм в Казани:
Оренбургский тракт, д.20, оф.217
тел./факс: (843) 277-60-40
kazan@dia-m.ru

www.dia-m.ru
заказ on-line

ХИМЭКСПЕРТ

Агентство «Химэксперт»
info@khimexpert.ru www.khimexpert.ru
(499) 973-92-80, 972-06-90
127006, г. Москва, ул. Краснопролетарская, д. 7

Официальный дилер AB Sciex и Life Technologies, основной поставщик реактивов и расходных материалов Applied Biosystems и Ambion.

Предлагаем оборудование для анализа ДНК и РНК, фундаментальных протеомных исследований, фармацевтики и биотехнологии, стандартного прикладного тестирования, включая идентификацию личности и установление родства в криминалистике и судебно-медицинской экспертизе, тестирование пищевых продуктов.

- **АМПЛИФИКАТОРЫ**. Applied Biosystems разрабатывает и производит термоциклеры с 1987 года. На нынешний день широкий модельный ряд включает как несложные и недорогие модели, так и новейшую систему, обладающую шестью независимыми температурными зонами, дающими возможность быстро оптимизировать ПЦР и проводить различные ПЦР в одном блоке. Все приборы надежны в использовании, точны и просты в управлении. Вы можете выбрать модель, наилучшим образом отвечающую потребностям вашей лаборатории.
- **СИСТЕМЫ ПЦР В РЕАЛЬНОМ ВРЕМЕНИ.** Applied Biosystems совершенствует технологии систем Real-time PCR уже более 10 лет. Любые современные задачи: анализ экспрессии генов и микроРНК, типирование SNP, выявление транслокаций, определение вирусной нагрузки и т.п. – могут быть решены с применением оборудования Applied Biosystems. Нынешнее пятое поколение систем (ViiA7 и QuantStudio) обладает еще более широким спектром возможностей и рядом уникальных характеристик. Система ViiA7 способна детектировать 21 флуоресцентный краситель, а система QuantStudio позволят проводить цифровую ПЦР (Digital PCR) и работать с OpenArray-слайдами.
- **ГЕНЕТИЧЕСКИЕ АНАЛИЗАТОРЫ**. Applied Biosystems является безусловным лидером в производстве генетических анализаторов (секвенаторов) – специализированных систем капиллярного электрофореза с оптической детекцией флуоресцентного сигнала. Все имеющиеся модели оптимизированы для решения полного спектра задач: определение структуры ДНК, различных типов фрагментного анализа (LOH, AFLP, поиск SNP, STR-генотипирование и др.). Компания предлагает приборы различной производительности: от монокапиллярной системы ABI Prism 310 до 96-ти капиллярной AB 3730xl. Все приборы могут применяться при проведении геномной дактилоскопии (идентификация личности, установление родства), микробиологических, научных и медицинских исследований, а также исследований в сельском хозяйстве.
- **ПРИБОРЫ НОВОГО ПОКОЛЕНИЯ.** Applied Biosystems производит приборы, позволяющие производить полногеномное секвенирование ДНК – SOLiD 5500 и 5500xl с производительностью 90 и 180 млрд. нуклеотидов за запуск соответственно. Это высокопроизводительные приборы, принцип действия которых основан на секвенировании методом последовательного лигирования флуоресцентных олигонуклеотидных проб с последующей регистрацией флуоресцентного сигнала 4-х различных цветов.
  В конце 2012 года появится прибор SOLID 5500W, для которого предусмотрена линейная изотермическая амплификация на поверхности чипа, позволяющая существенное упростить процесс пробоподготовки.
  Также к приборам нового поколения относятся системы Personal Genome Machine™ (PGM) или Ion Torrent и Ion Proton™ Sequencer – приборы для высокопроизводительного секвенирования ДНК без флуоресцентной детекции. Принцип работы: последовательное удлинение олигонуклеотидной затравки ДНК-полимеразой с одновременной регистрацией локального изменения pH в результате встраивания нуклеотидов. Расходный материал: полупроводниковые сенсорные микрочипы с разной производительностью за запуск. С помощью системы PGM можно секвенировать небольшие геномы или отдельные гены с высоким покрытием, система Ion Proton позволит в 100 раз повысить производительность и секвенировать геномы сложных организмов. Несмотря на меньшую, по сравнению с системами SOLiD, производительность, приборы имеют радикальное преимущество перед всеми полно-геномными системами за счет недорогих реактивов.
- **СИСТЕМЫ ВЭЖХ/МС/МС** производства AB Sciex. Спектр предлагаемых систем включает классические тандемные квадрупольные масс-спектрометры, гибридные системы с линейной ионной ловушкой, времяпролетные (TOF и TOF/TOF) «протеомные анализаторы», применяемые для широкого спектра задач – анализ объектов окружающей среды, продуктов питания, фармакологических исследований, криминалистической экспертизы, исследований пептидов и белков, поиска биомаркеров и т.д.

AB SCIEX life technologies™ Ambion® THE RNA COMPANY®

Contents

Computational and experimental genomics

Computational and experimental proteomics

Human genetics and personalized medicine: theoretical and experimental aspects

Synthetic biology

Bionanotechnology for diagnosis and therapy

Therapeutic proteins and nucleic acids

Abstracts received after dead-line

SYSTEMATIC ERRORS AND BIASES IN ILLUMINA SEQUENCING

Abnizova I., Leonard S., SkellyT., Jackson D.
Wellcome Trust Sanger Institute, Hinxton, UK

Key words: *NGS Illumina sequencing, systematic errors, context-dependency*

Motivation and Aim: Sequencing of individual genomes and the determination of rare variants across populations are enabled by whole genome sequencing at low cost [1]. However, whole genome sequencing is accompanied by higher error rates [2,3] than capillary sequencing. Improved methods that accommodate these high error rates are needed in the calling of heterozygous sites from low coverage data. The design of effective statistical methods requires precise characterization of error in high-throughput sequence data.

Together with well known error tendencies in Illumina sequencing: phasing inaccuracy and cross-talk,- there is considerable amount of evidence about context and positional dependency for Illumina errors now. In addition to context dependency, there was found [4] strong variability of error profiles and rates between different Illumina versions (v4 vs v5), machines, runs and even the first and second reads in a pair for the same run and machine.

The error rate for base calling is commonly measured by confidence value (QV Phred [5]) of a base call. The process of generating QV for a base call is done by so called 'calibration' method. There are several well known calibration methods accepted in NGS community, the most established is PHRED method, which is implemented by Illumina's (Gerald and Bustard), and by GATK re-calibration.

*Methods and Result***.** In this work we introduce simple, fast PHRED-inspired method of calibration, PB-calibration, which takes care of mentioned above error dependencies and variability. We compare common recalibration methods, and show how run-specific PB-calibration helps to distinguish systematic errors from possible variants. The PB-calibration of a base call, combined with mismatch recalibration, can reliably identify around 30-40% of high quality mismatches, which cannot be marked with any other calibration methods, so far as we know. In general, PB-calibration may be applied to any NGS platform. It is currently reference-based, but can be easily adjusted to contig-scuffold-based (or any other) assembly method. PB-based error correction can be applied even without any available reference.

*Availability***.** PB-calibration is currently implemented in the production pipeline for spiked runs at Sanger Institute, and is publicly available from authors by request.

References

1. Nielsen R: Genomics: In search of rare human variants. Nature 2010,467(7319):1050-1051.
2. Hoff K: The effect of sequencing errors on metagenomic gene prediction. BMC Genomics 2009, 10:520
3. Dohm JC, Lottaz C, Borodina T, Himmelbauer H: (2008)Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. Nucleic Acids Research, 36(16):e105.
4. McElroy et al. (2012) GemSIM: general, error-model based simulator of next-generation sequencing data, BMC Genomics, 13:74
5. Ewing, B. & Green, P.(1998) Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Res. 8, 186–194.

COVERAGE DEPTH ANALYSIS IN NEXT GENERATION SEQUENCING DATA

Amstislavskiy V.S., Sultan M., Kim K., Schrinner S., Lehrach H., Yaspo M.-L.
Max Planck Institute for Molecular Genetics, Berlin, Germany
e-mail: amstisla@molgen.mpg.de

Key words: *next generation sequencing (NGS), exome enrichment, whole-genome sequencing (WGS), coverage depth, DNA complexity, copy number variation*

Motivation and Aim: NGS is now a common and widely used approach for a comprehensive analysis of the genetic information in health and disease. Specialized protocols (e.g. WGS, RNA-Seq, Exome sequencing, ChIP-Seq etc.) allow scientists to investigate sequence information at the genome and transcriptome levels. But despite of expanding NGS applications and analysis tools, the biases introduced by the sequencing methods and/or by the context properties (GC-content, complexity) of the genomic regions of interest are not yet fully understood. Herein we developed an interactive web-based tool enabling the online visualization of the distribution of coverage along chromosomes, exons, transcripts or regions of interests for different sequencing applications. This tool allows the visualization of the distribution of sequence reads and helps to estimate the impact of context properties on the data quality through an intuitive web interface. It further helps to identify at a glance copy number variation in e.g. cancer related whole genome sequencing data.

Methods and Algorithms: Raw sequence data underwent quality controls (custom algorithm, fastqc) and mapped using BWA/samtools on the UCSC hg19 reference genome. The coverage profiles are being generated using pileup files and an annotation file (NCBI36, Ensembl 62). The Python matplotlid module was used for graphics and php/mysql - for web-programming.

Results: The tool developed herein was used to analyze the coverage depth of sequence data produced in the context of Oncotrack, a European consortium project aiming at identifying new markers for colon cancer. We used results from WG and mRNA sequencing experiments generated on the Illumina HiSeq2000 platform as well as exome enriched sequence data produced on the SOLiD 5500 platform. We built graphs with simultaneously representation of the average depth of coverage in a given bin as well as the fraction of bin positions covered at least ones. Additionally, the bin GC-content and informational entropy curves were overlaid. We observe that in general the coverage correlates with the informational entropy: it is lower in case of GC-content different than 50%. However, in some regions of WGS sharing the same complexity, we observe significant coverage differences, which are linked to copy number variation events occurring in the tumor sample. We further observe that increasing the number of sequences per experiment by additional sequencing has no drastic effect on the profile of the coverage curve. However enrichment rates in exome protocols and DNA quality significantly affects the average coverage.

Conclusion: The tool will be freely available on the web requiring no prior installation to be used. As an input it requires only a pileup file of the data to be analyzed. It is easy to use, and it is platform independent, which makes it useful for quick checking of the quality of sequencing data, coverage estimation and fast capturing of potential regions of interest for the further analysis.

PATHWAY SIGNAL FLOW ANALYSIS FOR HIGH-THROUGHPUT GENE EXPRESSION DATA

Arakelyan A A.*
Institute of Molecular Biology NAS RA, Yerevan, Armenia
e-mail: aarakelyan@sci.am
** Corresponding author*

Key words: *high-throughput gene expression, data analysis, molecular pathway mining*

Motivation and Aim: Molecular pathway sources (such as KEGG and Biocarta) may become the most useful tools for gene expression analysis. To date, in most cases pathway mining simply means mapping and coloring nodes in pathways without evaluation of changes of flows through the network. Herein, we introduce a new algorithm for functional annotation and biological interpretation of gene expression data based on in-depth pathway analysis.

Methods and Algorithms: The Pathway Signal Flow (PSF) algorithm evaluates how a signal from network inputs spreads downstream to the outputs depending on relative expressions of nodes (R) and interaction types between them. The signal flow between two connected nodes is calculated as:

$R_{(node1)} * R_{(node2)}$, *if node 1 activates node 2, and* $R_{(node\ 1)}/R_{(node\ 2)}$, *if node 1 inhibits node 2*.

The PFS value is the mean of signal flows at output nodes. Empirical probability distribution of PSFs is used to calculate significance of PSF values.

Results: We have reanalyzed data from several microarray datasets related to pulmonary sarcoidosis and compared with published results. Analysis of gene expression was performed using growing support sets algorithm [1] combined with PSF analysis. Results indicate that sarcoidosis is characterized by up-regulation of pathways related to pro-inflammatory response, such as Fc (IgG) receptor mediated phagocytosis, focal adhesion, chemokine signaling and T-cell receptor signaling. Eminently, among the different possible functional consequences of pathway activation, PSF was able to detect biologically meaningful outcomes. For example, chemokine signaling pathway may induce different responses in cells such as activation and migration of immune cells, expression of pro- or anti-inflammatory mediators, apoptosis. In sarcoidosis, we detected up-regulation of chemokine pathway branches related to cell proliferation and migration, which is in agreement with previously reported experimental results [2].

Conclusion: The PSF analysis is an attempt to achieve deeper level of biological interpretation of gene expression data and provide more biologically meaningful results compared to existing techniques. Moreover, our algorithm is suited for all current techniques for high-throughput gene expression experiments.

Availability: MATLAB code for PSF analysis is available upon request from authors.

References

1. V. Palchevskiy et al. (2011) Immune response CC chemokines CCL2 and CCL5 are associated with pulmonary sarcoidosis, *Fibrogenesis Tissue Repair,* **4:** 10.
2. A. Arakelyan et al. (2010) Growing support set systems in analysis of high-throughput gene expression data, In: *New Trends in Classification and Data Mining*, K. Markov et al (eds)**,** 47-53, ITHEA.

EVALUATION OF GENOMIC INSTABILITY IN SEVERAL SPECIES OF MAMMALS USING THE MICRONUCLEI TEST

Astafieva E.E.*, Karpushkina T.V., Kulikova K.A., Glazko T.T.
Russian State Agrarian University – Moscow Agricultural Academy named after K.A. Timiryazev (RSAU–MTAA), Moscow, Russia
e-mail: k-astafeva@mail.ru
** Corresponding author*

Key words: *micronuclei, genomic instability, species, environmental factors*

Motivation and Aim: Micronuclei test in somatic cells allows to evaluate the genomic instability and predict the reproductive "success" of humans and animals [1, 2]. It is widely used for bioindication of genotoxic environmental effects. However, the species specific traits of the micronuclei test results, the influence of duration of environmental factor action still remain insufficiently studied. In this regard the comparative analysis of the micronuclei test in erythrocytes of the peripheral blood of domesticated animals (cattle, sheep, goats, horses), semi domesticated (yaks) and wild species (musk ox) was carried out. Groups of cattle, sheep, goats, horses, musk ox reproduced in relatively favorable environmental conditions and in the area of high-risk animal breeding of the Southern Gobi Desert (Mongolia) had been compared.

Methods and Algorithms: Blood smears were prepared by mixing a drop of peripheral blood with a drop of saline solution (1:1) on a glass slide then spreading it over a glass slide. Preparations were fixed with methyl alcohol and then stained with Giemsa (Merk). The number of erythrocytes with micronuclei were calculated under a microscope Motik (DMBA300) with a built-in digital camera (x1000) in 3000 cells and expressed in ppm (‰). Statistical significance was evaluated by Student's test (t_S).

Results: The frequency of erythrocytes with micronuclei was significantly higher ($P < 0.05$) in domesticated species compared to the wild species (musk ox). The increase is observed in the following order: musk ox (0.3 ± 0.2 ‰), horses (2.4 ± 0.1 ‰), yak (3.2 ± 0.6 ‰), cattle (4.6 ± 0.7 ‰), goats (4.6 ± 0.4 ‰), sheep (5.3 ± 0.4 ‰). The lowest values of the frequency of erythrocytes with micronuclei were detected in musk oxen, the highest – in sheep (birliksky type of the edilbai sheep – 5.2 ± 0.2 ‰, the Kalmyk breed – 4.3 ± 0.3 ‰, Mongolian sheep – 5.3 ± 0.4 ‰). There were no significant differences in the interbreed micronuclei test results, but animals of different species reproduced in the conditions of zone of high-risk animal breeding (South Gobi, Mongolia) had the statistically significant lower levels of the frequency of red blood cells with cytogenetic anomalies (yaks – 0.3 ± 0.2 ‰, cattle – 1.8 ± 0.6 ‰, goats – 0.9 ± 0.2 ‰, sheep – 0.9 ± 0.1 ‰) compared to the other animals reproduced in relatively more favorable conditions.

Conclusion: A trend to relatively increased genomic instability in domesticated animals compared to the wild ones was revealed. Micronuclei test results indicated also that in the zone of high-risk animal breeding in Southern Gobi Desert the animals with relatively higher genomic stability achieve benefits for the reproduction in generations.

References

1. L. Migliore et al. (2006) Relationship between genotoxicity biomarkers in somatic and germ cells: findings from a biomonitoring study, *Mutagenesis*, Vol. 21. No. **2**: 149-152.
2. J. Rubes et al. (1991) Somatic chromosome mutations and morphological abnormalities in sperms of boars, *Hereditas*, **115**: 139-143.

BIOMARKER CHALLENGE: A CLOUD INSTEAD OF A SET OF THE VANTAGE POINTS

Baranova A.V.
Russian Center of Medical Genetics RAMS, Moscow, Russia,
School of Systems Biology, George Mason University, Fairfax, VA, USA
e-mail: abaranov@gmu.edu

Key words: *biomarkers, prognosis, cancer, whole transcriptome analysis, continuous prognosis model*

Motivation and Aim: To date, the quantification of the diagnostic and prognostic biomarker molecules in the human serum and tissues remains the primary means of enhancing the clinician's ability to predict and detect cancer before it spreads and to predict the outcome of treatment. Importantly, with innumerable molecular markers in development, the discovery of novel standalone markers with acceptable sensitivity and specificity is an extremely rare event. The conventional method to overcome the problem of relatively low sensitivity and specificity of newly discovered biomarkers is to combine them into biomarker panels. However, in many cases these biomarker panels suffer from relatively low reproducibility of results in independently collected sets of samples. This is especially true for the mRNA biomarkers identified by microarray experiments.

Methods and Algorithms: We challenged the biomarker paradigm by developing a distance measure between the entire gene expression profile of a tumor and the center of the space occupied by normal samples. This novel concept allows one to depart from the classical two-bin prediction model (e.g. "bad prognosis/good prognosis") as it produces a continuous prognosis model, where each sample is located in the neighborhood of other samples analyzed post-hoc and associated with known survival.

Results: Whole-transcriptome based distances calculated using Pearson correlation coefficients provide easy visualization of the relative degree of the malignancy characteristic for studied samples. In all studied datasets, on average, tumors were further away from the Normal Sample Space than the paired samples with normal histology. The distance analysis demonstrated remarkable behavioral invariance observed in eighteen independent tumor data sets and provided a robust validation of this approach. The concept of distance analysis is not limited to cancer as it could be generalized to quantify the departure of any given sample from its reference set, i.e. tissue sample of aged persons from reference of non-aged, samples of insulin resistant tissues from normally functioning tissues, and even model cell lines that drift away from the standard phenotype.

Conclusion: If successful, this unconventional approach will shift the tumor biomarker paradigm from expression biomarker panels associated with low reproducibility, to the distance analysis of robust molecular portraits. The proposed distance analysis is versatile in its application as it will be equally attributable to gene expression profiles collected both by microarrays and by RNA-seq platforms.

Availability: on collaborative request.

Acknowledgements: Dr.Ganiraju Manyam (MD Anderson Cancer Center, TX, USA), Dr. Alessandro Jiuliani (Istituto Superiore di Sanita, Roma, Italy), Dr. Boris Veytsman and Lei Wang (George Mason University, Fairfax, USA).

SMALL NON-CODING RNAs OF HUMAN BLOOD PLASMA OF HEALTHY DONORS AND PATIENTS WITH NON-SMALL CELL LUNG CANCER

Baryakin D.N.*[1], Semenov D.V. [1], Brenner E.V. [1], Kurilshikov A.M. [1], Kozlov V.V.[2], Narov Y.E. [2], Vasiliev G.V. [3], Bryzgalov L.O. [3], Chikova E.D. [1], Filippova J.A. [1], Kuligina E.V. [1], Richter V.A. [1]

[1] Institute of Chemical Biology and Fundamental Medicine SB RAS, Novosibirsk, Russia;
[2] Novosibirsk Regional Cancer Centre, Novosibirsk, Russia;
[3] Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia
e-mail: baryakindn@niboch.nsc.ru
** Corresponding author*

Key words: *high-throughput sequencing, circulating RNAs, miRNA, non-small cell lung cancer*

Motivation and Aim: Circulating nucleic acids are subject of numerous modern researches aimed at establishing of new mechanism for distant regulation of physiological processes, and at using of extracellular nucleic acids as diagnostic and prognostic markers of pathological processes in the organism.

The aim of this study is a detailed description of structure and forms of short extracellular RNAs of human plasma, aimed at identifying new forms of regulatory RNA and the establishment of the mechanism of their action, and the development of unique and complex diagnostic markers of human diseases.

Methods and Algorithms: In this study we analyze the diversity of short non-coding RNA blood plasma of 8 healthy volunteers and 8 patients with non-small cell lung cancer. In order to obtain cDNA libraries that encode the most full-scale set of circulating RNAs, short (n> 19) plasma RNAs were exposed to dephosphorylation followed by 5'-phosphorylation, ligation with adapters, reverse transcription and amplification. Individual cDNA libraries were sequenced on a platform SOLiD (V.3). The resulting data sets were analyzed using the Bowtie/Cufflinks software.

Results: It was found that human blood plasma contains fragments of rRNA, tRNA, mRNA transcripts of the mitochondria, the mature miRNA, scRNA, snRNA, snoRNA, as well as fragments of transcripts not annotated previously.

A detailed analysis of miRNAs and miRNA-like forms including determination of the most represented form of known human miRNAs in the blood plasma, finding new miRNA-like forms using the software mirDeep 2.0; identifying potential mRNA target of circulating miRNAs and candidate miRNAs.

Conclusion: Comparative analysis of plasma miRNA of healthy donors and patients with non-small cell lung cancer allowed us to characterize a set of changes in expression profile of extracellular human microRNAs at the origin and development of malignant tumors.

Acknowledgements The work was supported by the Russian Ministry of Science and Education (02.740.11.0715); RFBR grants № 10-04-01442-a; № 10-04-01386-a.

"RANDTRAN": RANDOM TRANSCRIPTOME SEQUENCE GENERATOR

Borzov E.A.*[1], Marakhonov A.V.[1], Baranova A.V.[1, 2], Skoblov M.Yu.[1, 3]
[1] Federal State Budgetary Institution "Research Centre for Medical Genetics" under the Russian Academy of Medical Sciences, Moscow, Russia;
[2] School of Systems Biology, College of Science, George Mason University, Fairfax, VA USA;
[3] State Budgetary Institution of Higher Education "Moscow State Medical and Dental University", Moscow, Russia
e-mail: eborzov@generesearch.ru
** Corresponding author*

Key words: *random sequence, transcriptome, generator*

Motivation and Aim: Random transcriptome generation is a frequently used method in modern bioinformatics researches. However algorithms of its creation are not widely discussed in the literature and the concept of making of random transcriptome is often limited to conserve mononucleotides frequency. Our aim was to make the program for random transcriptome generation that takes into account the structure of transcript (5'-UTR, CDS, 3'-UTR, or ncRNA) with its own dinucleotide and trinucleotide (codon) frequencies for different species.

Methods and Algorithms: We have used classic Monte-Carlo methods. The program is written in Perl using Tk Module for interface (http://www.cpan.org/).

Results: The program "RANDTRAN" has user-friendly interface and flexible options to create your own random transcriptome. It used a standalone .txt frequency file (2KB) that defined a lot of parameters for specific organism. For the moment the program is packaged with 23 files (for 11 eukaryotes and 12 bacteria and archea species). The frequency file contains a lot of parameters described special features, such as dinucleotide and trinucleotide frequencies for each transcript region, length distribution data of each part of transcript, quantity of coding and non-coding transcripts. The advantage of our program is to allow changing of above arguments or creating new frequency file that gives a user the opportunity to generate the transcriptome with necessary characteristics. The detailed manual is attached.

Availability: Download from our website http://www.generesearch.ru/research.html

MASSIVE PARALLEL EXON SEQUENCING AS FUNDAMENTAL APPROACH IN STUDYING SNPs THAT CAN LEAD TO ALZHEIMER DISEASES

Boulygina E.S.[1], Nedoluzhko A.V.[1], Tsygankova S.V.[1], Tchekanov N.N.[2], Mazur A.M.[2], Artemov A.V.[3], Prokhorchuk E.B.[1, 2], Skryabin K.G.[1, 2]

[1] *National Research Center "Kurchatov Institute", Kurchatov Sq 1, Moscow, Russia;*

[2] *Centre "Bioengineering" of the Russian Academy of Sciences, Moscow, Russia;*

[3] *T-gene LTD, Dubna, Moscow region, Russia*

The members of Russian family that had Alzheimer disease in three generations were used to study genetic variations that are caused of this disease. Analysis of 249 Alzheimer associated genes was performed using SureSelect approach capturing the coding regions of the genes (including 200 bp intron sites), 5' promoter regions (up to 1000 bp in 5' region from the start point of gene transcription) and 3' noncoding regions (up to 1000 bp in 3' region from a site of polyadenilation of mRNA). Trapped DNA was sequenced with SOLiD 4.0 («Applied Biosystems», the USA). The analysis of the received data has shown, that 80 % of all short reads mapped on the chosen sites of human genome (hg18) with average depth of a coverage 313. SNP calling revealed 2335 SNPs (663 (31 %) within exons) and 166 short inserts/deletions (33 within exons) in analyzed sites. Neither nonsynonymous SNP's nor short indels were found within crucial Alzheimer associated genes (APP, PSEN1, PSEN2 and APOE).

Using the software developed by us the polymorphisms which are characteristic for all family N members with diagnosed Alzheimer's disease have been selected. Also the absence of contradictions in the found genotypes between parents and their children was checked. Taking into account the above described restrictions only those polymorphisms which are in coding regions of genes and lead to amino acid replacement (missense) or to occurrence of a stop codon in a gene (nonsense) have been selected for the further analysis. Then each of selected has been checked up using SIFT and ANNOVAR programs to evaluate its influence on corresponding protein sequence. As a result only 16 SNPs, having the maximum probability of association with Alzheimer's disease have been chosen for the subsequent analysis. The confirmation of those SNPs were done by Senger sequencing on ABI 3730. The distribution of these SNPs within Russian population is currently studied.

Potential variants which can lead to development of this disease are revealed, and some of these variants are new and were not described previously. Information on variants of genes will allow to learn more about molecular etiology of disease, and also to use simple methods of diagnostics (PCR) for personal predisposition to the disease.

HIGH-THROUGHPUT SEQUENCING OF MYCOBACTERIUM STRAINS USED FOR STEROID COMPOUNDS BIOSYNTHESIS

Bragin E.Yu.*, Ashapkin V.V., Shtratnikova V.Yu., Schelkunov M.I., Dovbnya D.V., Donova M.V.

Innovative Technology Center "Biologically Active Compounds and Application", Russian Academy of Sciences, Moscow;

G.K. Skryabin Institute of Biochemistry & Physiology of Microorganisms, Russian Academy of Sciences, Pushchino, Moscow Region

e-mail: bragory@yandex.ru

** Corresponding author*

Key words: *steroid bioconversion, Mycobacterium, whole-genome sequencing*

Motivation and Aim: Strains of *Mycobacterium* spp VKM Ac-1815D, 1816D, 1817D are used for bioconversion of phytosterol to androst-4-en-3,17-dione (AD), androsta-1,4-diene-3,17-dione (ADD) and 9a-hydroxy androst-4-ene-3,17-dione (9-OH-AD) respectively. Aim of this study - investigation of genes and operons of steroid catabolism of these strains.

Methods and Algorithm: High-throughput sequencing was performed with Genome Analyser IIx (Illumina), pair-end reads, 72+72 b.p. Genomes assembling was performed with Velvet software. Search of genes was made by Blast (NCBI) and NCBI Genome Workbench software. Operons was determined with internet-service FgenesB.

Results: We find out 36-38 transcriptional units associated with steroid catabolism in strains VKM Ac-1815D and VKM Ac-1816D, and 64 in VKM Ac-1817D. The key role in steroid catabolism (start of destruction of steroid rings) play 3-ketosteroid dehydrogenases (kstD) and 3-ketosteroid-9-alpha-hydroxylases (consist of kshA and kshB). A single gene kstD is in VKM Ac-1815D and in VKM Ac-1816D, a one SNP between kstD of these strains probably inactivate this gene in VKM Ac-1815D. 4 genes kstD are in VKM Ac-1817D, although kstD activity in this strain expected blocked. 2 genes kshA and 1 gene kshB are in strains VKM Ac-1815D and VKM Ac-1816D and no mutations between them. Strain VKM Ac-1817D have 5 genes kshA and 2 genes kshB, which can define high accumulation of 9-OH-AD. So biochemical characteristics of this strain should depend on activity of another genes or regulated regions.

Conclusion: We obtain long genome sequences of strains VKM Ac-1815D, 1816D, 1817D, and this will allow to investigate SNP and regulatory regions for understanding of steroid catabolism in these strains. We determine genes and operones included in the steroid catabolism of strains VKM Ac-1815D, 1816D, 1817D. It allows to select targets for improvement of biotecnological characteristics of these strains.

A NEW APPROACH TO IDENTIFY THE rSNPs IN THE HUMAN GENOME BASED ON CHIP-seq DATA

Bryzgalov L.O.*, Antontseva E.V., Matveeva M.Yu., Kashina E.V., Shilov A.G., Bondar N.P., Merkulova T.I.
Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia
e-mail: leon_l@bionet.nsc.ru
** Corresponding author*

Key words: *rSNP, ChIP-seq, regulatory region, human genome*

Motivation and Aim: The functional interpretation of non-coding disease-associated single nucleotide polymorphisms (SNPs) identified by genome-wide association studies (GWAS) is challenging. Many of these SNPs are likely to be regulatory SNPs (rSNPs): variations which affect the ability of a transcription factor (TF) to bind to DNA. However, experimental procedures for identifying rSNPs are expensive and labour intensive. In silico methods are required for rSNP prediction usually based on weight matrix (PWM), and can determined which SNPs are likely rSNPs but its not very well. We report a new approach for rSNP prediction based on ENCODE ChIP-seq data analysis.

Methods and Algorithms: ChIP-seq data sets were downloaded from ENCODE project site UCSC. Potential rSNPs were searched in the ChIP-seq peaks overlappings. SNP localized in more than 8 ChIP-seq peaks were introduced into EMSA experiments. Protein nuclear extracts from the different human cell lines (HeLa S3, HTC, K562 and HepG2) were prepared for the EMSA.

Results: Genomic data on the locations of SNPs in the human genome were downloaded from database dbSNP NCBI http://www.ncbi.nlm.nih.gov/snp/ (Human Genome Build 37). Sample Sclinic consisting of 4046 clinical submitted SNPs was selected from dbSNP NCBI by limits criteria "organism: Homo sapiens" and "annotation: Clinical/LSDB Submissions" with following exception of SNPs having references on the proteins. Sample Somim composed of 3160 SNPs was also selected from database dbSNP NCBI in criteria "organism: Homo sapiens" and "annotation: OMIM" eliminating every SNPs having references on the proteins.This samples of SNPs were analized. 799 and 438 SNPs are localized in at least one ChIP-seq peak, accordingly. 110 and 64 SNPs fell at more then 8 ChIP-seq peak. For experimental verification 41 SNPs were selected. It was established that in the EMSA 31 SNPs influence on the binding of transcription factors. Subsequent analysis of the different samples of SNPs from dbSNP showed that samples Sclinic and Somim are enriched by SNP localized in ChIP-seq peak compared to samples of random SNPs. The relative enrichment was rising when we increased number of ChIP-seq peak overlaps in the SNP location.

Conclusion: We developed a new method for rSNP prediction. Experimental verification showed high efficiency of this method. We suggest that it may be used for regulatory SNP prediction.

POPULATION STUDY OF THE VARIATION IN TRIPLET DISTRIBUTIONS OBSERVED ALONGSIDE A CHROMOSOME, FOR YEAST SPECIES

Bushmelev Eu.Yu.
Siberian federal university, Krasnoyarsk, Russia
e-mail: eugenijbushmelev@gmail.com

Key words: *order, periodicity, correlation, function, taxonomy*

Motivation and Aim: The aim of the study is to identify, describe and visualize both the intragenomec, and intergenomec differences in triplet distributions observed alongside a DNA sequence, for a family of yeast genomes.

Methods and Algorithms: The distribution of the triplets alongside a sequence was developed. It was defined as a distance to the nearest neighbour, where two triplets ω_1 and ω_2 are as far, as n_l nucleotides one each other so that there is no other word ω_2 embedded somewhere inside the string of the length n_l. The longest distance to detect the nearest neighbour was as long, as 10^5 nucleotides. The distribution function was developed for all 4096 couples of triplets, for each chromosome. Then standard techniques of statistical analysis have been implemented to figure out the correlations and interdependencies between the distributions observed in the different chromosomes of the same species, and different species.

Results: A number of chromosomes of yeast genomes have been studied. All chromosomes exhibit a strong and extremely unusual structures in the distribution of the triplets (to the nearest neighbour). Thus, an explicit and strong periodicity in CCC – GGG triplets has been found, with the period of 13. Some other couples exhibit more complex and long-range correlations (up to 250 nucleotides). There is significant correlation in the distribution patterns observed within a genome, and quite sounding divergence in the correlations, when compared the chromosomes from different (while closely related) species, or stains.

Conclusion: a new character is figured out to study the evolution processes of DNA molecules. It is evident, that the found behaviour of the triplet distribution yields some universal patterns, and some specific ones. Such specific patterns can be reliably related to taxonomy of a genome.

SEQUENCE AND ANNOTATION OF THE CHROMOSOME OF PROBIOTIC STRAIN *LACTOBACILLUS RHAMNOSUS* 24

Danilenko V.N.*, Poluektova E.U., Klimina K.M., Kjasova D.H., Chervinetz J.V., Malko D.B., Makeev V.J., Gusev F.E., Tyajelova T.V., Reshetov D.A., Rogaev E.I.
N.I. Vavilov Institute of General Genetics RAS, Moscow, Russia
e-mail: valerid@vigg.ru
** Corresponding author*

Key words: *chromosome sequencing, Lactobacillus rhamnosus, toxin-antitoxin*

Motivation and aim: The goal of this work is to determine complete nucleotide sequence of the genome of probiotic bacterium *L. rhamnosus* 24 and to identify essential genes underlying probiotic properties of this bacterium. The genetic modules of our particular interest are toxin-antitoxin systems[1] which can be potential regulators of bacterial probiotic properties. These systems have not yet been studied in lactobacilli.

Methods and Algorithms: The strain *L. rhamnosus* 24 was isolated from the oral cavity of a healthy child (collaboration with microbiology and immunology chair of Tver Medical Academy (Russia)). The biochemical, probiotic and biotechnology properties of the strain have been characterized. To elucidate further the potential for probiotic activity of the strain we performed the complete genome sequencing of the strain. We have prepared the libraries of fragmented genomic DNA and proceeded with sequencing according to Illumina's HiSeq2000 manufacturer's paired-end protocol with some modifcations. Series of contigs were assembled into complete genome sequence with SOAP denovo. Glimmer program was used for *ab initio* gene prediction. Functional genome annotation was transferred from orthogous genes which were predicted using UniProtKB database.

The cloning of toxin gene was performed into the expression vector pET32a and the gene construct was transformed in *E.coli* D3 strain cells.

Results. The genome of *L. rhamnosus* strain 24 consists of one circular chromosome 2 647 976 bp with ~46,0% GC context. Apparently, the strain has no plasmid DNA. The genome of strain 24 was compared with nine reference genomes of *L. rhamnosus* strains from the GenBank (http://www.ncbi.nlm.nih.gov/genbank/). In addition to annotated *L. rhamnosus* genes, the genes unique for the strain 24 have been identified. In the genome of *L. rhamnosus* 24 seventeen genes for sugar utilization, two genes related to the process of adhesion, two probacteriocin genes, and one serine threonine protein kinase gene were found. Remarkably, two toxin-antitoxin systems (*PemK-A1*$_{Lrh}$ и *YefM-YoeB*$_{Lrh}$*)* and one toxin gene (*relE*$_{Lrh}$) were detected. Cloning and expression of YefM toxin in *E.coli* cells provided the evidence for functional activity of this protein, i.e., the capacity to kill bacterial cells.

Conclusion: L. rhamnosus strain 24 is the first lactobacilli probiotic strain from Russian resources/collections with the known genome nucleotide sequence. The data obtained in this study will be explored to evaluate the probiotic potential of this strain in comparison to other strains of geographically different origin.

Availability: At the moment the nucleotide sequence of *L. rhamnosus* strain 24 is available on request from the authors.

References:

1. A. A. Prozorov, V. N. Danilenko. (2010) Toxin–Antitoxin Systems in Bacteria: Apoptotic Tools or Metabolic Regulators? *Mikrobiologiya*, **79** (2):147-159.

APPLICATION OF REPAIR ENZYMES TO IMPROVE THE QUALITY OF THE DNA TEMPLATE IN PCR AMPLIFICATION OF DEGRADED DNA

Dovgerd A.P.*, Zharkov D.O.
Institute Chemical Biology and Fundamental Medicine, SB RAS, Novosibirsk, Russia
e-mail: antdovgerd@gmail.com
** Corresponding author*

Key words: *base excision repair, ancient DNA, forensic DNA, PCR, DNA damage*

Motivation and Aim: Although DNA is the main carrier of genetic information in living organisms, this molecule is inherently unstable. When an organism is alive, its repair systems resist DNA damage, but when these processes cease working after death, the accumulation of lesions in DNA becomes irreversible. Analysis of damaged DNA may be problematic. For example, the efficiency of PCR is sharply reduced if the template is subjected to oxidation or apurinization. This is particularly actual in studies of "ancient DNA" and DNA in the forensic practice. We are developing a system in which DNA repair enzymes are used to improve the quality of degraded DNA templates before PCR.

Methods and Algorithms: In most cases, lesions in postmortem DNA are located opposite an undamaged base in the complementary strand [1, 2] and, therefore, can be correctly repaired. We are creating a kit that includes several major DNA glycosylases, an AP endonuclease, a DNA polymerase and a DNA ligase, the combined effect of which leads to repairing much of the damage in DNA samples prior to their use as PCR templates.

Results: We have developed model systems of degraded high-molecular DNA with predominance of different types of lesions and have shown that the efficiency of PCR is sharply reduced if the template is subjected to oxidation or apurinization. We have reconstituted a repair system for a major oxidative DNA lesion, 8-oxoguanine, with *E. coli* 8-oxoguanine-DNA glycosylase Fpg, *E. coli* AP-endonuclease Nfo, human DNA polymerase β and bacteriophage T4 DNA ligase. We also include thermostable translesion DNA polymerase in the PCR to efficiently overcome the residual damage in the amplification process.

Conclusion: We have developing a system in which repair enzymes are used to improve the quality of degraded DNA templates before PCR. The primary use of our system may be found in the analysis of forensic samples, highly processed food and "ancient DNA".

Acknowledgements: This work was funded by the U.M.N.I.K. program (№06U/02-10)

References:

1. Hoss M., Jaruga P., Zastawny T.H., Dizdaroglu M., Paabo S. (1996) DNA damage and DNA sequence retrieval from ancient tissues. Nucleic Acids Res., 24, p. 1304–1307.
2. Mitchell D., Willerslev E., Hansen A. (2005) Damage and repair of ancient DNA. Mutat. Res., 571, p. 265–276.

SEMICONDUCTOR SEQUENCING FOR LIFE

Matt Dyer
Life Technology, USA
e-mail: matt.dyer@lifetech.com

Ion Torrent has pioneered an entirely new approach to sequencing that enables a direct connection between chemical and digital information and leverage decades of semiconductor technology advances. The result is the first commercial sequencing technology that does not use light, and as a result delivers unprecedented speed, scalability, accuracy, and low cost. In just the first year the Ion Torrent Personal Genome Machine (TM) has become the fastest selling sequencing platform. The throughput scaled 100X, from 10Mb to 1Gb, in just the first year and will scale another 100X in the next year with the new Proton (TM) sequencer, which will enable the single day $1000 human genome. Automated data analysis is driven by Torrent Suite, an open-source software suite that provides a simple and intuitive interface to streamline data analysis and provide results in minutes to hours, not days. Built on top of Torrent Suite is a flexible SDK that allows users to expand the analysis capabilities through the development and utilization of plugins and APIs.

IRAP-PCR MARKERS AND MICRONUCLEI TEST IN THE CHARACTERIZATION OF GENETIC STRUCTURE OF THE KALMYK SHEEP AND TYPES OF THE EDILBAY SHEEP

Elkina M.A., Astafieva E.E., Glazko T.T.
Russian State Agrarian University–MTAA named after K.A.Timiryazev, Moscow, Russia
e-mail: mariyaelkina@yahoo.com

Key words: *IRAP-PCR markers, micronuclei test, the Kalmyk sheep, the Edilbay sheep, the Suyunduk type, the Birlik type*

Motivation and Aim. The originality of the breed's gene pool is determined by the balance between the initial genetic diversity and populational and genetic responses to the factors of natural and artificial selection. In this regard, the selection of molecular genetic and cytogenetic methods that could identify the specificity of breed gene pool in order to control and optimize their use is of a particular interest. This is important especially for saving local breeds, such as the Edilbay and Kalmyk breeds of sheep.

Methods and Algorithms. The estimates of DNA polymorphism (IRAP-PCR markers) obtained using in polymerase cycle reaction (PCR) the terminal parts of retrotransposons: LTR SIRE-1 (GCA-GTT-ATG-CAA-GTG-GGA-TCA-GCA) , PawS 5 (AAC-GAG-GGG-TTC-GAG-GCC) and BARE-1 (CCA-ACT-AGA-GGC-TTG-CTA-GGG-AC) as the primers are used to study the genetic structure of populations, as well as the frequency of occurrence of erythrocytes with micronuclei (EMN in ‰) in the peripheral blood of two types of the Edilbay sheep (The Suyunduk and Birlik types) and the Kalmyk sheep.

Results. We obtained the following data. The Suyunduk type of the Edilbay sheep was the most homogeneous in comparison with the Birlik and Kalmyk sheep (P = 27%, PIC = 0,157; P = 29%, PIC = 0,221 and P = 47%, PIC = 0,199, respectively) according to the polymorphic information content (PIC) of the spectra of DNA amplification products (amplicons) obtained using PawS 5 primer in PCR. This type was also distinguished from the Birlic type and the Kalmyk sheep by its low frequency of EMN (0,2 ‰, 4,6 ± 0,3 ‰ and 4,3 ± 0,3 ‰, respectively). The Suynduk type differed from other groups of sheep by a high level of amplicon polymorphism in the case of use terminal fragments of retrotransposons LTR SIRE-1 and BARE-1 as primers in PCR. The relatively low frequency of EMN in peripheral blood of the Suyunduk type of the Edilbay sheep may be due to the fact that it was created in the nuclear testing area called the Azgirsky landfill, in contrast to other studied sheep groups reproducing in relatively more favorable environmental conditions. We could explain the relatively low frequency of cells with cytogenetic abnormalities in the Suyunduk type of sheep by chronic action of natural selection that promotes the reproduction of animals with increased resistance to environmental stress factors.

Conclusion and Availability. IRAP-PCR markers, using different terminal sites of retrotransposons as primers in PCR can reliably identify the specific features of gene pool of sheep breeds and intrabreed types. The characteristics of gene pool depend on the primers used in PCR. The Kalmyk sheep and the Birlik type of the Edilbay sheep did not differ from each other by micronuclei test. The relative reduction in the frequency of erythrocytes with micronuclei in a peripheral blood of the Suyunduk type suggests that a unique gene pool was created as a result of natural selection during chronic stressful environmental factors. That's why such populations have high potential for adaptation to adverse conditions of keeping and breeding.

COMPLETE MITOCHONDRIAL AND CHLOROPLAST GENOMES OF DIATOM ALGA *SYNEDRA ACUS*

Galachyants Y.P.
Limnological Institute, Siberian Branch of the RAS, Irkutsk, Russia
e-mail: yuri.galachyants@lin.irk.ru

Key words: *diatoms, complete mitochondrial and chloroplast genomes, next-generation DNA sequencing, comparative genomic analysis, phylogenetic analysis*

Motivation and Aim: Only several complete mitochondrial and chloroplast genomes of marine diatoms have been sequenced to date. Here we present the complete mtDNA and cpDNA sequences of freshwater araphid pennate diatom alga *Synedra acus* subsp. *radians* (Kьtz.) Skabitsch from Lake Baikal. Also we present the results of comparative genomic and phylogenetic analyses for available diatom genome sequences.

Methods and Algorithms: To sequence mtDNA and cpDNA we analyzed the shot-gun genomic library prepared from *S. acus* total DNA with Roche/454 GS FLX Titanium instrument. The mitochondrial- and chloroplast-specific contigs were identified in the assembly according to their similarity with known organellar diatom sequences. Finishing of mtDNA was performed using the primer-walking approach. For chloroplast genome, the reference-guided assembly of the short Illumina reads was used to enhance the quality of cpDNA sequence. At final step, the overlapped ends of the chloroplast contigs were merged into a circular molecule.

Results and discussion: Mitochondrial genome of *S. acus* has length of 46,657 bp. It encodes 2 rRNAs, 24 tRNAs, and 33 proteins. The mtDNA of *S. acus* contains three group II introns which seem to be polyphyletic. The compact gene organization contrasts with the presence of a 4.9 kb-long intergenic repeat region. Comparison of the three sequenced mtDNAs showed that these three genomes carry similar gene pools, but the positions of some genes are rearranged.

Chloroplast genome of *S. acus* possesses a canonical quadripartite structure and maps as a circular molecule of 116 251 bp. It encodes 160 genes including tRNAs, rRNAs, and 128 protein genes. Comparative analysis of diatom cpDNA reveals 154 common genes and the absence of an overlapping between *atp*D and *atp*F gene coding sequences in *S. acus* genome. The transfer of *acpp* genes to a host nuclear genome is hypothesized to occur independently in several lineages of diatoms.

Acknowledgements: This work was financially supported by the Program of the Presidium of RAS ‘‘Molecular and Cell Biology’’ (project #22.3). We thank the distributed computing group of IDSTU SB RAS in help to perform the computationally-extensive bioinformatic analyses.

References:

1. N.V. Ravin *et al*. (2010) Complete sequence of the mitochondrial genome of a diatom alga *Synedra acus* and comparative analysis of diatom mitochondrial genomes, Curr. Genet., **56:** 215-223.
2. Y.P. Galachyants *et al*. (2012) Complete chloroplast genome sequence of freshwater araphid pennate diatom alga *Synedra acus* from Lake Baikal, Intl. J. Biol., **1:** 27-35.

SEARCHING FOR DISTANT HOMOLOGS OF SMALL, NON-CODING RNAs

Giegerich R.
Center of Biotechnology and Faculty of Technology,
Bielefeld University
e-mail: robert@techfak.uni-bielefeld.de

Our appreciation of the functional repertoire of noncoding RNA has increased enormously in recent years. Present RNA-sequencing projects yield a large number of bonafide small RNA transcripts, calling for bioinformatics approaches to filter out promising candidates for experimental study. Potential interaction partners and phylogenetic conservation can provide additional evidence. However at present, there is no fully automated solution either problem.

In the SNAP project (Small Non-coding RNA in Alphaproteo-Bacteria), we identified about 1000 small ncRNA transcripts in *S. meliloti*. Aside from confirmation of small ncRNAs known with *S. meliloti* at that date, search of the Rfam data base with covariance models brought little extra information. 52 trans-encoded RNA transcripts were chosen, for which 39 new family models were build with a mixture of automated and hand-crafting methods.

The talk gives an overview of our findings and then concentrates on two candidate ncRNAs that are widely distributed in the *Rhizobiales*, and cannot be detected with standard methods. Given that we have Internet connection, the talk will end with the audience creating their own search program for finding distant ncRNA homologs.

The SNAP project is joint work with Jan Reinkensmeier, Anke Becker, and Jan-Philip Schl¨uter.

References:

1. Reinkensmeier, Jan and Schl¨uter, Jan-Philip and Giegerich, Robert and Becker, Anke: *Conservation and Occurrence of Trans-Encoded sRNAs in the Rhizobiales* in GENES, Pages: 925-956, 2011.
2. Schl¨uter, Jan-Philip and Reinkensmeier, Jan and Daschkey, Svenja and Evguenieva-Hackenberg, Elena and Janssen, Stefan and J¨anicke, Sebastian and Becker, J¨org and Giegerich, Robert and Becker, Anke: A *genome-wide survey of sRNAs in the symbiotic nitrogen-fixing alpha-proteobacterium Sinorhizobium meliloti* in BMC Genomics, 11(1) , Pages: 245, 2010.

DEVELOPMENT OF A NOVEL PYROSEQUENCING-BASED METHOD FOR STUDYING *E. COLI* DIVERSITY AND MICROBIAL SOURCE TRACKING

Goodman A.*, Montana A., Neal E., VanderKelen J., Black M., Kitts C., Dekhtyar A.
California Polytechnic State University, San Luis Obispo, CA, USA
e-mail: agoodman@calpoly.edu
** Corresponding author*

Key words: *microbial genome diversity, pyrosequencing, assay development, modeling*

Motivation and Aim: Fecal contamination of food and water supply frequently cause public health problems around the world. There is an urgent need for rapid and inexpensive method of identifying the sources of contamination. The presence of *E. coli* is typically used as an indicator of fecal contamination; however, identifying the source of specific strains of *E. coli* remains a major challenge. We have developed a novel method for the identification of *E. coli* strains by generating molecular fingerprints via simultaneous, multi-locus pyrosequencing of the ribosomal RNA (rRNA) operon.

Methods and Algorithms: Previously, rRNA genes have been eliminated from the list of potential targets in sequence-based assay development because rRNA operons are present in seven copies in each *E. coli* genome [1]. Our method takes advantage of the multiple copies of the rRNA operon to help discriminate between closely related strains. We designed a novel assay that uses PCR to amplify all seven copies of the ribosomal RNA intergenic regions and then sequences them together in a single reaction. The two polymorphic intergenic transcribed spacer (ITS) regions that reside between the rRNA segments (16S, 23S, and 5S) are used as targets to distinguish between similar strains. The raw pyrosequencing data from each reaction is a pattern of peaks. While these data cannot be used for sequence analysis due to the use of multiple templates, they may be used to differentiate between strains. These patterns are reproducible and characteristic of each strain, resulting in output that is analogous to a fingerprint; therefore, we refer to the patterns as pyroprints. By pyrosequencing multiple templates that differ in their sequences, the effect of single-nucleotide polymorphisms (SNPs) may be amplified through a "ripple effect": a difference at one of the seven loci resulting in changes of multiple subsequent signal peaks. Whether or not the "ripple effect" is observed depends on the genomic sequence and the dispensation order. To optimize assay parameters, we developed a pyroprint modeling program to model pyroprints from 38 finished *E. coli* genomes. We use Pearson correlation coefficient to compare pyroprints.

Results: Pyroprints are highly reproducible (>99%). Our collection currently contains nearly 3,000 pyroprints from *E. coli* isolated from 16 avian and mammalian hosts. Some strains appear to be shared by multiple hosts, while others appear host-specific. We are developing web based pyroprint database and analytical tools and extending this method to other bacterial species.

Conclusion: Pyroprinting is a novel method for identification of *E. coli* strains. Pyroprinting has the following advantages: simple protocol, reproducible, low cost, easy transfer from lab to lab, easy to scale up, and good discrimination between closely related strains.

References:

1. K.M. Ivanetich et al. (2006) Microbial source tracking by DNA sequence analysis of the Escherichia coli malate dehydrogenase gene. *J Microbiol Methods.* 67(3):507-26.

SEARCH FOR FUNCTIONAL PATHWAYS FOR INTRAMEMBRANE ASPARTIC PROTEASE IMPAS1/SPP

Grigorenko A.P.[1, 2], Moliaka Y.[1], Alexandrov I.[1], Rogaev E.I.[1, 2]
[1] *University of Massachusetts Medical School, USA;*
[2] *Vavilov Institute of General Genetics, Research Center of Mental Health, Moscow*
Correspondence should be addressed to Evgeny.Rogaev@umassmed.edu

Motivation and Aim: Previously, we and others have identified five genes in human genome for novel polytopic family of intramembrane aspartic proteases IMPAS/IMP (or H13, SPP/SPPL or PSH) homologous to presenilins (Grigorenko et al, 2002; Moliaka et al, 2004; Grigorenko et al, 2004). This type or proteases supposed to be critical for cleavage of type II transmembrane proteins (including signal peptides) in transmembrane domain. It has been predicted that IMPs substrates are type II membrane proteins with C-terminal part oriented into the lumen. *In vitro* assay demonstrated that SPP/IMP1 cleaves short signal peptide remnants tethered in ER membranes. This activity may generate short signal sequence that is essential for HLA-E epitope (Weihofen et al, 2002). However, the major functions *in vivo* and proteolytic substrates of IMP1 proteins *in vivo* are unknown.

Methods and Algorithms. We generated the knockout mouse model and cellular models (*mIMP1-/-)* for *IMPAS1/SPP* gene. Several approaches were taken to elucidate molecular pathways regulated by IMP1 gene: testing of apoptosis and autophagy events; comparative transcriptome profiles of all protein-encoding genes in *IMP1* knockout and control mouse brains; real-time PCR gene expression study and signaling pathway analysis with Cignal™ Reporter Assays in *IMP1-/-* and *IMP1-/+* or *IMP1+/+* cells.

Results. We have shown that *mIMP1* gene is crucial for early embryonic development and plays an important role in brain development (unpublished). We have identified alteration of certain transcriptional factors in *mIMP1* knockout mice cells suggesting that IMP1/SPP is essential for signal transduction regulation in early development. Empirically, we have excluded the number of type II protein receptors as putative substrates for IMP1 protease but confirmed the cleavage of HCV (viral) core protein and showed that both human IMP1/SPP and its *C.elegans* ortholog protease capable to cleave (in co-transfected cells) the C-terminal transmembrane domain of multipass transmembrane protein presenilin 1. The endogenous substrates for IMP1/SPP with single or multiple transmembrane domains involved in signaling regulation in development can be predicted and have to be identified.

Acknowledgements. Supported by NIH/NIA AG029360.

References:

1. Grigorenko AP, Moliaka YK, Korovaitseva GI, Rogaev EI. Biochemistry (Mosc) 2002; 67: 826-35.
2. GrigorenkoAP, Moliaka YK, Soto MC, Mello CC, Rogaev EI. Proc Natl Acad Sci U S A 2004; 101: 14955-60.
3. Moliaka Y., Grigorenko A., Madera D., Rogaev E. FEBS Letters (2004); **557**:185-192. 4.
4. Weihofen A et al, Science 2002; 296: 2215-8.

ANALYSIS OF SNP DISTRIBUTION AND INTER-SNP DISTANCE IN THE HUMAN GENOME

Ignatieva E.V.*, Levitsky V.G., Yudin N.S.
Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia
e-mail: eignat@bionet.nsc.ru
** Corresponding author*

Key words: *SNP distribution, the human genome, 1000 Genomes Project*

Motivation and Aim: Mutation frequencies are known to vary along a nucleotide sequence. Nucleotide positions with an exceptionally high mutation frequency are called hotspots [1]. The results of the pilot phase of the 1000 Genomes Project were approximately 15 million SNPs, 55% of which were previously undescribed [2]. But the genome-scale patterns of SNP distribution have been poorly investigated yet. Study of the SNP distribution and inter-SNP distance in the human genome is of great importance for experimental researches. Sensitivity and specificity of most experimental methods of SNP typing are strongly decreased from the presence of additional SNP variants in the neighborhood with the target SNP. Using the data from the 1000 Genomes Project, we have analyzed the pattern of SNP distribution among the different regions of the human genome and the DNA context features of nucleotide sequences with adjacent SNPs.

Results: It was found that about 1.3% of SNPs occur in neighboring positions (adjacent SNPs). It is 3 times more, than expected accident frequency for two adjacent SNPs. In 0.8 % of cases SNPs are separated by one nucleotide. SNPs density was dependent on their localization across the different parts of the gene. Low SNPs density was found in the vicinity of transcription start sites, 5'- and 3'- parts of introns. This observation is in good agreement with important functional roles of these regions. General DNA context features of the stretches with adjacent SNPs were detected: the frequency of CpG was found to be higher among this two adjacent SNPs and AT-content was found to be higher in 5' and 3' flanks (3-5 bp).

Conclusion: SNPs are localized non-uniformly along the human genome. Contrary to expected stochastic SNP distribution (one SNP occurs on average every 268 bp), more than half of all SNPs (69%) occurred at distances less than 250 bp. One of the most popular tools for screening of SNP variants, high-resolution melting curve analysis (HRM) is especially vulnerable, because for SNP screening, DNA fragments of 150–250 bp are usually used [3]. The possibility of existence of additional SNP variants in melting fragment requires careful consideration for improvement of new HRM assay specificity.

Acknowledgements: SB RAS project № 136, RAS projects № 6.8, 28, 30.29., Russia's President's project №-5278.2012.4.

References:

1. Rogozin I.B. et al. (2003) Computational analysis of mutation spectra, *Brief. Bioinform.* **4**:210-27.
2. 1000 Genomes Project Consortium et al. (2010) A map of human genome variation from population-scale sequencing, *Nature,* **467**:1061-1073.
3. Vossen R.H. et al. (2009) High-resolution melting analysis (HRMA): more than just sequence variant screening, *Hum. Mutat.*, **30:**860-866.

DE NOVO SEQUENCING, ASSEMBLY AND CHARACTERIZATION OF TRANSCRIPTOME IN TETRAPLOID PLANT *CAPSELLA BURSA-PASTORIS*

Kasianov A.S.*[1, 2], Logacheva M.D.[3], Oparina N.J.[1], Penin A.A.[3]
[1] *Engelhardt Institute of Molecular Biology RAS, Moscow, Russia;*
[2] *Vavilov Institute of General Genetics, RAS Moscow, Russia;*
[3] *Lomonosov Moscow State University, Moscow, Russia*
* *Corresponding author*

Key words: *plants, polyploidy, transcriptome, sequence assembly, Capsella bursa-pastoris*

Motivation and Aim: Transcriptome sequencing data is an essential component of modern genetics, genomics and evolutionary biology. Huge improvements of sequencing technologies allowed characterization of transcriptomes in many non-model species. However, de novo assembly of transcriptomes of flowering plants is still a challenging task due to the fact that many of them are recent polyploids and thus multiple paralogs very similar to each other are present in their genomes and this hampers the assembly.

We performed sequencing and analysis of transcriptome of *Capsella bursa-pastoris*. This plant is a tetraploid with uncertain origin, being a recent allotetraploid or more ancient autotetraploid. Its close relationship with model plant species, *Arabidopsis thaliana* (*Capsella* belongs to the same family) makes *C. bursa-pastoris* a perfect model for the studies of gene and genome evolution after genome duplication events.

Methods and Algorithms: cDNA corresponding to the genes expressed *C.bursa-pastoris* flowers and inflorescences was sequenced using Illumina and 454 sequencing platforms. Also, additional set of cDNA libraries corresponding to the genes expressed in various stress conditions (cold, over-illumination, wounding) was sequenced. In total, nearly 60 millions of reads were generated. Different programs were tested (MIRA, Velvet, CLC Genomics workbench) for assembly, but none of them demonstrated the capacity to assemble paralogous genes separately because of their high similarity. We developed the algorithm for partitioning of reads into subsets corresponding to each of the paralogs. After partitioning the subsets were assembled separately thus allowing to generate separate sequences for each paralog. Sequences were annotated using BLAST2Go; after annotation each pair of paralogs was analyzed in terms of sequence divergence, the presence of intact ORF and codon usage.

Results and conclusion: Transcriptome of *C. bursa-pastoris* was sequenced and assembled using newly developed algorithm for separation of reads into subsets corresponding to the paralogous genes. Patterns of molecular evolution in paralogous genes were inferred.

COMPLEX GENOME SEQUENCING: PRELIMINARY DATA OF SIBERIAN LARCH COMPLETE GENOME SEQUENCING

Krutovsky K.V.*[1, 2, 5], Vaganov E.A.[2], Chubugina I.V. [2, 3], Oreshkova N.V.[4], Tretyakova I.N.[4], Tyazhelova T.V. [5]

[1] *Department of Ecosystem Science and Management, Texas A&M University, College Station, TX, USA;*
[2] *Genome Research Center, Siberian Federal University, Krasnoyarsk, Russia;*
[3] *Center for Forest Protection, Krasnoyarsk, Russia;*
[4] *V.N. Sukachev Institute of Forest, SB RAS, Krasnoyarsk, Russia;*
[5] *N.I. Vavilov Institute of General Genetics, RAS, Moscow, Russia*
e-mail: k-krutovsky@tamu.edu
** Corresponding author*

Key words: *de novo sequencing, complex genomes, conifers*, *Larix sibirica Ledeb*

Motivation and Aim: The purpose of the study is to completely sequence, assemble and annotate *de novo* the Siberian larch (*Larix sibirica* Ledeb.) genome (1C = 12.03 Gbp), which is four times larger than human genome (1C = 3.20 Gbp) that remains the largest genome completely sequenced so far. Meanwhile, larch is one of the most important key elements of Siberian boreal forests that have great economic and ecological values. However, the study of larch and other closely related important conifer forest tree species is hindered by almost complete lack of data on its genome structure and genes that control important adaptive and selective traits. Complete larch genome sequence would allow us to obtain such data and effectively use them for studying conifer forests genetic variation, genetic adaptation to global climate change and for creating conservation and breeding programs.

Methods and Algorithms: The gigantic genome size of conifers and high allelic variation impede their complete genome sequencing and assembling. The conifer genomes are not only extremely large, but also contain a great number of repetitive elements and large gene families with high similarity in nucleotide sequences. To overcome these problems and facilitate assembling we use an innovative unique approach via using haploid tissue cultures developed from haploid immature megagametophytes (female gametophytic tissue).

Results: The haploid nature of tissue cultures or calluses obtained from megagametophytes was confirmed by genotyping their nuclear genomic DNA with informative SSR markers that are heterozygous in the diploid tissue of the parent tree. After fragmentation the fraction of nuclear genomic DNA within 550-600 nucleotide base pairs size range was used for paired-end sequencing with 101 cycles and four lanes of a flow cell of the *Illumina HiSeq 2000* sequencer that should give an expected ~12X genome coverage. The preliminary results based on these sequence data will be presented.

Conclusion: The obtained data represent the first step in the multi-disciplinary integrative innovative international project on complete *de novo* larch genome sequencing that is planned to be done in the Genome Research Center recently established at the Siberian Federal University (Krasnoyarsk, Russia).

Acknowledgements: This work was supported by grant from Russian government department of Science and Education to Siberian Federal University «The genetic researches of the Siberian larch».

WHOLE GENOME SEQUENCING AND PHYLOGENETIC ANALYSIS OF *VIBRIO CHOLERAE* O1 ELTOR INABA № 301 STRAIN

Kuleshov K.V.*[1], Shipulin G.A. [1], Markelov M.L. [1], Dedkov V.G. [1], Podkolzin A.T. [1], Vodop'ianov S.O. [2], Kermanov A.V. [2], Kruglikov V.D. [2], Mazrukho A.B. [2], Vodop'ianov A.S. [2], Pisanov R.V. [2]

[1]Federal Budget Institute of Science "Central Research Institute for Epidemiology", Moscow, Russia;
[2]The Federal Government Health Institution "Rostov-on-Don Plague Control Research Institute", Rostov-on-Don, Russia
e-mail: konstantinkul@gmail.com
** Corresponding author*

Key words: *whole genome sequencing, Vibrio cholerae, core genome phylogeny*

Aim: The aim of this research was to determine molecular genetic markers for the understanding of the possible origin of Vibrio cholerae O1 Eltor Inaba № 301 (#2011EL-301) strain isolated from sea water in the city of Taganrog in the summer of 2011.

Methods and Algorithms: Whole genomic sequencing protocol included the following steps: sequencing of fragment libraries using both Roche 454 GS Junior(Roche Diagnostics) and MiSeq (Illumina) sequencers; *de novo* assembling contigs from Roche's single reads using Newbler 2.5; mapping the Illumina pair-end reads to Roche's assembled contigs with CLC Genomics Workbench v5.5.

Results: Using Roche 454 reads 124 contigs were assembled *de novo* the N50 was 135 kb, average coverage was 33x. For validation and correction of significant SNPs on assembled contigs we mapped Illumina pair-end reads to this contigs. The consensus sequence was used for further analysis. The average coverage for 124 contigs was 100x. The isolate under analysis contains a hybrid prophage CTX localized in chromosome I. This prophage carries the classical type allele of ctxB and rstR El-Tor allele. VPI-1 region carries tcpA El-Tor allele. Loci (VSPI, VPI-1, VPI-2, SXT) have high similarity level with recently isolated strains (1). Comparison of assembled contigs with several complete and incomplete genomes of different *V. cholerae* strains from the GenBank database using Progressive Mauve algorithm revealed the most similar genomes which were isolated from Haiti, Africa and Asia (1). In compliance with earlier proposed scheme for SNP typing of Vibrio cholerae strains of 7th pandemic, which is based on 30 SNPs, our strain belongs to group V. 29 genomes were analyzed on the basis of common ortholog genes to understand in more detail the phylogenetic relations of 2011EL-301 strain among recent isolates of V. cholerae. As a result of the core genome phylogeny our isolate was included in one cluster with strain associated with cholera cases from South Africa in 2009 (# 2011EL-1137) and cases of transmission of cholera from Pakistan (#3582-05, #2009V-1116, #2009V-1046, #2010V-1014). In comparison with CRIS101 genome sequence in core genome phylogenetic tree and data from earlier published study (3) our strain refers to the 3th wave of the 7th pandemic. Whole Genome Shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession AJFN00000000.

References:

1. A.R. Reimer et al. (2011), *Emerg Infect Dis.*,**17(11)**:2113-21
2. C. Lam et al. (2011) *Emerg Infect Dis.,***16(7)**:1130-2
3. A. Mutreja et al. (2011), *Nature,* **477**, 462–465

GENETIC DIVERSITY IN EXTREMOPHILIC BACTERIAL COMMUNITY FROM HOT SPRING "URITSKY", BAIKAL

Kurilshikov A.M. [1], Babkin I.V. [1], Morozova V.V. [1], Bryanskaya A.V. [2], Tikunov A.Yu.[1], Lazareva E.V. [3], Zhmodik S.M. [3], Tikunova N.V. [1]

[1] Institute of Chemical Biology and Fundamental Medicine SB RAS, Novosibirsk,Russia;
[2] Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia;
[3] Institute of Geology and Mineralogy SB RAS, Novosibirsk, Russia
e-mail: alexa.kur@gmail.com

Key words: *metagenomics, extremophile, bacteria, 16s rRNA, Baikal*

Environmental sample for this study was obtained from Uritsky hot spring located in the Barguzin basin of the Baikal rift zone. DNA was isolated and then amplified with primers specific to 16s ribosomal RNA gene fragment with 1230 nucleotides of estimated length. Purified amplicons were cloned into E.coli plasmids. Since monoclonal libraries were grown, inserted fragments were sequenced by Sanger.

In total, 242 clones were sequenced. All sequenced clones were chimera checked by Mallard (Cardiff University, Great Britain) and Bellerophon v.3 (GreenGenes project) software packages. 14 sequences were identified as chimeras. To calculate richness of the studied environmental sample, rarefaction analysis was performed with MOTHUR package. Number of OTU (with 95% cutoff) in sample was evaluated as 180 by Chao estimator. Phylogenetic analysis was performed by RDP software (http://rdp.cme.msu.edu) with 80% confidence threshold. 47 sequences were indetified as *Deinococcus*, 20 – *Cyanobacteria* , 25 – *Bacteroidetes*, 77 – *Proteobacteria*, 17 – *Actinobacteria*, 30 – *Planctomycetes*, 16 – *Chloroflexi*, 2 – *Fusobacteria*, 1 – OD1, 6 – *Firmicutes*. In addition, there were several *Proteobacteria* phylum members: *Alpha* – 21, *Beta* – 19, *Gamma* – 30 and 2 *Deltaproteobacteria*. 46 sequences couldn't be determined on family level; 5 *Proteo-* and 5 *Eubacteria* couldn't be determined to any class. The greatest difference between sequenced bacteria and reference sequences from GenBank (NCBI), RDP and other databases was revealed in *Cyanobacteria* phylum. None of the sequenced *Cyanobacteria* can be related to any family. Sequences with highest homology to these *Cyanobacteria* were discovered in hot springs from different places: Yellowstone (USA), Kumamoto (Japan), Rincon de la Vieja (Costa-Rica).

THE ROLES OF THE MONOMER LENGTH AND NUCLEOTIDE CONTEXT OF PLANT TANDEM REPEATS IN NUCLEOSOME POSITIONING

Levitsky V.G.*[1], Vershinin A.V.[2]
[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia;*
[2] *Institute of Molecular and Cellular Biology, SB RAS, Novosibirsk, Russia*
e-mail: levitsky@bionet.nsc.ru
** Corresponding author*

Key words: *nucleosome positioning, tandem repeats, periodicity of dinucleotides*

Motivation and Aim: Tandem DNA arrays are consist of regularly alternating monomers, which have almost identical nucleotide sequences. Such organization makes these arrays especially interesting for clarifying the role of intrinsic DNA preferences in nucleosome positioning (NP). 10-11 –base periodicity of certain dinucleotides is a ubiquitous hallmark of NP.

Methods and Results: We have compared the nucleotide context of the monomers with length (ML) that is equal and unequal to the integer number of the nucleosome repeat length (NRL). 161 plant tandem repeat families from the PlantSat [1] were divided into two classes based on this criterion. We assessed the content of periodic dinucleotides (CPD) in the families of monomers of two classes. The excess of CPD in 1st class was significant according to the chi-square test. Then we applied wavelet transformation [2] to the Phase [3] prediction of NP. The Phase method compares the occurrences of PDs in a potential nucleosomal DNA with those known for approved nucleosomal DNA. Three criteria were used for classification of families of tandem repeats onto types of nucleosome arrangements (NA): (a) the ratio of ML to NRL; (b) the number of peaks in the profiles and their heights; (c) and the heterogeneity of these characteristics within a family. Three main types of NA in DNA tandem repeat arrays were distinguished: Regular (all nucleosomes are positioned in a context-dependent manner), Partial (nucleotide context influences the positioning of only a subset of the nucleosomes), and Flexible (the least effect of nucleotide context in determining NP).

Conclusion: We have demonstrated that the integer ratio of ML to the NRL is accompanied by an increased CPD and a greater influence of the nucleotide context on the NP. Based on the ML and the nucleotide context, three main types of NA in arrays of tandem repeats have been distinguished.

References:

1. J. Macas et al. (2002) PlantSat: a specialized database for plant satellite repeats, *Bioinformatics,* **18**: 28-35.
2. X.Q. Lu et al. (2004) Maximum spectrum of continuous wavelet transform and its application in resolving an overlapped signal, *J Chem Inf Comput Sci*, **44**: 1228-1237.
3. V.G. Levitsky et al. (2008) Nucleosome formation potential estimation via dinucleotide periodicity preferences, *Proceedings of the sixth international conference on bioinformatics of genome regulation and structure,* 140.

A DRAFT GENOME SEQUENCE OF TARTARY BUCKWHEAT, *FAGOPYRUM TATARICUM*

Logacheva M.D.*[1, 2, 3], Sutormin R.A. [2], Naumenko S.A. [2], Demidenko N.V. [2, 4], Vinogradov D.V. [3], Gelfand M.S. [2, 3], Penin A.A. [2, 3, 4]

[1] *A.N. Belozersky Institute of Physico-Chemical Biology, M.V. Lomonosov Moscow State University, Moscow, Russia;*

[2] *Faculty of Bioengineering and Bioinformatics, M.V. Lomonosov Moscow State University, Moscow, Russia;*

[3] *A.A. Kharkevich Institute for Information Transmission Problems, Russian Academy of Science, Moscow, Russia;*

[4] *Department of Genetics, Biological Faculty, M.V. Lomonosov Moscow State University, Moscow, Russia*

e-mail: maria.log@gmail.com

** Corresponding author*

Key words: *genomics, plant genomes, buckwheat, assembly, annotation*

Motivation and Aim: Despite significant progress in the technologies of DNA sequencing genomic data are lacking for many groups of living organisms, in particular, many plant taxa. Tartary buckwheat (*Fagopyrum tataricum*) belongs to the order Caryophyllales, a large and diverse group of flowering plants, none of which have their genome sequenced and characterized. *F. tataricum* is a close relative of common buckwheat, an important food crop, and a potential donor of favorable traits absent in common buckwheat, such as self-compatibility and improved stress resistance.

Methods and Algorithms: The sequences of five genomic DNA libraries with different insert length (from 100 bp to 3 Kb) and different read length were generated using the Illumina sequencing platform up to 180x genome coverage. In total, about 500 millions of paired-end reads were used for de novo assembly. To facilitate genome annotation and to lay the groundwork for functional genomics we also constructed and sequenced 16 transcriptome libraries corresponding to various stress conditions and developmental stages. They were used for cDNA-based gene prediction and as a training set for ab initio gene prediction.

Results: Draft genome sequence of *Fagopyrum tataricum* was generated. The total number of bases included in the assembly is 372 Mb, that corresponds to 70% of the genome. Nearly 30 thousands of genes were predicted using cDNA-based, homology-based and ab initio approaches. A number of taxon-specific gene duplications, as well as genes specific to the buckwheat genome were found. Using transcriptome sequence data, we also identified alternative transcript isoforms and characterized their expression in different plant organs and under different stress conditions.

Conclusion: We sequenced, assembled and annotated the genome of *F. tataricum* and performed genome-wide analysis of its genes. Given the close relationships of *F. tataricum* to cultivated buckwheat we expect that this sequence will be useful not only for evolutionary and comparative genomics, but also for practical aspects such as identification of genes responsible for agriculturally important traits.

ADVANCES IN GENOMIC AND METAGENOMIC STUDIES OF EXTREMOPHILIC MICROORGANISMS

Mardanov A.V., Kadnikov V.V., Gumerov V.M., Ravin N.V.*
Centre "Bioengineering" RAS, Moscow, Russia
e-mail: nravin@biengi.ac.ru
** Corresponding author*

Key words: *extremophiles, archaea, genome, microbial community, metagenomics*

Sequencing of microbial genomes is one of the departure points for researches in the fields of microbiology, molecular biology and evolution of living organisms, and also has practical importance for medicine and biotechnology. The objects of our research are microorganisms living in extreme environments, primary thermophilic archaea. They are evolutionary ancient organisms with relatively small genomes but possessing complete systems of autotrophic or heterotrophic metabolism. We determined complete genome sequences of twelve thermophilic microorganisms representing either new phylogenetic lineages or biotechnologically/environmentally important species. Genomic data were used for identification of the metabolic pathways of investigated microorganisms, studying the molecular mechanisms of genetic processes such as DNA replication, analysis of evolution of the genome and individual genes, structural and functional studies of particular proteins. Next generation sequencing technologies facilitated analysis of the structures of microbial communities based on deep sequencing of 16S RNA clones and studies targeted to the whole community metagenome. We investigated metagenomes of extreme environments, - methane hydrate bearing sediments of the Lake Baikal, deep subsurface thermal waters in Tomsk region and hot springs of Kamchatka region. The results obtained revealed the influence of extreme conditions on the diversity of microorganisms, the main biogeochemical processes in these extreme environments, and new groups of prokaryotes representing deep phylogenetic lineages.

DEVELOPMENT OF THE OPTIMAL ALGORITHM OF BACTERIAL WHOLE GENOME SEQUENCING ON MiSeq AND GS JUNIOR 454 SEQUENCERS

Markelov M.L., Gordukova M.A.*, Kuleshov K.V., Dedkov V.G., Alvarez Figueroa M.V.
FBIS Central Research Institute for Epidemiology, Moscow, Russia
e-mail: maria.gordukova@pcr.ru
** Corresponding author*

Motivation and Aim. We deal with two types of tasks for whole genome analysis: resequencing with analysis of SNPs and *de novo* assemble. Precision of SNPs analysis is important for interlaboratory networks for the investigation of outbreaks and clinical tasks related to the detection of mutations linked to microorganisms' resistance. We have evaluated relative performance of new platforms with the aim to offer the optimal algorithm for bacterial whole genome sequencing. We used two Personal Sequencing Platforms: GS Junior System 454 (Roche Diagnostics) and MiSeq (Illumina). We performed whole genome sequencing of two different clinically significant bacterial strains with different GC-content: the vaccine strain of *M. bovis* BCG Russia (65.4%) (BCG) and *V. cholerae* (47.6%) (VCh).

Methods and Algorithms. Both genomes were sequenced using two runs on 454 to get enough coverage for high quality *de novo* assemble. For MiSeq sequencer we realized one run with pooled libraries from BCG and VCh. For *de novo* assemble of 454 data we used Newbler v2.5 software, for MiSeq pair-end reads – CLC Genomics Workbench v5.0.1. To make investigation and correction of SNVs and indels appearing due to low coverage of *de novo* contigs obtained from 454 we mapped MiSeq pair-end reads to 454 *de novo* assembled contigs.

Results. We received 112 Mb for BCG and 87 Mb for VCh from 2 runs of each sample by using 454 and 550Mb for BCG and 508.2 Mb for VCh with pooled libraries by using MiSeq. Some data on sequencing results and *de novo* assembly contigs for each bacterial strain are summarized in the table.

Strain	Sequencer	Runs	Sequenced reads	Average length of reads	Average coverage of *de novo* contigs	*De novo* contigs	N50, kb
BCG	MiSeq	0.5	3 741 498	147x2	98	243	38
	454	2	275 865	477	27	146	101
Vch	MiSeq	0.5	3 741 498	147x2	102	202	55
	454	2	218 882	350	33	124	135

We detected incorrectly identified nucleotides and indels in homopolymer regions in VCh sequenced contigs and single nucleotide deletions in GC-rich regions from contigs of BCG strain by mapping MiSeq reads on contigs generated with 454.

Conclusion. On the basis of these data we offer the following algorithm of genome sequencing of bacteria and *de novo* genome assembly. If necessary to receive long contigs, the first step would be to use 454 due to its ability to generate long reads to obtain long contigs for subsequent scaffolding procedures. Further due to the previously described limitations of pyrosequencing technology when dealing with homopolymers it is necessary to confirm the accuracy of the received sequences. It is only possible with huge coverage reads through Sequencing-by-Synthesis usingIllumina instrument.

CORRELATION BETWEEN TRANSCRIPTION EFFICIENCY INITIATION AND TRANSLATION EFFICIENCY FOR *SACCHAROMYCES CEREVISIAE* AND *SCHIZOSACCHAROMYCES POMBE*

Matushkin Yu.G.*, Levitsky V.G., Orlov Yu.L., Likhoshvai V.A.
Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia;
Novosibirsk State University, Novosibirsk, Russia
e-mail: mat@bionet.nsc.ru
** Corresponding author*

Key words: *translation efficiency, nucleosome formation potential, elongation efficiency index correlations, Saccharomyces cerevisiae, Schizosaccharomyces pombe*

Motivation and Aim: The hypothesis to test is the following: effective gene expression needs coordinately optimized processes of transcription and translation, in particular – transcription initiation and translation elongation.

Methods and Algorithms: Elongation Efficiency Index (EEI) was suggested earlier in [1] to estimate gene expression efficiency by nucleotide context of coding sequence in unicellular organisms. We have analyzed association between EEI and nucleosome formation potential (NFP) in 5' regulatory regions upstream translation initiation site (TIS) from two yeast species. Theoretical estimations of NFP based on DNA sequence were obtained by Recon method [2].

Results: Elongation efficiency negatively correlates with nucleosome formation potential. Therewith the selection may lead both on NFP decrease for high-expressing sequences (for *S.pombe*), and on NFP increase for low-expressing ones (for *S. cerevisiae*). Apparently this is the cause of distinct distributions of poly(A) tracts in upstream regions for two yeast species. For both yeast species there are regions with significant negative correlation of elongation efficiency index and nucleosome formation potential in 5' regulatory regions for all genes and especially for high expression level genes estimated by EEI. This negative correlation is more significant for *S.pombe.* Therefore for genes with low expression level there are regions with significant positive correlation, and this positive correlation is more significant for *S.cerevisiae*. In addition, for *S.cerevisiae* there is strong significant correlation between elongation efficiency index and nucleosome formation potential in coding regions for all genes and especially for genes with high expression level.

Conclusion: We have shown inter-relation between nucleosome localization in promoter regions and elongation efficiency index for yeast species *S. cerevisiae* and *S. pombe*.

References:

1. N.V. Vladimirov, V.A. Likhoshvai, Yu.G. Matushkin (2007) Correlation of Codon Biases and Potential Secondary Structures with mRNA Translation Efficiency in Unicellular Organisms. *Mol Biol (Mosk)*., **41**(5): 843–850.
2. Victor G. Levitsky (2004) RECON: a program for prediction of nucleosome formation potential. *Nucleic Acids Res*., **32: 346–349.**

NOVEL APPROACHES TO RNA SEQUENCING

Mazur A.M., Artemov A.V.
ZAO "Genoanalytica", Moscow, Russia

Motivation and Aim: RNA-seq analysis became the most powerful tool to explore disease-related phenomena. On one hand, transcription, unlike genotype, describes a phenotype of a cell. On the other hand, compared to different approaches of massive phenotype analyses, e.g. proteomics, RNA-seq has become much more comprehensive, sensitive and unbiased after the development of modern sequencing methods.

In this work we performed a comprehensive RNA-seq analysis of a skin disease. We explore not only disease-associated gene expression changes, but also shifts in alternative splicing patterns (including appearance of novel splicing isoforms) and changes in the expression within intergenic genome regions.

In order to reduce the influence of between-individual variation, we took two samples of the same tissue for every individual: one affected by the disease and the other not affected. 4 individuals participated in the study giving the total of 8 samples analyzed.

After the removal of ribosomal RNA, a paired-end sequencing (50bp from one side of a fragment and 35bp from the other) of the total RNA was performed on a SOLiD4 machine. The obtained colorspace reads were mapped to the human genome (hg19).

Methods and algorithms

On the first step, we performed a differential gene expression analysis with in-house R scripts utilizing 'edgeR' package. Taken the number of reads mapped to all exons of all annotated genes in all the samples analyzed, we estimated a p-value indicating if a gene's expression level differs between the samples affected and not affected by the explored disease.

On the next step we compared splicing patterns between affected and not-affected tissue samples. Paired-end reads were mapped with respect to potential exon-exon junctions by tophat software. Particularly, two reads of a pair were allowed to be mapped to different positions of genome and moreover every read were split into two parts and each of those was mapped separately. This mapping was further used to search for the splicing events annotated in Gencode V7 database which significantly differ between the samples of skin affected and not affected by the disease. We also performed *de novo* splicing isoforms assembly (i.e. without taking any reference exon-intron annotation into account) to find novel splicing events.

On the third step we explored intergenic expression in the given samples. We first defined clusters (i.e., genomic regions) of RNA reads expressed in at least one sample. The mappings for all samples were merged together. Several approaches of clustering were implemented: peak calling analogous to the one performed while ChIP-seq data processing; defining a cluster by the minimum number of reads it contains and maximal gap length between adjacent reads in one cluster; defining a cluster by the minimal number of reads and the maximal length of a cluster. The third definition with minimal number of reads = 30 and maximal length = 500 bp appeared to be the best in terms of presence of the reads within a cluster in more than one sample. Next, for every cluster we quantified its presence in every sample by counting the reads mapped to the cluster. These data were then used for a differential expression analysis similar to the gene expression analysis described earlier.

Results: Within the top 100 significant genes we found over-represented Gene Ontology (GO) categories which were consistent with the current understanding of the studied disease. Among differentially spliced genes a particular category of receptors was found. The differentially expressed regions will be further annotated with respect to known genomic features.

Conclusions: RNA-seq proved to be a source of biologically relevant information in the study of a skin disease. Extraction of the splicing and intergenic expression patterns provides meaningful results in addition to the results of gene expression analysis.

GENOME-WIDE ASSOCIATION STUDY OF CARDIOVASCULAR DISEASE RISK FACTORS IN THE MOSCOW STUDY OF THE WESTERN DISTRICT

Meshkov A.N.*[1], Khasanova Z.B., Konovalova N.V., Kotkina T.I., Sergienko I.V., Karpov Iu.A., Kukharchuk V.V., Boytsov S.A.[1]
Russian Cardiology Research Center, Moscow, Russia;
[a]*National Research Center for Preventive Medicine Moscow, Russia*
e-mail: meshkov@cardio.ru
** Corresponding author*

Motivation and Aim: Cardiovascular Disease is the leading cause of mortality in the Russian Federation. To identify common genetic polymorphisms associated with cardiovascular disease risk factors - total cholesterol, LDL- and HDL-cholesterol, triglycerides, lipoprotein(a), modified LDL-level, hypertension, smoking, and type-2 diabetes, we performed a genome-wide association study (GWAS) in 1,200 patients (366 male/834 female) from Moscow study of the Western District cohorts.

Methods and Algorithms: We genotyped 1,200 samples using the Illumina Cardio-Metabo BeadChips for a total of 196725 SNPs passing QC and allele frequency filters. Genotypes were called using a clustering algorithm in Illumina's BeadStudio software.

Results: We report 2 SNP associated with triglycerides level rs7259004 (p=9.623e-08) and chr19:50124992 (p=2.834e-07), 1 SNP associated with LDL-cholesterol level chr19:50103919 (p=6.654e-09), one cluster of 67 SNP located on chromosome 6 associated with lipoprotein(a) and one cluster of 14 SNP located on chromosome 2 associated with modified LDL level.

Conclusion: With the GWAS in the Moscow study of the Western District we confirm the previously found genetic associations with LDL-cholesterol levels, triglycerides levels, lipoprotein(a) levels and levels of modified LDL.

Acknowledgements: It is supported by Moscow City Government (#8/3-280н-10 and #8/3-281н-10).

LARGE-SCALE AMPLICON TARGETING MASSIVE PARALLEL RE-SEQUENCING REVEALS NOVEL VARIANTS IN ALZHEIMER'S DISEASE GENES

Mikhaylichenko O.A.[1, 2], Goltsov A.Y.[1, 2, 3], Gusev F.E.[1], Reshetov D.A.[1, 2], Tyazhelova T.V.[1], Andreeva T.A.[1,3], Kaljina N.R.[1,3], Grigorenko A.P.[1,3,4], Rogaev E.I.*[1,2,3,4]

[1] Vavilov Institute of General Genetics, Moscow; [2] Lomonosov Moscow State University; [3] Research Center of Mental Health, Russian Academy of Medical Sciences, Moscow, Russia; [4] Brudnick Neuropsychiatric Research Institute, Department of Psychiatry, University of Massachusetts Medical School, 01604, MA, USA

e-mail: EVGENY.ROGAEV@umassmed.edu

** Corresponding author*

Key words: *multiplex PCR, targeted resequencing, SOLiD, Alzheimer's disease*

Motivation and Aim: Emerging next-generation sequencing (massive parallel sequencing, MPS) technologies may transform the field of personalized medicine. Currently, direct sequencing of individual genome or exomes (all protein encoding genes) is already a feasible task. The next challenging task is to develop a large-scale MPS analysis of targeted genes in an extra- large clinical or population cohort, which could not be performed by conventional sequencing approach. We developed here the time- and cost-effective experimental and bioinformatics methodology for MPS analysis of disease-related genes. To perform the proof-of-concept study we selected the presenilin (*PSEN1, PSEN2*) and Amyloid precursor protein (APP) genes bearing mutations in familial cases of early Alzheimer's disease (AD). The comprehensive analysis of these genes is required to elucidate their role in the most common "sporadic" AD. Thus, we employed the un-biased clinical cohort of AD patients regardless their familial history and age-of onset. In addition, several other genes interacting with functional pathway of presenilins and APP were included in the study. A pipeline for MPS analysis of pooled multiple gene amplicons, threshold-criteria for filtration of sequencing errors and ready-to-run solutions for variant predictions has been developed and tested in a large size clinical cohort sample.

Methods and Algorithms:(1) We developed methodology for "all exons" multiplex amplification in single reactions for encoding regions of *APP,* and γ-secretase complex genes (*PSEN1, PSEN2, PEN-2, APH1A, NICASTRIN*), risk factor of disease *APOE* and the cleaving APP enzyme *BACE1.* In total, the samples from 552 AD and control individuals were tested. (2) The optimal algorithms for pooling of individual genomic DNAs and bar-coded MPS libraries were generated. The bar-coded pooled amplicon sets were sequenced using SOLiD™ 4 MPS platform. The raw reads were aligned with BioScope software and processed with in-house created PrimerCut tool which we developed to discriminate between genomic and synthetic PCR primer sequences. (3) Aligned and processed reads were transformed into pileup format with SamTools (http://samtools.sourceforge.net/); SNP calling was based on predicted variation frequency in a pool of N individuals/n reads.

Results: We showed the feasibility of analysis of 80 exons (18696 bp overall length)/per individual for >500 individuals in a single SOLiD™ 4 run and identified novel mutations in AD genes. The proposed bioinformatics pipeline is a ready-to-use solution for pooled DNA targeted re-sequencing analysis for both the research and medical purposes.

Availability: The PrimerCut tool is available on request.

Acknowledgements: This work was supported by Grant № 16.512.11.2083 "Research and development on priority directions of scientific-technological complex of Russia for years 2007-2013".

THE GENOME OF THE CTENOPHORE *PLEUROBRACHIA BACHEI*: NEW INSIGHTS INTO EVOLUTION OF METAZOA AND ORIGIN OF NERVOUS SYSTEMS

Moroz L.L.[1, 9], Kohn A.[1], Grigorenko A.P.[2,3], Yu F.[1], Farmerie W.[1], Citarella M.[1], Tyazhelova T.V.[3], Reshetov D.[3], Gusev F.[2], Bostwick C.[1], Winters G.[1], Dabe E.[1], Povolotskaya I.[4], Kocot K.[5], Halanych K.[5], Gusev F.E.[2], Kondrashov F.[4], Solovyev V.[6], Ross J.[1], Rubakhin S.[7], Romanova E.[7], Daily C.[7], Sweedler J.[7], Berezikov E.[8], Rogaev E.I.*[2,3]

[1] University of Florida, Gainesville, USA; [2] University of Massachusetts Medical School (UMMS), BNRI, USA; [3] Vavilov Institute of General Genetics, Center of Genomics, Moscow; [4] Centre for Genomic Regulation, Barcelona, Spain; [5] Auburn University, USA; [6] University of London, London, UK; [7] University of Illinois, USA; [8] Hubrecht Institute, Netherlands; [9] University of Washington, Seattle and Friday Harbor, USA

e-mail: Evgeny.Rogaev@umassmed.edu

** Correspondence*

Key words: *ctenophore, Pleurobrachia, whole-genome*

Motivation and Aim: Our understanding of the animal origin is incomplete because of limited data from basal Metazoa. The phylum *Ctenophora* is one of the earliest animal lineages, but the phylogenic relationship of this group to Bilateral and non-Bilateral branches of metazoan animals remained controversial.

Methods and Algorithms: We have performed the complete genome sequencing for *Pleurobrachia bachei*. It has one of the most compact genomes within Ctenophore group. These holoplanktonic predators have sophisticated ciliated locomotion and well-recognized nervous and "true" mesoderm-derived muscular systems. Using 454/Roche, Illumina GA and HiSeq2000 sequencing platforms we achieved at least ~1,000x coverage of the genome. We performed RNA-seq profiling from major tissues to validated the genome assembly: 96% of predicted gene models are supported by transcriptome data.

Results: Our phylogenomic analysis demonstrates the most basal phylogenetic position of Ctenophores within the animal tree. This hypothesis is further supported by comparative analysis of selected gene families (including apparent lack of HOX genes and certain genes for canonical miRNA machinery). However, there is no evidence for elements required for serotonin, melatonin and histamine synthesis or conversion of phenylalanine to tyrosine. Our data of functional pathways reveal the Wnt, Notch classical MAPK, JNK and p38 MAPK signaling systems in *Pleurobrachia*. The genes for canonical proteolysis pathways including, e.g., intramembrane aspartic proteases, IMPAS/SPP and presenilin homologs, have also been identified. Apparently, there are no Toll-like receptors and cell-adhesion molecules, which are essential components in immunity. It is of great interest that our preliminary biochemical analysis and *in situ* hybridization showed no evidence for canonical neurotransmitters (e.g., dopamine and serotonin) in *Pleurobrachia*, implying that completely different neurosignaling molecules can be utilized in this branch of the Animal Kingdom. Our experimental data indicate that the nervous system in ctenophores is one of the most distinct in its morphological and molecular organization. Many "classical bilaterian neuron-specific" genes either not present or, if present, they are not expressed in neurons. Finally, we identified novel markers for ctenophores neurons. These data suggest that at least some of the ctenophore neural systems were evolved independently from those in other animals.

Acknowledgements. Supported by NSF, NIH, McKnight Brain Research Foundation, University of Florida, University of Washington and Friday Harbor, University of Massachusetts MS and Rostock group. Evgeny.Rogaev@umassmed.edu

DEEP SEQUENCING CHRYSANTHEMUM microRNA ON DIFFERENT STAGES OF PLANT DEVELOPMENT

Nedoluzhko A.V.*[1], Pantiukh E.S.[2], Rastorguev S.M.[1], Gruzdeva N.M.[1], Shulga O.A.[3], Prokhorchuk E.B.[1, 3], Skryabin K.G.[1, 3]

[1] *National Research Centre "Kurchatov Institute" (NRC "Kurchatov Institute"), Moscow, Russia;*
[2] *Lomonosov Moscow State University (MSU), Moscow, Russia;*
[3] *Centre "Bioengineering" Russian Academy of Sciences, Moscow, Russia*
e-mail: nedoluzhko@gmail.com
** Corresponding author*

Key words: *MicroRNA, Chrysanthemum morifolium, Next Generation Sequencing*

Motivation and Aim: MicroRNAs (miRNAs) are small endogenous non-protein-coding RNA molecules (18-24 bp), which regulate different processes in cells. Plants, animals, fungus and even viruses have various miRNA families, but sometimes they are very conservative among evolutionally distinct taxa. To understand the molecular basis of microRNA-induced processes it's important to accurately quantify known miRNAs expression profiles as well as to discovery novel miRNAs.

Methods and Algorithms: In this study we have determined miRNA profiles of Chrysanthemum morifolium in three stages of plant development: premature stage, generative stage (budding) and generative stage (full flowering). The miRNA sequencing was performed using SOLiD 4.0. Sequenced miRNA data were mapped on plant mirBase (version 18.0) using Bowtie software.

Results: We have found a broad range of known microRNAs in Chrysanthemum genome and identified miRNAs that are differentially expressed between different stages of the plant development. For example, for the generative stage we have described widely presented miRNA families: miR396 (14680), miR167 (7431), miR166 (6967), miR162 (5059), miR159 (1965) and etc.

Conclusion: Furthermore, we have identified potentially novel microRNAs, which will be proved using other technics such as Sanger sequencing, Northern blotting and (or) bioinformatic analysis.

UGENE ASSEMBLY BROWSER: A TOOL FOR NGS DATA VISUALIZATION

Novikov I.A.*, Fursov M.Y., Efremov I.E.
Novosibirsk Center of Information Technologies 'Unipro'
e-mail: inovikov@unipro.ru
** Corresponding author*

Key words: *bioinformatics, genome, next generation sequencing, visualization*

Motivation and Aim: Next generation sequencing methods has already become a very popular starting point of genome analysis. An important feature of NGS methods is the volume of produced data, which can exceed tens of gigabytes, thus presenting a challenge for developers of interactive visualization software. Users of such software typically need to get a complete overview of the whole data as well as to navigate between narrow regions of interest quickly. Moreover, it is usually needed to make further analysis of the data. Unipro UGENE [1] is a free bioinformatics suite incorporating popular algorithms and a convenient designer for organizing them into workflow. Assembly Browser is a quickly developing module of UGENE aimed for visualization and analysis of NGS data.

Methods and Algorithms: Assembly Browser stores aligned short reads in an embedded database. It makes possible to instantly access any region of short reads assembly/map without a need to load entire file into computer memory. Packing reads and computing overall coverage from scratch are time-consuming tasks, therefore they are done only once, during import of reads into the database. Reference nucleotide sequences are also imported into database when opened in UGENE instead of loading into memory entirely.

Results and Conclusion: We have developed an interactive software solution called Assembly Browser. It allows users to import data from SAM or BAM format and then get the overview with the most covered regions. Users can quickly navigate them even on computers with little memory. It is possible to view the reads, get their properties, highlight differences from reference, examine coverage graph and consensus sequence of a given region, as well as to export reads and consensus.

Availability: Assembly Browser is available as a part of UGENE suite. A ready to use version of UGENE as well as the complete source code is freely available under the terms of GPL license from the web site [2] or from the repositories of Ubuntu and Fedora Linux distributions. Binary packages are available for Linux, Windows and Mac OS X operating systems.

References:

1. Konstantin Okonechnikov, Olga Golosova, Mikhail Fursov, the UGENE team (2012) Unipro UGENE: a unified bioinformatics toolkit, *Bioinformatics,* doi: 10.1093/bioinformatics/bts091.
2. Unipro UGENE website: http://ugene.unipro.ru.

COMPARATIVE GENOME ANNOTATION OF TRYPANOSOMATIDS

Pintus S.S.*[1], Serrano M.G.[1], Alves J.M.[1,2], Matveyev A.[1], Sheth N.[1], Lara A.[1], Lee V. [1], Koparde V.N.[1], Rivera M.C.[1], Voegtly L.J.[1], Arodz T.J.[1], Maia da Silva F. [2], Camargo E.P. [2], Teixeira M.M.G.[2], Buck G.A.[1]

[1] Virginia Commonwealth University, Center for the Study of Biological Complexity and the Departmnent of Microbiology and Immunology, Richmond VA, USA;

[2] Departamento de Parasitologia, Universidade de São Paulo, São Paulo, SP, Brazil

email: sspintus@vcu.edu

** Corresponding author*

Key words: *Comparative genomics, trypanosomatids, rare tropical diseases, high throughput sequencing, genome annotation, Assembling the Tree of Life project*

Motivation and Aim: Trypanosomatids are a group of exclusively parasitic kinetoplastid protozoa, infecting primarily insects. Life cycles of some trypanosomatid species, e.g. *T. brucei*, *T. cruzi* and various species of *Leishmania* involve a secondary human host, resulting in major deseases in humans, such as sleeping sickness, Chagas disease and leishmaniasis. Here we present an approach and results for high throughput sequencing and comparative annotation of several genomes of trypanosomatids, including the genera *Trypanosoma*, *Leishmania* and *Crithidia*.

Methods and Algorithms: The whole genome sequencing was done using Roche 454 pyrosequencing and Illumina (Solexa) technology. The assemblies of the genomes were performed *de novo* and improved using various software, such as Newbler (Roche), CLC Assembly Cell (CLC Bio), Velvet (EMBL-EBI), Minimus2 (AMOS consortium), etc.

In addition to recently developed ASGARD software [1], we have developed an informatics pipeline to annotate genes from the genomes under study. The pipeline includes such steps as prediction of protein coding genes, annotation of the gene orthologs and protein families, tRNA genes prediction, etc.

To assess the completeness of the assemblies in terms of genes annotation, we estimated the presence of trypanosomatid orthologs in genomes under study. The sets of trypanosomatid orthologs were recently created applying OrthoMCL software (PCBI, University of Pennsylvania) on the annotated *T. cruzi* CL Brenner, and *L. major* genomes. The ortholog presence tests were performed using the recently developed tool which implements Blast searches against the ortholog sets.

Calling orthologs from different incomplete subsets of reads allows draft estimating of the completeness of the genome assembly, while comparison between ortholog calls from annotated genes and contigs, allows identification of certain misassembled genes, as well as giving insights for trypanosomatid genome evolution.

Acknowledgments: The work was supported by grant Assembling the Tree of Life: Phylum Euglenozoa NSF DEB 0830056 Buck (PI) 09/15/08 – 09/14/13

References:

1. J.M. Alves, G.A. Buck. (2007) Automated system for gene annotation and metabolic pathway reconstruction using general sequence databases, Chemistry & Biodiversity, 4:2593-2602.

EPIGENETIC STATUS AND QUANTITATIVE CHARACTERISTICS OF CIRCULATING DNA IN LUNG CANCER

Ponomaryova A.A.*[1], Rykova E.Y.[2], Cherdyntseva N.V.[1], Skvortsova T.E. [2], Dobrodeev A.Y.[1], Zav'yalov A.A.[1], Tuzikov S.A.[1], Vlassov V.V. [2], Laktionov P.P. [2]
[1] *Cancer Research Institute, SB RAMS, Tomsk, Russia;*
[2] *Institute of Chemical Biology and Fundamental Medicine, SB RAS, Novosibirsk, Russia*
e-mail: anastasia-ponomaryova@rambler.ru
** Corresponding author*

Key words: *circulating DNA, methylation, tumor suppressor genes, lung cancer*

Motivation and Aim: Analysis of concentration and tumor suppressor genes methylation of cirDNA in lung malignancies and chronic obstructive pulmonary disease.

Methods and Algorithms: Blood samples were taken from 32 healthy subjects (HS), 22 patients with chronic obstructive pulmonary disease (COPD) and 55 untreated patients with non-small cell lung cancer (NSCLC) before treatment. The cell-surface-bound cirDNA (csb-cirDNA) fractions were obtained by successive treatment with PBS/EDTA and trypsin solutions. The copy number of ACTB gene and LINE-1 repetitive element was measured by quantitative real-time PCR. Concentration of methylated and unmethylated RASSF1A and RARB2 tumor suppressor genes circulating in blood was quantified by methylation-specific PCR and methylation index (IM) was calculated as IM = 100 × [copy number of methylated /(copy number of methylated + unmethylated gene)] for cirDNA and csb-cirDNA.

Results: Using PCR assays the significantly decreased concentrations of csb-cirDNA were shown in the blood of NSCLC and COPD patients compared with HS ($P<0.05$ – for ACTB and $P<0.01$ – for LINE-1, Mann-Whitney U test). IM values for RASSF1A and RARB2 genes were significantly elevated in cirDNA from NSCLC patients compared with HS (39% vs 19% in csb-cirDNA, 45% vs 21% in plasma cirDNA for RASSF1A; 35% vs 11% in csb-cirDNA, 51% vs 17% in plasma cirDNA for RARB2; $P<0.05$). If at least one from RASSF1A or RARB2 IM exceeded the cut-off values NSCLC patients were discriminated from HS with sensitivity and specificity of 90% and 82% when both plasma cirDNA and csb-cirDNA were analyzed. Discrimination of NSCLS from COPD patients was characterized by 88% sensitivity and 80% specificity. Values of RARB2 IM significantly increased in csb-cirDNA and plasma cirDNA from COPD (23%) and NSCLC (35%) patients compared with HS (11%) ($P<0.05$). RASSF1 IM values of plasma cirDNA and csb-cirDNA did not differ between COPD and HS. RARB2 gene IM increase was associated with advanced stage of NSCLC.

Conclusion: Concentration changes of ACTB and LINE-1 fragments demonstrate a strengthening of the processes in lung cancer leading to unequal representation of the genomic DNA fragments in cirDNA of blood. Epigenetic alterations of tumor suppressor genes in the total cirDNA were found to be associated with lung cancer development and progression. Methylation status of two candidate epigenetic markers (RARB2 and RASSF1A genes) in the cirDNA from plasma and csb-cirDNA fractions in combination was found to be valuable for lung cancer diagnostics and tumor staging.

Acknowledgements: The research has been carried out with support of the grants from RFBR № 11-04-12105-offi-m-2011, SB RAS Program in collaboration with other scientific organizations № 65, RAS Program "Fundamental Science for Medicine" № 23, Federal Special-Purpose Program " Scientific, Academic, and Teaching Staff of Innovative Russia" 2009-2013 (№ P256).

NEW ALGORITHM FOR IDENTIFICATION OF INDIVIDUAL DIFFERENCES IN GENE EXPRESSION

Pošćić F. *[1], Khlopova N. [2]
[1] *Udine University, Udine, Italy, e-mail: filip.poscic@uniud.it;*
[2] *RSAU-MTAA named after K.A. Timiryazev, Moscow, Russia*
** Corresponding author*

Key words: *probabilistic similarity index, microarray, gene expression, Sus scrofa*

Motivation and Aim: A methodology for identification of individual differences without replicates in gene expression is described. Comparing two or more groups of individual is well known as analysis of variance. However, as the microarray requires a lot of processing and it is quite expensive, it is not always possible to have replicates.

A few articles concerning individual variability used replicates or simply reported genes with 2, 3, 4-fold change criteria as up or down-regulated genes [1]. This method ignores the fact that these differences may or may not account biological effects.

Housekeeping genes (hkgs) are widely used as internal controls in a variety of study types like microarrays and it is possible to choose appropriate internal controls according to procedures such as those from [2].

Methods and Algorithms: As there exist no *a priori* definition of similarity, the use of Goodall probabilistic similarity index using Manhattan distance (PSI) [3] seems to be appropriate. In the PSI an ordering relationship between individuals for each gene is established according with everyday concepts of similarity, but also taking into account the probability that a random sample of two would have the values in question.

Since the PSI procedure is computationally consuming, as an approximation one may divide the distribution into a convenient number of groups. Under the assumption of limited variance of hkgs, we developed a new algorithm for not-arbitrary grouping of individuals. Each individual is situated in a specific group according to its intra-group similarity. Variance of each group should not be larger than those in the hkgs.

For the validation of our model we analysed real data. The study was performed on *Sus scrofa* individuals from the same breed, of the same age and raised in the same sheds. Standard procedures for the microarray technology were applied. For accounting systematic variations occurred during experiments, sample values were normalized [4].

Results: Our results were than compared with those from simple fold change criteria. We clearly demonstrated the power of the model for searching the potentially differently expressed genes in individuals.

Conclusion: The PSI and our model were applied successfully on our data for identification of significantly differently expressed genes between individuals.

Availability: The whole procedure has been programmed in C++ for massive analysis of data on personal computer and is currently on check for optimization. It will be available upon request.

References:

1. Sasaki A. et al. (2003) Individual differences in gene expression in primary cultured renal cortex cells derived from japanese subjects, IPSJ-DC, 2: 710-715.
2. Lee S. et al. (2007) Identification of Novel Universal Housekeeping Genes by Statistical Analysis of Microarray Data, JBMB, 40: 226-231.
3. Goodall D.W. (1966) A New Similarity Index Based on Probability, Biometrics, 22: 882-907.
4. Smyth G.K., Speed T.P. (2003). Normalization of cDNA microarray data. Methods 31, 265-273.

COMPARATIVE ANALYSIS OF TRIPLETS FREQUENCY IN MITOCHONDRIAL GENOMES

Putintseva Yu.A.[1, 2]
[1] *Siberian Federal University, Krasnoyarsk, Russia;*
[2] *Krasnoyarsk state medical university, Krasnoyarsk, Russia*
e-mail: yuliya-putintseva@rambler.ru

Key words: *triplet frequency, codons, elastic maps*

Motivation and Aim: A study of statistical properties of nucleotide sequences may bring a lot towards the architecture of genome as well as about the relation between structure and function encoded in these former. A consistent and comprehensive investigation of the features and peculiarities is based on the study of frequency dictionary of a nucleotide sequence. Such approach answers the questions concerning the statistical and information properties of DNA sequences. A frequency dictionary, whatever one understands for it, is rather multidimensional entity.

Methods and Algorithms: 2461 mitochondrion genomes were retrieved from the page of European Bioinformatics Institute (http://www.ebi.ac.uk/genomes/organelle.html). The list of available genomes is inhomogeneous, from the point of view of the equity of the number of species of various genders enlisted into the database. To eliminate the effect of the possible bias associated with this heterogeneity, we hashed the databases: a single genome from a gender was selected randomly, while the other ones were eliminated from the database. It resulted in a decrease of the number of entries in the database up to 1651 ones.

A standard unsupervised classification technique was implemented to develop a classification of the genomes in 63-dimensional space of triplets and codon frequencies. We used ViDaExpert software [1] to do that.

Results: 1651 mitochondrial genomes were classified in 63-dimensional space of triplets and codon frequencies for four genes with the K-means method and method of elastic maps. Picture of the distribution of genomes in this space is now being studied. However, it is clear that the picture is considerably different for different genes and is not random.

Conclusion: Genetically, the mitochondrion genomes are rather conservative, thus providing a good raw for knowledge extraction. The distribution of mitochondrion genes in 63-dimensional space of triplets and codon frequencies is far from a random one.

Acknowledgements: This work was supported by grant from Russian government department of Science and Education to Siberian Federal University «The genetic researches of the Siberian larch».

References:

1. 1. http://bioinfo-out.curie.fr/projects/vidaexpert/

POPULATION GENETIC ANALYSIS OF CASPIAN STURGEONS (*ACIPENCER GUELDENSTAEDTII, ACIPENCER PERSICUS*) USING NEXT GENERATION SEQUENCING AND CUSTOMIZED ILLUMINA GOLDENGATE GENOTYPING ASSAY

Rastorguev S.M.*[1], Nedoluzhko A.V.[1], Mazur A.M.[1], Gruzdeva N.M.[1], Tsygankova S.V.[1], Boulygina E.S. [1], Barmintseva A.E.[2], Mugue N.S.[2], Prokhorchuk E.B. [1]

[1] *National Research Centre "Kurchatov Institute" (NRC "Kurchatov Institute")*, *Moscow, Russia;*

[2] *Russian Federal Research Institute of Fishery and Oceanography (VNIRO), Moscow, Russia*

e-mail: rastorgueff@gmail.com

** Corresponding author*

Key words: *Acipencer gueldenstaedtii, Acipencer persicus, population genetics, Next Generation Sequencing, SNP, GoldenGate*

Motivation and Aim: The population structure and taxonomic status of *A. gueldenstaedtii* and *A. persicus* require to be investigated for determining quotas of catching and to identify biological characteristics of these objects such as homing, migration etc. Moreover, that is interesting to analyze inheritance of molecular markers in objects such as tetraploid sturgeons.

Methods and Algorithms: The transcriptome sequencing of *A. gueldenstaedtii* was done using SOLiD3 system, and polymorphic loci were determined. 384 SNPs were selected using these data to make customized Illumina GoldenGate DNA microarray for genotyping of 96 samples. Then the sturgeons from different locations of Caspian Sea were genotyped using GoldenGate assay and analyzed with several software packages. At the same assay a family group of sturgeon – male, female and their offspring - were analyzed to determine inheritance of examined polymorphisms.

Results: Population structure of Caspian sturgeons was investigated, and molecular markers for species differentiation were found. Differences between southern and northern populations of *A. gueldenstaedtii* and *A. persicus* were observed. A part of examined SNPs were combined into linkage groups.

Conclusion: Next Generation Sequencing technology is a good approach for large scale molecular markers identification, which can be used in population genetics of even non-model species. These markers allow distinguishing populations with a high statistical support. The customized microarray assay is an easy way to perform large scale genotyping.

GENOME SEQUENCES OF CENTENARIANS PRODUCE A BASIS FOR GENOME SCALE LONGEVITY STUDIES

Reshetov D.A.[1,2], Shagam L.I.[1,2], Tyazhelova T.V.[1], Grigorenko A.P.[1,3,4], Andreeva T.A.[1,3], Mikhaylichenko O.A.[1,2], Protasova M.S. [1], Goltsov A.Y. [1,3], Zenin A.A.[5], Gusev F.E.[1], Rogaev E.I.*[1,2,3,4]

[1] Vavilov Institute of General Genetics, Moscow; [2] Lomonosov Moscow State University; [3] Research Center of Mental Health, Russian Academy of Medical Sciences, Moscow, Russia; [4] Brudnick Neuropsychiatric Research Institute, Department of Psychiatry, University of Massachusetts Medical School, 01604, MA, USA; [5] Institute of Functional Nuclear Electronics NRNU MEPHI

e-mail: rogaev@vigg.ru, EVGENY.ROGAEV@umassmed.edu

** Corresponding author*

Key words: *centenarians, longevity, whole-genome sequencing*

Motivation and Aim: Human longevity is known to run strongly in families. Its heritability estimates in twin studies range from 23% to 33%. To date, however, little progress yet has been made in identification of genes for longevity using the common molecular-genetic methods such as genetic association of longevity trait with variations in candidate genes or genome wide association studies (GWAS). Identification of specific allelic gene variations contributing to human variations in lifespan can predict presumable therapeutic targets for common diseases associated with aging.

As alternative approach to GWA, recently emerged direct deep sequencing of genome potentially can provide a new insight on biology of longevity. For example, we hypothesized that very rare genetic variations or mutations rather than combinations of common single nucleotide polymorphism (SNP) (tested in GWAS), may underlie exceptional longevity.

Methods and Algorithms: We have sequenced complete genomes of three centenarian individuals (100 years old and older) of Russian origin and three middle-age individuals with ~ 30-40 fold coverage using Illumina's HiSeq2000 platform. Each genome was processed by *Ngs-pipeline*, our bioinformatics tool developed for complete genome re-sequencing analysis, elimination of errors and prediction of the functional effects of single nucleotide and structural variations. In follow up comprehensive bioinformatics analysis of "3 Centenarian Genomes", we selected several genes bearing rare mutations in these centenarians and tested them in additional longevity population cohort of > 300 individuals (>85-90 years old), which included ~ 150 centenarians (from Russia and USA).

Results: We have identified about 3.5M of single nucleotide variations (SNV) in each genome and thousands of structural variations. For each SNV we checked it presence and minor allele frequency (MAF) in publically available datasets such as 1000 genomes project and Complete genomics. Among very rare (MAF < 0.1%) SNVs identified in each centenarian genome, several SNVs were notable in genes for specific biological pathway that highlight the importance of epigenetic regulation in longevity and elements of insulin/IGF-1 pathway linked previously to lifespan regulation in animal models.
Availability: Three centenarian genomes along with SNVs and structural variations will be available for browsing at our http://centenarian-browser.org/

Acknowledgements: This work was supported by Rostok Group.

NEUROGENOMICS: CHALLENGES IN DEEP-GENOME STUDIES

Rogaev E.I.*[1,2,3,4], Reshetov D.A.[1,2], Tyazhelova T.V.[1] , Mikhaylichenko O.A.[1,2], Goltsov A.Y. [1,3], Gusev F.E.[1], Andreeva T.A.[1,3], Kaljina N.R.[1,3], Zenin A.A.[1], Protasova M.S.[1], Kunijeva S.[3], Grigorenko A.P. [1,3,4]

[1] Vavilov Institute of General Genetics, Moscow; [2] Lomonosov Moscow State University, Dep. of Bioengineering and Bioinformatics; [3] Research Center of Mental Health, Russian Academy of Medical Sciences, Moscow, Russia; [4] Brudnick Neuropsychiatric Research Institute, Department of Psychiatry, University of Massachusetts Medical School, 01604, MA, USA

e-mail: EVGENY.ROGAEV@umassmed.edu

** Presenting author*

Key words: *deep sequencing, SOLiD, schizophrenia, epigenomics, Alzheimer's disease*

Motivation and Aim: The identifications of genes or specific exogenous factors contributing to neuropsychiatric and behavioral diseases have been challenged. Interactions between environmental and genetic factors may underlie chronic neuropsychiatric disorders. Despite reduced reproductive fitness, the rate of incidence for schizophrenia and autism is relatively high in worldwide populations. We postulate that the individual genetic constitution attenuate the programmed epigenomic modifications during puberty. The study of regulatory sequences, including genes for non-coding RNA and the epigenomic regions marked by specifically modified chromatin, is a promising field in psychiatric genetics. To date the genome wide association studies (GWAS, employing relatively common SNPs across the genome) produced the genetic data with relatively low predictive and diagnostic values. The recently emerged concept that rare genetic variations, rather than common population variations, underlie common diseases challenges the standard genetic association approach in neuropsychiatric genetics. Direct sequencing of all genes, or preferably whole genome sequences, will provide most complete genetic information of the patient. However, from our current knowledge on population genetic variability, we expect millions of SNPs (single nucleotide polymorphisms) and up to thousand of CNVs per individual in comparison to reference genome sequence. Thus, excluding genetic "background" and identification of disease – related variations in the individual genomes require testing on experimental battlefield.

Methods and Results: We determined: (1) to our knowledge, the first complete genome sequences of patients with Alzheimer's disease and schizophrenia; 2) the complete genomes of familial cases of some neurological diseases with unknown mutant genes. (3) We also made the effort for development of experimental and bioinformatics methodology for identification of somatic mutations in human tissues. The data obtained by HiSeq2000, Solid and Pacific Bioscience platforms will be presented. The threshed criteria for filtering and identification of biologically significant private mutations and rare polymorphisms across the whole genome, selected chromosome or genetic locus linked to the disease have been developed. The role of non-coding RNAs, epigenetic–genetic interactions and genetic alterations, uniquely specific for *Homo sapiens* or occurring in both extinct and extant hominids, across epigenomic landscape will be discussed.

Acknowledgements: This work was supported by The Ministry of Education and Science of Russian Federation Federal target programs № 02.740.11.0854, 16.512.11.2083, 16.512.11.2102, FP7-HEALTH-2009, № 242257 (ADAMS), NIH/NIA 1R01AG029360, RFBR. We are grateful to Rostock group and A. Chikunov for support.

CIRCULATING DNA IN CANCER PATIENTS BLOOD: GENERAL CHARACTERISTICS AND WHOLE – GENOME ANALYSIS

Rykova E.Y.*[1], Morozkin E.S.[1], Loseva E.M.[1], Skvortsova K.N.[1], Ponomaryova A.A.[2], Kurilshikov A.M., Morozov I.V., Bryzgunova O.E.[1], Bondar A.A.[1], Zaporozhchenko I.S.[1], Kapitskaya K.Y.[3], Azhikina T.L.[3], Cherdyntseva N.V.[2], Vlassov V.V.[1], Laktionov P.P. [1]

[1] *Institute of Chemical Biology and Fundamental Medicine, SB RAS, Novosibirsk, Russia;*

[2] *Cancer Research Institute, SB RAMS, Tomsk, Russia;*

[3] *Shemyakin and Ovchinnikov Institute of Bioorganic Chemistry, RAS, Moscow, Russia*

e-mail: rykova@niboch.nsc.ru

** Corresponding author*

Key words: *circulating DNA, methylation, cancer, microarray, next generation sequencing*

Motivation and Aim: DNA modification studies in circulating DNA (cirDNA) may lead to development of specific non-invasive cancer biomarkers. CirDNA pool complexity became evident indicating that discovery of extracellular DNA generation and circulation patterns in cancer is essential to develop the valid diagnostic markers.

Methods and Algorithms: Comparative study of plasma cirDNA methylation modifications from prostate cancer patients and two control groups was made using the epigenome-wide screening by hybridization to the Human CpG island microarrays (UHN, Toronto, ON). Methylation index of RARB2 gene changes in cirDNA from lung cancer patient blood were detected using methylated DNA fragment enrichment by Methylated CpG Island Recovery Assay (MIRA) followed by real-time PCR. Apoptotic DNA isolated from culture medium (cm-apoDNA) of human umbilical vein endothelial cells (HUVEC) induced to apoptosis was compared with genomic DNA (gDNA) from the same normal cells using SOLiD 3 platform (Applied Biosystems, USA).

Results: Using the Differential Methylation Hybridization method 39 prostate cancer-associated changes in cirDNA methylation were identified. Pyrosequencing analysis of 7 selected loci revealed aberrant CpG methylation of two novel candidate cancer markers (ZC3H4 and RNF219). Gene RNF219 methylation was evaluated by cloning and sequencing of individual cirDNA molecules which demonstrated its diagnostic potential. A modified techniques for the methylated CpG-containing DNA fragment enrichment was developed based on the methyl-CpG affinity binding with the recombinant protein containing human methyl-binding domain 2 (MBD2) fused with glutathione S-transferase. Using the developed techniques RARB2 gene methylation index changes were found in cirDNA from lung cancer patients. As far as apoptosis is the main source of cirDNA, apoptotic DNA from HUVEC cells induced to apoptosis and genomic DNA from untreated cells were compared using SOLiD 3 platform. The representation analysis of repetitive elements revealed that cm-apoDNA is significantly enriched with Alu-repeats and depleted with LINE-1 elements compared with genomic DNA from intact cells. These data and the location of Alu-repeats mainly in euchromatin regions demonstrate the enrichment of cm-apoDNA with transcriptionally active gene-rich DNA sequences.

Conclusion: CirDNA genomics and epigenomics study provide promising source for development of non-invasive cancer biomarkers.

Acknowledgements: The research has been carried out with support of the grants from RFBR № 11-04-12105-offi-m-2011, SB RAS Program in collaboration with other scientific organizations № 65, RAS Program “Fundamental Science for Medicine” № 23.

INTRIGUING STRUCTURES IN TRIPLET DISTRIBUTION ALONGSIDE A GENOME

Sadovsky M.G.*, Mirkes E.M.
Institute of Computational Modelling of SB RAS, Krasnoyarsk, Russia
e-mail: msad@icm.krasn.ru;
** Corresponding author*

Key words: *order, periodicity, correlation, function, sense*

Motivation and Aim: The aim of the study is to identify, decipher and explore the semantics of a new structure revealed from the nucleotide sequences. The structure is expected to have a deep connection to the functionality and semantics of the nucleotide sequences. Simultaneously, similar relation is expected to be revealed between this newly revealed structure, and taxonomy of the bearers of those nucleotide sequences.

Methods and Algorithms: The distribution of the triplets (or, wide, short strings) alongside a sequence was developed. The distribution was defined as a distance to the nearest neighbour, i.e., for two triplets ω_1 and ω_2 the number n_l of nucleotides between them was counted so that there is no other word ω_2 embedded somewhere inside the string of the length n_l. The longest distance to detect the nearest neighbour was as long, as 10^4 nucleotides. The distribution function was developed for all 4096 couples of triplets. Two kind of surrogate sequences have been developed to distinguish the biological effects from the combinatorial ones: the former was random non-correlated sequence with the same composition of nucleotides, and the latter was a Markov chain sequence of order 3 to 6.

Results: A number of chromosomes of human genome, and chimpanzee genome have been studied. All chromosomes exhibit a strong and extremely unusual structures in the distribution of the triplets (to the nearest neighbour). Thus, an explicit and strong periodicity in CCC – GGG triplets has been found, with the period of 13 and 36 nucleotides. Some other couples exhibit more complex and long-range correlations (up to 250 nucleotides). The pattern of the couples distribution strongly differs from similar observed for random non-correlated sequence, and from those observed over Markov chain surrogate sequence of the order varied from 2 to 6.

Conclusion: a new structure is found in nucleotide sequences. Very complicated patter of the couples distribution is peculiar for higher eukaryotic organisms, while unicellular eukaryotes, as well as bacteria exhibit very smooth pattern close to that one observed for Markov chains of relevant order.

A MAP OF ANAPHASE CHROMOSOMAL BREAKS INDUCED BY CONDENSIN LOSS

Samoshkin A.[2], Dulev S.[3], Loukinov D.[4], Rosenfeld J.A.[5], Strunnikov A.V.[1]
[1] Guangzhou Institutes of Biomedicine and Health, Molecular Epigenetics Laboratory, 190 Kai Yuan Avenue, Science Park, Guangzhou 510530, Guandong, China
[2] NIH NCI, Genome Structure and Function Section, Bethesda, MD, USA;
[3] Ludwig-Maximilians-Universitдt, Adolf Butenandt Institut, 80539 Munich, Germany;
[4] NIH NIAID, Laboratory of Immunopathology, Rockville, MD, USA;
[5] Division of High Performance and Research Computing, University of Medicine & Dentistry of New Jersey, Newark, NJ, USA

Abstract: Condensin complexes are essential for mitotic chromosome condensation and segregation, while condensin dysfunction leads to chromosomal bridging in mitosis, paving the way for rapid genome destabilization undetectable by checkpoints, similar to genome rearrangements found in many cancer genomes. To map potential double-strand breaks specifically occurring in late anaphase, human chromosomes depleted of condensin were analyzed by gamma-H2AX ChIP-seq. Condensin-depleted chromosomes from HeLa cells contained distinct gamma-H2AX enrichment zones, 75% of which overlapped with known hemizygous deletions in cancers. Furthermore, some tandemly repeated DNA sequences, analyzed by ChIP-seq and custom repeat array ChIP-chip using independent high-throughput and bioinformatic approaches, showed significant gamma-H2AX enrichment in condensin-depleted anaphases. The preferential targets of such an enrichment included simple repeats, centromeric satellites, and rDNA. The genomic regions that are specifically destabilized upon condensin dysfunction may constitute a quantifiable condensin-specific CDP (Chromosome Destabilization Pattern).

PROFILE OF THE CIRCULATING RNA IN APPARENTLY HEALTHY INDIVIDUALS AND NON-SMALL CELL LUNG CANCER PATIENTS OBTAINED WITH MASSIVELY PARALLEL SEQUENCING OF TOTAL BLOOD PLASMA RNA

Semenov D.V.*[1], Baryakin D.N.[1], Brenner E.V.[1], Kurilshikov A.M.[1], Kozlov V.V.[2], Narov Y.E.[2], Vasiliev G.V.[3], Bryzgalov L.O.[3], Chikova E.D.[1], Filippova J.A.[1], Kuligina E.V. [1], Richter V.A.[1]

[1] *Institute of Chemical Biology and Fundamental Medicine SB RAS, Novosibirsk, Russia;*
[2] *Novosibirsk Regional Cancer Centre, Novosibirsk, Russia;*
[3] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*
e-mail: semenov@niboch.nsc.ru
** Corresponding author*

Key words: c*irculating RNA, massively parallel sequencing, human blood plasma RNA, non-small-cell lung cancer*

Motivation and Aim: The understanding of circulating RNA structures and functions expands fundamental knowledge of cell communications and signaling pathways as well as allows developing new molecular diagnostic approaches. The aim of this study was to document profile of common and peculiar RNA species normally circulating in blood of healthy individuals and of patients with non-small cell lung cancer with massively parallel sequencing of human blood plasma RNA.

Methods and Algorithms: Total RNA was extracted from blood plasma samples of 8 apparently healthy individuals and 8 patients with non-small cell lung cancer. To obtain comprehensive cDNA libraries RNA was dephosphorylated and then 5'-phosphorylated. 5'-phosphorylated total plasma RNA was ligated with adapters, reverse transcribed and 16 personalized cDNA libraries were constructed. Libraries were amplified and sequenced with SOLiD™ system. The sequenced 35-nt-long reads were mapped to human transcriptome/genome, classified and quantified with Bowtie/Cufflinks software.

Results: Fragments of rRNA, mitochondrial transcripts, microRNAs, fragments of scRNAs, snRNAs and snoRNAs, fragments of several mRNAs as well as the set of newly discovered transcripts were found to be permanent representatives of human blood plasma RNAs. Comparison of circulating RNA profiles of healthy subjects and cancer patients allowed us to document diagnostically significant RNA species, including fragments of mRNA, miRNA and others non-coding regulatory RNAs.

Conclusion: Documented profile of circulating RNA of healthy individuals and patients with lung cancer provides the basis for development of new research of circulating regulatory RNAs and allows to construct new platforms for early diagnosis of human malignancy.

Acknowledgements This study was supported by Integration SB RAS Grant #18, RFBR grants #10-04-01386-a and #10-04-01442-a.

TOWARDS AN ANALYSIS OF THE STRUCTURE OF THE SHORT ARM OF 5B CHROMOSOME OF THE BREAD WHEAT *TRITICUM AESTIVUM* L.

Sergeeva E.M.*, Afonnikov D.A., Bildanova L.L., Koltunova M.K., Timonova E.M., Salina E.A.
Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia
e-mail: sergeeva@bionet.nsc.ru
** Corresponding author*

Motivation and Aim: *Triticum aestivum* (bread wheat) is the allohexaploid (genomic formula BBAADD, 2n = 6x = 42), with the large genome size of 17 000 Mbp. Approximately 90 % of wheat genome account for repetitive sequences of different origin and degree of reiteration that make the wheat genomic sequencing very difficult and expensive. At the present day the wheat genome sequencing is in progress and little is known concerning the structure of the short arm of chromosome 5B. The basic approach for wheat genomic sequencing is the sequencing of minimal set of overlapping clones from chromosome-specific BAC-libraries. The key stage is the anchoring of mapped molecular markers to the BAC-clones. The essential task is to obtain the markers distributed uniformly along all the chromosome length. The 454-sequencing technology (Roche) produces the reads long enough for development such kind of PCR markers named ISBP (Insertion Site Based Polymorphism). The ISBP-marker is the pair of primers flanking the site of insertion of one transposable element into another, and despite the abundance of different transposable elements, ISBP could give the unique PCR product. We analyzed the composition of 5.5 % of 5BS chromosome covered by 454-reads. To perform the work related with marker anchoring we've made the selection of ISBP-markers developed on the base of 454-reads, and tested some of them for the mapping and screening of 5BS-specific BAC-library.

Methods and Algorithms: The set of 454 reads covers the 5.5 % of 5BS chromosome (39695 reads, total length 16183252 bp). The sequence composition was analyzed by RepeatMasker and search against TREP database. Also we assayed the content of specific sequences: rDNA and histone genes, centromeric and subtelomeric repeats, simple repeats and low complexity regions, mapped wheat EST (http://wheat.pw.usda.gov) and gene sequences of rice (http://rapdb.dna.affrc.go.jp). For the selection of the ISBP-markers we took 1302 markers obtained with ISBPFinder program and performed the BLAST search of amplicons against the wheat genomic sequences at NCBI database. We mapped the markers at the deletion bins of chromosome 5BS and used them for screening of BAC-library. The 5BS chromosome-specific BAC library contains 43776 clones (15-fold chromosome coverage). For fast and efficient PCR screening we performed the pooling of the library.

*Result*s: The known repeats families account for 47 % of the 454 reads, also we identified the matches with 5S rDNA, low complexity regions and the matches with mapped wheat ESTs. 12 % of ISBP-amplicons showed >90 % identity with published sequences; among them 4 % strongly matched with 5BS chromosome, 58 % had no data on chromosomal localization, 38 % matched with another chromosomes. The screening of 5BS BAC-library with some of selected ISBP-markers proved to be successful.

Conclusion: We performed the preliminary analysis of composition of the short arm of 5B chromosome, and showed the suitability of our method of markers selection for mapping and BAC-library screening, that will further contribute to the successful sequencing of the wheat 5B chromosome.

NextGen SEQUENCING REVEALS EXTENSIVE RNA EDITING IN PLASMACYTOID DENDRITIC AND OTHER PRIMARY CELLS

Sharma A.[1], Alomair L.[1], Doyle K.[1], Sikaroodi M.[3], Cherepanova A.[4], Laktionov P.P. [4], Birerdinc A.[1, 2], Gillevet P. [3], Baranova P.*[1, 2, 3, 5]

[1] School of Systems Biology, George Mason University, Fairfax VA 22030; [2]Betty and Guy Beatty Center for Integrated Research, Inova Health System, Falls Church, VA 22042;[3] Microbiome Analysis Center, George Mason University, Manassas, VA 20100; [4]Institute of Chemical Biology and Fundamental Medicine, SB RAS, Novosibirsk, Russia; [5]Research Centre for Medical Genetics, Moscow, Russia

e-mail: abaranov@gmu.edu

** Corresponding author*

Key words: *TLR receptors, RNA editing, 454 technology*

The alteration of the nucleotide sequence of RNA, termed RNA editing (RNAE) in mammals is mainly the conversion of adenosine to inosine which is translated as if it were guanosine. This reaction is catalyzed by so-called ADARs (adenosine deaminases that act on RNA) enzymes. ADARs editing is predominant in the CNS where it is essential for correct functioning of certain neurons. However important RNAE is, only a handful of physiologically important target genes for RNAE have been confirmed in humans.

In our previous study, extensive RNAE of the A→ G (ADAR) type was revealed in human mRNA encoding Toll-like receptor TLR7 that plays a fundamental role in pathogen recognition and activation of innate immunity in plasmacytoid dendritic cells (PDCs), capable of on-demand production of type I interferons and TNF-α. RNAE of TLR7 mRNA was confirmed by confirmed by re-sequencing of TLR7 gene in the DNA prepared from the same individual' PDCs preparations. PDCs represent only 0.4 % of total peripheral blood monocytes (PBMC), and it is not surprising that theywere overlooked by other investigators of the RNAE phenomenon.

Editing of TLR7 mRNA introduce non-random changes to the coding sequence of its LRR domains responsible for molecular recognition and change its amino acid sequence. TLR7 editing was observed at high level close to 100%, edited sites result in three to four amino acid changes per individual RNA and differ betweem molecules. TLR7 mRNA editing was sample specific, i.e. it was present in PDCs of some individuals, and absent in others, thus indicating that RNA editing in PDCs is probably an inducible event and a druggable target. TLR7 editing negatively correlated with an ability of of PDCs to respond to stimulation with CpG-A. Absence of the mutations in the DNA of the same individuals was. 454 sequencing demonstrates that RNA editing is not limited to TLR7, but also occurs in TLR9 and other TLR mRNAs extracted from PDCs and the extent of RNAE differ between healthy donors. RNAE of TLRs could be responsible for mutant phenotype of their Pathogen-Associated Molecular Patterns (PAMPs)-recognizing domain, and thus individual susceptibility to common and emerging infections in otherwise healthy subjects.

Futher study will justify if the levels of TLR editing is a lifetime characteristic of the individual or it is a temporary condition reflect external influence, whether TLR are edited in DCs and influences the antigen-presenting function of DCs, how ADAR enzymes activities regulated and whether an antagonistic interaction between ADAR and RNAi machineries is pertinent to PDCs functioning.

SEARCHING AND CLASSIFICATION OF BINDING SITES OF SIGMA FACTORS OF *CLOSTRIDIUM DIFFICILE*

Shelyakin P.
Institute for Information Transmission Problems RAS, Moscow, Russia
e-mail: f.serval@gmail.com, gelfand@iitp.ru

Annotation: The aim of this study is to identify genes that are regulated by different sigma factors and to obtain positional weight matrices for the promoters of Clostridia difficile 630.

Problem statement: We aim to sonstruct positional weight matrices for each sigma factor based on our knowledge of the transcriptional start points for 1500 genes of Clostridia difficile 630 [1], on the genes that potentially encode 22 sigma factors and on some genes that are known to be regulated by these factors. We then want to use those matrices to obtain information about other genes that are regulated by each sigma factor.

Results and Discussion: To construct positional weight matrices we used two methods. If sufficient experimental data on genes regulated by a certain sigma factor was available, we built the matrix using SignalX [2]. If the data were not extensive enough, we used matrices of Bacillus subtilis from the DBTBS database[3] .

Using Genome Explorer [2] we searched for genes regulated by each sigma factor in the genome of the *C. difficile* strain, for which all transcriptional start points were known, and in two genomes of other *C. difficile* strains. In the first strain we searched within the range of 40 bp from transcriptional start points; in the other two we searched in the area of 100 bp before the starting point of translation.

Having obtained three sets of genes, we then selected only genes that were preceded by candidate promoters in at least two strains. We constructed multiple alignment of their upstream nucleotide sequences in Muscle [4], after which we used SignalX to obtain positional weight matrices and Weblogo [5] to visualize the consensus sequences.

This is joint work with M. Gelfand.

References:

1. Marc Monot, Caroline Boursaux-Eude, Marie Thibonnier, David Vallenet, Ivan Moszer, Claudine Medigue, Isa Martin-Verstraete and Bruno Dupuy. *"Reannotation of the genome sequence of Clostridium difficile strain 630".* J Med Microbiol 2011 vol. 60 no. 8 1193-1199
2. Mironov A., Vinokurova N.P., Gelfand M.S., "Software for bacterial genome analysis", Mol. Biol., 2000
3. Sierro N., Makita Y., de Hoon M.J.L. and Nakai K. *"DBTBS: a database of transcriptional regulation in Bacillus subtilis containing upstream intergenic conservation information."* Nucleic Acids Res. 2008, 36 (Database issue):D93-D96; doi:10.1093/nar/gkm910
4. Edgar, R.C. (2004) *"MUSCLE: multiple sequence alignment with high accuracy and high throughput"* .Nucleic Acids Res. 32(5):1792-1797.
5. Crooks GE, Hon G, Chandonia JM, Brenner SE *WebLogo: A sequence logo generator*, Genome
6. Research, 14:1188-1190, (2004).

COMPUTATIONAL TOOLS FOR ANALYSIS OF NEXT GENERATION SEQUENCING DATA

Solovyev V.[1], Seledtsov I.[2], Vorobyev D.[2], Kosarev P.[2], Molodsov V.[2], Okhalin N.[2]
[1] *Department of Computer Science, Royal Holloway, University of London, UK;*
[2] *Softberry Inc., 116 Radio Circle, Suite 400, Mount Kisco, NY 10549, USA*
e-mail: victor@cs.rhul.ac.uk

Next-generation sequencing (NGS) transforms today's biology research by providing fast sequencing of new genomes, genome-wide association studies (GWAS), sequencing personal genomes, variant discovery by resequencing targeted regions or whole genomes, de novo assemblies of bacterial and eukaryotic genomes, annotating the transcriptomes of cells, tissues and organisms (RNAseq), and gene discovery by metagenomics studies. To analyze next-generation sequencing data we advanced further our **OligiZip** assembler and **Transomics pipelines** that provide solutions to the following tasks: 1) de novo reconstruction of genomic sequence; 2) reconstruction of sequence with a reference genome; 3) mapping RNA-Seq data to a reference genome and identification of alternative transcripts with quantification of their abundance; 4) fast alignment of assembled contigs to the genomic sequence.

To test reconstruction of bacterial sequences we assembled genomic sequence of Methanopyrus kandleri TAG11 and Methanopyrus kandleri AV19. Solexa reads, about 6 million each for AV19 and TAG 11, were produced by sequencing lab of Harvard PCPGM. AV19 genome itself has been assembled perfectly in one contig. Time of AV19 assembling is ~ 1 min on one Linux node. **OligiZip** also has been successfully applied to assembly a model eukaryotic genome in Assemblathon 1 collaborative efforts where many research groups presented genome assembles to estimate the accuracy of various assembling software.

For RNAseq data the **Transomics** pipeline initially maps reads to the genomic sequence and identifies spliced and non-spliced reads coordinates. This information used by our FGENESH gene prediction program that includes an iterative procedure for predicting alternative splicing gene variants. We have developed a module to compute a relative abundance of predicted alternative transcripts solving a system of linear equations. The initial variant of Transomics pipeline has been successfully applied to Human, C.elegans and Drosophila data of the RGASP project. We also developed a powerful **Sequence assembling Viewer** to work with the reads data and assembling results interactively. As an example of Transomics application for identification of disease specific genes we analyzed RNAseq data for non-tumorigenic epithelial cell line and epithelial cells from infiltrating ductal carcinoma of the breast. Comparative analysis shows a set of genes that have different alternative splicing forms in these cell lines.

OligoZip, Genome Aligner and **Transomics** pipeline components and other software programs are available to run independently at www.softwberry.com or as a part of integrated environment of the **Molquest** software package that can be downloaded at www.molquest.com for Windows, MAC and Linux OS.

CONTEXTUAL DNA FEATURES SIGNIFICANT FOR THE DNA DAMAGE BY THE 193 NM ULTRAVIOLET LASER BEAM

Vtyurina N.N.[1, 2], Grokhovsky S.L.[1], Vasiliev A.B.[2], Titov I.I.[3], Ponomarenko P.M.[3], Ponomarenko M.P.*[3], Peltek S.E.[3], Nechipurenko Yu.D.[1, 2], Kolchanov N.A.[3, 4, 5]
[1] Engelhardt Institute of Molecular Biology RAS, Moscow, Russia; [2] Lomonosov Moscow State University, Moscow, Russia;[3] Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia; [4] Novosibirsk State University, Russia; [5]National Research Centre "Kurchatov Institute", Moscow, Russia
e-mail: pon@bionet.nsc.ru
** Corresponding author*

Motivation and Aim. Ultraviolet **(**UV-) lasers are used in bioluminescence, cancer therapy, cosmetology, microsurgery, footprinting and many other ways of working with living organisms. A bulk of experiments had shown that laser-induced UV-radiation damages DNA. The most common UV-damage of the DNA at wavelengths ≤ 290 nm is the formation of 7,8-dihydro-8-oxoguanine and other oxidation products of guanine cation (G^+), leading to DNA strand breaks or nucleotide substitutions, which is fundamentally different from the DNA damage caused by other effects (e.g. ultrasound). These more frequent UV-damage of guanine is commonly associated with the energy hν of the absorbed photon that is enough for the escape of an electron (e-) from DNA with the formation of a "hole" (DNA^+). The "hole" then "walks" along the DNA helix and usually stops with the formation of G^+ as a result of a lower ionization threshold of G compared with A, T and C. Than G^+ is immediately attacked by free radical-anions. It was shown that the frequency of UV-damages of different guanines within one nucleotide sequence can vary widely, indicating its dependence on the nucleotide context of the area surrounding each guanine. However, this dependence is still unknown. The question of a systematic *in silico* analysis of the characteristics of the DNA nucleotide context affecting the frequency of guanine UV-damages has not yet been raised.

Methods and Algorithms. We are the first to carry out a systematic in silico analysis of the contextual characteristics of the DNA that affect the frequency of guanines damages at UV-radiation [Vtyurina et al., Biophysics (Mosc) 2011. 56: 410–41].

Results. Characteristics of the local environment of G that we found in this paper and that are significantly affecting the frequency of UV-damaged G, include: 1) consensus ttaaagcHtcg-actgc that is the significantly rare, 2) Position-Weight Matrix, PWM; 3) amount of the tetranucleotide YNVW upstream of G, and 4) estimated frequency of contact with the histone-like protein HU *Escherichia coli*, defects in which increase the frequency of UV-damages. We are the first to construct a linear-additive estimate of the UV-damaged G frequency upon its environment {S(G)} in the DNA strand, such as:

$$\mathbf{f\{S(G)\}=0.69-0.07N_{consensus}\{S(G)\}+0.19PWM\{S(G)\}+0.22YNWV_F\{S(G)\}+0.07P_{Histone\text{-}like}\{S(G)\}.}$$

Conclusion. We are the first to obtain the significant correlation between prediction *in silico* and measurements *in vitro* of the UV-damages frequency in both analyzed [Vtyurina et al., Biophysics (Mosc) 2011. 56: 410–41] ($r=0.679$, $\alpha<10^{-6}$) and independent experiment [Melvin T. et al. NAR 1998. 26:4935-42] ($r=0.821$, $\alpha<0.005$).

Acknowledgments. We thank RFBR 11-04-02001, 11-04-01888, 12-04-01584, НШ-5278.2012.4, and, also, The Russian Academy of Sciences, Presidium Programs "Molecular and Cellular Biology", "Biodiversity", "Biosphere Origin and Evolution".

COMPUTATIONAL NEW SPLICE VARIANTS DISCOVERY USING SINGLE MOLECULE SEQUENCING TECHNOLOGY

Vyatkin Yu.V. [1], Shtokalo D.N.[1, 2, 3], Kapranov P.[3], St. Laurent G.C. III*[3]
[1] *AcademGene LLC, Novosibirsk, Russia;*
[2] *A.P.Ershov Institute of Informatics Systems SB RAS, Novosibirsk, Russia;*
[3] *St.Laurent Institute, Providence, USA*
e-mail: georgest98@yahoo.com
** Corresponding author*

Key words: *alternative splicing, splice junctions discovery, next generation sequencing*

Motivation and Aim: Alternative splicing is an essential post transcriptional regulation phenomenon in eukaryotes by which exons of pre-mRNAs are reconnected in multiple ways. The discovery of alternative splicing events is important for understanding of cell functioning. Recent growth of Next Generation Sequencing technologies [1] provides a huge stream of new data, which leads to new discoveries in genomics [2]. We present a computational method of discovering alternative splice variants based on single molecule sequencing data. The method applied to sequenced *Drosophila melanogaster* transcriptome allowed us to discover previously unknown 435 splice events.

Methods and Algorithms: All possible exon-exon junctions both canonical and putative non-canonical were created combinatorically on per gene basis from all described genes from FlyBase 5.17 database. A set of reads sequenced using Helicos single molecule sequencing technology was aligned to the reference of splice junctions by applying Helisphere indexDPgenomic aligner [3] and then filtered for ambiguity. The non-canonical splice junctions that showed good coverage were annotated by a list of already reported spliced ESTs (http://genome.ucsc.edu) and recent studies [2].

Results and conclusion: Being applied to *Drosophila melanogaster* genome with 21 753 described genes, the method allowed us to build around 250 000 putative non-canonical and known canonical splice junctions. About 240 million informative reads of *Drosophila melanogaster* transcriptome were used to find significant expression of 46 648 splice junctions either canonical or non-canonical. Finally, 769 non-canonical splice junctions from 480 genes with an appropriate coverage were discovered, out of which 435 were not found by us to be reported neither by EST nor by similar study [2]. New sequencing technologies provide an opportunity to find new genomic features even in intensively studied genomes like *Drosophila melanogaster*'s.

References:

1. Thompson JF and Milos PM (2011). The properties and applications of single molecule DNA sequencing. *Genome Biology*, **12**: 217.
2. Graveley BR et al (2011). The developmental transcriptome of Drosophila melanogaster. *Nature*, **471**:473–479.
3. Giladi El. et al (2010). Error tolerant indexing and alignment of short reads with covering template families. *Journal of Computational Biology*, **17**: 1279-1293.

HOW LONG SEQUENCED GENOME CAN REMAIN STABLE

Zakharenko L.P.*[1, 2], Bak T.P. [1], Ignatenko O.M. [2]
[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*
[2] *Novosibirsk State University, Novosibirsk, Russia*
e-mail: zakharlp@bionet.nsc.ru
** Corresponding author*

Key words: *transposable elements, sequenced genomes, genome instability*

Motivation and Aim. Many genomes are sequenced. But how long is it possible to consider they stable? Especially it concerns stability of an arrangement in a genome of transposable elements (TEs) owing to their high mobility. *Methods and Algorithms.* We analyzed by FISH distribution of some *Drosophila melanogaster* TEs 20 years later after receiving library of clones which were used for sequences. *Results.* We have shown that rate of TEs moving varies from 10^{-2} to 10^{-4} in a genome of the isogenic *y cn bw sp* strain. If total number of TEs makes about 1500 copies on a Drosophila genome, and average speed of TEs moving is 10^{-3}, in each generation at least 1.5 TE change the position. In 20 years passed about 200 generations, that is about 300 Drosophila TEs or the fifth part of TEs changed there's positions. Strangely enough, rate of moving of TEs doesn't correlate with level of their expression and with number of full-size copies. So transposable element *blood* is presented in a genome by two tens full-size copies with an open reading frames and according literature *blood* is expressed in generative tissue. Nevertheless *blood* is extremely stable in this line. At the same time *hobo*-element is extremely movable in *y cn bw sp* genome though only one full-size copy and two tens of defective *hobo* is found in it. Moreover full-size *hobo* transcript wasn't found in any experiment. Discrepancy of level of a transcription of mobile elements, and speeds of their introduction in a genome observe also in other works (1, 2). According our data recombination between different sorts of repeats has significant influence on the TEs distribution in Drosophila genome. *Conclusion.* Thus, first, it is impossible to substitute speed of moving of TEs for level of its expression. Secondly, the genome is much more unstable, than it is accepted to think. However, despite high mobility of mobile elements a phenotype and genetic characteristics of the line under analysis didn't change. That is the importance of moving of mobile elements in life of population is probably strongly exaggerated. The instability of a genome caused by TEs moving, changes genome structure, but affects a little or doesn't affect mainly at all a phenotype. *Availability*: http://flybase.org/static_pages/lists/dmel_te.html. *Acknowledgements*: This work was partially funded by program of Russian Academy of Sciences "Biodiversity" (grant B 27-29) and program "Wildlife" (grant 30-30).

References:

1. Spradling et al.(2011) Drosophila *P* elements preferentially transpose to replication origins. PNAS 108 :15948-15953
2. Esnault et al.(2011) Intrinsic characteristics of neighboring DNA modulate transposable element activity in *Drosophila melanogaster.* Genetics. 187:319-331.

ArchiP: DETECTOR OF ARCHITECTURES IN 3D PROTEIN STRUCTURES

Aksianov E.A.*, Alexeevski A.V.
A.N. Belozersky Institute and Faculty of Bioengineering and Bioinformatics, Lomonosov Moscow State University,
Scientific-Research Institute for System Studies, Russian Academy of Science (NIISI RAN)
e-mail: evaksianov@gmail.com
** Corresponding author*

Key words: β-*sheet, protein architecture, automated detector*

Motivation and Aim: Fold description and classification of 3D protein structures currently requires human judgment in certain cases. The lack of rigorous criteria in fold definitions leads to misunderstanding between different experts in structural classifications. The more difficult step in fold description is characterization of the domain architecture, i.e. spatial arrangement of β-sheets and α-helixes in 3D space. The aim of the work is to elaborate formal definitions of architectures and to develop an automated detector of architecture of all-β or β/α classes according SCOP classification.

Methods and Algorithms: β-sheets are detected by a previously developed program, named SheeP (http://mouse.belozersky.msu.ru/sheep), α-helixes are detected by DSSP algorithm. β-sheets and α-helixes are primary structural units of the architecture. Pair contacts of structural units are determined. Modifications of the set of structural units on the base of contact regions in β-sheets are implemented. Architecture of a protein domain is described in terms of graph, vertices of which are structural units and edges are pairwise contacts between them. The algorithm detects β-sandwiches, β-barrels (open or closed), βα-sandwiches, and six other architectures.

Results: A program ArchiP and web-service were created. To test ArchiP we select a set of domains (one domain from one SCOP family) for each tested architecture. ArchiP detected correctly

– 98.7 % (611 of 619) domains selected from folds, annotated in SCOP as αβα-sandwich folds,

– 93.8 % (245 of 261) domains selected from folds, annotated as β-sandwiches,

– 68.0 % (81 of 119) domains from TIM-barrel folds,

– 62.7 % (140 of 223) domains from β-barrel folds,

– 34.7 % (8 of 23) domains from ββα-sandwich folds; it was observed, that ArchiP adds additional elements into architecture in a number of cases, which actually exist in structure according to expert judgment.

Except ArchiP mistakes, in certain cases misannotations are due to the absence of clear architecture information in SCOP for particular domains and families.

Conclusion: The results of ArchiP demonstrate the possibility of more rigorous approach for protein domain fold description and are of practical usage for protein domain architectures and fold classification.

Availability: Web-service of ArchiP (beta-version) is available at http://mouse.belozersky.msu.ru/archip.

Acknowledgements: The work is partially supported by RFBR grants 10-07-00685-a and 11-04-91340.

DEEP METAGENOMICS AND METAPROTEOMICS OF HUMAN GUT: DRAMAS AND DELIGHTS

Alexeev D.G.*, Tyakht A.V., Popenko A.S., Belenikin M.S., Altukhov I.A., Pavlenko A.V., Kostryukova E.S., Selezneva O.V., Larin A.K., Karpova I.Y., Govorun V.M.
Research Institute of Physico-Chemical Medicine FMBA, Moscow, Russia;
Moscow Institute of Physics and Technology, Dolgoprudny, Moscow Region, Russia;
Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, Moscow, Russia;
Institute of Bioorganic Chemistry of the RAS, Moscow, Russia;
RRC Kurchatov Institute, Moscow, Russia
e-mail: alexeev@niifhm.ru
** Corresponding author*

Key words: *metagenomics, metaproteomics, human gut microbiota*

Motivation and Aim: Gut microbiota is an essential component of healthy human existence, influencing metabolism, immunity and other facets of organism homeostasis. Composed of multiple microbial species, its metagenomic composition allows to assess the total metabolic potential of such complex ecological system. The study targets examination of 132 metagenomic samples of gut microbiota across a wide range of Russian metropolitan and rural areas using next-generation whole-genome sequencing.

Methods and Algorithms: Phylogenetic and functional profiling of metagenomic samples is performed basing on coverage depth resulting from alignment of reads to catalog of reference sequences, as well as statistical analysis and visualization. The reference sets contain prevalent microbial genomes and genes of human gut microbiota. Samples were compared across various socio-geographic, age- and health-related groups by means of statistical analysis and visualization using R language.

Results: Comparative analysis of Russian samples together with existing large metagenomic data sets from MetaHIT and Human Microbiome Project revealed both significant similarities, as well as novel distinctions across continents on a world scale.

SYSTEM BIOLOGY ANALYSIS OF *HELICOBACTER PYLORI* VIRULENCE AND ADAPTATION BASED ON PROTEOGENOMIC, TRANSCIPTOMIC AND METABOLOMIC ANALYSIS

Alexeev D.G., Momynaliev K.T., Selezneva O.V., Demina I.A., Pobeguc O., Tvardovsky A., Altukhov I.A., Govorun V.M.
Research Institute for Physico-Chemical Medicine
e-mail: alexeev@niifhm.ru

Key words: *System biology, proteomics, bacteria, metabolomics*

Motivation and Aim: Helicobacter pylori is a Gram-negative, microaerophilic, helical-shaped bacterium that colonizes the human stomach of at least half of the world's population.*H.pylori* preferentially colonizes the antrum of the stomach, where acid-producing parietal cells are not present, and the environmental pH is higher than in the corpus. In most cases, *H. pylori*can persist in the human stomach asymptomatically, but in some cases, *H. pylori* may progress to symptomatic chronic gastritis, gastric or duodenal ulcers, or gastric cancer. H. pylori is an extra macro- and microdiverse bacterial species. The high level of macrodiversity is supported by the fact that up to 25% of H. pylori genes are dispensable in at least one strain. The most unusual characteristic of the nucleotide sequence diversity of H. pylori is very high number of unique sequences for a given gene across the different strains.

Methods and Algorithms: 3 laboratory strains and one clinical isolate were subject to proteogenomic profiling, with following analysis of transcription and translation by means of PCR and high throughput HPLC-MS/MS analysis. Additionally metabolic potential was measured by metabolite MS profiling.

Results: Clinical isolate was sequenced and compared to laboratory strains. Each strain possesses over 10000 SNP. Over 600 proteins were identified in each strain, however over 100 from each group were uniquely expressed in one of the strains while similar genes silent in the others. Some transcripts were found for the present proteins while for others there were no transcription. System model explaining genetic basis for variability in functional states was built.

Conclusion and Availability: Variable expression is controlled by multiple factors – such as genome SNPs, asRNA, methylation and others. Therefore it allows for high adaptability and diversity. Some insights into functional state associated with high virulence were made.

THEORETICAL STUDY OF STRUCTURAL FEATURES OF VARIOLA VIRUS CrmB PROTEIN

Antonets D.V.*, Nepomnyashchikh T.S., Shchelkunov S.N.
State Research Center of Virology and Biotechnology "Vector", Koltsovo, Russia
e-mail: antonec@yandex.ru
** Corresponding author*

Key words: *tumor necrosis factor, orthopoxvirus, viral immunomodulatory proteins, CrmB, molecular modeling*

Motivation and Aim: Orthopoxviral TNF-binding proteins and especially variola virus (VARV) CrmB may be used to develop novel medications for treatment of rheumatoid arthritis, Chron's desease and other pathologies driven by TNF overproduction. The aim of this study was the theoretical analysis of molecular mechanisms underlying interaction of orthopoxviral TNF-binding CrmB proteins with their ligans.

Methods and Algorithms: Models of TNF receptor domains of VARV- and CPXV-CrmB and their complexes with different TNFs were constructed using Modeller (9v2) software (http://salilab.org/modeller). All constructed models were then energy minimized using either NOC (http://noch.sourceforge.net) or FoldX (http://foldx.crg.es). Stability of ligand-receptor complexes was predicted either with FoldX or using residue-level pairwise potentials BETM990101. FoldX was used for designing mutant forms of VARV CrmB. Spatial structure of VARV-CrmB C-terminal domain was predicted with I-TASSER (http://zhanglab.ccmb.med.umich.edu/I-TASSER/).

Results: Analysis of produced ligand-receptor models with either FoldX or with BETM990101 pair potentials revealed that mTNF should bind to CPXV-CrmB with higher affinity than hTNF. VARV-CrmB was predicted to bind both cytokines with higher affinity than CPXV-CrmB; CPXV-CrmB was predicted to bind hTNF(R31Q) with significantly higher affinity than wild type hTNF. And both CrmBs were predicted to less efficiently bind to hTNF(E127Q), than to the wild type hTNF. All these findings were then qualitatively approved by experimental evaluation of VARV- and CPXV-CrmB proteins ability to inhibit cytotoxic action of mTNF, hTNF, hTNF(R31Q) and hTNF(E127Q) on L929 murine fibroblast cells. These models were then used for designing mutant forms of VARV-CrmB which should have higher affinity to hTNF. Several mutant forms of CrmB, predicted with FoldX to be the most affine to hTNF, were chosen for further theoretical analysis and experimental evaluation.

Using the I-TASSER the spatial structure of VARV CrmB C-terminal chemokine-binding domain (SECRET) was predicted and it was assumed to be the structural homologue of CPXV vCCI protein belonging to the family of poxviral type II chemokine-binding proteins despite weak homology of their amino acid sequences (12 %). We suggested that SECRET should be included into the family of poxviral type II chemokine-binding proteins and that it might have been evolved from the vCCI-like predecessor protein. Recently our predictions were confirmed by the X-ray structure of Ectromelia virus CrmD protein SECRET domain.

Acknowledgements: This work was supported by Russian Foundation for Basic Research grants #10-04-00479-a and #12-04-00110-a.

MOLECULAR MODELING OF CYTOSOLIC PART OF a2-SUBUNIT OF MOUSE V-ATPase

Bakulina A.*[1], Merkulova M.[2], Hosokawa H.[2], Phat Vinh Dip[3], Gruüber G.[3], Marshansky V.[2]

[1] *State Research Center of Virology and Biotechnology "Vector", Koltsovo, Russia;*
[2] *Center for Systems Biology, Program in Membrane Biology, Simches Research Center, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA;*
[3] *Nanyang Technological University, Division of Structural and Computational Biology, School of Biological Sciences, Singapore*
e-mail: bakulina@gmail.com
** Corresponding author*

Key words: *V-ATPase, a2-subunit, cytohesin-2, ARNO*

Motivation and Aim: V-ATPase is a multiprotein proton pump. It interacts with variety of proteins, including Arf-exchange factor cytohesin-2 (also known as ARNO). Recently it was found that cytosolic part of mouse a2 isoform of a subunit of V-ATPase and six a2-derived peptides interact with ARNO [1]. 3D model of N-terminal cytosolic part of a2-subunit (a2N) is required for understanding details of intermolecular interactions and their functional roles.

Methods and Algorithms: The model is based on the following experimental data: crystal structure of N-terminal part of a homolog of subunit a (PDB ID 3RRK), NMR structures of peptides derived from C-terminal region of a2N, cryo-EM map of *Thermius Thermophilus* V-ATPase [2]. MODELLER and I-TASSER software packages were used for building several protein models, which were fitted into the cryo-EM map with Chimera program, and finally a MODELLER model was selected for the best fit.

Results: The a2N model consists of three domains: the distal lobe (DL), the proximal lobe (PL) and the central bar domain (BD). DL and PL have α/β structures, BD contains two anti-parallel rows of α-helices and links DL and PL. Mapping ARNO interacting peptides shows that there are two distinct sites on a2N, one is formed by two β-strands in DL domain and the loop between them, the other is formed by three β-strands, the adjacent loops and an α-helix in PL domain. We also mapped residues which were determined as contacting with G/E-subunits of V-ATPase by cross-linking experiments with Vph1p, an yeast homolog of a-subunit [3]. These regions occur in close proximity to ARNO interaction sites.

Conclusion: We built the structural model of cytosolic part of mouse a2-subunit of V-ATPase. We identified two ARNO binding sites on distinct domains of a2N. These sites are close, but not overlapping with G/E-subunits binding sites. We hypothesize that binding of ARNO to a2N may modulate its interaction with G/E-subunits of V-ATPase, and could promote disassembly of V-ATPase complex, and, thus, regulate its function.

References:

1. Merkulova, M., Bakulina, A., Thaker, Y. R., Gruber, G., and Marshansky, V. (2010) Specific motifs of the V-ATPase a2-subunit isoform interact with catalytic and regulatory domains of ARNO, *Biochim Biophys Acta,* **1797:** 1398-1409.
2. Lau, W. C., and Rubinstein, J. L. (2012) Subnanometre-resolution structure of the intact Thermus thermophilus H+-driven ATP synthase, *Nature,* **481:** 214-218.
3. Qi, J., and Forgac, M. (2008) Function and subunit interactions of the N-terminal domain of subunit a (Vph1p) of the yeast V-ATPase, *J Biol Chem,* **283:** 19274-19282.

HIGH-THROUGHPUT SCREENING FOR THE DEVELOPMENT OF NOVEL SELECTIVE LIGANDS OF D2 DOPAMINE RECEPTORS

Barnaeva E.*[1], Free R.B.[2], Hu X.[1], Southall N.[1], Bryant-Genevier M.[1], Titus S.[1], Ferrer M.[1], Marugan J.[1], Sibley D.R.[2]

[1] National Center for Advancing Translational Sciences;

[2] National Institute of Neurological Disorders and Stroke, National Institutes of Health, 9800 Medical Center Drive, Rockville, Maryland 20850, United States

e-mail: barnaevaes@mail.nih.gov

** Corresponding author*

Key words: *High-throughput screening (HTS), D2 dopamine receptor, functional selectivity*

The G-protein couple receptor (GPCR) D_2 dopamine receptor (D2 DAR) is an important therapeutic target for the treatment of a number of neuropsychiatric disorders, including Parkinson's disease. As most drugs targeting the D2 DAR are non-selective, there is great interest in the development of novel selective ligands of D2 DAR to further clarify the pharmacological role of this receptor for the treatment of different diseases. A novel approach for attaining greater selectivity of drugs targeting GPCRs is to identify small molecule ligands that exhibit "functional-selectivity". The phenomenon of "functional selectivity" can occur when activation of a GPCR transduces signals through different intracellular pathways, such as the traditional G-protein and second messengers (cAMP and Ca^{2+}), or β-arrestin and pERK. It has been shown that for some GPCRs efficacy and toxicity effects of ligands might be driven by activation of different signaling pathways, and compounds that are functionally selective might provide a better therapeutic window in clinic.

We recently have completed high throughput-screening (HTS) of 400,000+ small molecules in a variety of D2 DAR assays to identify functionally selective agonists and antagonists of this receptor. One assay measured D2 DAR signaling through Ca^{2+} by using a fluorescent calcium flux assay; the second assay measured signaling through β-arrestin by using a D2 DAR β-arrestin Enzyme Complementation cell line and PathHunter Assay from DiscoveRx. Both assays were miniaturized to 1536-well format and shown to be robust for automated HTS.

Data will be presented demonstrating the quality of screens implemented and analysis of the results focusing on selectivity between the two assays.

APPLICATION OF CONFORMATIONAL PEPTIDES FOR ANALYSIS OF ALLERGENIC PROTEINS

Bragin A.O.*, Demenkov P.S., Ivanisenko V.A.
Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia
e-mail: ibragim@bionet.nsc.ru
** Corresponding author*

Key words: *allergy, allergen, conformational peptides*

Motivation and Aim: Nowadays more than one third of the world population is affected from various allergic diseases. One of the most effective ways to prevent development of allergy is the elimination therapy. For this reason, development of methods for assessing allergenic properties of proteins is a major challenge. Currently existing *in silico methods of assessment* for protein allergenicity use only information about the amino acid sequences of proteins. Information about three-dimensional structure of proteins by such methods usually is not taken into account.

Methods and Algorithms: We have developed a method for representing a surface of protein molecule as a set of short amino acid sequences. These short sequences were called conformational peptides. Conformational peptides were calculated according to the following rules: 1) Two amino acids were accepted as bound in a conformational peptide, if the distance between their C-alpha atoms in protein 3D structure was not greater than 5E. 2) The average residue solvent accessibility for amino acids for a conformational peptide was not less than 50%. 3) The length of the conformational peptides and the linear peptides were the same (8 amino acid residues).

Allergenicity was predicted through search of such peptides from allergenic proteins in query protein.

Results, Conclusion and Availability: We have developed a method for allergenicity prediction, which uses information about the conformational peptides. This method is able to predict the allergenicity using only information about the primary sequence of a query protein. The allergenicity prediction method was integrated into the Protein Structure Discovery System (http://www-bionet.sscc.ru/psd/cgi-bin/programs/Allergen/allergen.cgi). A database of conformational peptides, calculated for three-dimensional structures of allergenic proteins, was created. It was shown that improvement of the allergenicity prediction method can be achieved by using conformational peptides.

Acknowledgements: work is supported by Russian Ministry of Education and Science, contract No. 07.514.11.4003.

COMPREHENSIVE ANALYSIS OF UNIDENTIFIED LC-MS FEATURES FOR INVESTIGATING PROTEINS DIVERSITY IN HIGH-THROUGHPUT PROTEOMICS EXPERIMENTS

Chernobrovkin A.L.*, Zgoda V.G., Lisitsa A.V., Archakov A.I.
Institute of Biomedical Chemistry RAMS, Moscow, Russia
e-mail: chernobrovkin@gmail.com
** Corresponding author*

Key words: *single amino-acid polymorphisms; lc-ms; proteins identification*

Motivation and Aim: More than 65 thousands nsSNP are known to exist in human genome, and more than 20% of them associated with different diseases. However, the vast majority of annotated nsSNP have not been observed at protein level yet. Investigation of diseases-related nsSNP at protein level can shed light on the molecular nature of diseases and provide additional information for molecular biomarkers discovering.

Methods and Algorithms: According to recent estimation only a small proteomes can be analyzed properly using high-accuracy LC-MS without using MS/MS for peptide identification [1]. Within the human proteome only 20% peptides can be properly identified using only accurate parent mass and retention time data. Here we propose the new strategy for unidentified LC-MS features analysis, which allows significantly increase the sequence coverage of proteins, identified using MS/MS data and reveal protein variants caused by translation of non-synonymous nucleotide polymorphisms. The method uses accurate m/z and retention time data analysis for assigning theoretical peptides of identified using MS/MS proteins to the unidentified LC-MS features. As an additional resource for removing the ambiguity in features annotating we use quantitative data of protein abundance changes during cells differentiation.

Results: There were 1370 proteins identified in HL60 cells using LC-MS/MS (LTQ Orbitrap Velos, Thermo Scientific) analysis of triptically digested cell lysates. Quantitative analysis was performed using Progenesis-LC-MS software and allows us to reveal 300 proteins that have changed their abundance more than 3 times during cells differentiation process. LC-MS chromatograms were reanalyzed to select those features that could be matched to the triptic peptides of selected proteins and their variants. Such procedure allows two to three fold increase in the sequence coverage of selected proteins. Additionally we observed 38 features that match 17 SAP-specific proteotypic peptides of identified proteins.

Conclusion: Proposed approach makes it possible to decrease number of unsigned features in LC-MS based proteomics experiments. Assigning of additional features to previously identified proteins allows increasing protein sequence coverage and revealing variant-specific proteotypic peptides.

References:

1. P. Bochet et al. (2010) Fragmentation-free LC-MS can identify hundreds of proteins, Proteomics, 11(1): 22-32.

MODELING OF PLANT KINESIN-8 MOTOR DOMAIN AND RECONSTRUCTION OF ITS L2 AND L11 LOOPS

Demchuk O.M.*, Karpov P.A., Blume Ya.B.
Institute of Food Biotechnology and Genomics NAS of Ukraine, Kyiv
e-mail: demom79@gmail.com, demom@univ.kiev.ua
** Corresponding author*

Key words: *Kinesin-8, modeling, bioinformatics*

Motivation and Aim: Among various classes of kinesines the kinesines-8 are minor class associated with microtubules [1]. The kinesines-8 play an important role in regulation of MT length [2]. In case of disruption of kinesin-8 mitotic functions leads to formation of abnormally long microtubules. Therefore it causes a disturbance during the divergence of homologous chromosomes [3]. The three-dimensional structure modeling of plant kinesin-8 motor domain is an important step in research of influence of tubulins posttranslational modifications on the interaction of mentioned proteins *in silico.*

Methods and Algorithms: Kinesin-8 motor domain three-dimensional models were constructed by I-TASSER (IT) and Swiss-Model Workspace (SMW) servers. Superimposition of kinesin structures and homology modeling of L11 loop using as a template a similar loop of chain A from crystal 2P4N were carried out using SwissPDBViewer 4.0.1. 3D-models were optimized with amber3 ff using conjugate gradient algorithm. Models were validated by MolProbity server.

Results and Discussion: Using Swiss-Model Workspace we generated kinesin-8 model from the primary sequence of *Arabidopsis thaliana* F25I16.11 by application chain B from crystal 3LRE (2.2 E) as matrix. I-TASSER generated also 5 models for the same sequence. MolProbity test of optimized models help to identify the best SMW-structure in comparison with IT-models. All constructed 3-D structures were superimposed with each other and with kinesines from X-Ray PDB-structures 3LRE (chain B) and 2P4N (chain A). It revealed some faults in the models generated *in silico*. The most problematic areas were L2 and L11 loops, which were also missed in X-Ray structure of human kinesin-8 (3LRE). However, L11 loop is available in the structure of kinesin in crystal 2P4N. We have constructed by homology a region S243-L267 from kinesin-8 *A. thaliana* (corresponds L11 loop) using as matrix the tertiary structure of S235-S259 region from 2P4N. In order to replace the coordinates generated by the server, the next step was to insert coordinates of constructed region S243-L267 into the SMW-structure. The area L36-V53 corresponding to the kinesin-8 L2 loop had a two β-strands. This area was replaced by a similar coiled region from one of IT-models. This region had the best results on «MolProbity» server.

Conclusion: The reconstructed model of kinesin-8 from *A. thaliana* has been optimized and demonstrated good score values on «MolProbity» server. These data suggest an acceptability of given model for subsequent bioinformatics studies

References:

1. A. Unsworth, H. Masuda, S. Dhut, T. Toda (2008) *Mol. Biol. Cell*, 19: 5104–5115.
2. G. Goshima, R.D. Vale (2006) *Nat. Cell Biol*., 8: 913–923.
3. M.M. Wargacki, J.C. Tay, E.G. Muller et al. (2010) *Cell Cycle*, 12: 2581–2588.

PROTEOMIC OF MYCOPLASMAS: NANOFORMING *MYCOPLASMA GALLISEPTICUM*

Demina I.A.*, Serebryakova M.V., Ladygina V.G., Rogova M.A., Kondratov I.G., Renteeva A.N., Govorun V.M.
Research Institute for Physical_Chemical Medicine of Ministry of Public Health of Russian Federation, Moscow, Russia
e-mail: idemina@mail.ru
** Corresponding author*

Key words: *Mycoplasma gallisepticum*, *adaptation, starvation, reversion, proteomics, 2D electrophoresis, mass spectrometry, real-time PCR*

The goal of this work was the creation of a model for the long persistence of *Mycoplasma gallisepticum* in depleted medium and under low growth temperature followed by proteomic study of the model. Nanoforms and revertants for *M gallisepticum* were obtained. Proteomic maps were produced for different stages of the formation of nanoforms and revertants. It is shown that proteins responsible for essential cellular processes of glycolysis, translation elonga”

tion, and DnaK chaperone involved in the stabilization of newly synthesized proteins are crucial for the reversion of *M. gallisepticum* to a vegetative form. Based on the current, data it is assumed that changes in the metabolism of *M. gallisepticum* during nanoforming are not post”mortal, thus *M. gallisepticum* does not transform into an uncultivable form but remains in a reversible dormant state during prolonged unfavorable conditions.

PROTEIN-PROTEIN AND PROTEIN-MEMBRANE RECOGNITION: A COMPUTATIONAL VIEW

Efremov R.G.[1], Polyansky A.A.[1], Chugunov A.O.[1], Nolde D.E.[1], Pentkovsky V.M.[2]
[1] *Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry RAS, Moscow, Russia;*
[2] *Moscow Institute of Physics and Technology (State University), Dolgoprudny, Moscow region, Russia*
e-mail: efremov@nmr.ru
** Corresponding author*

Motivation and aim. Cell membranes attract a growing attention as very perspective pharmacological targets. Rational design of new efficient and selective compounds modulating activity of biomembranes, requires atomic-scale information on their spatial structure and dynamics under different conditions. Because such details resist easy experimental characterization, important insight can be gained *via* computer simulations.

Methods and Algorithms. The work describes the results of computer simulations of structural/dynamic properties of membrane proteins and peptides with diverse fold, mode of membrane binding, and biological activities. Among the objects under study are: antimicrobial and cell-penetrating peptides, cardiotoxins from snake venom, transmembrane domains of receptor tyrosine kinases. The computational approach combines Monte Carlo simulations in implicit membranes, molecular dynamics in full-atom lipid bilayers, and molecular hydrophobicity potential analysis.

Results. Regardless different structure, dynamic behaviour, and mechanism of membrane permeation, in all cases the polypeptide-membrane recognition reveals a prominent "self-adapting" character. Namely, the membrane active agents employ a wide arsenal of structural/dynamic tools in order to insert into the lipid bilayer and to accomplish their function. Importantly, lipid bilayer of biological membranes plays an essential role in the recognition and binding events. In particular, the membrane surface reveals highly dynamic lateral heterogeneities (clusters), which differ in their packing and hydrophobic properties from the bulk lipids. Such a mosaic nature of membranes is tuned in a wide range by the chemical nature and relative content of lipids, presence of ions, etc.

Conclusion. Polepeptide-bilayer interactions represent a fine-tuned process, which requires the two active players – the polypeptide and the membrane. Interplay of the factors determining such a process assures efficient and robust binding of peptides and proteins to cell membranes. Understanding of such effects creates a basis for rational design of new physiologically active molecules and/or artificial membranes with predefined properties.

Acknowledgements: This work was supported by the Ministry of Education and Science of the Russian Federation, by the Russian Foundation for Basic Research, and by the RAS Programme "Basic fundamental research for nanotechnologies and nanomaterials". Access to computational facilities of the Joint Supercomputer Center RAS (Moscow) and Moscow Institute of Physics and Technology is gratefully acknowledged.

References:

1. Polyansky A.A. et al., (2009). ***J. Phys. Chem. B.*** 113, 1107-1119.
2. Pyrkova D.V. et al. (2011). ***Soft Matter*** 7, 2569-2579.
3. Konshina A.G. et al. (2011). ***PLoS ONE*** 6, e19064.

BASE EXCISION REPAIR OF TRIPLET REPEAT SEQUENCES ASSOCIATED WITH NEURODEGENERATIVE DISORDERS

Endutkin A.V. *, Derevyanko A.G., Zharkov D.O.
Institute of Chemical Biology and Fundamental Medicine, SB RAS, Novosibirsk, Russia
e-mail: Aend@niboch.nsc.ru
** Corresponding author*

Key words: *8-oxoguanine, DNA glycosylases, base excision repair, trinucleotide repeats, dynamic mutations, neurodegenerative disease*

Motivation and Aim: Trinucleotide repeat expansion provides a molecular basis of several neurodegenerative diseases. One of the main reasons of triplet repeat expansion in somatic cells is base excision repair (BER), involving damaged base excision and DNA repair synthesis that may be accompanied by expansion of the repaired strand due to DNA looping. Expansion of CAG triplets characteristic of Huntington disease is initiated in the course of normal repair of the damaged base 8-oxoguanine (oxoG). Yet it is unclear how its repair might cause the expansion and whether the repair of other DNA lesions can lead to this expansion. Little is known about the efficiency of BER enzymes and their specificity when the DNA substrate contains trinucleotide repeats.

Methods and Algorithms: Using a number of CAG-substrates with the damaged triplet in different positions we have determined the rate constants of oxoG excision (k_2) and product release (k_3) by human oxoG-DNA glycosylase (OGG1) under conditions of single turnover kinetics and burst phase kinetics, respectively. Using fractionated cell extracts we have obtained the repair initiation rates for the same substrates. In the case of CAG-run substrates containing uracil in different position we have obtained the values of Michaelis constant and catalytic constants for the reaction of uracil excision by human uracil-DNA glycosylase (UNG). To determine the ability of the minimal BER enzyme set to repair oxoG and uracil in CAG-runs, we reconstituted the BER process *in vitro* with human DNA glycosylase OGG1 or UNG, AP endonuclease APEX1, and DNA polymerase β.

Results: We have analyzed kinetics of excision of ubiquitous oxidized bases, oxoG and uracil, by OGG1 and UNG from the substrates containing a CAG run. The values of k_2 rate constant for the removal of oxoG from triplets in the middle of the run were higher than for oxoG at the flanks of the run. The value of k_3 rate constant dropped starting from the third CAG-triplet in the run and remained stable until the 3'-terminal triplet in the run where it decreased even more. In nuclear extracts, the profile of oxoG removal rate along the run resembled the profile of k_2 constant, suggesting that the reaction rate in the extracts is limited by base excision. For uracil containing substrates, a strong dependence of both constants (k_{cat}, K_M) on the position of the damaged triplet was also observed.

Conclusion: The efficiency of initiation of repair of oxoG located in trinucleotide repeat runs depends on position of the damaged base in the run, namely, it is lower on its 5'- and 3'-borders. This dependence concerns both the rate constant of oxoG excision by pure DNA glycosylase OGG1 and the rate of cleavage of damaged DNA in cell extracts and efficacy of initial repair stages in the reconstituted BER system.

Acknowledgements: This work was supported by RFBR (10-04-91058-PICS_a) and Russian Ministry of Education and Science #14.740.11.1195.

Gp39, A NOVEL PHAGE-ENCODED INHIBITOR OF BACTERIAL RNA POLYMERASE

Esyunina D.M.*[1, 2], Miropolskaya N.A.[1], Minakhin L.S.[3], Kulbachinskiy A.V.[1]
[1] *Institute of Molecular Genetics, Russian Academy of Sciences, Moscow, Russia;*
[2] *Molecular Biology Department, Biological Faculty, Moscow State University, Moscow, Russia;*
[3] *Waksman Institute of Microbiology, Piscataway, New Jersey 08854, USA*
e-mail: es_dar@inbox.ru
** Corresponding author*

Key words: *RNA polymerase, transcription initiation, open promoter complex, bacteriophage, transcription inhibitor*

Motivation and Aim: Bacterial RNA polymerase (RNAP) is an important target for antibacterial therapy. Bacteriophages often encode regulatory proteins that can bind and inhibit bacterial RNAP during phage infection. Such proteins and their derivatives can therefore be used for development of novel transcription inhibitors. Recently, we isolated a novel RNAP-binding protein, gp39, encoded by phage P23-45 that infects thermophilic bacterium *Thermus thermophilus*. This small 16 kDa protein was shown to inhibit transcription inihiation by RNAP. The purpose of this study was to characterize the mechanism(s) of transcription inhibition by gp39.

Methods and Algorithms: The effects of gp39 on different steps of transcription were characterized using *in vitro* transcription approaches with highly purified RNAP preparations from several bacteria. To localize the gp39-binding determinants, mutant gp39 and RNAP variants were obtained by site-directed mutagenesis followed by their expression, purification and functional analysis.

Results: Gp39 was found to affect different steps of transcription by bacterial RNAP. During initiation, gp39 specifically inhibits recognition of the -10/-35 class of bacterial promoters with high efficiency. During elongation, gp39 stimulates RNA synthesis and suppresses transcription termination. Gp39 was shown to target the flap domain of RNAP which is involved in both promoter recognition and RNA binding during transcription elongation. The observed effects of gp39 on initiation and elongation were shown to be independent of each other. Thus, gp39 is a bi-functional protein that exerts both inhibiting and activating effects on transcription. Analysis of the mechanism of inhibition of transcription initiation by gp39 revealed that it prevents the open complex formation, likely by disrupting recognition of the -35 element by RNAP. We propose that the interaction between gp39 and the RNAP flap domain can be used for development of fluorescence-based assays for selection of transcription inhibitors targeting the flap domain of RNAP.

Conclusion: Gp39 is a potent RNAP inhibitor that disrupts its interactions with promoters during transcription initiation. Gp39 can be used for design of small oligopeptide inhibitors of transcription and for development of highly sensitive assays for screening of novel antibacterial compounds targeting bacterial RNAP.

Acknowledgements: This work was supported by Federal Targeted Program "Scientific and scientific-pedagogical personnel of innovative Russia 2009-2013" (state contract 02.740.11.0771), by the Russian Foundation for Basic Research grant 10-04-00925 and by the grant of the President of Russian Federation MK-7156.2012.4.

MOSS PROTEOMICS AND PEPTIDOMICS. NEW INSIGHT IN THE OLD STORY. PEPTIDES IN THE STRESS ADAPTATION PROCESS

Fesenko I.A.*[1], Slizhikova D.K.[1], Seredina A.V.[1], Mageyka I.S.[1], Govorun V.M. [1, 2]
[1] *Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry RAS, Moscow, Russia;*
[2] *Institute of Physico-Chemical Medicine, Moscow, Russia*
e-mail: fesigor@gmail.com
** Corresponding author*

Key words: *Physcomitrella patens, proteome, peptidome, stress*

Motivation and Aim: The moss *Physcomitrella patens* is a new model system in plant science, that offers several advantages for studying of plant physiology, biochemistry and genetics. Phylogenetically, *P. patens* is situated in a key position between the green algae and the seed plants. The ancestors of mosses and seed plants separated shortly after the transition from water to land at least 500 million years ago. In addition, *P. patens* is the only land plant which has an exceptionally high rate of homologous recombination. We utilized the *P. patens* protoplast as a model to explore the mechanisms involved in stress adaptation in plant.

Methods and Algorithms: We used proteomic analysis to measure changes in the protein composition of freshly isolated protoplasts from the moss protonema. For this purpose, we compared results of 2D electrophoregrams of proteins from protoplasts and protonema using specific fluorescent dyes (DIGE) for identification of proteins specific to different living forms of *P. patens*. Besides, using a combination of high performance mass spectrometry with a bioinformatic analysis we described peptidome of the moss protoplast and green tissue.

Results: The DIGE demonstrate difference in the protein compositions of protoplasts and protonema. Thus, we have detected protein spots on two-dimensional electrophoregrams that were identified by mass spectrometry as Rubisco fragments in MW range from 10 to 20 kDa. The analyses of peptidome showed that the amount of peptides identified in protoplasts is almost six times greater than in the protonemata from which they are isolated and five times greater than in gametophores.

Conclusion: The isolation of moss protoplasts is accompanied by the degradation of proteins most of which are the proteins that belongs to the system of photosynthesis. Those processes of protein degradation lead to the generation of endogenous peptides, which is peculiar to stress responses of higher plants.

"GOLDEN TRIANGLE" FOR FOLDING RATES OF GLOBULAR PROTEINS

Garbuzynskiy S.O.*[1], Ivankov D.N.[1, 2], Bogatyreva N.S.[1], Finkelstein A.V. [1]
[1] *Institute of Protein Research, Russian Academy of Sciences, Pushchino, Moscow Region, Russia;*
[2] *Department of Genome Oriented Bioinformatics, Technische Universitaet Muenchen, Wissenschaftszentrum Weihenstephan, 85354 Freising, Germany*
e-mail: sergey@phys.protres.ru
** Corresponding author*

Key words: *protein folding rate; folding time; protein size; protein stability*

Motivation and Aim: An ability of protein chains to form their spatial structures spontaneously is a long-standing puzzle of molecular biology. Experimentally measured rates of spontaneous folding of single-domain globular proteins range from microseconds to hours: the difference (11 orders of magnitude!) is like that between the life spans of a mosquito and the Universe.

Results: We show that physical theory with biological constraints outlines a "golden triangle" limiting the range of folding rates possible for single-domain globular proteins of various size and stability, and that the experimentally measured folding rates fall within this narrow triangle (built without any adjustable parameters), filling it almost completely. In addition, the "golden triangle" predicts the maximal size of protein domains that fold under solely thermodynamic (rather than kinetic) control (this size is about 90 amino acid residues). It predicts also the maximal allowed size of the "foldable" protein domains (it is predicted to be 500 amino acid residues for spherical globules and 600 amino acid residues for oblong or oblate globules); the size of domains found in known protein structures is in a perfect concordance with this limit.

Acknowledgements: This work was supported by Molecular and Cell Biology (#01200957492) program, grants from Howard Hughes Medical Institute (#55005607), RFBR (#10-04-00162a), FASI (#02.740.11.0295), by grant of the Dynasty Foundation and the Russian Young Scientists' grants (#MK-4894.2009.4, #MK-5540.2011.4).

INTEGRATION OF – OMICS

Govorun V.M.
Research Institute of Physico – Chemical Medicine, Moscow, Russia

Recently rapid development allowed for explosion in amount of data and sensitivity of detection methods in Genomics, Transcriptomics, Proteomics and Metabolomics. However while the interdependence of –Omics in vivo is of no doubt, integration of experimental data mostly results in unpreceded complexity – whereas analysis and further predictions seem almost impossible.

Promising is the approach of Omics integration in bacterial studies, ta decade ago bacterial cells seemed to be tremendously simple compared to eukaryotic ones. The range of studies aimed at bacterial analysis however demonstrated comparable complexity in Omics result in bacteria, where the difference is only in sizes of genomes and numbers of variable proteins, RNAs and metabolites.

We have developed several experimental and bioinformatics pipelines allowing for data integration in experimental results. We demonstrate the emergence effects appearing from integration of Omics for several different organisms: Mollicutes including spiroplasma, mycoplasma and acholeplasma, H.pylory. Further we demonstrate the scalability of the approaches developed for data integration in human microbiome analysis, and variety of eukaryotic organisms.

STRUCTURAL AND FUNCTIONAL PROTEOMICS OF THE HUMAN PROTEIN SYNTHESIZING SYSTEM

Graifer D.M.*[1], Bulygin K.N.[1], Khairulina Yu.S.[1], Sharifulin D.E.[1], Ven'yaminova A.G.[1], Frolova L.Yu.[2], Karpova G.G.[1]

[1] *Institute of Chemical Biology and Fundamental Medicine SB RAS, Novosibirsk, Russia;*
[2] *Engelhardt Institute of Molecular Biology RAS, Moscow, Russia*
e-mail: graifer@niboch.nsc.ru
** Corresponding author*

Key words: *eukaryotic ribosome; decoding site; mRNA analogues; photoaffinity cross-linking; eukaryotic ribosomal protein S15; stop codon recognition; eRF1*

Protein synthesis is one of the basic events of the cell life. It takes place on ribosomes, very complicated cellular ribonucleoprotein machineries translating genetic information incoming as mRNA, and involves a number of assistant proteins (translation factors). Specific interactions of proteins during translation process underlie the work of the ribosomal machinery, and knowledge of the structural basis of these interactions is of principal importance for life sciences. These interactions in prokaryotes are known at the atomic level due to X-ray crystallography. However, structural and functional proteomics of the human protein synthesizing system is much less studied since ribosomes from higher organisms were not yet crystallized for the X-ray analysis. Here, we applied a site-directed cross-linking approach to study fine structure of mRNA binding site of the human ribosome and of a site of stop codon recognition in translation termination factor eRF1. Using a set of labeled mRNA analogues bearing cross-linker at designed locations, we determined peptides of ribosomal proteins involved in the formation of the mRNA binding site, and peptides of eRF1 recognizing mRNA stop codons. We found that eukaryote-specific peptide of ribosomal protein S26e is a key player in accommodation of mRNA region 5' of the codons interacting with tRNAs suggesting that it is involved also in the interaction with eukaryote-specific initiation factor eIF3 [1]. Eukaryote/archaea-specific decapeptide of ribosomal protein S15e was revealed at the ribosomal decoding site and most probably interacts with eRF1 [2]. We clarified the keystone aspect of protein synthesis termination related to the recognition of stop codon purines by eRF1. We found out that A and G are recognized by different N domain conformations of eRF1, which provides its ability to recognize all three stop codons, and discovered that the universally conserved dipeptide 31-GT-32 is the key player in this process [3,4]. The data obtained provide new insights into molecular basis of mammalian protein synthesis and are of great importance for understanding the nature of pathologies related to disturbances of any step of this process.

This study was supported by RFBR grant # 11-04-00597 to G.K. and by the Russian Academy of Sciences Presidium program "Molecular and cell biology" (grant to G.K.).

References:

1. Sharifulin D. et al. (2012) Nucleic Acids Res. In press (doi: 10.1093/nar/GKR1212).
2. Khairulina J. et al. (2010) Biochimie 92:820-825.
3. Bulygin K.N. et al. (2010) RNA 16:1902–191.
4. Bulygin K.N. et al. (2011) Nucleic Acids Res. 39:7134–7146.

PROTEIN FOLDING TURBULENCE

Kalgin I.V.[1], Chekmarev S.F.*[1, 2]
[1] *Novosibirsk State University, Novosibirsk, Russia*
[2] *Kutateladze Institute of Thermophysics, SB RAS, Novosibirsk, Russia*
e-mail: chekmarev@itp.nsc.ru
** Corresponding author*

Key words: *molecular dynamics, protein folding flows, turbulent flows, eddies, self-similarity, structure functions, cascades of the structural transformations*

Motivation and Aim: Protein folding and hydrodynamic turbulence are two long-standing challenges, in molecular biophysics and fluid dynamics, respectively. The theories of these phenomena have been developed independently and used different formalisms [1,2]. However, as has recently been observed [3,4], folding flows of a protein can also be filled with vortices, similar to turbulent flows of a fluid. Here, we characterize the folding flows in terms accepted in hydrodynamic turbulence [5] and examine how far the similarity between the protein folding flows and turbulent flows of a fluid extends.

Methods and Algorithms: Using molecular dynamics methods, two model proteins are studied: a fyn SH3 domain (C-alpha model) and beta3s miniprotein (all-atom model). Folding fluxes are calculated in a reduced (3D) space of orthogonal collective variables, which is characterized either by the numbers of native contacts between protein sections (a SH3 domain) or hydrogen bonds distances (beta3s).

Results: We have found that at and below the glass transition (melting) temperature, the folding flows are surprisingly similar to turbulent flows of a liquid. The flows have fractal nature and are filled with 3D eddies. The eddies contain strange attractors, at which the tracer flow paths behave as saddle trajectories. Two regions of the space increment exist, in which the folding flux variations are self-similar with the space correlation (structure) functions being in close agreement with those in the Kolmogorov theory of turbulence [2]. In one region, the cascade of protein structural transformation is directed from larger to smaller scales (net folding), and in the other, it is oppositely directed (net unfolding).

Conclusion: The protein folding turbulence has many properties of hydrodynamic turbulence. The cascade mechanism of protein transformations plays a key role for this similarity, unifying this new phenomenon with the wave, superfluid and market turbulences.

References:

1. K.A. Dill, S.B. Ozkan, M.S. Shell, T.R. Weikl (2008) The protein folding problem, Annu. Rev. Biophys., 37: 289-316.
2. L.D. Landau, E.M. Lifshitz (1987) Fluid Mechanics (Pergamon, New York).
3. S.F.Chekmarev, A.Yu.Palyanov, M. Karplus (2008) Hydrodynamic Description of Protein Folding, Phys. Rev. Lett, 100: 018107.
4. I.V.Kalgin, M. Karplus, S.F.Chekmarev (2009) Folding of a SH3 Domain: Standard and "Hydrodynamic" Analyses, J. Phys. Chem. B, 113: 12759-12772.
5. I.V.Kalgin, S.F.Chekmarev (2011) Turbulent Phenomena in Protein Folding, Phys. Rev. E, 83: 011920.

THE ROLE OF CASEIN KINASES 1 IN PLANT CYTOSKELETON REGULATION

Karpov P.A.*, Raevsky A.V., Sheremet Ya.A., Blume Ya.B.
Institute of Food Biotechnology and Genomics NAS of Ukraine, Kyiv
e-mail: karpov.p.a@gmail.com, karpov_pavel@univ.kiev.ua
** Corresponding author*

Key words: *Casein kinases 1, D4476, docking, MD, isoforms, D4476, bioinformatics*

Motivation and Aim: Casein kinases 1 (CK1) are ubiquitously expressed in eukaryotic organisms and yeast. Animal CK1 isoforms (α, β, γ1-3, δ, and ε) from different organisms demonstrate a high level of conservativeness in catalytic domains. Their phosphate-binding site is a target for ATP-competitive inhibitors. Phosphate binding regions of animal CK1 showed their isoform complete identity. The functions and structures of plant CK1-like kinase isoforms in plant cell, and regulation of cytoskeleton are still unclear.

Methods and Algorithms: Plant homologs were identified based on blastp-scanning of the UniProt database. NJ-clustering was performed in MEGA5. 3D-models were optimized with amber3 ff and subjected MD in Amber99 ff. Molecular docking was performed in CCDC GOLD. Spatial structure analysis was performed in PyMol. Experiments were performed on GFP-labeled microtubules (MT) in cells of *Arabidopsis thaliana*

Results and Discussion: blastp-scanning of the UniProt database against catalytic domains of human and animal CK1δ revealed presence of 34 CK1 homologous in *A. thaliana*. Comparing of gene loci (based on *Tair* data) we confirmed the existence of only 18 CK1-like kinases, and uniqueness of their catalytic domains was confirmed cladistically. According to the BLAST protocol, the maximum "Score" (456) belongs to *A. thaliana* KC1D (UniProt: P42158, identity = 78 %, similarity = 92 %).

We have performed a reconstruction of the spatial structure of all catalytic domains from CK1s from *R. norvegicus* and its homologs from *A. thaliana*. After optimization in Amber3 force field and molecular dynamics (10 ns in Amber99 force field), the models were structurally superimposed. The high similarity between folding of CK1s from *R. norvegicus* and 13 (out of 18) CK1-like kinases from *A. thaliana* was confirmed. Recently, it was shown that D4476 inhibit CK1δ at 10 μM concentrations by more than 90 % and had almost no effect on the other protein kinases [1]. In our experiments on *A. thaliana* we have observed strong effect of D4476 on primary root growth and MT organization. In view of the similarity of spatial structures and amino acid compositions of catalytic domains, we specify casein kinases KC1D (At4g26100.1) and CKL2 (At1g72710.1) as the most likely targets of D4476. We have identified that plant homolog CKL6 contain 382-VSEKGRNTSRYG-394 motive in C-end region and has homology to mammalian MT-associated protein Eml4. These data confirm G. Ben-Nissan et al. experimental data [2]. It is known, that mammalian CK1δ binds and phosphorylates EB1 (MAP1) [3].

Conclusion: It seems like that in *A.thaliana* D4476-acts as inhibitor on KC1D, CKL6 and possibly CKL2 isoforms. Therefore the effects on MT organization in *A.thaliana* are likely associated with complex inhibition of mentioned CK1 isoforms. Thus, CK1s play important role in the plant cytoskeleton regulation.

References:

1. G. Rena, J. Bain, M. Elliot, P. Cohen (2004) *EMBO Rep*., 5(1): 60–65.
2. G. Ben-Nissan, W. Cui, D.-J. Kim, et al. (2003) *Plant Physiol.,* 148: 1897–1907.
3. D. Zyss, H. Ebrahimi, F. Gergely. (2011) *J. Cell Biol.*, 195(5): 781–797.

ALTERNATIVE HYDROGEN BONDING IN MOLECULAR DESIGN OF THERMOSTABLE ANTIOXIDANT PROTEIN

Kondratyev M.S.*, Kabanov A.V., Novoselov V.I., Samchenko A.A., Komarov V.M., Khechinashvili N.N.
Institute of Cell Biophysics RAS, Pushchino, Moscow region, Russian Federation
** Corresponding author: e-mail: ma-ko@bk.ru*

Key words: *protein structure, thermostability, molecular dynamics, peroxiredoxin VI, amino acid substitutions, homology*

Motivation and Aim: On of the goal of modern bioengineering is concerned with the development of molecules with predefined properties. Mainly it concerns the increase in stability of macromolecules, including thermostability of enzymes. That allows to speed-up of biocatalysis and protection the macromolecule from unfolding. Our work use newest theory of thermostabilization of small globular proteins [1, 2], developed in the our Laboratory of Structure and Dynamics of Biomolecular Systems ICB RAS. It is based on the alternative hydrogen bonding between side chains of amino acids on protein surface. Object of our investigations was human Peroxiredoxin VI (PRX6) – new promising antioxidant for burn treatment, discovered and described in our Institute [3, 4].

Methods and algorithms: homology analysis, molecular dynamics (GROMACS)

Results: analysis of homology, known spatial structure PRX6 and calorimetric data [4] give us information about localization and importance amino acid substitutions in human PRX6. Evolutionary changes of this protein lead us to search the most probable sites of mutations only in variable areas of amino acid sequence, because changes in stable regions can affect the functional properties of this antioxidant enzyme. It is important to notice that human and rat PRX6 have the highest homology (91,5 %, i.e. 19 residues) and rat protein possesses the greatest thermostability of studied PRX's. The structures of native human Peroxiredoxin 6 and it homologs has been studied by long full-atomic molecular dynamics simulation with solvent at various temperatures on GPU NVIDIA. We counted the amount of hydrogen bonds (salt bridges) on the surface of protein globules in each frame of an MD-trajectory and propose only 4 amino acid substitutions (V10E, N107D, I165D, D183K), which we believe will lead to increased thermostability of human PRX6 without distortion of the 3D-structure and antioxidant activity.

References:

1. Khechinashvili N.N., Volchkov S.A., Kabanov A.V., Barone G., Thermal stability of proteins does not correlate with the energy of intramolecular interactions. // Biochim. Biophys. Acta (BBA) Proteins & Proteomics 2008, 1784 (11), P.1830.
2. Khechinashvili N.N., Fedorov M.V., Kabanov A.V., Monti S., Ghio C., Soda K., Side Chain Dynamics and Alternative Hydrogen Bonding in the Mechanism of Protein Thermostabilization. // J. Biomol. Struct. & Dyn., 2006, 24(3), P.255-262.
3. Peshenko I.V., Novoselov V.I., Evdokimov V.A., Nikolaev Yu.V., Shuvaeva T.M., Lipkin V.M., Fesenko E.E. // Novel 28-kDa secretory protein from rat olfactory epithelium. FEBS Lett. 1996, 381, P.14-19.
4. Sharapov M.G., Novoselov V.I., Ravin V.K. Cloning, expression and comparative analysis of peroxiredoxine 6 from different species // Mol Biol (Mosk). 2009, May-Jun; 43(3). P.505-11.

MASS-SPECTROMETRY-BASED IDENTIFICATION OF ENDOGENOUS PEPTIDES IN BLOOD SERUM

Kovalchuk S.I.*, Ziganshin R.H., Arapidi G.P., Azarkin I.V., Govorun V.M., Ivanov V.T.
Institute of Bioorganic Chemistry RAS, Moscow, Russia
e-mail: xerx222@gmail.com
** Corresponding author*

Key words: *endogenous peptides, mass-spectrometry, serum*

Motivation and Aim: In 2011 the Human Proteome project was started, however most of the well-known and reliable protocols (both experimental and computational) are designed to study large protein molecules. The classic way to study proteome is, for example, to make 1D or 2D electrophoresis, to treat different parts of a gel containing moderate number of proteins with trypsin and to analyse the tryptic peptide fragments by LC-MS/MS. Another way is to use different immunochemical approaches. However both these approaches miss such a huge group of small proteins as peptides being the products of natural degradation of the host proteins while peptidome studies could give invaluable data concerning internal human body processes. Most of the present papers on human peptidomics work with body liquids such as saliva, cerebrospinal and synovial liquids, urea, and most of all with blood serum and plasma. And this is the place where the troubles begin: nonpeptide contaminants (proteins, hydrocarbons, low molecular weight impurities etc.) impeding normal chromatographic peptide separation, incredible native peptide diversity with very high concentration dynamic range… The only known way to study peptidomes consists of prefractionation of native peptide mixtures with subsequent analysis by high-mass-accurate LC-MS/MS with modern fast and precision mass-spectrometers such as Q-TOF, Orbitrap and TripleQuad machines.

Methods and Algorithms: Our group works on the human serum peptidome. As a first step we separate serum samples on magnetic beads with weak cationic exchange surface followed by several alternative separation methods: SAX-HPLC, RP-HPLC and IEF. The resultant samples are analysed by RP-LC-MS/MS using Q-TOF mass-spectrometer and the mass-lists are searched against SwissProt database.

Results: We also add one special step while making samples – we heat the eluates after magnetic beads separation at 95^{o}C for 15 min. This leads to peptide desorption from the major serum proteins particularly from albumin. This step utterly increases the number of analyzable peptides – based on the LC-MS analysis this number rises up to many thousand individual compounds. This quantity clearly demonstrates the necessity of the fast and sensitive MS equipment to analyse more peptides beyond the most abundant ones. For now we have gathered a database containing about 3000 thousand unique peptides which significantly surpasses all the published results.

Conclusion: However it is important to understand that exhaustive analysis of peptidome cannot be an end in itself. The serum peptide database we are creating contains not only peptides from healthy donors, but also from the people with different socially-significant diseases and can be used for potential peptide biomarker search. Besides, the knowledge about changes in degradome activity could be of great importance for studying molecular mechanisms of different human diseases developing.

DIRECT COMPUTER SIMULATION OF PROTEIN-PROTEIN INTERACTION

Kovalenko I.B.
Biological faculty of Moscow Lomonosov State University, Moscow, Russia
e-mail: ikovalenko78@gmail.com

Key words: *Brownian dynamics, diffusion, electrostatic interaction, protein*

Motivation and Aim: The goal of this work is to study the kinetics of protein-protein interactions between the electron transport proteins involved in photosynthesis by means of computer simulation.

Methods and Algorithms: We present a new method for computer simulation of formation of protein-protein complexes in a cell environment (1). The method makes it possible to simulate association reactions of several hundreds of protein pairs in sub-cellular compartments, and to obtain the real-time dynamics of protein-protein interactions. The method allows us to explore the effect of electrostatic forces on the protein-protein complex formation and evaluate the kinetic rate constants.

Results: Calculations correctly reproduce binding interactions between the electron transport proteins involved in photosynthesis for different values of ionic strength in the solution and in the chloroplast thylakoid lumen, while taking into account electrostatic interactions between proteins and the thylakoid membrane (2, 3). The model demonstrates non-monotonic dependences of complex formation rates on the ionic strength as the result of long-range electrostatic interactions (4). The developed method can also be used as a predictive tool to resolve the binding sites and to describe complex structures for a range of proteins (5).

Conclusion: The simulation method presented in this work serve to reveal the molecular interactions (diffusion, electrostatic interactions) underlying the arrangement of photosynthetic electron transport regulation.

Availability: Available on request from the authors.

Acknowledgements. The work is supported by grants from the Russian Foundation of Basic Research (11-04-01019 and 11-04-01268).

References:

1. I.B. Kovalenko et al. (2006) Direct simulation of plastocyanin and cytochrome f interactions in solution, *Phys. Biol.* **3:** 121-129.
2. I.B. Kovalenko et al. (2010) Direct computer simulation of ferredoxin and FNR complex formation in solution, *Phys. Biol.* **7:** 26001.
3. O.S. Knyazeva et al. (2010) Multiparticle computer simulation of plastocyanin diffusion and interaction with cytochrome f in the electrostatic field of the thylakoid membrane, *Biophysics* **55:** 221-227.
4. I.B. Kovalenko et al. (2011) Computer simulation of interaction of photosystem 1 with plastocyanin and ferredoxin, *BioSystems* **103:** 180-187.
5. I.B. Kovalenko et al. (2009) A novel approach to computer simulation of protein–protein complex formation, *Dokl. Biochem. Biophys.* **427:** 215–217.

CHANGES IN PROTEIN COMPOSITION OF HUMAN URINE AFTER PROLONGED ORBITAL FLIGHTS

Larina I.M.*[1], Nikolaev E.N.[2], Pastushkova L.H.[1], Valeeva O.A.[1],
Kononihin A.S.[2], Kireev K.S.[3], Tiys E.S.[4], Ivanisenko V.A.[4], Kolchanov N.A.[4]
[1] Institute of Medicobiologic Problems Russian Federation State Scientific Research Center RAS, Moscow, Russia;
[2] Emanuel Institute of Biochemical Physics RAS, Moscow, Russia;
[3] Gagarin Center of Cosmonauts Training, Star City, Russia;
[4] Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
e-mail: irina.larina@gmail.com
** Corresponding author*

Motivation and Aim: Changes in the human body during space flights affect all physiological systems, including changing the protein composition of body fluids.

Methods and Algorithms: To investigate the adaptive plasticity of the protein composition of urine we analyzed its samples obtained from six Russian cosmonauts at the age of 35 to 51 years who have completed space missions lasting from 169 to 199 days on the International Space Station (ISS). The analysis of mixtures peptides derived from the samples was carried out by chromato-mass spectrometric method, using the accurate mass and time tag retention of peptides in the chromatographic column (TMMVU). Of the total number of detected peptides (430) twenty-one peptide was presented in the samples of the cosmonauts urine at all stages of the survey. After the establishment of their matching proteins (as UniProt), an analysis of their cellular localization, tissue specificity and functions was carried out.

Results: It was found that the presence of the peptides in the groups of samples (preflight, post-flight) is different, and thus, it was detected a certain drift of the protein-peptide composition of urine, caused by prolonged space flight. 209 peptides, based on a database Tissue-specific Gene Expression and Regulation (TiGER), were characterized as belonging to different tissues. Proteins of liver tissue, bone and soft tissues were significantly more represented in cosmonauts urine in comparison with those in the database TiGER ($P<0,05$) according to Bonferroni adjusted Fisher's exact test for multiple comparisons. Also on the basis of analysis of composition of proteins specifically expressed in the kidney, the Gene Ontology processes were identified that are specific to the background period, for the first and seventh days after the flight.

Conclusion: Proteomic study of the composition of urine, performed by a highly sensitive proteomics methods have provided new data required to clarify the origin of changes in the human body, occurring under the influence of space flight.

IN SILICO STRUCTURAL 3D MODELLING OF NOVEL *CRY1I* AND *CRY3A* GENES FROM LOCAL ISOLATES OF *BACILLUS THURINGIENSIS*

Mahadeva Swamy H.M.[1], Asokan R. [1], Riaz Mahmood[2]
[1] *Division of Biotechnology, Indian Institute of Horticultural Research (IIHR), Hessarghatta lake post, Bangalore 560089 INDIA;*
[2] *Post-Graduate Department of Studies and Research in Biotechnology and Bioinformatics, Kuvempu University, Jnanasahayadri, Shankaraghatta, Shimoga 577451, Karnataka, INDIA*
Corresponding author e-mail: clintonbio@gmail.com

Key words: *3D models, homology modeling, coleopteran insects, domains*

Motivation and aim: Determining the structure and function of a novel protein is a cornerstone of many aspects of modern biology.

Methods and algorithms: The 3D structures was predicted using phyre2 server (http://www.sbg.bio.ic.ac.uk/phyre2/), Conserved Domains and Protein Classification (http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml, version CDDv2.32-40526 PSSMs) and the predicted structure was validated using protein structure validation software suite (PSVS) tool. Determination of protein functional analysis obtained from databases using InterProScan (http://www.ebi.ac.uk/Tools/pfa/iprscan/) and ProFunc (http://www.ebi.ac.uk/thornton-srv/databases/profunc/index.html).

Results: Three-dimensional (3D) models for the 79.2-kDa activated *Cry*1I and 77.4kDa activated *Cry*3A δ-endotoxins from *Bacillus thuringiensis* native isolates that are specifically toxic to Coleopteran insect pests were constructed by homology modeling. They were structurally similar to the known structures, both derived 3D models displayed a three domain organization: the N-terminal domain (I) is a seven helix bundle, while the middle and C-terminal domains are primarily comprise of anti-parallel β-sheets. Significant structural differences within domain II in this model among all Cry protein structures indicates that it is involved in recognition and binding to cell surfaces. Comparison of Coleopteran-active cry toxins predicted structure with available experimentally determined Cry structures reveals identical folds.

Conclusions: The collective knowledge of Cry toxin structures will lead to a more critical understanding of the structural basis for receptor binding and pore formation, as well as allowing the scope of diversity to be better appreciated. Taken together, these studies provided promising evidence that domain swapping, epitope-mapping and protein-engineering under the guidance of molecular modeling can serve as a rational and useful tool in understanding the mode of action of Cry toxins, and ultimately in producing better toxins. Structural insights from these molecular modeling studies would therefore increase our understanding of the mechanic aspects of these two closely related Coleopteran-active insecticidal crystal proteins. These proteins are of interest for agriculture, as they offer a means for control of beetles and other insect crop pests.

Availability: Academic, Acknowledgments: The authors are grateful to Indian Council of Agricultural Research (ICAR), New Delhi for funding this study under Network project on Application of Microbes in Agriculture and Allied Sectors (AMAAS). Infrastructure facility and encouragement by The Director, Indian Institute of Horticultural Research (IIHR) are duly acknowledged.

TOWARDS AN UNDERSTANDING OF THE ROLE OF HUMAN RIBOSOMAL PROTEINS IN VARIOUS CELLULAR PROCESSES RELATED TO HEALTH AND DISEASES

Malygin A.A.*, Ivanov A.V., Babaylova E.S., Karpova G.G.
Institute of Chemical biology and Fundamental Medicine SB RAS, Novosibirsk, Russia
e-mail: malygin@niboch.nsc.ru
**Corresponding author*

Key words: *human ribosomal proteins, gene expression regulation, HCV, IRES*

Motivation and Aim: Ribosomal proteins are constitutive components of the ribosome and, therefore, they are involved in the work of translation machinery. However, apart from their "ribosomal" functions, many ribosomal proteins are also implicated in a variety of other cellular processes related to health and disease.

Results: We found that human ribosomal protein (rp) S13 can regulate expression of its own gene at the splicing step by a feedback mechanism. According to this mechanism, rpS13 was demonstrated to inhibit excision of intron 1 from its own pre-mRNA due to the binding to the intron nearby the splicing sites that interferes interaction of the conventional splicing factors with these sites [1]. The resulting mRNA retains the intron and therefore is nonsense. Similar mechanism was found to be realized with rpS16 [2] and rpS26 [3]. The feedback regulation of ribosomal protein genes expression at the splicing step may provide fine tuning of the level of each ribosomal protein in cell. Ribosomal protein SA is known not only as a component of the ribosome, but also as a precursor of the cell-surface laminin receptor, LAMR. RpSA is homologous to eubacterial rpS2, but in contrast to it rpSA is not a constant ribosomal component. It has a eukaryote-specific C-terminal domain that was found to be responsible for the protein binding to the 40S ribosomal subunit involving mainly the 18S rRNA helix 40. The C-terminal domain of rpSA was shown to contain a receptor domain for Venezuelan equine encephalitis and tick-borne encephalitis viruses. We examined the internal ribosome entry site (IRES) of the hepatitis C virus (HCV) RNA for its ability to bind to 40S ribosomal subunits (i) deficient in rpSA, (ii) saturated with recombinant rpSA, or (iii) pretreated with monoclonal antibodies whose epitops are located in the C-terminal domain of rpSA. Binding of HCV IRES to 40S subunits was shown to depend largely on the rpSA content in the subunits and to be blocked by the antibodies [4]. The results obtained imply that eukaryote-specific C-terminal domain of rpSA is implicated in binding of the HCV IRES to the ribosome and, therefore, in translation initiation of HCV RNA.

Acknowledgements: The work was supported by grant from RFBR 11-04-00672-a.

References:

1. A.A. Malygin et al. (2007) Human ribosomal protein S13 regulates expression of its own gene at the splicing step by a feedback mechanism, *Nucleic Acids Res.,* **35:** 6414-6423.
2. A.V. Ivanov et al.. (2010) Human ribosomal protein S16 inhibits excision of the first intron from its own pre-mRNA, *Molecular Biology,* **44:** 82-88.
3. A.V. Ivanov et al. (2005) Human ribosomal protein S26 suppresses the splicing of its pre-mRNA, *Biochim. Biophys. Acta,* **1727:** 134-140.
4. A.A. Malygin et al. (2009) Binding of the IRES of hepatitis C virus RNA to the 40S ribosomal subunit: role of p40, *Molecular Biology,* **43:** 997-1003.

STATISTICAL ANALYSIS OF DATABASE DERIVED INTER-RESIDUE CONTACT POTENTIALS

Mavropulo-Stolyarenko G.R.
Saint-Petersburg State University, Saint-Petersburg, Russia
e-mail: gm2124@mail.ru

Key words: *protein structure prediction, comparative modelling, statistical contact potentials*

Motivation and Aim: So-called, statistical contact potentials, have proven to be a valuable tool of modern days protein structure modelling. By applying Boltzmann formalism, to probabilities of amino acids being in contact in real protein structures and in random "reference state", energy-like measures are derived that allow filtering decoy from a native-like structures [1,2]. Thus our understanding of how to construct more precise and sensitive potentials is the key to a possible breakthrough in protein structure prediction. The purpose of this study is to statistically assess "database derived contact potentials" model in order to find ways to improve it.

Methods and Algorithms: We had performed a statistical analysis of distance dependent contact potentials derived from the fresh (PDBselect 25% Mar. 2012) non redundant protein structure list [3]. As a random reference state, both: contact map (amino-acid sequence) shuffle and Monte Carlo random structure generation models were implemented. We used binomial distribution statistics with multiple comparisons correction for testing contact probabilities for equality.

Results: Careful statistical analysis of statistical contact potentials had shown that most of the currently known methods to construct them, oversimplify underlying model thus missing some of its statistically significant parts. Although it is hard to estimate the amount of the error introduced by these simplifications, our findings suggest that existing statistical potentials could be improved.

Conclusion: Results presented in this study show possible paths of improving statistical potentials, to account for newly discovered effects like N-C pair asymmetry or low intra-sequence separation.

Availability: software used in this research will be available on request from the authors since the fall of 2012.

References

1. Hendlich M, Lackner P, Weitckus S, Floeckner H, Froschauer R, Gottsbacher K, Casari G, Sippl MJ. Identification of native protein folds amongst a large number of incorrect models. The calculation of low energy conformations from potentials of mean force. J Mol Biol. 1990; 216:167–80.
2. Sippl MJ. Boltzmann's principle, knowledge-based mean fields and protein folding. An approach to the computational determination of protein structures. J Comput Aided Mol Des. 1993; 7:473–501.
3. Sven Griep and Uwe Hobohm. PDBselect 1992–2009 and PDBfilter-select. Nucleic Acids Research, 2010, Vol. 38, Database issue D318–D319.

MOLECULAR DYNAMICS SIMULATION OF Nip7 PROTEINS FROM HYPERTHERMOPHILIC ARCHAEA AT HIGH TEMPERATURE AND PRESSURE

Medvedev K.E.*[1], Afonnikov D.A.[1, 2], Vorobyev Y.N.[3]
[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia;*
[2] *Novosibirsk State University, Novosibirsk, Russia;*
[3] *Institute of Chemical Biology and Fundamental Medicine, SB RAS, Novosibirsk, Russia*
e-mail: kirill-medvedev@yandex.ru
** Corresponding author*

Key words: *Nip7 protein, high pressure, molecular dynamic simulation, specificity determining positions, water-protein interaction*

Motivation and Aim: To investigate mechanisms of protein structure adaptation to high pressure environment we compare the dynamic properties of Nip7 proteins from hyperthermophilic archaea, Pyrococcus abyssi (deep sea habitat) and Pyrococcus furiosus (shallow water habitat) at different pressures (0.1-300 MPa) and temperatures (300, 373K). We perform search for Specificity Determining Positions (SDPs) in homologous archaeal proteins with respect to the organism's habitat depth (deep/shallow water) and compare results with molecular dynamics simulation analysis.

Methods and Algorithms: The structure of P.furiosus protein was obtained by homology modeling using Nip7 from P.abyssi as template (PDBID 2p38). MD simulations and structure analysis were performed using GROMACS [1]. Search for SDPs was performed using SDPPred [2] and GroupSim [3] software.

Results: It is shown that the structure of Nip7 N-terminal domain is more stable and has smaller structure fluctuations than C-terminal (RNA-binding) domain. We demonstrated that protein from P.abyssi is more stable under extreme conditions than from P.furiosus. P.abyssi protein has larger polar solvent accessible surface area in comparison with P.furiosus protein. Most of detected SDPs in Nip7 homologs display significant changes in side chain polarity between deep/shallow water organisms.

Conclusion: In general, these data demonstrate the importance of water-protein interactions for the protein stability under high pressure.

Acknowledgements: The work supported by RFBR (11-04-01771-a), SB RAS (projects 130, 39, 93), RAS (project 6.8 and program 28), State contract 82/201 and Scientific school 5278.2012.4.

References:

1. D.Van der Spoel et al. (2005) GROMACS: Fast, Flexible and Free. J. Comp. Chem. 26:1701-1718.
2. Kalinina OV, Mironov AA, Gelfand MS, Rakhmaninova AB. (2004) Automated selection of positions determining functional specificity of proteins by comparative analysis of orthologous groups in protein families. Protein Sci 13(2): 443-56
3. Capra JA and Singh M. (2008) Characterization and Prediction of Residues Determining Protein Functional Specificity. Bioinformatics, 24(13): 1473-1480.

MASS-SPECTROMETRIC MEASUREMENT OF LEVELS AND ENZYMATIC ACTIVITY OF CYTOCHROMES P450

Moskalyova N., Zgoda V.G.*, Tikhonova O., Novikova S., Kopylov A., Archakov A.I.
Orekhovich Institute of Biomedical Chemistry, Russian Academy of Medical Sciences
e-mail: vic@ibmh.msk.su
** Corresponding author*

Key words: *cytochrome P450 (CYP), multiple reaction monitoring (MRM), mass-spectrometry, enzymatic activity, drug oxidation*

Motivation and Aim: Personalized medicine requirements dictate the possibility to study metabolism of several drugs in individual patient. It is well known that cytochromes P450 (CYP) play a crucial role in oxidation of most medicines. Thus, a need is in new multiplied methods to measure a signature of levels of these enzymes and their activities.

Methods and Algorithms: Levels of several members of mammalian CYP subfamilies 1A, 3A, 1E, 2C, 2D were measured by multiple reaction monitoring in triple quadrupole mass spectrometer. Method was developed and validated using samples of murine liver microsomes taken from intact controls and mice induced by phenobarbital and methylcholanthrene xenobiotics.

Results: The method allowed reliable measurements of levels of several CYP isoforms without use of isotopic molecular mass labels and specific derivatizing agents. The results correlated with values of enzymatic activity determined using marker substrates specific for CYP isoforms of interest.

Conclusion: The new method to measure cytochrome P450 signature by state-of-the-art mass-spectrometry is developed which may be easily translated to human healthcare.

Acknowledgement: The work was funded by Russian Academy of Medical Sciences.

PROF_PAT, THE DATABASE OF PROTEIN FAMILY PATTERNS – AN EFFECTIVE TOOL FOR SEQUENCES ANNOTATION

Nizolenko L.Ph.*, Bachinsky A.G.
FBIS NRC of Virology and Biotechnology "Vector", Koltsovo, Russia
e-mail: nizolenko@vector.nsc.ru
** Corresponding author*

Key words: *protein families; patterns; similarity search; data banks; amino acid sequences*

Motivation and Aim: In the case of distant proteins, the search for global similarity of complete sequences may fail to show positive result, because conservative blocks responsible for their special functions may prove to be relatively short and scattered all over the sequence. So, when analyzing novel protein sequences, in addition to routine searches of the primary data sources, it is essential to extend search strategies to include a range of "secondary" databases, representing protein families. The level of noise is lower, because comparison is usually made with patterns that represent conservative intervals of positions.

Methods and Algorithms: Prof_Pat is a sample of "secondary" database, created for the greatest possible number of proteins of the UniProt. Flexible program for fast searching is supplied. It can investigate individual amino acid sequence, as well as large set of them in one pass. Technology of construction of Prof_Pat is described in detail in [1].

Results: The main field of application of Prof_Pat is an annotation of the new amino acid sequences. Such as amino acid sequences, translated from complete genomes of micro organisms. An example of this type of the investigation is an analysis of the open reading frames of *M. tuberculosis* described in [2]. A brief summary of the study and results of the investigation of amino acid sequences of some other microorganisms are shown.

Microorganism name	Number of open reading frames	Results of comparison with Prof_Pat		
		Similarity not founded	Recognized with high significance level	New[a] similarity
M. tuberculosis	3924	2	2777	44
Bacillus subtilis	4105	4	3284	22
Salmonella typhi	4767	4	4578	16
Brucella abortus	1033	5	992	36

[a]This protein's functions were not predicted earlier.

Conclusion: Despite the existence of multiple databases and algorithms to search distant similarity of amino acid sequences, Prof_Pat provides innovative results and, therefore, serves as a useful tool in the study of new sequences.

Availability: http://wwwmgs.bionet.nsc.ru/mgs/programs/prof_pat/.

References:

1. A.G. Bachinsky et al. (2000) PROF_PAT 1.3: updated database of patterns used to detect local similarities, *Bioinformatics*, 16: 358-366.
2. L.F. Nizolenko et al. (2005) Investigation of the amino acid sequences of open reading frames of the Mycobacterium tuberculosis complete genome with the use of the protein family pattern bank Prof_Pat, *Biofizika*, 50: 986-992.

PUTracer: A NOVEL METHOD FOR IDENTIFICATION OF CONTINUOUS-DOMAINS IN MULTI-DOMAIN PROTEINS

Parsa M.[1], Pashandi Z.[2], Mobasseri R.[2], Arab S.S.*[2]
[1] *Institute for Research in Fundamental Sciences IPM, Tehran, Iran;*
[2] *Tarbiat Modares University TMU,Tehran, Iran*
e-mail: sh.arab@modares.ac.ir
** Corresponding author*

Key words: *proteins, continues-domains, accessible surface area, energy of hydrogen bonds, top-down approach, recursive*

Motivation and Aim: Computer-assisted assignment of protein domains is considered as an important issue in structural bioinformatics. Exponential increasing in the number of known three-dimensional structure of proteins and the significance role of proteins in biology, medicine and pharmacology still illustrate the necessity of a reliable method to detect automatically structural domains as protein units.

Methods and Algorithms: For this aim, we develop a program based on the accessible surface area (ASA) and the energy of hydrogen bonds in backbone residues (EHB). PUTracer builds on the features of a fast top-down approach to cut a chain or domain into the new domains with minimal change in ASA as well as EHB, and without destruction of disulfide bonds at every recursive step.

Results: Performance of program was assessed by a comprehensive benchmark dataset of 124 protein chains which is based on agreement among experts (e.g., CATH, SCOP) and was expanded to include structures with every type of domain combinations. Equal number of domain and at least 90 % agreement in critical boundary accuracy were considered as correct assignment conditions. PUTracer assigned domains correctly in more than 82% of proteins.

Conclusion: Although, low critical boundary accuracy in 18 % of proteins leads to the incorrect assignments, adjusting the scales is likely to improve the performance up to 92 %. We discuss here the successes or failure of adjusting the scales with provided evidences.

Availability: PUTracer is available at http://bioinf.modares.ac.ir/software/PUTracer/

COMPUTATIONAL EVALUATION OF IMPACT OF AMINO ACID SUBSTITUTION p.W172C ON STRUCTURE AND FUNCTION OF GAP-JUNCTION PROTEIN CONNEXIN 26 AND ITS ASSOCIATION WITH HEARING IMPAIRMENT

Pintus S.S.[1], Bady-Khoo M.S.[1, 2], Posukh O.L.*[1, 3]
[1] Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia; [2] Republican Hospital #3, Kyzyl, Russia; [3] Novosibirsk State University, Novosibirsk, Russia
** Corresponding author: posukh@bionet.nsc.ru*

Motivation and Aim: Mutations in the *GJB2* gene, encoding the gap-junction protein connexin 26 (Cx26), are the most common cause of non-syndromic deafness. Transmembrane protein Cx26 forms intercellular channels that permit the exchange of ions and small molecules between adjacent cells. Defects of Cx26 lead to the disturbance in ion homeostasis of inner ear endolymph which results in hearing impairment. To date, about 200 different pathogenic mutations, polymorphisms and changes with unknown relation to the disease in *GJB2* gene have been reported («Connexins and Deafness Homepage»: http://davinci.crg.es/deafness/). We analysed nucleotide sequences of the *GJB2* gene entire coding region in 90 deaf patients of Tuvinian ethnicity (the Tuva Republic, Russia) and revealed sequence alterations: c.516G>C (p.W172C), c.235delC (p.L79fs), c.109G>A (p.V37I), c.299_300delAT (p.H100fs), c.79G>A (p.V27I), c.341A>G (p.E114G), and c.571T>C (p.F192L). Most of them are known pathogenic mutations or polymorphisms. Interestingly, we observed high prevalence of a nonsynonymous substitution c.516G>C (p.W172C) (79.4% out of all mutant chromosomes) which previously rarely detected in deaf patients (Posukh et al., 2005; Tekin et al., 2010). The aim of this study is *in silico* evaluation of the p.W172C impact on structure and function of protein Cx26 and its association with hearing impairment.

Methods and Algorithms: The PolyPhen (http://genetics.bwh.harvard.edu/pph/) and PolyPhen2 (http://genetics.bwh.harvard.edu/pph2/) are the tools which predict possible impact of an amino acid substitution on the structure and function of a human protein using straightforward physical and comparative considerations. Both programs were used for prediction of the p.W172C effect on structure and function of human Cx26 (http://www.uniprot.org: CXB2_HUMAN). Multiple sequence alignment of the Cx26 amino acid sequences of different species was performed by ClustalX.

Results: The p.W172C change is located in the second extracellular domain of Cx26. Its effect on the protein Cx26 was assessed using the PolyPhen where a PSIC (position-specific independent counts) score difference between W and C was found to be 2.669 that classified W172C as probably damaging mutation. According to the PolyPhen2 prediction model HumVar the W172C was also classified as probably damaging mutation (p=0.901). Multiple sequence alignment of the Cx26 amino acid sequences in different species (*H. sapiens, O. anatinus, M. domestica, H. glaber, C. porcellus, M. gallopavo*) by ClustalX revealed high conservative W at position 172 in connexin 26.

Conclusion: Based on evolutionarily conservation of W at position 172 in connexin 26 in many species and the PolyPhen analysis, p.W172C (c.516G>C) was considered to be probably pathogenic alteration and associated with hearing impairment. Substitution of an aromatic non-polar tryptophan at position 172 on small polar cysteine probably results in impairment of connexin protomers connection or opening-closing mechanism of intercellular Cx-channels.

Acknowledgements: The study was supported by RFBR grant #11-04-01221-a.

MULTISTATE ORGANIZATION OF TRANSMEMBRANE HELICAL PROTEIN DIMERS GOVERNED BY THE HOST MEMBRANE

Polyansky A.A.*[1, 2], Volynsky P.E.[1], Efremov R.G.[1], Shemyakin M.M.[1], Ovchinnikov Yu.A.

Institute of Bioorganic Chemistry, Russian Academy of Sciences, Moscow, Russia;
[2] Department of Structural and Computational Biology, Max F. Perutz Laboratories, University of Vienna, Campus Vienna Biocenter 5, Vienna, AT-1030, Austria
** Corresponding author*

Keywords: *prediction of protein structure, molecular dynamics simulations, free energy calculations, molecular hydrophobicity potential, protein-membrane interactions, receptor tyrosine kinases, epidermal growth factor receptors*

Motivation and Aim: Association of transmembrane (TM) helices taking place in the cell membrane has an important contribution into the biological function of bitopic proteins, among which receptor tyrosine kinases represent a typical example and very potent target for medical applications. Since this process is driven by many factors (i.e. primary structures of TM domains and juxtamembrane regions, composition and phase of the local membrane environment, etc.), it is still far from being fully understood.

Methods and Algorithms: We have used original modeling approach (so-called PREDDIMER [1]), which allows prediction of TM helical oligomers from their primary sequences based on quantitative estimations of complementarity between geometrical and polar properties of helical surfaces. We have estimated the free energy of association of several predicted dimers in in full-atom explicit lipid bilayers composed of phosphocholine lipids with different acyl chains, using umbrella sampling techniques with the mean force integration.

Results: We present a computational modeling framework, which we have applied to systematic consideration of dimerization for 18 TM helical homo- and heterodimers of different bitopic proteins, including the family of epidermal growth factor receptors. For this purpose, we have developed a novel surface-based modeling approach, which is able not only to predict some particular conformations of TM dimers displaying good agreement with the experiment, but also provides screening of their conformational heterogeneity together with simple estimation of the dimerization efficiency. To elucidate a putative role of the environment in a selection of a particular conformation, we have employed full-atomic MD simulations of several of the predicted dimers in different model membranes. Analysis of about 20 μs of MD statistics clearly shows that each particular bilayer preferentially stabilizes one of possible conformations of a dimer, and that the energy gain depends on interplay between structural properties of the protein and the membrane.

Conclusions: Our results suggests a multistate organization of TM helical dimers in heterogeneous membranes, and emphasizes an importance of consideration of their conformational variability to design potent selective modulators of dimerization acting on pharmaceutically relevant targets in the natural medium of cell membranes.

Refereces:

1. A. A. Polyansky, P. E. Volynsky, R. G. Efremov. (2011) *Adv. Protein Chem. Struct. Biol.*, **83**, 129-161.

INTEGRATED APPROACH TO MOLECULAR DYNAMICS STUDY OF PROTEINS AND PROTEIN-DNA COMPLEXES

Popov A.V.*, Zharkov D.O., Vorobyev Y.N.
Institute of Chemical Biology and Fundamental Medicine SB RAS, Novosibirsk, Russia
e-mail: apopov@niboch.nsc.ru
** Corresponding author*

Key words: *molecular dynamics, trajectory analysis, base excision repair, glycosylase*

Motivation and Aim: Within the bounds of computer systems designed to predict new pharmacological targets, an integrated approach to modelling and analysis of proteins and DNA-protein complexes should be developed. Molecular dynamics (MD) simulations of biopolymers on meaningful time scales produce large trajectories rarely amenable to manual analysis; therefore a convenient analysis tool is required. The approach should be applied to model and analyze structure and dynamics of some vital enzymes to affirm its suitability and topicality.

Methods and Algorithms: BioPASED, a molecular dynamics modelling program, is a part of BISON complex and uses mathematical model of molecule in the terms of classic molecular mechanics. GUI-BioPASED, a graphical web interface to the BioPASED, was developed to assist in error-free task formation and preliminary structure validation [1]. MDTRA, a molecular dynamics trajectory analyzing program, was developed to facilitate the process of trajectory analyzing, meaningful data extraction and representation. All the tools work in close liaison and form a conveyor from initial PDB structure to numeric data and plot images.

Results: The approach described was applied to analysis of structure and dynamics of DNA-*N*-glycosylases involved in the process of DNA base excision repair. Three different wild-type enzymes were modeled (*Lactococcus lactis* Fpg, *Bacillus stearothermophilus* MutY, human OGG1) with a DNA helix fragment containing a lesion 8-oxoguanine (oG) which is their common substrate. Fpg was modeled in four different charge states, two states of two amino acids of the active site to discover the best charge pair. MutY was modelled with either oG or G opposite to A to study its substrate discrimination. The same was done for OGG1 (oG:C and oG:A pair). Analysis of trajectories shed some light to how these enzymes function. To investigate a mutation impact on structure stability and catalytic activity, three OGG1 mutants were also modeled (C253I, C253L, Q315W). A noticeable difference between these mutants, and between the mutants and the wild type, was shown; it coincided well with the steady state kinetics results.

Conclusion: The integrated approach to molecular dynamics study was developed and proved useful in studies of structure, dynamics and function of some DNA repair enzymes. It can be used to perform high-quality molecular dynamics experiments, including prediction of new pharmacological targets, and analyze their outcome.

Availability: The programs are available on the website (bison.niboch.nsc.ru).

Acknowledgements: This research was supported by the Presidium of the Russian Academy of Sciences (6.14) and by Russian Foundation for Basic Research (11-04-00807, 09-04-00136).

References:

1. A.V. Popov, Yu.N. Vorobjev. (2010) GUI-BioPASED: A Program for Molecular Dynamics Simulations of Biopolymers with a Graphical User Interface, *Molecular Biology (Moscow)*, **44**: 648-654.

A SCREENING OF G-QUADRUPLEX MOTIFS AS A STRUCTURAL BASIS OF APTAMERS TO TICK-BORNE ENCEPHALITIS VIRUS GLYCOPROTEIN

Potapova U.V. *[1], Potapov V.V. [1], Kondratov I.G.[1], Solovarov I.S., Belikov S.I.[1], Vasiliev I.L. [2]

[1] *Limnological Institute, SB RAS, Irkutsk, Russia*

[2] *Institute of System Dynamics and Control Theory, SB RAS, Irkutsk, Russia*

e-mail: potapova@lin.irk.ru

** Corresponding author*

Key words: *G-quadruplex, aptamer structure, tick-borne encephalitis virus*

Motivation and Aim: A replacement of specific antibodies by DNA-aptamers is a modern trend in a therapy of viral infections. Anti-viral substances for a hepatitis C and other viruses are currently developed. A transfer of aptamers inside a cell to achieve an interaction with target proteins is a main problem of aptamer therapy. A purpose of the research is to obtain aptamers to surface protein E of tick-borne encephalitis virus, as this protein is available for aptamers outside a host cell. From a pool of 171 aptamer sequences, the sequences should be found which could selectively bind to glycoprotein E and prevent a penetration of a virion to a cell. The common structural motifs should be determined in the selected sequences.

Methods and Algorithms: A frequency analysis was used as a method of an aptamer screening. The complete enumeration gave 8 most frequent segments with a length not less than 10 nucleotides. Web-service QGRS Mapper [1] shows motifs of G-quadruplexes in the selected segments.

Results: G-quadruplexes are the structural backbone of aptamers and are quite stable relative to simple structural motifs of hairpins and pseudo-knots. G-quadruplexes are most often in DNA aptamers [2].

The possible tertiary structures of the selected aptamers are predicted from a known tertiary structures of G-quadruplexes.

Conclusion: The results obtained could lead to a development of a safe antiviral medicine on a base of highly specific DNA aptamers.

Acknowledgements: This work was funded in part by the Russian Ministry of Science and Education (State Contract No. 16.512.11.2258), the interdisciplinary integration project of SB RAS № 141.

References:

1. Kikin O, (2006). QGRS Mapper: a web-based server for predicting G-quadruplexes in nucleotide sequences, *Nucleic Acids Research,* **34:** 676–682
2. А.В. Кульбачинский (2006). Методы отбора аптамеров к белковым мишеням, *Успехи биологической химии,* **46:** 193-224

STUDY OF CONFORMATIONAL FLEXIBILITY OF *E. COLI* RNA POLYMERASE ALPHA SUBUNIT INTERDOMAIN LINKER

Purtov Yu.A.*, Kondratyev M.S., Ozoline O.N., Komarov V.M.
Institute of Cell Biophysics RAS, Pushchino, Moscow region, Russian Federation
e-mail: yapurtov@yahoo.com
** Corresponding author*

Key words: *alpha-subunit RNA-polymerase, interdomain linker, structure, dynamics*

Motivation and Aim: In spite of the fact that process of transcription initiation in prokaryotes is well understood some of its aspects are complicated for experimental study. In particular it can be attributed to positioning of RNA polymerase alpha subunits on DNA. This protein consists of two domains (N- and C-terminal) joined with linker preventing protein crystallization and determination of its spatial structure. Earlier it was shown that changes in amino acid sequence of linker led to decrease of promoter recognition efficiency [1]. This work is dedicated to study of structural and dynamical aspects of interdomain linker flexibility by use of computational full-atomic modeling.

Methods and algorithms: homology analysis, sequence alignment, quantum-chemical PM3 method (MOPAC2009), molecular dynamics (GROMACS), molecular docking (HEX, Autodock Vina).

Results: analysis of spatial structure and dynamic flexibility of *E.coli* RNA polymerase alpha subunit linker as well as its artificially selected homologues allowed to determine significant role of several amino acid residues in linker topology. We assume that charged amino acids and prolines to be essential for this process. Charged residues (Arg, Lys, Asp, Glu) affect the linker structure forming salt bridges, while prolines provide appropriate bending and strength of peptide chain. Molecular docking of linker on DNA showed its potential capability to interact with both major and minor DNA groove. Comparison of amino acid sequence of linker and its homologues from other prokaryotes allowed to determine natural variability of this region and suggest a hypothesis that linker composition is dependent on ecology of bacterial species.

Conclusion: investigation of structure of *E.coli* RNA polymerase alpha subunit interdomain linker gives a possibility to state that amino acid sequence of this protein region directly affects alpha subunit C-terminal domain positioning on recognized DNA sequence. We consider this linker flexibility preventing alpha subunit crystallization as adapting mechanism for interaction with variety of promoter sequences and transcription factors.

References:

1. Fujita N, Endo S, Ishihama A. (2000) Structural requirements for the interdomain linker of alpha subunit of Escherichia coli RNA polymerase. Biochemistry. 39, 6243-6249.

IN SILICO STUDIES OF POTENTIAL PHOSPHORESIDUES IN THE HUMAN NUCLEOPHOSMIN/B23: ITS KINASES AND RELATED BIOLOGICAL PROCESSES

Gioser Ramos-Echazábal*[1], Glay Chinea[2], Rossana Garcia-Fernández[3], Tirso Pons[4]

[1] *Department of Animal and Human Biology, Faculty of Biology, University of Havana, Havana 10400, Cuba;*

[2] *Biomedical Division, Center for Genetic Engineering and Biotechnology, P.O. Box 6162, Havana 10600, Cuba;*

[3] *Center for Protein Studies, Faculty of Biology, University of Havana, Havana 10400, Cuba;*

[4] *Structural Biology and Biocomputing Programme, Spanish National Cancer, Research Centre (CNIO), C/Melchor Fernőndez Almagro 3, Madrid E-28029, Spain*

e-mail: gio@fbio.uh.cu, gioser.echazabal@gmail.com

** Corresponding author*

Key words: *phosphorylation; prediction; kinases; protein–protein interactions; cellular localization; signaling; pathway; conservation*

Motivation and Aim: Human nucleophosmin/B23 (32 kDa /pI 5.1) is a phosphoprotein involved in ribosome biogenesis, centrosome duplication, and apoptosis. Its function, localization, and mobility within cells, are highly regulated by phosphorylation events (1). Up to 21 phosphosites of B23 have been experimentally verified even though the corresponding kinase is known only for seven of them (Phosida code: P06748 and Phospho.ELM code: P06748). In this work, we predict the phosphorylation sites in human B23 using seven public servers.

Methods and Algorithms: Of these, six were kinase-specific servers (KinasePhos 2.0, PredPhospho, NetPhosK 1.0, PKC Scan, pkaPS, and MetaPredPS) and one was not (DISPHOS 1.3). The results were integrated with information regarding 3D structure and residue conservation of B23, as well as cellular localizations, cellular processes, signaling pathways and protein–protein interaction networks involving both B23 and each predicted kinase.

Results: Thus, all 40 potential phosphosites of B23 were predicted with significant score (>0.50) as substrates of at least one of 38 kinases. Thirteen of these residues are newly proposed showing high probability of phosphorylation considering their solvent accessibility. Our results also suggest that the enzymes CDKs, PKC, CK2, PLK1, and PKA could phosphorylate B23 at higher number of sites than those previously reported. Furthermore, PDK, GSK3, ATM, MAPK, PKB, and CHK1 could mediate multisite phosphorylation of B23, although they have not been verified as kinases for this protein.

Conclusion: Finally, we suggest that B23 phosphorylation is related to cellular processes such as apoptosis, cell survival, cell proliferation, and response to DNA damage stimulus, in which these kinases are involved. These predictions could contribute to a better understanding, as well as addressing further experimental studies, of B23 phosphorylation.

References:

1. Okuwaki M. The structure and functions of NPM1/Nucleophsmin/B23, a multifunctional nucleolar acidic protein. J Biochem 2008;143:441-448.
2. Gnad F, Gunawardena J, Mann M. PHOSIDA 2011: the posttranslational modification database. Nucleic Acids Res 2011;39:D253-D260.
3. Dinkel H, Chica C, Via A, Gould CM, Jensen LJ, Gibson TJ, Diella F. Phospho.ELM: a database of phosphorylation sites--update 2011. Nucleic Acids Res 2011;39:D261-D267.

FAMILY OF KCTD PROTEINS: STRUCTURAL AND FUNCTIONAL PECULIARITIES

Skoblov M.Yu.*[1, 2], Marakhonov A.V. [1], Baranova A.V. [1, 3]
[1] *Federal State Budgetary Institution "Research Centre for Medical Genetics" under the Russian Academy of Medical Sciences, Moscow, Russia;*
[2] *State Budgetary Institution of Higher Education "Moscow State Medical and Dental University", Moscow, Russia;*
[3] *School of Systems Biology, College of Science, George Mason University, Fairfax, VA USA*
e-mail: mskoblov@generesearch.ru
** Corresponding author*

Key words: *KCNRG, KCTD proteins, function, tumor suppressor*

Motivation and Aim: Previously in our laboratory the novel potential tumor suppressor gene *KCNRG* had been discovered during the study of the cause of the B-cell chronic lymphocytic leukemia. The protein appeared to contain a single conserved domain, T1 potassium channel tetramerization domain. T1 domain is required for the protein interaction of proteins voltage-gated potassium channels subunit. KCNRG has been shown to negatively regulate potassium currents *in vitro* as well as suppress the proliferation and activate the apoptosis in cancer cell lines. It is known that activation and proliferation of lymphocytes depends on potassium channels. It was supposed to investigate an opportunity of an induction of apoptosis in B-CLL cells by the restoration of potassium channels inhibition with the help of low-molecular weight compounds. We also carried out i*n silico* whole-genome searching which revealed the whole family of only T1 domain-containing proteins in human genome called KCTD family. The main goal of this study was to functionally characterize the family of KCTD proteins.

Results: We selected 25 low molecular weight compounds for study its ability to inhibit potassium channels. Preliminary testing were performed on B-cell line Raji. Cells from 15 B-CLL patients isolated and maintained in culture. Ten most perspective candidates were chosen out of tested panel for further investigation. We also analyzed the participation of KCTD proteins in different cellular pathways. KCTD family showed to be a diversified group. Moreover the reported data for the interaction of KCTD family with other proteins support the hypothesis of heterogeneous functions of different KCTD proteins.

Conclusion: It is generally accepted that presence of the conserved domain in protein sequence would predominantly determine the function of the protein. This study stresses that, after initial *in silico* analysis, experimental verification of the molecular function is warranted.

References:

1. B. Biderman et al. (2010). Inhibition of potassium currents as a pharmacologic target for investigation in chronic lymphocytic leukemia. *Drug News & Perspectives*. **23** (10):625-31.
2. A Birerdinc et al. (2010). Pro-apoptotic and antiproliferative activity of human KCNRG, a putative tumor suppressor in 13q14 region. *Tumour Biol*. **31** (1):33-45.

SEARCH OF PLASMA PROTEIN BIOMARKERS FOR SCHIZOPHRENIA

Smirnova L.P.*[1], Koval V.V. [2], Loginova L.V. [1], Fedorova O.S.[2], Ivanova S.A.[1]
[1] *Mental Health Research Institute SB RAMSci, Tomsk, Russia;*
[2] *Institute of Chemical Biology and Fundamental Medicine SB RAS, Novosibirsk, Russia*
e-mail: lpsmirnova@yandex.ru
** Corresponding author*

Key words: *schizophrenia, biomarkers, proteomic analysis*

Motivation and Aim. Schizophrenia is a complex mental disorder with fairly high level of heritability. Pathogenesis of schizophrenia is still unclear but disturbances of protein metabolism in schizophrenia are known. Nevertheless, protein marker, inherent in only this illness, still has not been detected. Objective is the proteomic analysis of blood plasma in patients with schizophrenia and healthy persons.

Methods. Object of investigation was the blood of 10 healthy persons and 16 patients with schizophrenia. Patients were under therapy at clinics of Mental Health Research Institute SB RAMSci, Tomsk. Diagnosis was conducted according to current classification ICD-10. Plasma proteins were separated with gel-electrophoresis, digested by trypsin, and analyzed by MALDI-TOF mass-spectrometry (Autoflex II, Bruker Daltonics). The proteins were identified using Mascot software (Matrix Science).

Results. Our study reveals that protein of metabotropic glutamate receptor (mGlurR6) is detected in plasma of patients with schizophrenia. Long activation of metabotropic glutamate receptors results in reinforcement of NMDA-dependent generation of active forms of oxygen what entails damage of receptors. Research of microchips has revealed decrease of expression of regulator of transmission of signal in synapses of G-protein-4 (RGS4) in schizophrenia. RGS4 is a negative regulator of receptors connected with G-proteins, including metabotropic glutamate receptors and plays an important role development of nervous system. Thus, obtained results allow to suppose that glutamatergic synapses are the basic place of action in pathogenesis of schizophrenia.

Support by Grant (N5) from SB RAS for collaborative research.

PROTEOMICS AND METABOLOMICS OF THE RAT LENS: ANALYSIS OF AGE AND CATARACT-SPECIFIC CHANGES

Snytnikova O.A.*[1,2], Kopylova L.V.[1,2], Cherepanov I.V.[1,2], Duzhak T.G.[1,2], Kolosova N.G.[3], Sagdeev R.Z.[1], Tsentalovich Y.P. [1,2]
[1] *International Tomography Center SB RAS, Novosibirsk, Russia;*
[2] *Novosibirsk State University, Novosibirsk, Russia;*
[3] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*
e-mail: koa@tomo.nsc.ru
** Corresponding author*

Key words: *cataract, lens proteomics, post-translational modifications, rats*

Motivation and Aim: During the last ten years a big surgical breakthrough was made in the curing of the eye diseases, but cataract is still the most frequent cause of impairment and loss of vision in elderly people. The developed cataract is characterized by numerous post-translational modifications (PTMs) of the major lens proteins, crystallins. The transparency of the eye lens strongly depends on the crystallin solubility and structure, while the solubility is mediated by protein modifications accumulating with age. At present, it is not completely clear which modifications are cataract-specific, and which are just a part of the normal maturation and aging processes. The vast majority of experimental data on the biochemical content of cataractous human lenses corresponds to lenses with developed cataract surgically removed from patient eyes. The studies of early stages of human cataract are limited. One approach used for studies of etiology and pathogenesis of human diseases and for development of new methods for their treatment is the use of biological models. Recent studies have shown that the OXYS rat strain meets the main requirements for the model of senile cataract. The purpose of this study was to determine the age-related and the cataract-specific changes in the crystallin composition in lenses of accelerated-senescence OXYS (cataract model) and Wistar (control) rats.

Methods and Algorithms: The water-soluble (WS) and urea-soluble (US) fractions of the lens proteins were separated; the identity and relative abundance of each crystallin were determined by 2-DE and MALDI-TOF/TOF mass spectrometry.

Results: This report provides the data on the proteomic and metabolomic analysis of two rat strains of different ages: OXYS and Wistar. The interstrain differences in the crystallin compositions appear at 3 months of age. One of the most pronounced effects is the insolubilization of gamma crystallins, and this process proceeds faster in OXYS rat lenses. The main established PTMs are oxidation, N-term acetylation and asparagines deamidation.

Conclusion: The major age-related changes in proteomic composition of the rat lens are insolubilization of gamma and alpha crystallins. The major PTMs which lead to significant interstrain difference between OXYS and Wistar lenses, and, presumably, have cataract-specific character, are determined.

Acknowledgements: We appreciate: RFBR projects 11-04-00143, 11-0300296, FASI state contract 14.740.11.0758 and grant № 11.G34.31.0045, grant NSh-2429.2012.3, RAS № 21.13, CCU.

LOOKING FOR MEANINGFUL SIGNS: THE EXPERIENCE WITH COMPARATIVE ANALYSIS OF NON-ALIGNED PROTEIN SEQUENCES

Sobolev B.*[1], Oparina N.J.[1, 2], Veselovsky A.[1], Filimonov D.[1], Poroikov V. [1]
[1] Orekhovich Institute of Biomedical Chemistry of the Russian Academy of Medical Sciences, Moscow, Russia;
[2] Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, Moscow, Russia
e-mails: borissobolev-5@yandex.ru
** Corresponding author*

Motivation and aim Large protein families frequently contain evolutionary related but functionally divergent proteins. The correct identification of amino acid positions, which can be used to discriminate functional groups within the family (Group-Specific Positions, GSP) is one of the most actual problems of Bioinformatics. Numerous approaches were developed and applied for comparative analysis of proteins. All of them require multiple alignment of amino acid sequences.

Methods and algorithms We propose a new fundamentally different approach suitable for GSP selection in non-aligned proteins. We implement the local similarities estimation, classification based on the training set and calculation of the probabilities for each region of the analyzed protein to be specificity determining.

Results We have shown the applicability of developed algorithm for detection of various types inter-subfamilies functional variations, including even a weak signals associated with protein-nucleic acid and protein-protein interaction.

Conclusion We have shown the applicability of alignment-free approach for protein families analysis and group-specific positions detection.

Availability The developed software is freely available as a web service at http://195.178.207.160/spros/.

THE VESICLE BUILDER – A PLUGIN FOR THE CELLmicrocosmos 2 MembraneEditor

Sommer B.*, Yan Zhou
Bio-/Medical Informatics Department, Bielefeld, Germany
e-mail: bjoern@CELLmicrocosmos.org
** Corresponding author*

Key words: *Membrane Modeling, CELLmicrocosmos2 MembraneEditor, 3D Membrane Packing Problems, Vesicle Generation*

Motivation and Aim: The CELLmicrocosmos 2.2 MembraneEditor (CmME) is a Java WebStart program to solve heterogeneous, two-and-a-half-dimensional Membrane Packing Problems [1]. It imports and exports lipids and proteins based on the PDB format [2]. Originally, CmME was developed for generating rectangular membranes which are also often used in conjunction with molecular simulations. A demanding task is the generation of lipid-based spheres or vesicles.

Methods and Algorithms: Here, the first version of a new CmME-plugin is presented, enabling the solution of three-dimensional Lipid Packing Problems. An algorithm as implemented which enables the generation of ellipsoid vesicular mono-layer or bilayer membranes.

Results: The generation of ellipsoid vesicles is now a simple task by using the new *Vesicle Builder*.

Conclusion: The new methodology extends the capabilities of the MembraneEditor to solve now additionally three-dimensional Lipid Packing Problems. The next steps will be to extend the framework of CmME to support its well-known features also for vesicles, for example: the definition of microdomains, the automatic computation of lipid densities and the semi-automatic placement of proteins.

Availability: http://Cm2.CELLmicrocosmos.org

Acknowledgements: This work was supported in part by: DFG Graduate College for Bioinformatics (GK635).

References:

1. B. Sommer et al. (2011) CELLmicrocosmos 2.2 MembraneEditor: A Modular Interactive Shape-Based Software Approach To Solve Heterogeneous Membrane Packing Problems. *Journal of Chemical Information and Modeling,* **5**:1165–1182.
2. H.M. Berman et al. (2000) The Protein Data Bank. *Nucleic Acids Research,* **28**:235–242.

NOVEL APPROACH FOR IDENTIFICATION OF DNA-BINDING PROTEINS OF BLOOD CELL SURFACE

Tamkovich S.N.*[1,2], Duzhak T.G. [3], Starikov A.V.[4], Vlassov V.V.[1], Laktionov P.P. [1]
[1] *Institute of Chemical Biology and Fundamental Medicine SB RAS, Novosibirsk, Russia;*
[2] *Novosibirsk State University, Novosibirsk, Russia;*
[3] *International Tomografic Center SB RAS, Novosibirsk, Russia;*
[4] *Novosibirsk Regional Oncology Dispensary, Novosibirsk, Russia*
e-mail: s.tamk@niboch.nsc.ru
** Corresponding author*

Key words: *cell-surface-bound DNA, DNA-binding proteins, blood cells, proteomics, breast cancer*

Motivation and Aim: It was shown, that endogenous extracellular DNA (cirDNA) circulate in blood and other biological fluids as free DNA, in the complexes with biopolymers (nucleosomes, etc.) and being bound with surface of blood cells. Cell-surface bound DNA may be composed of free cirDNA molecules or supramolecular complexes of cirDNA. As far as circulation biological effects and elimination of cirDNA from blood are mediated by their interaction with cells, binding of cirDNA with cellular surface is of enormous importance. Earlier we found that in blood of healthy donors more then 90% cirDNA are bound to the surface of blood cells (csbDNA), whereas in blood of breast cancer patients less then 10% cirDNA are bound to cell surface [1]. Redistribution of cirDNA in blood of breast cancer patients allows to execute a comparative study of circulating nucleoprotein complexes bound with surface of blood cells in healthy women and BC patients.

Methods and Algorithms: The proteins mediating cirDNA binding with cell surface were identified by mass-spectroscopy after isolation of nucleoprotein complexes eluted from blood cell surface by DNA-affinity chromatography. In the pilot experiments the following proteins were identified: C3 complement component, haptoglobin, fibrinogen, serum albumin, apolipoprotein and 59,5 kDa histidine-riched glycoprotein.

Results and conclusion: Identification of proteins from circulated nucleoprotein complexes will allow to reveal the characteristic features of generation and circulation cirDNA in blood of breast cancer patients. The data obtain are demanded for isolation of tumor-specific DNA, identification of the mechanisms of binding and penetration of cirDNA into cells, provide a new data regarding molecular pathology of breast cancer and could potentially help to identified new protein markers of breast cancer.

References:

1. S.N. Tamkovich et al. (2005) Level of extracellular nucleic acids bound with blood cell surface in diagnostics of breast cancer, *Mol. Med. (Russian),* **2:** 46-50.

SYSTEM ANALYSIS OF HUMAN CELL LINE: TRANSCRIPTOME, PROTEOME

Zgoda V.G.*, Tikhonova O., Novikova S., Kurbatov L., Kopylov A., Moskalyova N., Archakov A.I.
Orekhovich Institute of Biomedical Chemistry, Russian Academy of Medical Sciences
e-mail: vic@ibmh.msk.su
** Corresponding author*

Key words: *human leukemia 60 cell line (HL60), retinoic acid, transcriptome, microarray, proteome, mass-spectrometry*

Motivation and Aim: A differentiation of HL60 human promyelocyte leukemia cell line after addition of retinoic acid was used as a model for systems pathway analysis of transcriptome and proteome expression data.

Methods and Algorithms: HL60 transcriptome in several endpoints of cell differentiation was estimated using full human genome microarray platform by Agilent. Proteomes were analyzed using LC-MS/MS platform (LTQ-Orbitrap, Thermo). The results were analyzed to find potential node regulators, e.g., transcription factors responsible for cell differentiation with GeneXplain software platform (www.genexplain.com).

Results: Full genome transcriptomics showed significant differences in expression of more than 1500 genes after treatment of HL60 cell line by retinoic acid. Of them, after differentiation launch, 134, 207, 364, 393 and 1197 genes changed their expression at least 2-fold upon 30 min, 60 min, 3 h, 24 h and 96 h, respectively.

By proteomics with LC-MS/MS, about 1370 proteins were identified in HL60 cell line. Label-free quantitation allows detection of differential expression for 65, 103 and 331 proteins upon 3, 24 and 96 h after retinoic acid addition, respectively.

Data from transcriptomics and proteomics were united and considered using GeneXplain to find pathways and transcription factors responsible for HL60 differentiation under retinoic acid.

Conclusion: GeneXplain platform allows at least general explanation of cell processes during HL60 differentiation using transcriptome and proteome data.

Acknowledgement: The work was funded by Russian Academy of Medical Sciences.

Human genetics and personalized medicine: theoretical and experimental aspects

THE REVIEW OF EXISTING SERVICES IN THE FIELD OF PERSONALIZED GENETICS

Anashkin S.S.
Company executive, "Genoanalytica" JSC, Moscow, Russia
e-mail: sa@genoanalytica.ru

The demand for study of personal DNA in its infancy. Most of the companies involved in the promotion of DTCGT, rely on the curiosity of consumers and social networks on the Internet. Physicians are traditionally more conservative in everything new, they know not very much about genetics. They have no experience in the application of the results for research DNA in their clinical practice.

Market DTCGT is being born for now. At this stage it is important to assess the willingness of society to consume directly, bypassing a doctor. On the other hand, the question is whether this expediency?

With the use of PCR methods. It's almost all the major network laboratories Center for Molecular Genetics. The products of these laboratories in different combinations and with different names are in the range of services many commercial hospitals. Available analyzes of polymorphisms of 1 to several, batch studies on core areas: cardiovascular, oncological, endocrinology, neurodegenerative, autosomal. Prices range from 150 rubles for the polymorphism of up to tens of thousands of batch studies. Some medical centers integrate such research into their own programs "check-up" diagnostics.

Microarrays. There was about 3 years ago, which allowed to talk about that a few years ago, VS Baranov called the "genetic passport". Screening study on several thousands of SNP, which is the main feature of the method of chip, as a rule, artificially divided into the following packages:

-Monogenic disease (carrier status)
-multifactorial (polygenic) diseases (predisposition)
-Pharmacogenetics (reaction to medications)
-physical characteristics/specifications (+ sports genetics)

The cost of such "wholesale" method of research (thousands of SNP at once) as a result is cheaper PCR method.

However, managers of laboratories that perform conventional PCR assays, insist that the patient does not need it all at once, the entire screening. Consequently, consumers do not need to pay from 30 000 to 75 000rub - a range of cost studies of DNA on the microarray. From this we can agree, provided the consumer-patient, but rather, his doctor knows exactly - what specific SNP should be analyzed. In all other cases it is advisable to conduct a one-time screening of the maximum number of SNP.

We should also provide a completely new high-tech area - sequencing of individual exons or entire genes, or even a few genes at once. And although so far the cost of such studies is still high, but there are technological solutions that able significantly affect the price. In this regard, very soon, probably, it will be possible to note the appearance of the new market of DNA testings, which will make adjustments to the current structure of the market.

What motivates consumers of DTCGT:

- Cognitive, curiosity (fun)
- Responsible for managing health
- Planning for pregnancy, IVF (carrier status of monogenic diseases)
- Prescription of doctor to confirm or refute the diagnosis, or for adjustments of treatment
- Common sense (the possession of the knowledge available today to help with good reason and consciously choose a lifestyle that is harmonious own genes)

About genetic discrimination (in Russia there is no law); on ethical issues.

MUTATIONS IN *K-Ras* AND *EGFR GENES* AND THE SEARCH FOR SNPs, ASSOCIATED WITH THEIR OCCURRENCE

Antontseva E.V.*[1], Bryzgalov L.O.[1], Matveeva M.Yu.[1], Ponomaryova A.A.[2], Ivanova A.A.[2], Rykova E.Y.[3], Cherdyntseva N.V.[2], Merkulova T.I.[1]
[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia;*
[2] *Cancer Research Institute SB RAMS, Tomsk, Russia;*
[3] *Institute of Chemical Biology and Fundamental Medicine SB RAS, Novosibirsk, Russia*
e-mail: schly@mail.ru
** Corresponding author*

Key words: *K-Ras, EGFR, somatic mutations, single nucleotide polymorphism (SNP) , lung cancer*

Motivation and Aim: The purposes of this research were to obtain data on the frequency of somatic mutations in *EGFR* and *K-Ras* genes and inherited SNPs in *K-Ras* among patients with non-small cell lung cancer (NSCLC) in the Western Siberian region of Russia, the development of minimal invasive diagnostics for these mutations by the use of extracellular DNA, analysis of potential rSNP in *K-Ras* gene, and the search for new rSNP, which may be relevant to the development of NSCLC.

Methods and Algorithms: DNA for genotyping was isolated from 40 NSCLC tumors, 14 of them identified as adenocarcinoma (AC), and 26 as squamous cell carcinoma (SCC) Detection of the somatic mutation Leu858Arg in *EGFR*, was carried out by allele-specific RT-PCR with SybrGreen. Identification of 9-18 bp microdeletion of *EGFR* gene (the region of 746-750 amino acids of the protein) was performed by PCR with further separation by gel electrophoresis in agarose gel. Mutations at the codon 12 in *K-Ras* gene was determined by sequencing after enrichment of DNA template with mutant allele. A search for potential rSNPs in *K-Ras* intron 2 was carried out by sequencing of patient's DNA to identify markers of genetic susceptibility to lung cancer. The electrophoretic mobility shift assay was used to study the predicted rSNPs.

Results: Mutations in the *EGFR* gene were found in 10 % of ACs and in 7 % of SCCs. Mutations at codon 12 in *K-Ras* were identified in 29 % of ACs and in 8 % of SCCs. Tumors with simultaneous presence of mutations in both genes were not detected in this sample. We identified two SNPs in intron 2 in *K-Ras*, affecting the binding sites of transcription factors (TF) NF-Y and GATA-6 and characterized by an increased frequency of occurrence (3.3-3.5-fold) relative to the random sample.

Conclusion: Despite the low volume of the studied sample, we can make a preliminary conclusion that in patients with NSCLC of the West Siberian region, the frequency of mutations studied do not differ from the average, and cases of joint detection of two mutations are very rare. We have identified two rSNPs in intron 2 of *K-Ras* that affect binding sites for the same TF as in the case of mice [1] in which similar SNPs have a *cis*- effect on the formation of mutations at codon 12 K-ras.

References:

1. E.V. Gorshkova et al. (2006) Lung cancer-associated SNP at the beginning of mouse K-ras gene intron 2 is essential for transcription factor binding, *Bull Exp Biol Med,* **6:** 731-733.

CIRCULATING microRNAs AS POTENTIAL BIOMARKERS OF LUNG CANCER

Aushev V.N.*[1, 2], Akselrod M.E.[1], Zborovskaya I.B.[1], Krutovskikh V.A. [2]
[1] *Laboratory of Oncogenes Regulation – Institute of Carcinogenesis of N.N. Blokhin Russian Oncological Science Center RAMS – Moscow, Russia;*
[2] *Epigenetics group – The International Agency for Research on Cancer – Lyon, France*
e-mail: vaushev@gmail.com
** Corresponding author*

Key words: *microRNAs, biomarkers, cancer*

Motivation and Aim: The work was devoted to the research of microRNAs as potential cancer biomarkers in blood. We supposed to investigate the limited and possibly homogenous panel of samples from patients with squamous cell carcinoma. In order to further decrease individual variations, design of our study was based on the comparison of miRNA expression profiles of the same patient before and after removal of tumour.

Methods and Algorithms: Patients diagnosed with squamous cell carcinoma underwent surgical resection of the tumour, and blood samples were taken at 2 points: 1 day before surgery, and 7-10 days after. Blood samples were taken in EDTA-containing tubes and centrifuged to separate plasma fraction. RNA was isolated using "NucleoSpin miRNA Plasma" columns from Macherey-Nagel (Germany).

First step included global analysis of a wide set of microRNAs. This initial list of candidate miRNAs (90 miRNA species) was composed based on known data from literature: we included all miRNAs that were reported as possible circulating biomarkers of non-small cell lung cancer, most of miRNAs assigned for other types of lung cancer, and part of miRNAs reported for non-lung cancers. MiRNA profiling was performed by qPCR analysis on custom 384-well microRNA PCR panels from Exiqon (Denmark). Most promising candidates were further validated by qPCR with individual miRNA probes on wider set of samples.

Results: Most of miRNAs from the designed list were successfully detected (with Ct values between 18 and 35) in almost all samples. Highest signal was detected for miR-451, -223, -15a, -486-5p, -16, -21, which is in agreement with literature data. Small part of miRNAs (including miR-206, -518b, -422a, -202, -566) could not be detected and probably are not presented in plasma. We selected miRNAs which displayed the most significant difference between pre- and post-operative samples. Namely, expression of miR-205, -19a, -19b, -451 and -30b was decreased after removal of tumour. Part of these miRNAs were further validated using wider panel of samples. Results of the validation confirm indicated changes of miRNA expression. Additional experiments revealed that most of detected miRNAs are presented in exosomes-enriched fraction of plasma.

Conclusion: Our results suggest that tumour-related miRNAs can be successfully detected in plasma of lung cancer patients and can thus serve as potential biomarkers for this disease.

Acknowledgements: Authors are grateful to colleagues in Oncological Science Center for their assistance in blood samples collection.

THE COMET-FISH TECHNIQUE FOR MONITORING CANCER TREATMENT RESPONSE AT THE GENOMIC LEVEL

Babayan N.S.*, Gevorkyan A.L., Aroutiounian R.M., Hovhannisyan G.G.
Yerevan State University, Yerevan, Armenia
e-mail: babayannelly@yahoo.com
** Corresponding author*

Key words: *personalized medicine, Comet-FISH, telomere, peptide nucleic acid, cancer, cisplatin, bleomycin*

Motivation and Aim. Nowadays, new molecular methods and techniques are developing which can contribute to the objective of personalized medicine [1]. Although telomerase and telomeres have become an attractive therapeutic cancer targets, since most human cancer cells (85-90%) typically express high levels of telomerase [2]. The aim of the present study is the development of a method for comparative investigation of action of widely applied anticancer preparations: cisplatin (cis-DDP) and bleomycin (BLM) on total DNA and telomeres in human blood cells.

Methods. The "Comet-FISH technique" - single cell gel electrophoresis ("comet assay") in combination with fluorescent *in situ* hybridization (FISH) was used for this purpose. This newly applied combined approach permits to detect on the same specimen the total DNA damage in individual cells and evaluate specific DNA sequences as well. Telomere - specific - PNA (peptide nucleic acid) probes were used for the localization of telomeres in the comet's head and their migration to the tail.

Results. By comparing the slopes of the linear regressions of the BLM and the BLM-cis-DDP treatment a slope of 1.07 was found for the BLM treatment alone, which indicates that the telomeres are of the same sensitivity as the average DNA. In contrast the treatment with the combination of BLM and cis-DDP reduces telomere migration more than the migration of total DNA and results in a slope smaller than 1 ($b = 0.77$) and hints to an enhanced cross-linking effect on the telomere sequences. Thus, preferentially telomeric action of the cis-DDP can be concluded.

Conclusion. The presented Comet-FISH approach with telomere PNA permits direct and precise detection of the telomere migration from the former cell nucleus to the comet tail in cells treated with cytostatics, with a direct analysis of correlation to the overall DNA fragmentation. That can be important for monitoring the application of clinical relevant cytostatics during therapy, especially in combinatory approaches, where more than one substance is used at a time.

Acknowledgements. This study was greatly enhanced by the cooperation with prof. E. Gebhart (Institute of Human Genetics, Erlangen, Germany) and prof. A. Rapp (Institute for Molecular Biotechnology, Jena, Germany).

References

1. C.T. Caskey. (2010) Using genetic diagnosis to determine individual therapeutic utility, Annu. Rev. Med., 61: 1–15.
2. A. Rцth *et al*. (2010) Imetelstat (GRN163L) - telomerase-based cancer therapy, In: Small Molecules in Oncology (Recent Results in Cancer Research), U.M. Martens (Edr.), 221-234 (© Springer-Verlag).

A SIMPLE PERSONAL GENOME VIEWER

Bakulina A.*[1], Diakonov A.[1], Zagrivnaya M.[2]
[1] *State Research Center of Virology and Biotechnology "Vector", Koltsovo, Russia;*
[2] *Pavlov State Medical University of St. Petersburg, St. Petersburg, Russia*
e-mail: bakulina@gmail.com
** Corresponding author*

Motivation and Aim: Genetic diagnosis allows to improve the quality of life. It helps to implement more effective treatments of various diseases and to choose the optimal lifestyle. The role of a genetic testing rises with dramatical cost reducing of personal genome sequencing. However, most of people are not familiar with the capabilities of modern methods of genetic testing and therefore can not benefit from them. Development and distribution of software for personal genome analysis can contribute to solving this problem.

Results: We have developed the OhMyGenes software for storage and unprofessional analysis of user's personal genome. It contains the list of genetic polymorphisms with known effects in the human genome and information where they can be tested, and provides view of variety of the user's genetic test results. The user can view polymorphisms and mutations by category, or perform full-text search of effects or genes descriptions. For some multifactor diseases a post-test risk value is calculated. The database of the software package includes the most important polymorphisms, associated with the risk of monogenic and multifactorial diseases drugs metabolism, athletic status.

Conclusion: Our program helps to solve following problems: promotion of genetic testing among population, informing people about important genetic polymorphisms, storage of testing results in a convinient way.

Availability: The demo-version of the software is available for Windows and Android upon request.

The final version will be available as soon as the database of known polymorphisms, mutations and medical centers is completed enough and the server for automatic update is running.

BIOINFORMATICS IN TRANSLATIONAL RESEARCH

Baranova A.V., Chandhoke V.
School of Systems Biology, George Mason University, Fairfax, VA, USA
e-mail: abaranov@gmu.edu, vchandho@gmu.edu

Key words: *biomarkers, prognosis, cancer, whole transcriptome analysis, continuous prognosis model*

Translational bioinformatics hold a number of promises, namely, an ability to derive pathogenic mechanism from the gene interactions, to ascertain effectiveness of prevention strategy for a given genotype, compare effectiveness of treatments, infer patient-specific drug doses and the possibilities for the development of given side effect. One day, these promises may amount to a truly "tailored", individualized treatment. However bright this perspective might be, it is important to remember that clinical data analysis routinely encounters various data biases that substantially influence the accuracy of the derived conclusions. In this presentation, we will sort through typical biases present in the clinical datasets and wil use specific examples how to overcome these hidden perils.

STUDY OF INTERINDIVIDUAL VARIABILITY OF WARFARIN DOSAGE AMONG POPULATION OF THE WESTERN SIBERIAN REGION OF RUSSIA

Belozertseva L.A.*, Voronina E.N., Koh N.V., Cvetovskaya G.A., Lifshits G.I., Filipenko M.L.
Institute of Chemical Biology and Fundamental Medicine SB RAS, Novosibirsk, Russia
e-mail: white_lilia89@mail.ru
** Corresponding author*

Key words: *warfarin, VKORC, CYP2C9*

Motivation and Aim Warfarin is an oral anticoagulant widely used around the world. The molecular target of warfarin is enzyme vitamin K epoxide reductase (VKOR), which reduces the oxidized form of vitamin K to hydroquinone. The reduced vitamin K is cofactor of gamma-glutamyl carboxylase (GGCX), which provides the carboxylation and, thereby, activates coagulation factors II, VII, IX, and X and proteins C, S, and Z. Warfarin inhibits VKOR that slows hemostasis. In the appointment of warfarin, patients show a wide range of interindividual differences in drug doses needed to achieve the desired therapeutic effect. Thus, the aim of our study was to investigate the influence of polymorphic variants of *VKORC1* (-1639 G> A, rs9923231), *CYP2C9* (* 2,* 3), *GGCX* (12970 C> G, rs11676382), *PROC* (2583 A> T, rs1799810), *FVII* (10976 G> A, rs6046), *CYP4F2* (23454 G> A, rs2108622) genes, and non-genetic parameters on the variability of warfarin dose among the population of the West Siberian region of Russia.

Methods and Algorithms The group of patients taking warfarin (N=113) contain patients treated in the CNMT in Novosibirsk. DNA was extracted from venous blood using standard procedures. Determination of genotypes of polymorphic loci was performed by Real-time PCR. Statistical analysis was performed using the program Statistica for Windows, v.8 (StatSoft, Inc.).

Results Warfarin dose significantly varied between carries of different genotypes polymorphic loci of *VKORC1* (-1639 G>A) and *CYP2C9* *3 (p=0,00001 и p=0,018 respectively). The media dose amounted 6,25 mg/day for G/G-carries (N=47), 4,4 mg/day for G/A-carries (N=57), 2,5 mg/day for A/A-carries (N=9) of *VKORC1* -1639 G>A; 5 mg/day for *1/*1-carries (N=101), 3,75 mg/day for *1/*3-carries (N=12). According linear regression results the -1639 G>A *VKORC1* and *CYP2C9*3* variants accounted for 20,7 and 5,7% of the warfarin dosage variability in the patients studied (p=0,000001 и 0,00629 respectively). None of the other polymorphic loci was statistically significantly associated with dose (*CYP2C9*2* (p=0,3579), *GGCX* 12970 C>G (p=0,3745), *PROC* 2583 A>T (p=0,7262), *FVII* 10976 G>A (p=0,4045), *CYP4F2* 23454 G>A (p=0,2844)). According to multiple linear regression results, -1639 G>A *VKORC1* and *CYP2C9*3* together could explain about 22,3 % of the total warfarin dose variation in our group of patients.

Conclusion According to our results, polymorphic variant -1639 G>A in *VKORC1* gene can explain 20,7 %, allele *CYP2C9*3* - 5,7 % of warfarin dose variation among the population of Western Siberian.

CONTRIBUTION OF GENOTYPE VARIATION TO WARFARIN PHARMACOKINETICS

Chernonosov A.A.*, Koval V.V., Kokh N.V., Tsvetovskaya G.A., Lifshits G.I., Fedorova O.S.
Institute of Chemical Biology and Fundamental Medicine SB RAS, Novosibirsk, Russia
e-mail: sandy@niboch.nsc.ru
** Corresponding author*

Key words: *warfarin, pharmacokinetics, mass-spectrometry*

Motivation and Aim. Warfarin is the one of the is one of the most widely prescribed oral anticoagulants worldwide; it is used to prevent and treat venous or arterial thrombi and emboli associated with atrial fibrillation or cardiac valve replacement. For the effective warfarin therapy patients have to take the lowest warfarin dose required to maintain the target international normalized ratio (INR). However, the individual doses required to each patients could be very different. It is known that cytochrome P450 (CYP), mainly CYP2C9, activity is an essential source of dosage variability. In addition, vitamin K epoxide reductase complex subunit 1 gene (VKORC1) plays a major role in the optimization of warfarin dose. The aim of this study was to assess the pharmacokinetics of warfarin in relation to cytochrome P450 (CYP2C9) and VKORC1 genotypes in residents of the West Siberian region.

Methods. The method was applied to determine the plasma concentrations of warfarin from a clinical trial in which 11 healthy volunteers received a single 5 mg oral dose of warfarin. All volunteers gave their signed informed consent to participate in the study. They have been genotyped for CYP2C9*2, CYP2C9*3 alleles and 1173C>T VKORC1 polymorphism. Blood samples were collected before and 1, 2, 3, 8, 24, 48 and 72 h post-dosing. Samples were centrifuged and plasma was separated and stored at −20 °C until analyzed. The concentrations of warfarin in the all plasma samples were analyzed by MS method using mass-spectrometer Agilent 6410 QQQ (Agilent Technologies, USA). Every samples were analyzes 3 times.

Results. All volunteers have *1/*1 genotype CYP2C9. Two of them have CC, five have TC and four have TT genotype VKORC1. For the TT and CT genotypes it was noted that $T_{1/2}$ of the warfarin is about 25% longer in compare with the CC VKORC1 genotype. Area under the curve (AUC) of warfarin concentration increased in the order CC, CT, TT genotypes VKORC1. Ratio of AUC between CC and CT genotypes was 1.65 and between CC and TT genotypes - 2. We have also observed that the clearance was increased in the order TT, CT, CC. It should be noted that this effect is also additive, and that heterozygotes respond to an intermediate warfarin dose, and homozygous carriers of the T allele respond to the lowest dose of warfarin. The results indicated that pharmacodynamic response to warfarin is highly variable between subjects. The association between warfarin concentration in plasma and 1173C>T VKORC1 genotypes suggests that the VKORC1 genotype significantly affects the personal dosage of warfarin.

Acknowledgements. Supports by grants from Ministry of Education and Sciences (no. 16.512.11.2073, NS-64.2012.4).

VALIDATION OF THE *PPP1R12B* AS A CANDIDATE GENE FOR CHILDHOOD ASTHMA SUSCEPTIBILITY

Freidin M.B.*[1], Polonikov A.V.[2]
[1] *Research Institute of Medical Genetics, Tomsk, Russia;*
[2] *Kursk State Medical University, Kursk, Russia*
e-mail: mfreidin@medgenetics.ru
** Corresponding author*

Motivation and Aim: Genome-wide association studies (GWAS) are a powerful tool for revealing positional candidate genes of complex diseases and traits. A number of such the studies have been performed in childhood and adult bronchial asthma (BA), which allowed disclosing dozens of novel candidate genes. However, given the high genetic heterogeneity of BA, a validation of the GWASs in independent populations is of critical importance. Recently, the *PPP1R12B* gene was revealed as a new candidate gene for childhood asthma in GWAS in Russians of West Siberia using Illumina 610QUAD chip (1). We now set out to validate this finding in Russian population of Kursk.

Methods and Algorithms: One hundred and fourteen patients with childhood BA (age-of-onset up to 16 years) and 279 healthy controls were genotyped using same Illumina chip. After quality assessment, total of 26 markers embracing the *PPP1R12B* gene region were analyzed using linear models approach to measure the associations between the disease and markers.

Results: Four markers (rs17438212, rs12734001, rs3767423, and rs3817222) were found to be significantly associated with childhood BA after correction for multiple testing. They represent two relatively distinct linkage disequilibrium blocks with two markers in each. The odds ratios for the associations varied between 1.78 and 2.04 (Padj = 0.029 – 0.03). The direction of association and the magnitude of the effect for the rs12734001 and rs3817222 markers were in accordance with initial finding in Russians of West Siberia, providing the validation of its verity.

Conclusion: Thus, the present data confirmed that the *PPP1R12B* gene is associated with childhood BA in Russians. Further studies are required to provide the functional basis for the association.

References:

1. M.B. Freidin et al. (2011) Genome-wide association study of allergic diseases in Russians of West Siberia, Molecular Biology, 45: 421-429.

IDENTIFICATION OF RARE VARIANTS AND POLYMORPHISMS OF THE *IL12RB1* GENE AND ANALYSIS OF THEIR ASSOCIATIONS WITH TUBERCULOSIS

Garaeva A.F.*, Rudko A.A., Bragina E.Yu., Babushkina N.P., Freidin M.B.
Research Institute of Medical Genetics, SB RAMS, Tomsk, Russia
e-mail: gaf_1986@mail.ru
** Corresponding author*

Key words: *atypical familial mycobacteriosis; IL12RB1, nucleic acids sequencing; rare variants, polymorphisms*

Motivation and Aim: Interleukin-12 acting through specific receptor encoded by the *IL12RB1* gene plays a key role in the immune response to *M. tuberculosis*. Several mutations in the *IL12RB1* gene lead to the development of rare syndrome of atypical familial mycobacteriosis (OMIM # 209950). This suggests that rare variants and polymorphisms of this gene predispose to tuberculosis.

The aim of this study is to find rare variants of the *IL12RB1* gene and evaluate their prevalence in TB patients and healthy Russians of Tomsk region.

Methods and Algorithms: The search for rare variants of the *IL12RB1* gene was carried out by Sanger's method using Applied Biosystems 3130xl Genetic Analyzer in 10 individuals suffered from aggressive forms of TB. The analysis of the results of sequencing was performed using BioEdit and Sequence Scanner v1.0 software. On the next stage of the study the identified nucleotide substitutions were genotyped in 310 patients and 250 healthy individuals using PCR-RFLP analysis. Association analysis was done using the http://ihg2.helmholtz-muenchen.de/ihg/snps.html on-line resource.

Results: Sequencing of intron-exon regions of the gene *IL12RB1* identified seven previously described polymorphisms, including four in exons (synonymous substitutions rs11086087, rs17852635; missense-mutations rs11575934 and rs401502), two in introns (rs12461312, rs17882555), and one in 3'-UTR (rs3746190). Synonymous substitutions were excluded from further analysis. Other variants were genotyped in patients and healthy individuals. The polymorphisms demonstrated no statistically significant differences between the studied groups.

Conclusion: No rare variants in the *IL12RB1* gene in patients with severe forms of tuberculosis were revealed by direct sequencing. The detected known polymorphisms in the *IL12RB1* gene are not associated with tuberculosis.

GENETIC BARCODE AS A PERSONAL IDENTIFIER OF EACH INDIVIDUAL

Garafutdinov R.R.*, Chubukova O.V., Sakhabutdinova A.R., Mashkov O.I., Shakirov I.G., Chemeris A.V.
Institute of Biochemistry and Genetics Ufa Science Centre RAS, Ufa, Russia
e-mail: garafutdinovr@mail.ru
** Corresponding author*

Key words: *DNA identification, single nucleotide polymorphism, tetraallelic SNPs, digital encoding, genetic barcode*

Motivation and Aim: Identification of individuals has become an urgent problem of mankind. Recently the DNA identification of personality is actively evolved along with traditional biometric methods [1, 2]. Previously we have developed an approach to identification of individuals by genetic barcoding based on SNPs. The aims of this work are the creation and comparative analysis of genetic barcodes of specific individuals with different ethnicity.

Results: According to approach proposed each possible pair of polymorphic nucleotides in particular SNP is digitized as a unit and received the graphic symbol. Computer software we have written creates genetic barcode in digital and graphic formats. For DNA identification purpose the genetic barcode of each person must be unique. So it is very important to determine the number of SNPs to be analyzed. The feature of our approach is the use of tetraallelic SNPs, which theoretically provide 10^{24} possible combinations of polymorphic nucleotides. We have selected 24 tetraallelic SNPs from SNP database (http://www.ncbi.nlm.nih.gov/snp) and found polymorphic nucleotides by allele-specific PCR. SNP analysis was performed on genomic DNA of more than 100 people – citizens of Russia ethnicities (Russian, Tatar, Bashkir, Yakut, Buryat, Mordovian, Mari, Udmurt, Komi) followed by genetic barcodes creation. For none of the individuals analyzed the genetic barcode does not resemble the barcode of another person despite the fact that all SNPs turned out biallelic. However it is possible for relatives analysis of 24 SNP will be insufficient to provide a uniqueness of genetic barcodes. Therefore we will increase the number of tetraallelic SNPs analyzed up to 48. Earlier it was shown at least 50 biallelic SNPs should be taken for DNA identification [3].

Conclusion: So we have demonstrated the possibility of proposed approach to DNA identification by genetic barcoding based on SNPs. It was found tetraallelic SNPs can provide the uniqueness of genetic barcodes which can be used as a personal identifier of each individual.

Acknowledgements: This work was supported by the Ministry of Education and Science of Russian Federation (contracts No. 14.740.11.1059, 16.518.11.7047).

References:

1. M.A.Jobling, P.Gill. (2004) Encoded evidence: DNA in forensic analysis, *Nat Rev Genet,* 5(**10**): 739-51.
2. R.A.Oorschot, K.N.Ballantyne, R.J.Mitchell. (2010) Forensic trace DNA: a review, *Investig Genet*, 1(**1**): 14.
3. P.Gill. (2001) An assessment of the utility of single nucleotide polymorphisms (SNPs) for forensic purposes, *Int J Legal Med*, 114(**4-5**): 204-10.

SYSTEMS BIOLOGY ANALYSIS OF COMPLEX DISORDERS

Gilliam C.*[1], Balasubramanian S.[1], Xie B.Q.[1], Sulakhe D.[1], Berrocal E.[1], Maltsev N.[1], Boernigen-Nitsch D.[2], Chitturi C.[3], Paciorkowski A.[4], Dobyns W.[4]
[1] *The University of Chicago, IL, USA;*
[2] *Harvard School of Public Health, Boston, MA, USA;*
[3] *Amrita University, Amritapuri, Kerala, India;*
[4] *University of Washington, Seattle, WA, USA*
e-mail: cgilliam@bsd.uchicago.edu
** Corresponding author*

Key words: *translational medicine, biological networks, gene prioritization*

We present a bioinformatics approach and supporting computational platform - GEDI (http://gedi.ci.uchicago.edu/) - for systems-level analysis of complex heritable disorders such as autism, schizophrenia and diabetes. Our approach is based on a large-scale integration of genomic and clinical data provided by our collaborators, as well as various classes of biological information from over 35 public databases and private collections. This data is used for identification of genes and molecular networks contributing to the phenotypes of interest, as well as for the prediction of additional high-confidence disease genes to be tested experimentally. Our analytical strategy is three-fold and includes (a) the enrichment analysis of high-throughput genomic data (e.g. the results of GWAS and CNV analysis), (b) feature-based gene prioritization and (3) the development of the networks-based disease models for identification of molecular mechanisms involved in pathogenesis of disease of interest. Bayes factor and P-value estimate are used for enrichment analysis; support vector machine algorithm for feature-based prioritization of the candidate genes. Networks-based gene prioritization leverages our previous work [1] and utilizes Heat Kernel diffusion, Random Walk, PageRank with priors, HITS with priors and K-step Markov model algorithms. These algorithms were modified to accommodate variety of weighted data types to be used for gene prioritization (e.g. ranked gene to phenotype associations, weighted canonical pathways data). We will illustrate our approach using analysis of brain connectivity disorders (e.g. agenesis of corpus callosum, autism, schizophrenia) as an example. Our analysis allowed uncovering some of the common molecular mechanisms that underlie these disorders. This knowledge will eventually lead to the development of efficient diagnostic and therapeutic strategies.

[1] Nitsch D, et al. PINTA: a web server for network-based gene prioritization from expression data. Nucleic Acids Res. 2011 Jul;39(Web Server issue):W334-8. Epub 2011 May 20. PubMed PMID: 21602267; PubMed Central PMCID: PMC3125740.

EXPERIENCE OF THE PERSONALIZED ANTIPLATELET THERAPY: THE EFFECTS OF *CYP2C19* GENE

Knauer N.Yu.*, Voronina E.N., Lifshits G.I.
Novosibirsk State University, Novosibirsk, Russia;
Institute of Chemical Biology and Fundamental Medicine SB RAS, Novosibirsk, Russia
e-mail: knauern@mail.ru
** Corresponding author*

Key words: *clopidogrel, personalized therapy, CYP2C19*

Motivation and Aim: The clopidogrel-based prevention of thromboses and rethromboses is the important aspect of cardiological practice. However, clinical and laboratorial response on clopidogrel of different patients is known to differ from the predicted one thus stipulating the use of personalized approach for search of the optimal dose. One of the principal instruments of this approach is the revelation of the allelic variants of genes influencing the drug metabolism. Particularly, the protein CYP2C19 of the cytochrome P450 family was shown to play the important role in the clopidogrel metabolism, with its activity depending of the allelic variants of the gene.

The aim of this work was to examine frequencies of occurrence of CYP2C19 allelic variants *1, *2, *3 among the patients receiving clopidogrel by medical indications (n=158) and to determine their contribution to the clopidogrel laboratorial efficacy

Methods and Algorithms: The assessment of the laboratorial response on clopidogrel was conducted by light transmission aggregometry using ADP (20 μM). The identification of the CYP2C19 allelic variants was performed by real-time PCR-HRM assay using specific primers and also by RFLP assay. The results were compared with the aggregometry data to elucidate the effect of polymorfisms on the platelet aggregation after the clopidogrel taking.

Results: Depending on the platelet aggregation change after clopidogrel taking, the following groups were assigned: responders (58.9%), semi-responders (20.9%) and non-responders (7.6%). In addition, the group of patients (12.7%) was assigned, with the platelet aggregation increasing up to 3.5 times after the taking of clopidogrel. According to the data of CYP2C19 allelic variants identification, the whole group of patients contained allelic variants CYP2C19*1 (74.7 %), CYP2C19*2 (24.7%), CYP2C19*3 (0.6%). The response on clopidogrel was shown to correlate with the allelic variant CYP2C19*2 and not to correlate with the variant CYP2C19*3.

Conclusion: The results obtained are of the practical interest and can be used for the optimization of the clopidogrel-based antiplatelet therapy.

GENETICS AND DISEASE PROGRESSION OF FAMILIAL MULTIPLE SCLEROSIS IN NOVOSIBIRSK REGION OF RUSSIA

Korobko D.S.*[1], Malkova N.A.[2], Kudryavtseva E.A.[3, 4], Filipenko M.L. [3]
[1] State Novosibirsk Regional Clinical Hospital, Regional MS Centre; [2] Novosibirsk State Medical University, [3] Institute of Chemical Biology and Fundamental Medicine, Siberian Division, Russian Academy of Sciences; [4] Novosibirsk State University, Novosibirsk, Russia
e-mail: denn007@ngs.ru
** Corresponding author*

Key words: *multiple sclerosis, familial cases, single nucleotide polymorphisms, TNFα, recurrence risk*

Motivation: Multiple sclerosis (MS) is a chronic disorder of the central nervous system, characterized by inflammation, demyelination, development of plaque lesions and episodes of neurologic dysfunction that frequently leads to progressive degeneration [1]. MS is a multifactorial disease in which both genetic and environmental factors intervene. In some studies it is observed that familial multiple sclerosis has more benign course in compared to sporadic cases. For neurologists it is important to predict course of such heterogeneous disease as MS.

Aim: To investigate clinical and genetic features of familial MS.

Materials and methods: It is ongoing retrospective-prospective study. Up to date 248 patients with MS according to the diagnostic criteria of McDonald et al. (2005) were recruited to study in Novosibirsk Regional MS Center. Single nucleotide polymorphisms (SNPs) in *CD40* (rs6074022, rs1883832, rs1535045, rs11086998), *TNF-a* (rs1800629) and *TNFRSF1A* (rs4149584) and rs3135388 (genetic marker of allele HLA-DRB1*15) were genotyped by real time PCR, using competing TaqMan probes. The control group (n=567) comprised of people without inflammatory CNS disease living in Novosibirsk. The study was approved by local ethics committees and all participants signed informed consent forms. To assess disease progression MS severity score (MSSS) [2] and rate of progression [3] was calculated. MSSS was determined in 238 patients with disease duration > 1 year.

Results: 17 familial cases with relapsing-remitting MS who belonged to 15 different families were identified. Recurrence risk in families of this group was 6,0 %, total recurrence risk was calculated (6,9 %). Age at MS onset in familial cases was lower than in sporadic (23,03 ± 9,22 vs 27,28 ± 8,59, p = 0,09). The rate of disease progression and MSSS within family group was significantly lower (0,33 vs 0,69, p < 0,001; MSSS: 3,22 vs 4,4, pANOVA = 0,05). The association between SNP rs3135388 and development of MS was found (OR = 3.04, 95 %CI 2.33-3.96, p = $1.7*10^{-17}$), but this polymorphism has not influence on disease progression. The association between *GA*, *AA* TNF*a* genotypes and higher average annual relapse rate was revealed (OR 6,07, 95 %CI 1,19-30,9, p = 0.032). These genotypes were more frequent (23 % vs 10 %) in subjects without family history of MS.

Conclusion: This study of Siberian cohort confirmed more benign course of familial MS. The further research is needed to understanding the genetic basis of susceptibility in MS.

References:

1. Compston A, Coles A. Multiple sclerosis. Lancet. 2008 Oct 25; 372(9648):1502-17.
2. Roxburgh RHSR, Seaman SR, Masterman T. et al. Multiple Sclerosis Severity Score: Using disability and disease duration to rate disease severity Neurology April 12, 2005 64:1144-1151
3. Verjans E., Theys P., Delmotte P. et al. Clinical parameters and intrathecal IgG synthesis as prognostic features in multiple sclerosis. Part I. J Neurol. 1983; 229(3): 155-65.

MS/MS ANALYSIS OF METABOLIC DISORDERS

Koval V.V.*, Alekseeva I.V., Chernonosov A.A., Fedorova O.S.
Institute of Chemical Biology and Fundamental Medicine, SB RAS, Novosibirsk, Russia
e-mail: koval@niboch.nsc.ru
** Corresponding author*

Key words: *metabolic disorders, tandem mass-spectrometry*

Motivation and Aim. Metabolic disorders are typically accompanied by the accumulation of corresponding metabolites in blood. Using modern tandem mass spectrometry (MS/MS) more than 20 inherited metabolic disorders can be detected simultaneously from a single blood spot by quantifying concentrations of up to 50 metabolites. Many disorders of fatty acid oxidation, organic acid metabolism, and amino acid metabolism can be detectable by use of MS/MS. The use of MS/MS to detect metabolic disorders in newborns is one of the most important advancements in neonatal screening. In spite of errors of metabolism are a rare group of genetic disorders (0.001–0.01 %) they can produce serious clinical consequences. Early medical intervention in many cases provides the only possibility to avoid physical and mental retardation, or even death.

Methods. The sample preparation utilizes ultrasonic extraction of the blood spot from paper (Whatman 903, Whatman Inc., NJ, USA) in methanol. The extraction is followed by butyl-esterification of the free carboxyl groups of amino acids (requiring hydrochloric acid plus butanol). Esterification converts zwitterionic amino acids to amino esters similar to simple amines, dramatically enhancing the ionization efficiency in triple-quadrupole mass spectrometers Agilent 6310 Triple Quard LC/MS (Agilent Technologies, USA). Most amino acid butyl esters show a characteristic loss of neutral butyl-formate (102 Da) in collision-induced dissociation (CID). Monitoring this process by the constant neutral loss scan technique allows simultaneous detection of these compounds.

Results. The blood-spot samples from 255 patients of Novosibirsk clinics, including inborn, were analyzed. Phenylketonuria (PKU) was confirmed in 28 cases. Besides, 18 cases of urea cycle disorders, 104 cases of short- and medium-chain fatty acid catabolism defects, 12 cases of homocysteine/methionine synthesis defects were detected. Often the diseases have complex character and represent the superposition of few disorders.

Acknowledgements. Supports by grants from SB RAS (no. 5), RAS "Fundamental sciences to medicine" (no. 5.6 and Ministry of Education and Sciences (no. 16.512.11.2073).

IS THE ASSOCIATION BETWEEN -308G->A TNF-α AND MULTIPLE SCLEROSIS INDEPENDENT OF HLA-DRB1*15?

Kudryavtseva E.A.*, Rozhdestvenskii A.S., Kakulya A.V., Khanokh E.V., Malkova N.A., Korobko D.S., Platonov F.A., Aref'eva E.G., Zagorskaya N.N., Alifirova V.M., Titova M.A., Smagina I.V., El′chaninova S.A., Zolovkina A.G., Puzyrev V.P., Tsareva E.Y., Favorova O.O., Boiko A.N., Filipenko M.L.
Institute of Chemical Biology and Fundamental Medicine SB RAS, Novosibisk;
Novosibirsk State University, Novosibirsk; Omsk State Medical Academy, Omsk;
Novosibirsk Oblast State Clinical Hospital, Novosibirsk;
Republican Hospital No. 2, Ministry of Health, Sakha Republic, Yakutsk;
State Health Facility Kemerovo Oblast Clinical Hospital, Kemerovo;
Siberian State Medical University, Tomsk;
Research Institute of Medical Genetics, SD, RAMS, Tomsk;
Altai State Medical University, Barnaul; Moscow center of multiple sclerosis on the basis of hospital № 11
e-mail: kudryavtseva_ekaterina@ngs.ru
** Corresponding author*

Key words: *multiple sclerosis, TNFa, HLA-DRB1*

Motivation and Aim The tumor necrosis factor α (TNF-α) is known as proinflammatory cytokine implicated in the pathogenesis of autoimmune and infectious diseases. The gene of TNF-α are located in the region HLA class III. TheTNF-α promoter polymorphism-308 G->A in the human has been most often used in the associated studies. In view of gene location, it has been speculated that polymorphism of this locus might contribute to HLA association with multiple sclerosis (MS). The aim of this study is an assessment of chosen factors of genetic susceptibility to MS such as HLA-DRB1 and TNF-α polymorphism-308(G/A) in residents RF. More over we would like to estimate whether possible relationship of TNF-α-308(G/A) is due to a primary association or mediated by LD with the susceptible DRB1 alleles.

Methods and Algorithms A group of 1650 MS patients and 992 healthy peoples were randomly selected as the subject of this study. Determination of genotypes of TNF-α-308(G/A) was performed by Real-time PCR. Determination of genotypes of HLA-DRB1 was performed by sequencing. Statistical analysis was performed using the R-language.

Results The G allele of -308G->A TNFa is associated with MS (OR=1.34 [1.10-1.64] p= 0.004). A complete meta-analysis of all analogous studies published to date showed that the G allele is a risk (OR=1.29, 95%C.I.=[1.13-1.46], $p=9*10^{-5}$). The HLA-DRB1*15 is associated with MS (OR=2.40, 95%C.I.=[1.60-3.61], $p=3*10^{-5}$). The -308G->A TNFa and HLA-DRB1*15 are in linkage disequilibrium (D'=.78, $r^2=0.02$, $\chi^2=5.53$). Using LRT we have shown the HLA-DRB1*15 is key risk factor between -308G->A TNFa and HLA-DRB1*15.

Conclusion Association -308G->A TNFA and MS is induced by *HLA-DRB1*15.*

ASSOCIATION OF *ITGB3* AND *GNB3* VARIANTS WITH THE DEVELOPMENT OF VASCULAR COMPLICATIONS IN PATIENTS WITH ACUTE CORONARY SYNDROME

Kulish E.V.*[1], Makeeva O.A.[1, 2], Golubenko M.V. [1, 2], Zykov M.V. [2], Kashtalap V.V.[2]
[1] *Research Institute of Medical Genetics, SB RAMS, Tomsk, Russia;*
[2] *Research Institute of Complex Problems of Cardiovascular Diseases, SB RAMS, Kemerovo, Russia*
e-mail: elena-kulish@medgenetics.ru
** Corresponding author*

Key words: *acute coronary syndrome, single nucleotide polymorphism, arrayed primer extension-based genotyping method, DNA microarray*

Motivation and Aim: Acute coronary syndrome is a common disorder and is a significant cause of morbidity and mortality worldwide.

The aim of the study was to examine association of genetic polymorphisms in *ACE, AGT, AGTR1, CEPT, EDN1, F2, F5, GNB3, ITGB3, LIPC, LPL, MTHFR, NOS3, PON1, PPARG, TCF7L2, MTND2, TNF, CDKN2A/B* with cardiovascular complications in patients with acute coronary syndrome (n=171) in one year observational period after acute event.

Methods and Algorithms: The choice of SNPs was based on results of association studies, analysis of electronic databases, most of markers were confirmed on local population. Arrayed primer extension based genotyping with Genorama™ Imaging System (Asper Biotech Ltd) was used for SNP detection.

Results: A comparative analysis of allele frequencies in patients with acute coronary syndrome with and without vascular complications showed statistically significant higher *ITGB3* gene (rs5918) allele C frequency (p=0.044). Differences in genotype distribution of *ITGB3* (rs5918) and *GNB3* (rs5443)had also been detected in patients with and without vascular complications(p=0,007 and p=0,042, respectively).

Conclusion: ITGB3 (rs5918) and *GNB3* (rs5443) may play a role in development of vascular complications in patients with acute coronary syndrome.

GENETIC SUSCEPTIBILITY PROFILE FOR COMORBIDITY VARIANTS OF MULTIFACTORIAL DISEASES

Puzyrev V.P.*[1], Makeeva O.A.[1], Barbarash O.L.[2], Sleptcov A.A.[1], Markova V.V.[1], Polovkova O.G.[1]

[1] *Research Institute of Medical Genetics SB RAMS, Tomsk, Russia;*
[2] *Research Institute of Complex Issues of Cardiovascular Diseases SB RAMS, Kemerovo, Russia;*
[3] *Genoanalytica, LLC.*
e-mail: valery.puzyrev@medgenetics.ru
**Corresponding author*

Key words: *multifactorial diseases, genetic susceptibility, cardiovascular continuum, syntropy.*

Motivation and Aim: The "Cardiovascular Continuum" (CVC) was described in 1991 by Dzau and Braunwald to explain progression of coronary heart disease through other complications and diseases to inevitable end stage of heart failure. Concept of CVC includes several diseases such as coronary artery disease (CAD), arterial hypertension (AH), metabolic syndrome, and diabetes mellitus type 2 (DM2) (Dzau, Braunwald, 1991). In 1921 Pfaundler and von Seht used the term 'syntropy' to designate diseases, which tend to co-occur with each other in patients (or in families) more often than it could be expected by chance (Pfaundler, von Seht, 1921). Based on these two concepts, the term "syntropy genes" (SG) was proposed to designate a set of functionally interacting, co-regulated genes involved in common biochemical and physiological pathways leading to syntopy (Puzyrev, 2008). Thus, the aim was to explore the genetic profile of CVC and to identify SG.

Methods and Algorithms: A large sample of patients with ischemic heart disease and population sample were analyzed to select subgroups for present study. A total of 309 patients out of 800 were selected according to the following criteria: "syntropy" subgroup diagnosed with CAD, AH, DM2, and dyslipidemia in each patient (N=68); comorbidity of CAD and AH with other cardiovascular pathology excluded (n=180), and a subgroup of patients with CAD only (other diseases excluded, N=61). Healthy subjects with normal cardiovascular endophenotypes (N=131) were selected out of a sample of 1600 individuals. Genotyping was done using Illumina Human custom chip microarrays with a panel of markers used for direct-to-consumer genomic service "My Gene" (www.i-gene.ru) (Genoanalytica, LLC). For statistical analysis R v2.14.0 software environment was used, including specialized packages "GenABEL", "snpStats" and "genetics". Predictive value of each candidate SNP was tested using AUC (area under curve).

Results and conclusion: Syntropy group differed significantly from other samples analyzed. Pathway of *ITGA4, KLF7,* and *TAS2R38* genes was involved in the development of this comorbidity. Advanced classifier analysis yielded that *KLF7* rs7568369 reached AUC of 63% and three other SNPs *(LDLR* rs2738446, rs688*,* and *CDKN2A* rs1333048) reached maxAUC of 63%. GG genotype of the rs6501455 located in a region between *KCNJ2* and *SOX9* genes yielded most substantial risk effect in reference to CVC syntropy (OR 3,91; 95% CI 1,56-10,33; P<0.0016). GG genotype of the rs7568369 in *KLF7* had highest protective effect in reference to CVC syntropy (OR 0.34; 95% CI 0,16-0,68; $P<7*10^{-4}$). Cluster analysis which involved 90 SNPs, related to different cardiovascular phenotypes, showed that syntopy forms a separate cluster, while other subgroups are close to each other in genetic characteristics.

The study demonstrates that CVC syntropy differs significantly in genetic characteristics from other forms of cardiovascular parthology and has specific genes (SG) involved.

TARDIVE DISKINESIA AND POLYMORPHISM OF PHOSPHATIDYLINOSITOL- 4-PHOSPHATE 5-KINASE IIA GENE IN RUSSIAN SCHIZOPHRENIC PATIENTS

Rudikov E.V.*[1], Gavrilova V.A. [1], Fedorenko O.Y. [1], Boyarko E.G. [1], Semke A.V. [1], Sorokina V.A. [2], Govorin N.V. [3], Ivanova S.A. [1]

[1]*Mental Health Research Institute SB RAMSci, biological psychiatry, Tomsk, Russia;*

[2] *Kemerovo Regional Clinical Psychiatric Hospital, psychiatry, Kemerovo, Russia;*

[3] *Chita State Medical University, psychiatry, Chita, Russia*

e-mail: korvin_w@mail.ru

** Corresponding author*

Background: Pharmacogenetic studies of tardive dyskinesia are very promising direction to develop individualized antipsychotic treatment. Phosphatidylinositol-4-phosphate-5-kinase IIa (PIP5K2A) is one of the key enzyme in phosphatidylinositol-4,5-bisphosphate biosynthesis, plays an important role in membrane transduction of neurotransmitter signals and in intracellular signaling. PIP5K2A gene is located in schizophrenia candidate region on chromosome 10p14-11. Polymorphisms of this gene have been shown to be associated with schizophrenia in European and Chinese populations, but there were no such studies in Russia.

Objective: Our study aimed to investigate association of (N251S)-PIP5K2A (rs10828317) polymorphism with tardive diskinesia in Russian schizophrenic patients.

Materials and methods: Blood samples from 355 Russian Caucasian patients with clinically established schizophrenia (with an age of 43±15.8 years) were taken in four different psychiatric departments in West Siberia. Abnormal Involuntary Movement Scale (AIMS) was used to assess tardive dyskinesia cross-sectionally. Control group consisted of 100 healthy volunteers. Genotyping of (N251S)-PIP5K2A (rs10828317) was performed on ABI StepOne Plus with TaqMan1 Validated SNP Genotyping Assay (Applied Biosystems). The program SPSS11.5 was used for statistical analysis. Hardy–Weinberg equilibrium (HWE) and differences in genotype frequencies were tested using a chi-square test. Comparisons of AIMS-score in different groups were carried out with Kruskal Wallis test and Mann-Whitney test with Bonferroni correction.

Results: The genotype distribution of (N251S)-PIP5K2A (rs10828317) polymorphism was in agreement with HWE ($\chi 2$ =0.32, p=0.6481) in control group, but there was a disequilibrium in group of schizophrenic patients ($\chi 2$ =9.06, p=0.0028). 40% of patients and 45% of healthy volunteers were homozygous for the T-allele, 40.3% of patients and 46% of control persons were heterozygous, and 19.7% of patients and 9% of healthy volunteers were homozygous for the C-allele ($\chi 2$ =6.25, p=0.044). We found an association of rs10828317 with schizophrenia (p=0.04, Odds Ratio=2.48, 95%CI=1.19-5.17 for the CC genotype). CC-carriers with schizophrenia also had a higher mean AIMS score (6(2-9)) (median (25%-75% percentiles)) in comparison to those with the CT (3(0-6)) or the TT (2 (0-4)) genotype (Kruskal Wallis test – 24.74, p<0.0001, Mann-Whitney test with Bonferroni correction - p(CC/TT)<0.0001, p(CC/CT)=0.0009, p(TT/CT)=0.09). Subsequently, frequency of CC-carriers was significantly higher in group of schizophrenic patients with tardive diskinesia compared with the group of schizophrenic patients without tardive diskinesia (34.4% and 15.7% respectively, $\chi 2$ =15.4, p=0.0004, and OR=2.81 95%CI=1.61-4.91 for the CC genotype, p=0.0005).

Discussion: (N251S)-PIP5K2A (rs10828317) is known to be a functional mutation. Previous studies show that mutant kinase inefficient to activate the KCNQ channels that may lead to lack of dopaminergic control in schizophrenic patients [1]. Moreover, (N251S)PIP5K2A decreased membrane abundance of excitatory amino acid transporter EAAT3 in study on EAAT3-expressing oocytes and human embryonic kidney cells [2]. Taken together, these facts may act as biological explanation of association of (N251S)-PIP5K2A (rs10828317) with tardive diskinesia and schizophrenia.

Conclusions: The significant association of (N251S)-PIP5K2A polymorphism with tardive diskinesia has been found. CC-carriers with schizophrenia had higher risk of tardive diskinesia and more severe symptoms as evaluated by AIMS. Further studies are needed to support our findings.

CYTOKINE PROFILE AND CIRCULATING DNA IN THE BLOOD OF PATIENTS WITH TICK-BORNE BORRELIOSIS

Shkoda O.S.*, Chikova E.D., Fomenko N.V., Laktionov P.P., Rykova E.Y.
Institute of Chemical Biology and Fundamental Medicine SB RAS, Novosibirsk, Russia
e-mail:olgashkoda@sibmail.com
** Corresponding author*

Key words: *Lyme borreliosis, circulating DNA, immune response, cytokines*

Motivation and Aim: Tick-borne borreliosises are widespread, characterized by high disease incidence and clinical presentations polymorphism. Their study is of high social importance due to the possibility of Lime disease chronic form development accompanied by autoimmune complications. Extracellular circulating DNA (cirDNA) were shown to be present in blood plasma in low concentrations normally, whereas their concentration has been increased in cancer and autoimmune patients blood. The aim of the study was to reveal the association of immune response development and cirDNA changes in Lime disease patients.

Methods and Algorithms: Blood samples were taken from 60 healthy subjects (HS) and 61 patients with a tick bite, which were hospitalized into Novosibirsk City Infectious Clinics №1 with signs of infection disease diagnosed as Lime disease. Blood was fractionated into plasma and cells, the cell-surface-bound cirDNA (csb-cirDNA) fraction was obtained by successive treatment of cells with PBS/EDTA and trypsin solutions. Anti-borrelia IgG and IgM antibodies titers and cytokine concentrations (IL-10, IL-6, IL-4, αTNF, γIFN) were estimated using ELISA Kits. The cirDNA concentration was evaluated using was measured by quantitative real-time PCR specific for LINE-1 repetitive elements.

Results: Reliable increase is shown of the csb-cirDNA concentration in blood from patients with Lime disease compared with healthy subjects (98 versus 13 ng/ml, Mann-Whitney U test, $p<0,0001$). Csb-cirDNA and plasma cirDNA concentration increase is associated with absence of unspecific inflammatory reaction (erythema) development (195 and 19 ng/ml in patients without erythema versus 72 and 7 ng/ml in patients with erythema, Mann-Whitney U test, $p<0,05$). Patients with Lime disease demonstrated increase of pro-inflammatory cytokine concentration (IL-6, γ-IFN, α-TNF) compared with healthy subjects (14 versus 0,1; 19 versus 1; 3 versus 0,8 pg/ml, respectively, Mann-Whitney U test, $p<0,05$). Otherwise, anti-inflammatory cytokine concentration (IL-10 and IL-4) was decreased in Lime disease patients blood compared with healthy subjects (5 versus 10; 1 versus 2 pg/ml respectively, Mann-Whitney U test, $p<0,05$).

Conclusion: The study has shown a reliable association of the cirDNA concentration increase with pro-inflammatory cytokine secretion and with the absence of unspecific inflammatory reaction during Lime disease development.

Availability: Obtained data demonstrate the cirDNA further study availability as the potential complementary factor, applied for diagnostics, and possibly, for participating in the Lime disease pathogenesis.

Acknowlegements: This work was supported by grants from RFBR № 11-04-01066-a, RAS Program “Fundamental Sciences for Medicine” № 5.25.

PREVENTING COMMON HEREDITARY DISORDERS THROUGH TIME-SEPARATED TWINNING

Tchourbanov A.
Beijing Institute of Genomics (BIG), Building G, No.7 Beitucheng West Road, Chaoyang District, Beijing 100029, P.R.China
e-mail: alexander@big.ac.cn

Key words: *Artificial twinning, in vitro fertilization, complex diseases, Mendelian disorders, prevention, preimplantation genetic screening*

Motivation and Aim: Biomedical advances have led to a relaxation of natural selection in the human population in developed countries. In the absence of strong purifying selection, spontaneous and frequently deleterious mutations tend to accumulate in the human genome and gradually increase the genetic load; that is, the number of potentially lethal genes in the gene pool. It is not possible to assess directly the negative impact of the genetic load on modern society because it is influenced by many factors such as constantly changing environmental conditions and continuously improving medical care. However, if modern medicine loses its effectiveness, significantly higher than normal mortality is expected before equilibrium with the environment is re-established. Recent advances in *in vitro* fertilization (IVF) combined with artificial twinning and improved embryo cryoconservation offer the possibility of preventing significant accumulation of genetic load and reducing the incidence of hereditary disorders.

Discussion: Many complex diseases such as type 1 and 2 diabetes, autism, bipolar disorder, allergies, Alzheimer disease, and some cancers show significantly higher concordance in monozygotic (MZ) twins than in fraternal twins (dizygotic, DZ) or parent-child pairs, suggesting their etiology is strongly influenced by genetics. Preventing these diseases based on genetic data alone is frequently impossible due to the complex interplay between genetic and environmental factors. We hypothesize that the incidence of complex diseases could be significantly reduced in the future through a strategy based on time-separated twinning. This strategy involves the collection and fertilization of human oocytes followed by several rounds of artificial twinning. If preimplantation genetic screening (PGS) reports no aneuploidy or known Mendelian disorders, one of the MZ siblings would be implanted and the remaining embryos cryoconserved. Once the health of the adult MZ sibling(s) is established, subsequent surrogate parenthood with the cryoconserved twins could substantially lower the incidence of hereditary disorders with Mendelian or complex etiology.

Summary: The proposed method of artificial twinning has the potential to alleviate suffering and reduce the negative social impact induced by dysgenic effects associated with known and unknown genetic factors. Time-separated twinning could deliver more accurate health predictions for children of surrogate parents compared to estimates based on the health records of sperm and egg donors.

Acknowledgements: Supported by grant 2011Y1SA09 from the Chinese Academy of Sciences Fellowship for Young International Scientists and by grant 31150110466 from the National Natural Science Foundation of China (NSFC).

DEVELOPMENT AND APPLICATION OF THE GENOMIC CONTROL METHODS FOR GENOME-WIDE ASSOCIATION ANALYSIS USING NON-ADDITIVE MODELS

Tsepilov Y.A.[1]*, Read J.[2], Strauch K.[2], Axenovich T.I.[1], Aulchenko Y.S.[1]
[1]*Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2]*Helmholtz Centre Munich, Germany*
e-mail: drosophila.simulans@gmaol.com
[*]*Corresponding author*

Key words: *genome-wide association analysis, genomic control, non-additive models of inheritance*

Motivation and Aim: Genome-wide association (GWA) analysis is a powerful tool for mapping genes of complex traits. GWA analysis assumes that the samples are randomly ascertained from the same genetic population. In this case, the phenotypes of individuals correlate only with the genotypes of loci that are involved in the control of the trait. In reality, correlations are also caused by confounders that are in correlation both with the phenotype and the genotypes of various loci. Among most prominent factors, genetic (sub)structure of the sample can work as such confounder, leading to false-positive genetic associations. However, in the framework of GWA analysis the availability of genotypes of large set of markers allows to adjust the results of the analysis using the genomic control (GC) method. At present stage, the GC is formulated and implemented for additive models. The aim of this work was to develop, validate and implement methods of genomic control for various models of inheritance.

Results: We have derived analytical expressions for adjustment factors for association test statistics that depend on the allele marker frequency and the parameters describing the population-genetic structure of sample data. We also proposed a new method of correcting the test statistics by the polynomial depending on the allele frequency. Obtained expressions, procedures of parameter estimation and procedures for the correction of the analysis results are implemented as a computer program written using R language. The software developed was used to characterize the statistical properties of the implemented methods - described earlier, and introduced us. Methods were applied to study real data.

Conclusion: Results for non-additive genome-wide association analysis cannot be adjusted using GC methods proposed for the additive model of inheritance. We proposed several methods and strategies that can be used for correction of non-additive genome-wide association analysis results. The use of non-additive models of inheritance can improve the power of GWA analysis and help discovering new loci involved in control of complex traits.

NEW CANDIDATE GENES FOR SCHIZOPHRENIA DISORDER

Zakharyan R.V.*, Boyajyan A.S.
Institute of Molecular Biology NAS RA, Yerevan, Armenia
e-mail: r_zakharyan@mb.sci.am
** Corresponding author*

Key words: *schizophrenia, genetic polymorphisms, proinflammatory and chemotactic cytokines, genotyping, polymerase chain reaction with sequence-specific primers*

Motivation and Aim: Schizophrenia is a polygenic and multifactorial disease with strong involvement of the inflammatory component in etiopathogenic mechanisms [1, 2]. However, only limited studies have explored the association of polymorphisms in genes encoding inflammatory mediators with schizophrenia. The present study aimed to investigate the possible association of functional genetic polymorphisms of proinflammatory and chemotactic cytokines including interleukin (IL)-6, tumor necrosis factor-α (TNF-α), monocyte chemoattractant protein 1 (MCP-1), and IL-8 with schizophrenia.

Methods and Algorithms: Genomic DNA samples were isolated from fresh blood of schizophrenia-affected (n=225) and healthy subjects (n=225) according to the standard phenol–chloroform method. All DNA samples were genotyped for targeted single nucleotide polymorphisms (SNP) by polymerase chain reaction with sequence-specific primers (PCR-SSP). All primers for the PCR-SSP were designed using the genomic sequences in the "GenBank" database. The genotypes were assessed for the presence/absence of PCR amplicons specific to the particular alleles using a standard 2% agarose gel electrophoresis followed by ethidium-bromide staining. Distributions of genotypes for the studied polymorphisms were checked for correspondence to the Hardy–Weinberg equilibrium. In order to find potential relevance of targeted SNPs to schizophrenia, the allele and phenotype frequencies in patients and control subjects were compared. The significance of differences between allelic and phenotype frequencies was determined using Pearson's Chi-square test. The odds ratio (OR), 95% confidence interval (CI), statistical power, and Pearson's p-value were calculated in each case.

Results: According to the data obtained, the *IL-6* -174G/C, *TNF-α* -308G/A, *MCP-1* -2518A/G, and *IL-8* +293G/T polymorphisms were significantly associated with schizophrenia.

Conclusion and Availability: The *IL-6* -174G/C, *TNF-α* -308G/A, *MCP-1* -2518A/G, and *IL-8* +293G/T minor alleles seems to be a risk factors for the development of schizophrenia.

References:

1. M. Baron. (2001) Genetics of schizophrenia and the new millennium: progress and pitfalls. Am. J. Hum. Genet., 68: 299-312.
2. A. Monji, T. Kato, S. Kanba. **(**2009) Cytokines and schizophrenia: microglia hypothesis of schizophrenia. Psychiatry Clin. Neurosci., 63: 257–265.

SYSTEM PROFILING (GENOMIC, TRANSCRIPTOMIC, METABOLOMIC) OF MYCOPLASMAS. NEW INSIGHTS ON MINIMAL STYLE OF LIFE

Alexeev D.G.*, Gorbachev A., Demina I.A., Ladygina V.G., Vanushkina A., Govorun V.M.
Research Institute of Physico Chemical Medicine
e-mail: Alexeev@niifhm.ru
** Corresponding author*

Motivation and aim: A great effort has been made in the past years towards the identification of the minimal gene complement necessary for life, especially with the boom of synthetic biology. In fact the task has been driven by the considerable sequencing projects in place, particularly in the prokaryote kingdom.

Such computation analysis has been very powerful, however comparative genomics, aiming at finding the common gene set among bacteria, can lead to underestimations given the fact that different organisms can use different genes to perform the same function.

Furthermore, forward genetics approaches have provided us with lists of essential genes that are believed to represent the basic core for a living cell, whereas the rest of the genes, though apparently dispensable, must be there for a reason. Intriguingly many of those essential genes revealed to have unknown functions (albeit been conserved in evolution), functions that remain obscure after this high-throughput analysis. This highlights how far we are of unveiling the biology of even "simple" unicellular organisms.

Methods and Algorithms: We have developed and applied several novel approaches to genomic, proteomic, transcriptomic and metablolomic analysis using conventional MS/MS, PCR and novel methods of NGS.

Results: We have extensively studied genomes, proteomes, transcriptomes and metabolomes – starting from description of steady state cocentrations and going deeper into functional comparison.

Conclusion: Unprecedented complexity demonstrated in Omics profiling of mycoplasma conforms to the idea that however small mycoplasmas can virtuously adapt to variety of conditions using simple and interconnected mechanisms.

MELANOGENESIS HELPS HUMAN ADIPOSE TISSUE WITHSTAND LOW-GRADE SYSTEMIC IINFLAMMATION

Baranova A.V.
School of Systems Biology, George Mason University, Fairfax, VA, USA
e-mail: abaranov@gmu.edu

Purpose: To curtail secondary morbidities in obesity and overweight individuals without the need of the weight loss.

Methods: Molecular pathways specifically activated in adipose of morbidly obese individuals were studied by microarray and by various biochemical assays.

Results: The melanin, common skin pigment, was discovered in human adipose tissue. A marked heterogeneity of melanin content was observed in individual adipose tissue extracts. Melanogenesis was shown to be excessively stimulated in morbid obesity. Positive correlation between fasting glucose levels and total outputs of the melanogenic pathway in adipose tissues was observed. Novel hypothesis stating that ectopic synthesis of melanin may serve as a compensatory mechanism that utilizes its anti-inflammatory and its oxidative damage absorbing properties was developed. With the progression of obesity and an increase of the cellular fat deposition, adipocytes become more exposed to endogenous apoptotic signals, especially ROS. To counteract pro-apoptotic ROS effects, the adipocytes in turn may ectopically activate the genetic program of melanogenesis, thus neutralizing excessive ROS. Adipocytic melanin would also suppress the secretion of pro-inflammatory molecules, thereby decreasing the pro-inflammatory background in obese subjects and alleviating the metabolic syndrome. High polymorphisms of human genes regulating melanin biosynthesis may account for the differences in propensity to develop secondary complications of obesity.

Conclusion: Molecular compounds stimulating melanogenesis, particularly, the synthetic agonists of α-MSH receptors, have already been proven safe in human trials for therapeutic tanning. These compounds shall be tested as the preventive medications aimed at curtailing the development of devastating metabolic complications in obese and overweight populations.

MULTIPLEX ASSEMBLY OF HIGH-FIDELITY DNA MOLECULES FROM COMPLEX MIXTURES OF SYNTHETIC OLIGONUCLEOTIDES

Borovkov A.
Synbuild, LLC
e-mail: aborovkov@synbuild.com

To meet the growing demand of post-genomics biology for synthetically assembled genes we developed an approach for mass production of protein encoding genes from complex mixtures of microarray synthesized oligonucleotides. We overcame the key limitations associated with use of such starting material for gene assembling, namely the poor quality and low amounts of individual components, by introducing an intermediate block assembling step coupled with a hybridization-based self-elimination process that disallowed incorporation of incorrectly synthesized oligonucleotides into block assemblies. During this step an original oligonucleotide mixture converted into a plurality of partially overlapping gene building blocks, which was subsequently used for gene assembly. The developed protocol found to be simple and robust. It was successfully tested on oligonucleotides from three independent vendors. Quality of the assembled products, assessed by direct DNA sequencing and their functional activity in a coupled transcription/translation system, was found to be indistinguishable from that of the genes assembled by a traditional method from high quality column synthesized oligonucleotides.

MULTIPLEXED REGULATORY ELEMENT SYNTHESIS AND CHARACTERIZATION

Kosuri Sriram
Wyss Institute For Biologically Inspired Engineering, Harvard University, USA
e-mail: sri.kosuri@wyss.harvard.edu

In this talk I will discuss our ongoing efforts using DNA microchips to synthesize large numbers of designed regulatory elements in *E. coli*. We synthesized ~27,000 combinations of promoter, ribosome binding site, and peptide leader sequences in an effort to look at the composability of regulatory elements governing gene expression. We clone the library of regulatory elements in a plasmid library to drive expression of GFP. The transcription levels of individual members of the library are quantified using RNA-Seq and protein levels using a new method called Flow-Seq. In Flow-Seq, the library of cells is sorted into bins based on the level of GFP fluorescence. The binned cells are then deep-sequenced to determine which bins individual constructs were contained within. In this manner, we can reconstruct GFP expression levels of all members of the library. Thus, we are able to quantify both the transcription and translation levels of each member of the library independently, leading to insights on how regulatory elements in combination affect transcription and translation rates.

STRUCTURE–FUNCTION STUDIES OF MODULAR, ACTIN–REGULATING PROTEINS BY HETERONUCLEAR ($^{1}H/^{15}N/^{13}C$) NMR

Smirnov S.L.
Western Washington University, Bellingham, WA, USA
e-mail: smirnov@chem.wwu.edu

Key words: *Modular Protein, NMR, Actin, Calcium, Regulation*

Motivation and Aim: Villin is a modular, potentially therapeutically relevant protein of 826 residues and capable of switching between F-actin bundling and severing in response to changes in levels of calcium and other signals [1]. Structure/function studies of villin (actin interface, calcium sites etc) focus on its fragments as no crystal or solution structure of the intact protein is available. Villin polypeptide consists of six gelsolin-like repeats (D1 – D6) linked with the C-terminal headpiece (HP). The solution structure of domain D1 and crystal structure of D6 were reported in calcium-free environments. The solution and crystal structures of the isolated headpiece, which forms a calcium-independent actin-binding site, were also determined. Our previous study of D6-HP polypeptide (headpiece linked with the adjacent core domain D6) showed that in high calcium monomeric D6-HP forms the smallest known, two-domain (D6 & HP) actin-bundling villin fragment [2]. This led us to propose a calcium-sensitive, F-actin-binding site within D6. In the present work we test this hypothesis by the structure/function investigation of D6 domain in high calcium.

Methods and Algorithms: The solution structure of D6 domain was determined by heteronuclear ($^{1}H/^{15}N/^{13}C$) NMR. The calcium-binding locus of D6 was deciphered by monitoring its ^{15}N-HSQC NMR spectra recorded at varying Ca^{2+} levels. F-actin binding by D6 was assessed by the ultracentrifugation pull-down assay quantified with HPLC.

Results: Here we report the solution NMR structure of villin D6 domain at high calcium (PDB 2LLF) and identify the main differences with the no-calcium crystal structure. We also show that the D6 residues whose NMR chemical shift values are changing in response to calcium back-titration belong to the major helix and nearby loops thus localizing the calcium binding site(s). Lastly, we show that isolated D6 domain does not bind F-actin.

Conclusion: Our findings indicate that calcium-binding D6 domain of villin does not form a functional F-actin-binding site in isolation. Therefore, we hypothesize that the 40-residue intrinsically disordered linker connecting gelsolin core and headpiece must be a part of the cryptic F-actin binding site within D6-HP and possibly within intact villin.

Acknowledgements: The NMR experiments were conducted at the National High Magnetic Field Laboratory (Tallahassee, Florida, USA) supported by cooperative agreement DMR 0654118 between the NSF (USA) and the State of Florida (USA).

References:

1. S.Khurana, S.P.George (2008) "Regulation of cell structure and function by actin-binding proteins: villin's perspective", *FEBS Letters*, **582**(14)**:** 2128-2139
2. S.L.Smirnov et al. (2007) "The isolated sixth gelsolin repeat and headpiece domain of villin bundle F-actin in the presence of calcium and are linked by a 40-residue unstructured sequence", *Biochemistry,* **46**(25)**:**7488-7496.

DNA-MICROARRAY SYNTHESIZER

Sinyakov A.N.*, Kamaev G.N.[1], Bessmeltsev V.V.[2], Shelkovnikov V.V.[3]
Institute of Chemical Biology and Fundamental Medicine SB RAS, Novosibirsk, Russia
[1] Rzhanov Institute of Semiconductor Physics SB RAS, Novosibirsk, Russia
[2] Institute of Automation and Electrometry SB RAS, Novosibirsk, Russia
[3] Novosibirsk Institute of Organic Chemistry SB RAS, Novosibirsk, Russia
e-mail: sinyakov@niboch.nsc.ru
**Corresponding author*

Key words: *oligonucleotide, DNA, microarray synthesizer.*

Motivation and Aim: The goal of this work was to develop a multichannel synthesizer of a new type that would allow for a concurrent parallel synthesis of a large number of oligonucleotides as a material for genetic engineering design and diagnostic purposes.

Methods and Algorithms: A silicon microarray equipped with communications for inputting and outputting reagents as well as with reaction cells for generating a strong acid and synthesizing individual oligonucleotides was selected as a reactor for oligonucleotide synthesis.

As the main method for solid-phase oligonucleotide synthesis, we selected the well-reputed phosphoramidite synthesis.

Note that since the number of cells for performance of oligonucleotide synthesis is great, it is impossible to provide each cell with an individual input line for the acid used for deblocking. Therefore, we decided to generate a strong acid for deblocking from a neutral photoacid via exposure to ultraviolet radiation either in the reaction cell itself or in a cell adjacent to the reaction cell.

Results: To implement such generation of the acid, we designed an optical projection system involving a controllable micromechanical mirror array.

Conclusion: Thus, the microarray oligonucleotide synthesizer comprises a conventional DNA synthesizer, used as a pump for inputting reagents; optical projection system involving a controllable micromechanical mirror array; and the microarray for synthesis of target oligonucleotides itself.

References:

1. Sinyakov A.N., Ryabinin A.A., Maksakova G.A., Shelkovnikov V.V., Loskutov V.A., Vasil'eva E.V., Shekleina N.V. Sulfonium derivatives of thioxanthenone, a new class of photodetritylating agents for microarray oligonucleotide synthesis. Russian Journal of Bioorganic Chemistry. 2010. T. 36: 130-132.
2. V. V. Shelkovnikov V. A. Loskutov, E. V. Vasil.ev, N. V. Shekleina, V. A. Ryabinin, and A. N. Sinyakov. New acid photogenerators based on thioxanthen-9-one sulfonium derivatives for detritylation in the oligonucleotide synthesis. Russian Chemical Bulletin, International Edition, 2011, 60: 1-9.

ANTIVIRAL ACTIVITY OF NANOBIOCOMPOSITES BASED ON PEPTIDE NUCLEIC ACIDS AND TITANIUM DIOXIDE NANOPARTICLES

Amirkhanov N.V.*[1,2], Mazurkova N.A.[2,3], Amirkhanov R.N.[1,2], Zarytova V.F.[1,2]
[1] *Institute of chemical biology and fundamental medicine, Siberian Branch of RAS, Novosibirsk, Russia;*
[2] *Novosibirsk State University, Novosibirsk, Russia;*
[3] *Vector State Scientific Center for Virology and Biotechnology, r.p. Kol'tsovo, Novosibirsk oblast, Russia*
e-mail: nariman@niboch.nsc.ru
** Corresponding author*

Key words: *TiO_2 nanoparticles, PNA Delivery, influenza A virus, TiO_2•Polylysine•DNA/PNA bionanocomposites*

Motivation and Aim: Various viral diseases can be targeted and treated with nucleic acids designed to inhibit viral amplification by interacting with viral RNA. To ultimately achieve clinical utility, any nucleic acid-based antiviral therapeutic will need to possess a number of favorable pharmacologic qualities, including *in vivo* stability, low toxicity, and ability to penetrate through cell membranes and to reach viral RNA targets within relevant cell populations. In order to create an effective antiviral drug, we have developed nanobiocomposites based on: 1) peptide nucleic acids (PNA) which are able to interact with viral RNA, and 2) non-toxic titanium dioxide nanoparticles which play a role of a powerful vehicle to deliver of PNA molecules into the cells.

Methods: PNA, with high binding affinity to RNA of influenza A virus (FLUAV), was immobilized in the form of hybrid DNA/PNA duplexes on the surface of TiO_2 nanoparticles (size of 5-6 nm) covered with polylysine (PL). Attaching of DNA/PNA duplex on TiO_2•PL nanoparticles occurred due to electrostatic interactions between the negatively charged DNA chain and the positively charged amino groups of PL. Binding of the PNA with the nanocomposite achieved through noncovalent *Watson-Crick* interactions between the PNA and complementary DNA. The antiviral activity of nanobiocomposites has been investigated in *MDCK* cell culture infected with human influenza A virus (subtype H3N2).

Results: The capacity of obtained TiO_2•PL•DNA/PNA nanocomposites was 10-30 nmol PNA per 1 mg of TiO_2. By method of confocal laser scanning microscopy it has been shown that TiO_2•PL•DNA/PNA bionanocomposites effectively penetrate in eukaryotic cells. Bionancomposites containing the complementary PNA sequence have a high antiviral activity against human influenza A virus. The Virus Neutralization Index for investigated bionanocomposites at 200 and 2000 nM concentration of PNA was 0.3 and 2.7 log (degree of viral suppression was 50 and 99.8%, correspondingly). The TiO_2•PL•DNA/PNA bionanocomposites exhibited minimal cytotoxicity in cell viability assays (with a CC_{50} of approximately 13000 nM). Thus, the Selectivity Index for antisense PNA was over 60.

Conclusion: PNA molecules can be immobilised on the TiO_2 nanoparticles through the hybrid DNA/PNA duplexes. The resulting TiO_2•PL•DNA/PNA bionanocomposites consisting of TiO_2 nanoparticles are able to deliver antisense PNA in eukaryotic cells and exhibit high biological antiviral activity.

Acknowledgements: This work was supported by RFBR grant *N⁰* 11-04-01408-a; Federal Target Program, Government Contract grant *N⁰* 16.512.11.2267.

IMMOBILIZATION OF PEPTIDE NUCLEIC ACIDS IN THE FORM OF HYBRID DNA/PNA DUPLEXES ON TiO_2 NANOPARTICLES COVERED WITH POLYCATIONS CONTAINING QUATERNARY AMINO GROUPS

Amirkhanov R.N.*[1,2] , Amirkhanov N.V. [1,2], Zarytova V.F. [1,2]
[1] *Institute of Chemical Biology and Fundamental Medicine, Siberian Branch of RAS, Novosibirsk, Russia;*
[2] *Novosibirsk State University, Novosibirsk, Russia*
e-mail: rinat@nibcoh.nsc.ru
** Corresponding author*

Key words: *PNA, titanium dioxide, drug delivery, nanocomposites, PDDA*

Motivation and Aim. Design and construction of nanocomposites based on nanoparticles as vehicles for drug delivery of antisense oligonucleotides or their analogues into cells is relevant problem to molecular biology and modern medicine. Earlier, non-toxic titanium dioxide nanoparticles have been proposed for delivery of antisense peptide nucleic acids (PNA) into cells [1]. In this case PNA immobilization was carried out through hybrid DNA/PNA duplexes on TiO_2 nanoparticles, coated with polylysine. Capacity of the constructed nanocomposites was not high enough and essentially depended on the immobilization buffer medium. In order to find methods to increase the immobilization efficiency of PNA in DNA/PNA duplexes on TiO_2 nanoparticles, we propose to use polycations containing quaternary amine groups instead of polylysine.

Methods and Algorithms. TiO_2•PDDA Nanocomposite containing polycation with quaternary amine groups, PDDA (polydiallyldimethylammonium chloride), was prepared by mixing TiO_2 nanoparticles (size 5-6 nm) with different molar excess of PDDA in the different buffer systems. In the next step TiO_2•PDDA•DNA/PNA nanocomposite was prepared by mixing TiO_2•PDDA complex with preformed DNA/PNA duplex. Immobilization capacity was determined by the amount of radioactively labeled 5'-[32] P-DNA oligonucleotide immobilized on the nanoparticles or by the amount of PNA containing fluorescence group or by the measuring of the optical density of free DNA in solution. The binding between TiO_2•PDDA nanocomposite and DNA/PNA duplex was provided by electrostatic interactions between positively charged amino groups of PDDA and negatively charged DNA phosphates.

Results. It was shown that using of PDDA instead of polylysine as a linker allows to increase the immobilization efficiency by 3-4 times. Also, the immobilization efficiency did not depend on buffer composition and pH.

Conclusion. Thus, it was created new TiO_2•PDDA•DNA/PNA nanocomposites with essentially increased PNA capacity. This feature potentially allows to decrease toxicity of the bionanocomposites and allows to create more effective drugs for targeting pathogenic nucleic acids in the living cells.

Acknowledgements. This work was supported by RFBR grants No. 11-04-01408-a; Federal Target Program, Government Contract grant No. 16.512.11.2267.

References:

1. 1. R.N. Amirkhanov, V.F. Zarytova. (2009) Titanium dioxide nanocomposites containing peptide nucleic acids for their cellular delivery, "Rusnanotech" Nanotechnology International Forum, 748-749.
2. nology International Forum, 748-749.

FLUORESCENT PEG-MODIFIED SINGLE-WALLED CARBON NANOTUBES AS PLATFORM FOR THE CONSTRUCTION OF THERANOSTIC TOOLS

Apartsin E.K.*[1], Buyanova M.Yu. [1,2], Novopashina D.S. [1], Ven'yaminova A.G. [1]
Institute of Chemical Biology and Fundamental Medicine, Novosibirsk, Russia;
Novosibirsk State University, Novosibirsk, Russia
e-mail: eka@niboch.nsc.ru
** Corresponding author*

Key words: *carbon nanotubes, oligonucleotides, pyrene, hybrid material*

Motivation and Aim. Single-walled carbon nanotubes (SWNTs) possessing unique chemical and physical properties and compatibility with biomacromolecules and cells are investigated actively in nanobiotechnology and biomedicine as prospective intracellular transporters of biologically active compounds, e.g. theurapeutic nucleic acids. Such biohybrid materials should contain reporter groups to permit the monitoring of transfection and elimination processes. PEGylation, i.e. attachment of polyethylene glycol (PEG), was reported to be a powerful tool to increase the hydrophilicity of SWNT thus increasing their solubility and biocompatibility [1]. Such a method permits to functionalize the ends and defect sites in SWNT covalently, with the sidewall remaining unmodified and available for nucleic acid attachment.

Results. Here we propose the approach to the preparation of multifunctional non-covalent hybrids of pyrene-modified oligonucleotides with fluorescently labeled PEGylated SWNTs. SWNTs containing fluorophore moieties introduced by amino-terminated PEG (M.w. 1500) were prepared. Conjugates of oligodeoxyribonucleotides with pyrene residues attached to 5'-phosphate group directly or via hexa(ethyleneglycol) phosphate linker were synthesized and used for the preparation of hybrids. Non-covalent functionalization of the two types of fluorescent SWNTs by 5'-pyrene conjugates of oligonucleotides was performed according to [2]. Structural and biophysical properties of non-covalent hybrids were investigated. The capacity of carbon nanotubes as potential oligonucleotide transporters was estimated from adsorption isotherms and amounted to about 50-100 μmol/g depending on oligonucleotide structure.

Conclusion. The results obtained demonstrate the availability of proposed approach for rational design of oligonucleotide hybrids with carbon nanotubes. Such non-covalent hybrids can be applied to making the prospective theranostic tools.

Acknowledgements. This work was supported by the RFBR grant 11-04-01014-a, PICS 12-04-91053, State Contract MES RF P1334, RAS Presidium Program №24.

References:

1. M. Bottini, N. Rosato, N. Bottini. (2011) PEG-modified carbon nanotubes in biomedicine: current status and challenges ahead, *Biomacromolecules,* **12:** 3381-3393.
2. E.K. Apartsin, D.S. Novopashina, Yu.V. Nastaushev, A.G. Venyaminova. (2012) Fluorescently labeled single-walled carbon nanotubes and their hybrids with oligonucleotides, *Nanotechnol. Russ*. **7:** 99-109.

BIOMOLECULE SPECIFIC MIPs ON THE BASIS OF SIMPLE NYLON-6

Dmitrienko E.V.*[1], Pyshnaya I.A.[1], Lomzov A.A. [1], Bulyshev R.D. [1], Kosolobov S.S. [2], Latyshev A.V. [2], Pyshnyi D.V.[1]

[1] *Institute of Chemical Biology and Fundamental Medicine, SB RAS, Russia;*

[2] *A.V. Rzhanov Institute of Semiconductor Physics, SB RAS, Russia*

e-mail: elena.dmitrienko@niboch.nsc.ru

** Corresponding author*

Key words: *nylon-6, molecular imprinted polymers*

Motivation and Aim: Polyamides and nylon 6, in particular, are widely used in various fields of life sciences, biotechnology and medicine. Nylon is the biocompatible, mechanically, and thermally stable material with characterized in detail properties. To date, the preparation of molecularly imprinted nylon designed for recognition of low-molecular weight templates, for example, amino acids is described and when formic or acetic acids are used as the solvents. However, this approach can not be widely applied due to, probably, too stringent conditions of the MIP preparation. The goal of our studies was to find mild conditions for nylon MIPs preparation which are compatible with biomolecules imprinting.

Methods and Algorithms: We used various techniques to prepare and to analyse designed MIPs.

Results: A new simple approach for the preparation of molecularly imprinted polymers (MIPs) based on polyamide (nylon-6) was developed. The polymer matrix formation occurred during the transition of nylon from the dissolved to the solid state in the presence of template molecules in original solution. Fluorinated alcohol was chosen as a main solvent for the polyamide. It provided high solubility of nylon and did not significantly change the spatial structure of biopolymers in alcohol/water medium (e.g. simple proteins and DNA duplexes). The alteration of the polymer matrix structure after the addition of different types of porogens in the nylon solution was investigated. MIPs with prints of different bimolecular templates (low molecular weight ligands and proteins) were obtained. The specific binding of print molecules with corresponding washed up MIPs was monitored using various detection approaches. The selectivity coefficients of the prepared MIPs were estimated to be 1.4-4.6 depending on the type of templates and conditions of the polymer matrix formation.

Conclusion: Thus we suggest very convenient approach to obtain nylon MIPs for wider spectra of template molecules including structured biopolymers.

Acknowledgements: This work was supported by Russian Government (ГК № П1073) and Programs of SB RAS №86, №133.

RUPTURE EVENT SCANNING FOR BIOLOGY AND MEDICINE

Dultsev F.N.
Institute of Semiconductor Physics SB RAS, Novosibirsk, Russia
e-mail: fdultsev@isp.nsc.ru

Key words: *quartz crystal microbalance, biosensor, rupture event, bacteria*

Motivation. Direct measurement of bond rupture forces in biological systems may render exclusively important data for identification or diagnostic purposes. The energy of hydrogen bonds varies within a broad range; the data on the strength of these bonds are helpful in evaluating the state of macromolecules. An ideal method for measuring bonding in biological systems would be that excluding the disturbing action of the electromagnetic radiation. One of these methods is Rupture Event Scanning developed by us on the basis of the quartz crystal microbalance (QCM) used in the active mode.

Methods. Precision possibilities of QCM-based measurements have been increased by several orders of magnitude with QCM used not only as a passive sensor but also as an active element. Biochemical modification of the QCM surface allows us to detect the bodies specifically adsorbed on the oscillating surface by measuring the acoustic signal generated at the moment of rupture. The method is highly sensitive and applicable to measuring rupture forces either for weak bonds, for example hydrogen ones, or for strong covalent bonds. The method works as follows. A nanoparticle (in particular a biological body: a phage, a bacterium or a virus) is attached to the QCM surface, and then the amplitude of shear oscillations is smoothly increased to detect the moment of particle detachment. This generates an acoustic wave, detected in the form of noise with the help of the same QCM operating as a sensitive microphone. The amplitude of this noise is proportional to the amount of torn bonds; the amplitude of shear oscillations (or the alternating voltage applied to QCM) is used to determine rupture force. The characteristics of the measured signal allow us not only to detect specifically bound analytes and their affinity to the receptor but also to determine the number of bodies ruptured from the surface.

Results. The proposed technique confidently records the rupture of a particle (virus or bacterium) down to 30 nm in size. It was demonstrated experimentally that the dependence on concentration is linear within the range covering 6 orders of magniture. The sensitivity of the method allows determination of down to a single virus or bacterium. A special cell was designed and built in order to carry out measurements in liquid. After pumping the liquid under investigation through the cell, the bacteria were let to precipitate onto the surface during incubation time. Then scanning over the voltage was carried out. The bacteria of different kinds differ in their affinities to the surface, so they get detached at different voltage values. Thus, for example, it was possible to distinguish between *Escherichia coli* and *Staphylococcus aureus* present together. So, using the proposed procedure it is possible to detect the bacteria of different species in the case of their joint presence in the liquid sample under analysis.

Conclusion. The QCM-based technique is proposed for measuring the bond rupture force. The technique is suitable for selective determination of different types of biological objects during a single run and may be a reliable basis for biological sensing.

References

1. M. Cooper, F.N. Dultsev, V. Ostanin, D. Klenerman (2011) Separation and Detection of Bacteria Using Rupture Event Scanning, Anal. Chim. Acta 702 233– 238
2. F. Dultsev, E. Kolosovsky (2009) QCM Operating in Threshold Mode as Gas Sensor, Langmuir, 25, 12195–12200

ELECTROOSMOTIC PUMP BASED ON SILICON MICROCHANNEL MEMBRANE FOR BIOANALYTICAL PLATFORMS

Filippov N.S.*[1], Paraschenko M.A.[1], Vandysheva N.V. [1], Vashutina T.S.[2], Romanov S.I.[1]

[1] *A.V. Rzhanov Institute of Semiconductor Physics of SB RAS, Novosibirsk, Russia;*
[2] *Novosibirsk State Technical University, Novosibirsk, Russia*
e-mail: filippov@isp.nsc.ru
** Corresponding author*

Key words: *Electroosmotic pump, silicon membrane, electroosmotic flow, microfluidics*

Motivation and Aim: Electroosmotic pump is the preferred transport system in microfluidic devices [1]. It has no moving parts, could be as small as a few micrometers and allows electroosmotic flow (EOF) to be precisely controlled over a wide range of values.

Aim of this investigation is to develop micro- and nanofluidic systems based on silicon microchannel membranes with controllable electrokinetic flows of biological fluids.

Methods and Algorithms: Silicon membranes used as driving element of electroosmotic pump have had channels with square cross-section and pitch from 4 to 15 μm [2]. Membranes surface has been covered by 100 nm thermally grown SiO_2. The pump has been inserted into a two-chamber electrokinetic cell between two platinum electrodes spaced by 5 mm. Inlet chamber has been connected to a reservoir with 1 mM $Na_2B_4O_7$ buffer solution and outlet chamber has been connected either to a vertical capillary with measuring scale to monitor pressure capacity or to a reservoir placed on a balance to measure flow rate.

Results: We have investigated silicon microchannel membrane pumping performance during transport of buffer solutions through it. Dependence of the pressure capacity and volumetric flow rate on electric field magnitude and membrane structure parameters (such as microchannel width and length) has been observed. It has been found that EOF has directed from anode to cathode due to negative charge density of silica surface. Generated pressure measured by relative meniscus position in the vertical capillary has been directly proportional to the electric field value and inversely proportional to microchannel width. It has been shown that achievable pressure considerably depends on microchannel length, for example, pressure measured 30 seconds after applying voltage to the electrodes has been 290 Pa and 750 Pa for membranes with channels length 80 and 180 μm respectively. Also flow rate has been found to depend linearly on applied voltage. The maximum achieved pressure and flow rate per unit area and unit applied voltage have been 1.1 kPa and 62 μl/cm^2/V, respectively.

Conclusion: Obtained data shows that silicon microchannel membranes could be used in microfluidic systems intended for biomaterial analysis and manipulation. Performance of such devices can be accurately tuned by choosing their structural parameters and simply by varying an applied voltage that is more important.

References:

1. D.J.Laser, J.G.Santiago (2004) A review of micropumps, J. Micromech. Microeng., **14:** R35-R64.
2. S.I.Romanov et al. (2012) Sensors and filters based on nano- and microchannel membranes for biomedical technologies, Journal of Physics: Conference Series, **345:** 012001.

GOLD AND COMPOSITE NANOPARTICLES FOR BIOMEDICAL APPLICATIONS

Khlebtsov B.N.*, Khlebtsov N.G.
Institute of Biochemistry and Physiology of Plants and Microorganisms, Russian Academy of Sciences
e-mail: bkhl@ibppm.sgu.ru
** Corresponding author*

Key words: *gold nanoparticles, theranostics*

Motivation and Aim. Multifunctional nanoparticles that combine therapeutic, diagnostic, and sensing modalities are a new trend in nanobiotechnology. One attractive option for theranostic applications are composites that combine the unique optical properties of metal nanoparticles and the mesoporous silica functionalized with an appropriate photosensitizer.

Results. In this work, we fabricated nanocomposites based on silica-coated gold-silver nanocages [1] functionalized with the photodynamic sensitizer Yb2,4-dimethoxyhematoporphyrin. The hybrid nanoparticles combine several promising theranostic modalities: (i) an easy tunable plasmon resonance; (ii) a mesoporous silica shell that preserves the plasmon resonance from an aggregation; (iii) a combination of singlet oxygen generation with IR-luminescence band of photosebnsitizer [2]. Thus, these multifunctional nanocomposites seem an attractive theranostic platform for simultaneous IR-luminescence diagnostic, photothermal, and photodynamic therapy.

The other possible applications of gold nanocages is multiplexed solid-phase dot immunoassay [3]. Although current synthesis protocols provide a great variety of nanoparticle sizes, shapes, and structures, the Au/Ag alloy nanoparticles seem a convenient platform for the multiplexed dot immunoassay. Indeed, the galvanic replacement reaction with silver nanocube templates [1] allows subtle and robust tuning of the Ag/Au conversion ratio and the nanoparticle suspension color from yellow to blue. With these nanoparticles in hand, we were able to conduct a multicolor dot immunoassay in a proof-of-concept experiment with three types of molecular targets and biospecific probes. Without any instrumental detection, the assay sensitivity was about 20 fmol under naked eye examination.

References.

1. S. E. Skrabalak et. al. Gold nanocages: synthesis, properties, and applications. (2008) Acc. Chem. Res. 41. 1587–1595.
2. B. Khlebtsov et. al.Nanocomposites containing silica-coated gold-silver nanocages and Yb-2,4-dimethoxyhematoporphyrin: Multifunctional capability of IR-luminescence detection, photosensitization, and photothermolysis (2011) *ACS Nano.* 5. 7077–7089.
3. E. Panfilova et. al. Multiplexed dot immunoassay using Ag nanocubes, Au/Ag alloy nanoparticles, and Au/Ag nanocages (2012) *Nano Res*. (in press).

FLUORESCENT DNA PROBES: RECENT ADVANCES

Korshun V.A.*[1], Prokhorenko I.A. [1], Stepanova I.A.[1], Ryazantsev D.Y.[1], Zavriev S.K.[1], Tsybulsky D.A.[2], Kvach M.V. [2], Shmanai V.V. [2]
[1] *Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, Moscow, Russia;*
[2] *Institute of Physical Organic Chemistry, Minsk, Belarus*
e-mail: korshun@ibch.ru
** Corresponding author*

Key words: *real-time PCR, fluorogenic DNA probes, 'click' chemistry*

Motivation and Aim: A widely used type of the probe is a TaqMan, an oligonucleotide containing one fluorescent dye and one fluorescence quencher; the probe is complementary to a part of the amplified sequence capable of the 5'→3' degradation by the exonuclease activity of DNA polymerase thus amplifying the fluorescence intensity. On the other hand, some fluorophores (including fluorescein, FAM) are known to be prone to self-quenching when located near each other.

Methods: We employed modern conjugation chemistry methods (including 'click chemistry' postmodification [1]) to construct qrtPCR probes containing two fluorescent dyes (FAM and FAM or FAM and JOE [2]) and quencher(s), to improve fluorogenic properties of TaqMan probes.

Results: A series of fluorogenic probes containing one and two fluoresceins (FAM), or FAM–JOE pair, and one or two BHQ1 quenchers was designed and tested in real-time PCR using a ~300 bp-long natural DNA fragment as a template [3].

Conclusion: The structural prerequisites for lowering the probe background fluorescence and increasing the end-plateau fluorescence intensity were evaluated and discussed.

Acknowledgements: The research was supported by the Molecular and Cellular Biology Program of the Russian Academy of Sciences and Russian Foundation for Basic Research (project no. 10-04-00998).

References:

1. A.V. Ustinov et al. (2010) Modification of nucleic acids using [3+2]-dipolar cycloaddition of azides and alkynes, *Russ. J. Bioorg. Chem.,* **36:** 401-445.
2. D.A. Tsybulsky et al. (2012) 4',5'-Dichloro-2',7'-dimethoxy-5(6)-carboxyfluorescein (JOE): synthesis and spectral properties of oligonucleotide conjugates, *J. Org. Chem.,* **77:** 977-984.
3. D.Y. Ryazantsev et al. (2012) Two dyes and one or two quenchers DNA probes for real-time PCR assay: synthesis and comparison with a TaqMan™ probe, *submitted*.

NEW PYRENE EXIMER OLIGO(2'-O-METHYLRIBONUCLEOTIDE) PROBES FOR RNA DETECTION

Krasheninina O.A., Novopashina D.S., Ven'yaminova A.G.*
Novosibirsk State University, Novosibirsk, Russia
Institute of Chemical Biology and Fundamental Medicine SB RAS, Novosibirsk, Russia
e-mail: ven@niboch.nsc.ru
** Corresponding author*

Key words: *2'-bispyrene-labeled probe, eximer fluorescence, MDR1 mRNA*

Motivation and Aim: At present, the most appropriate techniques for identification of sequence, structure, quantity and function of NAs are various fluorescent assays. An application of these techniques implies the use of fluorescent-labeled probes which exhibit hybridization-induced changes in fluorescence spectra, possess high affinity to the NA targets and nuclease resistance. Here we present a new type of 2'-bispyrene-labeled oligo(2'-O-methylribonucleotides) as fluorescent probes for detection of RNA.

Methods: Series of novel 2'-bispyrene-modified oligo(2'-O-methylribonucleotides) (19 nt) were synthesized as described in [1]. The conjugates contained one insertion of 2'-bispyrenylmethylphosphorodiamidate derivative of ribonucleoside within the chain and "inverted" thymidine at the 3'-end. The sequences of probes were complementary to two accessible regions of MDR1 mRNA. The structures of 2'-bispyrene-modified probes were confirmed by mass-spectrometry, UV- and fluorescent spectroscopy. Thermal stability of duplexes of probes with model NA targets corresponding to MDR1 mRNA nucleotides 113-137 and 315-336 was investigated by thermal denaturation method. Steady state fluorescence emission spectra of the 2'-bispyrene-modified probes and their duplexes with corresponding NA targets were recorded and analyzed. Fluorescence emission quantum yields of probes and their duplexes with model short RNAs were determined. The hybridization of the most sensitive probes to model short RNA/DNA targets and 5'-terminal fragment of MDR1 mRNA (nucleotides 1-678) was investigated by fluorescence-monitored titration with an increasing target concentration.

Results: We found that the properties of the new probes varied depending on position of 2'-bispyrene-modified nucleoside. The UV melting studies indicate that the conjugates obtained form stable complexes with the model RNA/DNA. The fluorescence spectra of the probes displayed significant increase of eximer fluorescence intensity (480 nm) upon binding with both complementary short and extended (678-mer fragment of MDR1 mRNA) RNA targets.

Conclusion: The results obtained in this study indicate that the 2'-bispyrene-labeled oligo(2'-O-methylribonucleotide) probes can be useful tools for detection of specific RNA.

Acknowledgements: We appreciate M.A. Zenkova for supply of pMDR1 plasmid, A.A. Lomzov for carrying out the thermal denaturation experiments and V. V. Koval for recording mass spectra.

References:

1. O.A. Krasheninina, D.S. Novopashina, A.G. Venyaminova. (2011) Oligo(2′-O-methylribonucleotides) containing insertions of 2'-bispyrenylmethyl-phosphorodiamidate nucleoside derivatives as prospective fluorescent probes for RNA detection, Russ. J. Bioorg. Chem., 37: 273-277.

NEW ASPECTS OF OPTICAL METHODS FOR THE STUDY OF BLOOD IN DIAGNOSIS OF COLORECTAL CANCER

Kruchinina M.V.*[1], Voevoda M.I.[1], Peltek S.E.[2], Kurilovich S.A.[1], Kruchinin V.N.[3], Spesivtsev E.V. [3], Rykhlitskii S.V. [3], Volodin V.A.[3], Generalov V.M.[4], StarikovA.V. [5]
[1] *Institute of Internal Medicine SB RAMS, Novosibirsk, Russia;*
[2] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia;*
[3] *Institute of Semiconductor Physics SB RAS, Novosibirsk, Russia;*
[4] *SRC VB "Vector", Koltsovo, Novosibirsk region, Russia;*
[5] *Novosibirsk Regional Cancer Center, Novosibirsk, Russia*
e-mail: kruchmargo@yandex.ru
** Corresponding author*

Key words: *erythrocytes, optical methods, colorectal cancer*

Motivation and Aim. Aim of this work was to assess the possibility of optical methods for studying red blood cells and serum of patients with colorectal cancer (CC) for use in diagnosis. *Materials and Methods*. A total of 14 persons (8 women and 6 men) aged 49 to 72 years with CC, six of them with a tumor in the terminal stage (T3-4) after therapy (4 - after chemoradiotherapy, 2 - after combination therapy) has been surveyed. The metastases (in liver area) were detected in two patients, the remaining patients showed no metastasis. The degree of lymph node involvement in most of the patients had not been determined, the four matched the N1. The comparison group consisted of 12 people (48 to 69 years) without cancer and other internal diseases. Electric and viscoelastic parameters of erythrocytes (Er) were investigated with aid of electro-optical detection system of cells by dielectrophoresis. Optical properties of blood serum were researched by methods of ellipsometry, FTIR- and Raman-spectroscopy.

Results. In patients with CC rigidity, viscosity, electrical conductivity of Er, the index of aggregation, destruction were significantly higher and the average diameter of the cell, polarizability, capacitance of membranes, the dipole moment, the amplitude of deformation of Er – were lower than the control group ($p < 0{,}001$-$0{,}05$). The correlation of electrical conductivity with the presence of metastases ($r = -0{,}62$, $p < 0{,}03$), polarizability - with damage to the lymph nodes ($r = -0{,}57$, $p < 0{,}048$) were identified. A tendency of increasing the refractive index of thin films obtained from the serum in patients with CC was observed in considering the ellipsometric parameters ($p < 0{,}058$). The degree of heterogeneity, the thickness of the films were significantly higher in patients with CC compared with controls ($p < 0{,}032$-$0{,}05$) in correlation with the last stage of the disease ($r = 0{,}448$, $p < 0{,}05$). The intensities of the peaks of the IR spectra regions amide III (1250, 1310 cm^{-1}) and amide I (1685cm^{-1}) were greater in patients with CC compared with controls ($p < 0.005$-0.05). The area of the intermediate peak of 1418 cm^{-1} was, however, lower in CC ($p < 0.05$). The areas of peaks (1005-1520 cm^{-1}) in Raman spectra were significantly lower in patients with CC compared with healthy ($p < 0{,}006$-$0{,}045$), correlating with the stage of the process ($r = -0{,}82$, $p < 0{,}002$), and carried out anti-tumor therapy ($r = 0{,}905$, $p < 0{,}013$).

Conclusion. There were significant differences of the optical parameters of red blood cells and serum of patients with CC compared with healthy, in establishing the correlation of these parameters with disease stage, lymph node involvement, presence of metastases. Using a combination of optical methods for studying blood prospectively for the diagnosis of this cancer.

PRACTICAL APPROACH FOR SYTHESIS OF NON-NUCLEOTIDE PHOSPHORAMIDITE SYNTHONS WITH DESIRED FUNCTIONAL GROUPS CONNECTED BY ACHIRAL BACKBONE

Kuprushkin M.S.*, Pyshnyi D.V.
Institute of Chemical Biology and Fundamental Medicine, SB RAS, Novosibirsk, Russia
e-mail: kuprummax@gmail.com
** Corresponding author*

Key words: *non-nucleotide insertion, modified oligonucleotides*

Motivation and Aim: Nowadays modified oligonucleotides are common tools in a different research work and practical application. The majority of modifications can be routinely introduced in to DNA sequences during synthesis using non-nucleotide phosporamidite monomers bearing necessary functional groups. Although preparation of such synthons is individual and labor-consuming task in some cases. Thus development of universal methods for preparation of non-nucleotide synthons with required properties presents an actual challenge.

Methods and Algorithms: Combining of two previously published helpful techniques of application lactone moieties [1] and activated oxalate chemistry [2] in creation variety of phosporamidite monomers we develop new practical method of preparation non-nucleotide phosphoramidite synthons with desired functionality [3]. Our method involves individual syntheses of building blocks with distinct functions: a "Universal" block responsible for incorporation into DNA, and "Functional" block, bearing oligonucleotide modifier's residue(s). The "Universal" block is common reagent for various synthons preparation and allows sets of non-nucleotide monomers to be obtained with the same backbone but carrying different functional group(s). "Functional" block has to be an appropriate aliphatic amine bearing desired oligonucleotide modifier group(s).

Results, Conclusion: Using developed method we synthesized a row of different synthons: (a) branching points synthons, (b) Synthon with group for post-synthetic modification, (c) "Simple" modification unit, using commercially avaible aliphatic amine bearing pyrene group, and (d)"Complex" modification unit, using aliphatic amine substituted with two different groups – acridine moiety and hydroxyl group, blocked with TBDMS. Besides, we performed synthesis and characterization of linear and branched oligonucleotides with modification introduced in different positions.

References:

1. N.N. Dioubankova, A.D. Malakhov, D.A. Stetsenko, V.A. Korshun, M. J. Gait. (2002) (R)-2,4-Dihydroxybutyramide seco-Pseudonucleosides: New Versatile Homochiral Synthons for Synthesis of Modified Oligonucleotides, Org. Lett., **4**: 4607–4610.
2. A.M. Morocho, V.N. Karamyshev, N.N. Polushin. (2004) Methoxyoxalamido Chemistry in the Synthesis of Novel Amino Linker and Spacer Phosphoramidites: A Robust Means for Stability, Structural Versatility, and Optimal Tether Length, Bioconjugate Chem., **15**: 569–575.
3. M.S. Kupryushkin, D.V. Pyshnyi (2011) *Patent application №* RU 2011107337

HIGH BIOLOGICAL ACTIVITY OF DNA FRAGMENTS DELIVERED INTO CELLS WITH NANOMATERIALS. I. ANTIVIRAL PROPERTIES OF NANOCOMPOSITES CONSISTING OF TiO_2 NANOPARTICLES AND POLYLYSINE-DNA CONJUGATES

Levina A.S.*[1], Repkova M.N. [1], Ismagilov Z.R. [2], Shikina N.V. [2], Mazurkova N.A.[3], Zarytova V.F.[1]

[1] *Institute of Chemical Biology and Fundamental Medicine SB RAS, Novosibirsk, Russia;*
[2] *Institute of Catalysis SB RAS, Novosibirsk, Russia;*
[3] *GSC of Virology and Biotechnology"Vector", Novosibirsk region, Koltsovo*
e-mail: Levina@niboch.nsc.ru
** Corresponding author*

Key words: *nanocomposites, DNA-fragments, antiviral activity*

Motivation and Aim. Delivery of DNA-based drugs into cells is an important problem to be solved. Nanoparticles capable of penetrating through cell membranes may be used as transporters of DNA fragments into cells. The goal of this work was the study of antisense effect of nanocomposites consisting of TiO_2-nanoparticles and polylysine-DNA conjugates with an example of inhibition of influenza A virus.

Results. The prepared **TiO_2-PL-DNA** nanocomposites were studied for their antiviral activity against influenza A virus Aichi/2/68 (H3N2). Nanocomposites containing DNA fragments complementary to different regions of vRNA or complementary cRNA were examined. The most efficient nanocomposite bearing DNA fragment aimed to the 3'-noncoding region of vRNA reduced the virus production by >99.9% at 0.1 mkM and 0.2 mkM concentrations of DNA fragment in nanocomposite (neutralization index was 3.2 lg and 3.9 lg, respectively). This nanocomposite was more efficient by a factor of ~7 than the same oligonucleotides delivered with lipofectamin. Antiviral action of nanocomposites containing DNA fragment complementary and noncomplementary to viral RNA differed by a factor of ~1000. The proposed nanocomposites have a low cytotoxicity with the CC_{50} value being ~700 μg/ml. The selectivity index of the TiO_2-PL-DNA nanocomposite, i. e. the ratio of toxic and effective inhibitory doses, was evaluated to be ~140.

Conclusion. The proposed **TiO_2-PL-DNA** nanocomposites can be considered as an efficient tool for the delivery of DNA fragments into cells to reveal their biological activity as nucleic-acid-based drugs.

Acknowledgements. The work is supported by RFBR grant № 11-04-01408a and State contract № 16.512.11.2267 («2011-1.2-512-090-069») in Federal Program.

SYNTHESIS AND PROPERTIES OF MODIFIED OLIGODEOXYRIBONUCLEOTIDES CONTAINING NUCLEOTIDE RESIDUE WITH β-DIKETOGROUP AT C2' ATOM

Naberezhnov D.S.[1], Sekerina S.A.[2], Monakhova M.V.*[3], Romanova E.A.[2], Kubareva E.A.[2], Oretskaya T.S.[2,3]

[1] *Institute of Carcinogenesis, Blokchin Cancer Research Center;*
[2] *A.N. Belozersky Institute of Physico-Chemical Biology;*
[3] *Chemistry Department, Lomonosov Moscow State University, Moscow, Russia*
e-mail: monakhovamv@gmail.com
** Corresponding author*

Key words: *modified oligonucleotides, cross-linking, β-diketone, Nt.BspD6I, MutS*

Motivation and Aim: Recognition, formation and dissociation of highly specific nucleic acid-protein complexes take place during various biochemical processes. Oligonucleotides and their analogs have great potential as multi-purpose tools in defining these processes. Oligonucleotide decoys with intercross-linking activity are of especial interest. They contain a recognition site of a protein under study as well as an especially "built-in" chemically reactive group, which directs a cross-linking of functionally essential protein to the decoy. The strategy for the introduction of reactive β-diketogroup at selected positions of DNA strand was suggested. This work is devoted to the development of the methods of arginine affinity modification in proteins.

Results: During research 2'-amino-2'-deoxyribooligonucleotides were synthesized. For introduction of β-diketogroup 2'-amino-2'-deoxyribooligonucleotide was treated by 4,6-dioxoheptanoic acid at the presence of carbodiimide. The yield of the reaction product was close to the quantitative. The structure of modified oligonucleotide was confirmed by MALDI TOF analysis. Reaction ability of DNA duplexes containing β-diketogroup was demonstrated for two DNA binding proteins: nicking endonuclease BspD6I and MutS from methyl-directed mismatch repair system. Formations of covalent complexes of studied proteins with DNA duplexes containing β-diketogroup were revealed by Laemmli gel-electrophoresis. Optimal conditions of cross-linking were selected.

Conclusion: The method of DNA-recognizing protein cross-linking to DNA duplexes containing β-diketogroup was developed for investigation of DNA - protein complexes.

Acknowledgements

We thank Prof. L.A. Zheleznaya for the Nt.BspD6I isolation and Prof. P. Friedhoff for kindly provided MutS plasmid.

References:

1. N.G. Dolinnaya, E.M. Zubin, E.A. Kubareva, T.S. Zatsepin, T.S. Oretskaya (2009) Design and synthesis of 2'-functionalised oligonucleotides. Their application for covalent trapping the protein–DNA complexes, *Curr. Org. Chem.*, **13:** 1029–1049.
2. This work is supported by program "International Research Training Groups" (RFBR-DFG 11-04-91336 and DRK 1384) and RFBR grants (10-04-01578, 12-04-01399).

NANOWIRES AS BIO-CHEMICAL SENSORS

Naumova O.V.*, Fomin B.I., Nasimov D.A., Ilnitsky M.A., Popov V.P.
A.V. Rzhanov Institute of Semiconductor Physics, Novosibirsk, Russia
e-mail: naumova@isp.nsc.ru
** Corresponding author*

Key words: *nanowires, nanotechnology, biosensor*

Nano bio-chemical sensors are ones of the major topics in the research of fusion technologies combining bio- and nanotechnologies. Our safety and health are two reasons why we deal with such devices and require them to: 1) high sensitivity, 2) selectivity, and 3) possibility of mass production in the form of portable diagnostic systems.

This report is a review of the research results that focus on nanowires sensors.

The specific aspects addressed in this report are:

- advantages and disadvantages of nanowire sensors in comparison with other sensors on the base of nanotechnology,
- fundamental properties of NW sensors and
- limits of the detection for NW sensors to the different types of bio-particles.

DESIGN OF MULTICOMPONENT CONCATAMER-LIKE DNA NANOSTRUCTURES

Pyshnyi D.V.*[1,2], Vinogradova O.A.[1], Shevelev G.J.[1,2], Kuprushkin M.S. [1], Filippov N.S. [1,3], Lomzov A.A. [1,2]

[1] *Institute of Chemical Biology and Fundamental Medicine SB RAS, Novosibirsk, Russia;*
[2] *Novosibirsk State University, Novosibirsk, Russia;*
[3] *Institute of Semiconductor Physics SB RAS, Novosibirsk, Russia*
e-mail: pyshnyi@niboch.nsc.ru
** Corresponding author*

Key words: *modified oligonucleotides, DNA complexes, DNA nanostructures*

Motivation and Aim: Investigation of principles of multicomponent DNA complexes formation is highly important for DNA nanoarchitectonics. Rational design of DNA nanostructures is impossible without both detailed thermodynamic description of association process and structural features of binding DNA blocks. The aim of this work is the development of fundamental basis which allows one to construct vaious types of concatamer-like DNA duplexes.

Methods and Algorithms: Various approaches were used to characterize DNA nanostuctures and initial DNA blocks and multicomponent associates in terms of thermodynamic stability, spatial organization, size and topology.

Results: We have developed synthetic approach to obtain modified oligonucleotides which are capable to form bended and branched DNA duplexes. With the use of native and modified DNA blocks the features of concatamer-like duplexes formation were investigated. Equilibrium scheme of concatamers formation consist two different oligonucleotides has been developed. It was shown that in the general case, the stability of concatamer duplexes can not be obtained in explicit form. The algorithm of numerical solution of the problem is developed. A number of approaches that simplify the analysis of heat denaturation curves for concatamer complexes were considered. It was shown that stability of multicomponent DNA complexes as well as their structure can be predicted based on the properties of monomeric DNA blocks.

The work was supported by MCB program of RAS, RFBR and by grant from SB RAS.

CONVERGENCE OF BIOTECHNOLOGY AND NANOTECHNOLOGY AND ITS POTENTIAL APPLICATIONS IN AGRICULTURE, FOOD AND HEALTH CARE

Ramaswamy N.M.
Nanobiotechnology Programme
Kumaraguru College of Technology, Coimbatore, India
e-mail : nmrfrd@yahoo.co.in

Key words: *Agriculture, Bionanotechnology, Nanomedicine, Food Industry, Health Care, Regenerative medicine*

Introduction: The convergence of biotechnology and nanotechnology (Bionanotechnology / Nanobiotechnology) in the field of life sciences holds a tremendous potential to change the way we live, to revolutionize agriculture, food industry and health care. The main purpose of this presentation is to provide an overview of the recent advances and distinctive features of bionanotechnology and its applications in agriculture, food industry and health care.

Nanobiotechnology in Agriculture : Our first global challenge is to eradicate hunger and improve food security (FAO,2012). Nanobiotechnology have the potential to change the entire agriculture sector and food industry chain from production, conservation, processing, packaging, transportation etc. Nanobiotechnology concepts have proved its ability in modifying the genetic constitution of the crop plants for improvement in yield and quality produce. Nano capsules provide a more efficient means to distribute pesticides and fertilizers. Identification and tracking through commerce using implanted nanochips is important. Nanoparticles deliver growth hormones or vaccines to livestock. Nanobiotechnology support precision farming. In seed science, major approaches are nanopolymer for seed hardening, nanosensors, nanobarcodes, use of nanomagnetic particles, and use of carbon nanotubes.

Food and Processing Industry : Nanobiotechnology techniques are used during cultivation, production, processing or packaging of food. R & D in nanobiotechnology in agrifood sector includes the development of functional food, nutrient delivery systems and methods for optimizing food systems and methods for optimizing food appearance.

Health care and Nanomedicine: One of the great promises of nanobiotechnology is increased control over our personal health. Nanomedicine is the comprehensive monitoring, repair, construction and control of human biological system working from the molecule level using engineered nanodevices and nanostructures. The rapidly growing new fields of nanomedicine include molecular imaging, lab-on-a chip, quantum dots and other novel diagnostic tools, nanobiosensors, regenerative medicine, tissue engineering, drug targeting and delivery systems. Major therapeutic areas include cancer, cardiovascular diseases, Hepatitis.B, etc. Other scopes for nanomedicine are nano-ophthalmology and nano-dentistry. Nanobiotechnology also plays an important role in siddha, ayurvedha and homeo, systems of health care.

Conclusion: Future research and development in nanobiotechnology have high potential to benefit society through applications in agriculture, food industry and health care which will lead to predictive, preventive and personalized health care.

Reference :

1. Ramaswamy, N.M., Bridging Nano and Agrotechnology. Nanodigest, Vol.I, Issue 12, 2010, 42-43.

HIGH BIOLOGICAL ACTIVITY OF DNA FRAGMENTS DELIVERED INTO CELLS WITH NANOMATERIALS. II. ANTIVIRAL PROPERTIES OF NANOCOMPOSITES CONSISTING OF POLYLYSINE-TiO_2 CONJUGATES AND DNA FRAGMENTS

Repkova M.N.*[1], Levina A.S.[1], IsmagilovZ.R.[2], Shikina N.V. [2], Mazurkova N.A.[3], Zarytova V.F. [1]

[1] *Institute of Chemical Biology and Fundamental Medicine SB RAS, Novosibirsk, Russia;*
[2] *Institute of Catalysis SB RAS, Novosibirsk, Russia;*
[3] *GSC of Virology and Biotechnology"Vector", Novosibirsk region, Koltsovo*
e-mail: repk@niboch.nsc.ru
** Corresponding author*

Key words: *nanocomposites, DNA-fragments, antiviral activity*

Motivation and Aim. Nucleic acid fragments are promising candidates for drug therapy. Poor cellular uptake of DNA-based drugs necessitates designing delivery systems to facilitate cellular internalization. The ability of nanoparticles to penetrate through cell membranes is a promising way to solve the current problem of nucleic acid-based drug transfer into cells. The aim of this work is the study of biological activity of nanocomposites consisting of TiO_2-nanoparticles and DNA fragments with an example of inhibition of influenza A virus.

Results. We prepared **TiO_2-PL•DNA** nanocomposites consisting of TiO_2-nanoparticles covered with polylysine and DNA fragments bound to TiO_2-PL due to electrostatic interaction between amino groups of polylysine and phosphate groups of DNA. MDCK cells used in this work were infected with influenza A virus Aichi/2/68 (H3N2). Cytotoxicity of the proposed nanocomposites was evaluated (CC_{50}=600–700 μg/ml). Nanocomposite bearing DNA fragment aimed to segment 5 of viral RNA showed the high antiviral activity. Suppression of virus reproduction was 99.8% and 99.9% at 0.1 and 0.2 mkM concentrations of DNA fragment in nanocomposite, respectively (neutralization index was 3 lg and 3.2 lg, respectively). Nanocomposite containing random DNA sequence was much less efficient (neutralization index was 0.4 lg). Control samples of DNA fragments unbound to nanoparticles showed the insignificant antiviral activity.

Conclusion. High antiviral efficacy of the designed nanocomposites with an example of the inhibition of influenza A virus replication shows their potency as antisense drugs.

Acknowledgements. The work is supported by RFBR grant № 11-04-01408a and State contract № 16.512.11.2267 (Шифр «2011-1.2-512-090-069») in Federal Program.

NANOBIOCOMPOSITES SiO_2~DNTP CONSISTING OF SiO_2 NANOPARTICLES AND IMMOBILIZED dNTP

Shatskaya N.V.*[1], Vasilyeva S.V.[1], Levina A.S.[1], Repkova M.N.[1], Baiborodin S.I.[2], Silnikov V.N.[1], Zarytova V.F.[1]

[1] *Institute of Chemical Biology and Fundamental Madicine SB RAS, Novosibirsk, Russia;*

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

e-mail: shatskaya@googlemail.com

** Corresponding author*

Key words: *dNTP, SiO_2, nanocomposites, delivery*

Motivation and Aim: Existing approaches to chemotherapeutics for diseases associated with dysfunctional DNA replication classically involve nucleoside analogues that inhibit polymerase activity due to modification in the nucleobase and/or ribose moieties. These compounds must undergo multiple phosphorylation steps in vivo, converting them into triphosphates, in order to inhibit their targeted DNA polymerase. Little attention has been paid to triphosphates as drug platforms due to their expected poor deliverability. We suggest the use of SiO_2 nanoparticles as a vehicle for delivery of triphosphates into cells. Advantages of SiO_2 are there ability to penetrate cell membrane and low toxicity.

Methods: Commercial nanoparticles with the amino groups were treated with succinimide ester of ε-azidocaproic acid to introduce the azido groups. Nucleoside triphosphates were modified by introducing the triple bond at γ-phosphate. Modified dNTPs were attached to nanoparticles by the click-reaction between modified dNTPs and azido groups on the nanoparticle surface. The substrate properties were investigated using a synthetic primer and template as a model system.

Results: The method of dNTPs immobilization on SiO_2 nanoparticles was developed. Capacity of the prepared nanobiocomposites was evaluated to be 0.14 mkmol/mg. It was shown that nanocomposites SiO_2~$dNTP^{Flu}$ bearing the fluorescent label are able to penetrate through the cell membrane. Substrate properties of the proposed nanobiocomposites were tested using Klenov fragment and HIV reverse transcriptase. It was shown that nanocomposite SiO_2~dNTP was a substrate for these enzymes and provided DNA chain growth. The rate of polymerization with the use of SiO_2~dNTP or common dNTP differed insignificantly.

Conclusion: Results show a possibility of utilization of SiO_2 nanoparticles as a vehicle for delivery of triphosphates analogs into the cell. The fact that polymerase can recognize and incorporate such nanobiocomposite into growing DNA chain is of great importance. This makes it possible to use nanobiocomposites bearing nucleoside triphosphate analogs as promising therapeutic drugs.

VISUALISATION OF DNA NANOSTRUCTURES USING ATOMIC FORCE MICROSCOPY

Shevelev G. J.*[1,2], Vinogradova O. A. [1], Kuprushkin M.S. [1], LomzovA.A. [1,2], Sheglov D.V. [3], Latyshev A.V. [3], Pyshnyi D.V. [1,2]
[1] *Institute of Chemical Biology and Fundamental Medicine SB RAS, Novosibirsk, Russia;*
[2] *Novosibirsk State University, Novosibirsk, Russia;*
[3] *Institute of Semiconductor Physics SB RAS, Novosibirsk, Russia*
e-mail: shevelev@niboch.nsc.ru
** Corresponding author*

Key words: *AFM, modyfied oligonucleotides, DNA complexes, mica surface*

Motivation and Aim: Visualization of supramolecular DNA-structures formed by oligonucleotides is an important analytical goal. Atomic Force Microscopy (AFM) is a very perspective technique of biomolecular nanostructures analysis. The actual problem is the development of convenient visualization protocols of various DNA structures with high resolution. The aim of this work is the AFM analysis of spatial organization of concatemeric dsDNA stuctures formed by native and modified DNA blocks.

Methods and Algorithms: Visualization of DNA was carried out using an Atomic Force Microscope Multimode 8 (Bruker, Germany). All images were obtained in tapping mode on the mica (NT-MDT, Russia).

Results: The protocols of visualization of supramolecular DNA structure on mica surface by AFM were optimized. It is established that the addition of Ni^{2+} cations to the layer of freshly air-cleaved mica surface is minimized the perturbation of DNA associates structure. AFM images of dsDNA concatemers formed by native, bended or branched [1,2] oligonucleotides were obtained. The influence of structure of simple dsDNA-blocks on spatial organization of supramolecular DNA associates was demonstrated. The presence of non-nucleotide insert into oligonucleotide chain resulted in bending dsDNA-block and led to formation of self-limited circular DNA concatemeric structures. Interaction of branched oligonucleotides caused the formation of branched filaments as well as dense DNA network. The Work was supported by MCB program of RAS, RFBR and by grant from SB RAS.

References:

1. Vinogradova O.A. et al. Biorgan. Khim. 2009. V. 35, P. 384–396.
2. Kupryushkin M. S., Pyshnyi D.V. Patent application № RU 2011107337, 25.02.2011.

STUDY OF THE MECHANISMS OF GOLD NANOPARTICLES INTERACTION WITH HELA CELLS

Spitcyna Yu.E.*, Pyshnaya I.A., Lomzov A.A., Ryabchikova E.I.
Institute of Chemical Biology and Fundamental Medicine SB RAS, Novosibirsk, Russia
e-mail: fabaceae@yandex.ru
** Corresponding author*

Key words: *gold nanoparticles, clathrin-dependent endocytosis, caveolin-dependent endocytosis, nystatin, surface plasmonic resonance*

Gold nanoparticles (GNPs) are widely used in biology and medicine. Their chemical stability, biological inertness, absence of toxicity for cells and organisms make GNP promising candidates for targeted delivery of drugs into the cells. An understanding of "fine" mechanisms of the interaction of GNPs and a cell could result in the development of new sensing, diagnostic and treatment capabilities such as targeted drug delivery, gene therapy, magnetic resonance imaging contrast, and others. The aim of this research was to study the "fine" mechanisms of GNPs interaction with HeLa cell culture.

Methods: Spherical GNPs (14-15 nm) were prepared by citrate method and used for the study. MTT-assay was applied to determine the non-toxic concentration of GNPs suspension for studies with HeLa cell culture. Mechanisms of GNPs internalization and accumulation were examined by light- and electron microscopy. Role of caveolin-dependent endocytosis was evaluated using the nystatin which is a specific blocker of caveolin-dependent endocytosis. The samples were fixed in 4% paraformaldehyde, and routinely processed for transmission electron microscopy. Ultrathin sections were examined in JEM 1400 electron microscope (Jeol, Japan). Digital images were collected by Veleta camera (Olympus SIS, Germany).

Results: There were no visible changes of HeLa cell ultrastructure after treatment by GNPs (in absence of serum) and subsequent incubation during 5-30 min and 1-72 h. GNPs internalization occurs after 5 min of incubation at 37C: nanoparticles were detected inside clathrin-coated pits and vesicles as well as in caveolae. The GNPs were observed in complexes of fused caveolae and in early endosomes after 10-20 min. of incubation. The amount of these structures containing GNPs inside increased and reached the maximum at 30 min of incubation, GNPs were also found inside late endosomes (multivesicular bodies, MVB). The number of MVB containing GNPs as well as number of GNPs inside them increased during the incubation time. After 5h of incubation GNPs were detected in lysosomes and in MVB. The number of GNPs inside lysosomes increased and after 72h most of the nanoparticles were located in lysosomes. Nystatin blocked caveolin-dependent endocytosis in HeLa cells. The GNPs were detected only in clathrin-coated pits and early endosomes after 30 min of incubation, single particles were detected in MVB. The GNPs were not found in caveolae. The obtained results evidence for significant role of caveolin-dependent endocytosis in GNPs internalization by HeLa cells.

The ability of surface plasmon resonance induction in HeLa cells treated with GNPs was studied. Laser with power 100mW and wavelength 450 и 532nm was used.

Conclusion: The GNPs are able to enter HeLa cells by two different endocytosis mechanisms: clathrin- and caveolin-dependent.

Acknowledgements: The work was conducted in frames of the Integration Project № 9 SB RAS, and State Task of Russian Ministry of Education (Project # 4.3924.2011).

PATHWAYS OF GOLD NANOPARTICLES INTERNALIZATION BY DIFFERENT TYPES OF EUCARYOTIC CELLS

Spitcyna Yu.E., Pyshnaya I.A., Marchenko A.K., Razum K.V., Pyshnyi D.V., Ryabchikova E.I.*
Institute of Chemical Biology and Fundamental Medicine SB RAS, Novosibirsk, Russia
e-mail: lenryab@niboch.nsc.ru
**Corresponding author*

Key words: *gold nanoparticles, HeLa and hepatoma A1 cells, peritoneal macrophages, electron microscopy*

Nanoscale gold particles show great potential as imaging agents in living systems, drug delivery and photothermal therapy agents. Many applications of gold nanoparticles (GNPs) propose that the particles will be introduced into living organisms and will contact with living cells. The significant number of studies devoted to the interaction of GNPs and cells have been published, however the general understanding of the interactions between GNPs and cells remains poor, in particular early steps of the interaction. The aim of our work was examination of the pathways of GNPs entry into different types of eukaryotic cells.

Methods: Spherical GNPs (14-15 nm) were prepared by citrate method and used in all experiments. HeLa cells (continuous cell culture), hepatoma A1 cells (tumor cells), and peritoneal macrophages ("normal" organism cells) were grown in monolayer and treated by 10^4 GNPs/ml during 0, 10, 20, 30, 60 min., 2, 5 and 24 h. Hepatoma A1 cells and peritoneal macrophages were also treated by GNPs directly in mice peritoneal cavity, and the cells were sampled at 10, 20, 30, 60 min. and 2, 3, 5 and 24 h after GNPs injection. All cell samples were fixed in 4% paraformaldehyde, postfixed by osmium tetraoxide, and routinely processed for transmission electron microscopy. Ultrathin sections were examined in JEM 1400 electron microscope (Jeol, Japan). Digital images were collected by Veleta camera (Olympus SIS, Germany).

Results: All types of examined cells readily internalized GNPs however the pathways of the uptake differed. HeLa cells demonstrated clathrin- and caveolin-dependent endocytosis of GNPs which were accumulated in late endosomes and lysosomes. Hepatoma A1 cells *in vivo* trapped GNPs by macropinocytosis, phagocytosis and raft-dependent endocytosis, while peritoneal macrophages *in vivo* showed mainly phagocytosis and clathrin-dependent endocytosis. All examined cell types entrapped GNPs inside endosomes and lysosomes, no entry of the particles in cytoplasm and nucleoplasm was detected. The most active uptake of GNPs was observed in hepatoma A1 cells and peritoneal macrophages *in vivo*. Hepatoma A1 cells *in vitro* showed the same pattern of GNPs internalization, however total amount of the particles engulfed by cells was reduced in comparison with *in vivo* study. Macrophages kept high uptake of GNPs *in vitro*.

Conclusion: The obtained results show that different cell types possess different internalization pathways of the same GNPs. These data should be taken in account in biomedical researches and particularly in development of drug-delivery systems.

Acknowledgements: The work was conducted in frames of the Integration Project № 9 SB RAS, and State Task of Russian Ministry of Education (Project # 4.3924.2011).

PERSPECTIVES OF NUCLEIC ACID BIOSENSORS FOR MEDICAL APPLICATIONS; ELECTRICALLY READABLE BIOCHIPS FOR RAPID RNA ANALYSIS

Sprinzl M.
Laboratorium für Biochemie, Universität Bayreuth, Germany

Transduction of biochemical events arising from nucleic acid hybridisation to electrically readable signals is the objective of the presented research. The aim is to develop analytical devices (biochips) for use in diagnostics, biotechnology and environmental analysis.

Nucleic acid hybridisation directs a thermostable esterase (1) for binding to gold electrodes where an electrochemically detectable p-aminophenol is enzymatically synthesized and detected as a redox reaction – dependent current.

All reactions are performed directly on chips presenting 128 electrodes in a volume of ten microliters, or lower. Technical solutions and optimized biochemical procedures improving the sensitivity and specificity will be presented. Applications for detection of bacteria and microRNA will be discussed (2, 3).

Supported by German Federal Ministry of Education and Research in frame of "The Medical Valley, Europäische Metropolregion Nürnberg".

DNA ARCHITECTONICS: VISION OF THE FUTURE

Stetsenko D.A.*[1], Pyshnyi D.V. [2]
[1] *Department of Chemistry, University of Reading, P. O. Box 224, Whiteknights, Reading RG6 6AD, UK;*
[2] *Laboratory of Bionanotechnology, Institute of Chemical Biology and Fundamental Medicine, Novosibirsk, Russia*
e-mail: dmitry.stetsenko@ntlworld.com
** Corresponding author*

Key words: *DNA architectonics, bionanotechnology, dendrimer, RNA polyhedra*

Last five years have witnessed a starburst of DNA architectonics, a branch of DNA nanotechnology that employs DNA as a material to build nano- and microscale objects rather than in a common biological sense as a carrier of genetic information.[1] Those years have seen some spectacular examples of the use of double-stranded DNA as a futuristic construction material. In 2006 Rothemund has described a "DNA origami" technique of assembling complex DNA nanostructures.[2] Using that method, researchers from Denmark created a box with closeable lid made entirely of DNA.[3] A series of DNA polyhedra like DNA tetrahedron, dodecahedron, icosahedron and a DNA buckminsterfullerene were obtained.[4] Very recently, practical attempts were made of the use of DNA architectures as nanodevices.[5,6]

Therefore, one may consider DNA architectonics as perhaps the most vigorously evolving area of DNA chemistry nowadays. Yet, a striking feature of vast majority of the published works is the use of virtually unmodified DNA duplexes joined into higher order modules (Holliday junctions, double-crossover motifs, lattices etc). Thus, self-assembly of DNA molecules into a plethora of shapes and structures is driven exclusively by Watson-Crick complementary hydrogen bonding between pre-organised DNA modules as well as auxiliary oligonucleotide "staples". A consequence of that is the concomitant increase in size and complexity of DNA modules, where unmodified DNA is used. Enrichment of the methodology of DNA architectonics by elements from the arsenal of solid-phase oligonucleotide synthesis, chemoselective ligation and chemical DNA functionalisation in order to create new revolutionary DNA technologies will be discussed in the present talk. A particular emphasis is made on the use of dendrimers in template-directed assembly of DNA and RNA polyhedra, and the implications for nanorobotics and rational design of artificial life forms.

References:

1. O. A. Vinogradova and D. V. Pyshnyi (2012) *Russ. Chem. Rev.*, 81: 130-157.
2. P. W. Rothemund (2006) *Nature*, 440: 297-302.
3. E. S. Andersen *et al*. (2009) *Nature*, 459: 73-76.
4. Y. He *et al*. (2008) *Nature*, 452: 198-201.
5. S. Modi *et al*. (2009) *Nature Nanotechnol.*, 4: 325-330.
6. D. Bhatia *et al*. (2011) *Nature Commun*., 2: 1-7.

INORGANIC NANOPARTICLES AS DELIVERY VEHICLE FOR NUCLEIC ACIDS-BASED REAGENTS

ZarytovaV.F.*[1], Levina A.S.[1], Repkova M.N.[1], Ismagilov Z.R.[2], Shikina N.V.[2], Mazurkova N.A.[3], Evdokomov A.A.[3], Netesova N.A.[3], Baiborodin S.I.[4], Pavlova A.S.[1], Shatskaya N.V.[1]

[1] Institute of Chemical Biology and Fundamental Medicine SB RAS, Novosibirsk, Russia;
[2] Institute of Catalysis SB RAS, Novosibirsk, Russia;
[3] GSC of Virology and Biotechnology"Vector", Koltsovo, Novosibirsk region, Russia;
[4] Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia
e-mail: zarytova@niboch.nsc.ru
** Corresponding author*

Key words: *nanocomposites, DNA-fragments, antiviral activity*

Motivation and Aim. Nucleic-acid-based reagents are promising drugs for various diseases including infectious, cancer, neurological, cardiovascular, etc., based on defective nucleic acids. The selectivity of these drugs is provided by the complementary interaction between a nucleotide-based drug and a target nucleic acid. The goal of this work is the development of innovative technologies for the efficient and targeted action on intracellular genetic material.

Results. A new technology has been developed for the delivery nucleic-acid-based reagents into cells. It is based on the use of inorganic nanoparticles as transporters. Methods of immobilization of nucleic acid fragments onto titanium or silicium dioxide nanoparticles have been elaborated resulting in the formation of TiO_2~DNA or SiO_2~DNA nanocomposites. These nanocomposites were shown to deliver DNA into the cell cytoplasm and nucleus where they showed specific biological activity. DNA fragments fixed on nanoparticles are, apparently, resistant to intracellular enzymes. The used TiO_2 nanoparticles (4–6 nm) were not shown to be toxic for eukaryotic cells in a wide range of concentrations.

Conclusion. The proposed nanocomposites based on inorganic nanoparticles can be considered as an efficient tool for the delivery of DNA fragments into cells. These nanocomposites appeared to be effective agents for targeted action on intracellular genetic material.

Acknowledgements. The work is supported by RFBR grant № 11-04-01408a and State contract № 16.512.11.2267 (ШИфр «2011-1.2-512-090-069») in Federal Program.

BASE EXCISION REPAIR ENZYMES AS INSTRUMENTS FOR POSTGENOMIC BIOTECHNOLOGY

Zharkov D.O.
Institute of Chemical Biology and Fundamental Medicine SB RAS, Novosibirsk, Russia
e-mail: dzharkov@niboch.nsc.ru

Key words: *base excision repair, DNA glycosylases, DNA damage detection, genotyping, PCR, molecular cloning, bionanoengineering*

Motivation and Aim: Base excision DNA repair (BER) is a vitally important process that is responsible for removal of damaged bases from DNA, thus preventing mutagenesis, carcinogenesis, and cell death. Disturbances of BER are associated with cancer, aging, neurodegenerative and immune disorders. BER is initiated by one of a series of enzymes called DNA glycosylases, which recognize and excise specific types of damaged bases. The abasic (AP) site that remains in DNA is then nicked by AP endonucleases, and firther processed by DNA polymerases and DNA ligases. The unique substrate specificities and activities of BER enzymes make them promising for extension of a current repertoire of enzymatic tools used in molecular biology and biotechnology.

Methods and Algorithms: BER enzymes have been used in several practical applications, such as: detection of specific DNA lesions by crosslinking/immunoblotting, improving the quality of damaged DNA before PCR, highly efficient construction of covalently bound DNA/protein modules for bionanoengineering, detection of single-nucleotide polymorphisms, and molecular cloning without the restriction enzymes. Libraries of inhibitors have been screened for their ability to inhibit human and bacterial DNA glycosylases.

Results: Cross-linking of DNA glycosylases (Fpg, OGG1) specific for 8-oxoguanine, a widespread premutagenic base lesion, followed by immunodetection with antibodies against Fpg/OGG1 can be used to reveal 8-oxoguanine in DNA with high specificity and sensitivity. Treatment of model degraded DNA templates with a kit consisting of several DNA glycosylases, AP endonuclease Nfo, DNA polymerase β and T4 DNA ligase enhances the sensitivity of PCR. $NaBH_4$ crosslinking chemistry with bifunctional DNA glycosylases and their appropriate DNA substrates allows one to obtain DNA-protein cross-links that can later be ligated into multimodular nanoconstructions. DNA glycosylases specific for mismatches has been used to reveal single-nucleotide polymorphisms associated with susceptibility to several human diseases. Finally, a procedure has been designed for subcloning PCR fragments obtained with uracil-containing primers after generation of cohesive ends through cleavage with uracil-DNA glycosylase.

Conclusion: BER enzymes, in particular, DNA glycosylases and AP endonucleases, provide powerful tools for many areas of postgenomic biotechnology and medicine.

Acknowledgements: This work was funded by RAS Presidium (MKB 6.14) and RFBR (11-04-00807-a).

THE RESULTS OF THE GENE PRODUCT FOR THERAPEUTIC ANGIOGENESIS "NEOVASCULGEN" IN THE COMBINED THERAPY IN PATIENTS WITH CHRONIC ISCHEMIA OF LOWER EXTREMITIES

Deev R.V. [1], Dgavanadze N.M. [2], Filin A.A. [1], Nersesyan E.G. [3], Chervyakov U.V. [3], Staroverov I.N. [3], Kalinin R.E. [2], Kiselev S.L. [1], Isaev A.A. [1]

[1] *Human Stem Cells Institute , Moscow;*

[2] *Ryazan State Medical University, named by academician I. P. Pavlov, Ryazan;*

[3] *Yaroslavl State Medical Academy , Yaroslavl*

Chronic ischemia of lower extremities (CYLE) is still an unsolved problem in the structure of cardiovascular diseases. The prevalence is 5-8 % among the population older than 50 years and reaches 30 % for the population with risk factors (Hyperlipidemia, smoking, arterial hypertension or diabetes mellitus). Despite opportunities of the Surgery about 20-30 % cases remain inoperable. For this group of patients an arsenal of therapies can be replenished with a new drug – Neovasculgen (Human Stem Cells Institute). Its effect is based on the principle of therapeutic angiogenesis. The product is a nonviral, supercoiled, plasmid DNA, encoding human vascular endothelial growth factor. Purpose of the work : improve treatment outcomes of patients with chronic ischemia of lower extremities. Patients with obliterating atherosclerotic diseases of lower extremities arteries with II-a, II-b, III degrees of ischemia (according to a classification of A. V. Pokrovsky – Fontein) , participated in the first-second and second-b of the open randomized 1-2a and 2b-3 phases clinical trial (clinical groups – 125 patients ; control group – 25) . The drug was administered intramuscularly as closely as possible to the area of ischemia twice at intervals of two weeks. Patients were treated in three clinical centers : ГОУ ВПО «Ryazan State Medical University, named by academician I. P. Pavlov" ; Гуз «Yaroslavl Regional Hospital"; ФГУ «Russian Scientific Center of Surgery ", named by academician B. V. Petrovsky. Safety assessment has been carried out by fixation of subjective data, registration of adverse events and serious adverse events, and also by definition of clinical and laboratory measures of health status. The changes in the distance of painless walking (DPW) were determined as the main criterion, thus it's a fundamental element of CILE- classification and refers to the international recommendations of the first class of evidence. This figure was determined during the treadmill test. The effectiveness of secondary criteria was recorded and evaluated; they are: ankle-brachial index (ABI); percutaneously determined the partial pressure of oxygen (PDPO mm. Hg.). It was found that after 6 months of this treatment DPW increased to +149,47 m. from an average 135,3+/-102,37 m. to 284,7+/-242,02 m.; ABI increased from 76,69 m+/-9,96 mm. Hg. to 85,42+/-10,87 mm.Hg. Control of patients was continued for 6 months after the end of the observation period. During dynamic monitoring of patients of this group statistically significant results on the improvement of blood circulation in the low extremities in accordance the DPW-, ABI- and PDPO-results for the first year were obtained. Clinical improvement was observed in 77,8 % of cases. The majority of patients (61,1 %) moved to a lighter stage of the disease – Iib and 16,7 % moved to II stage, according to the clinical data. The data allow to conclude about effectiveness of drug therapy. It's noted marked improvement in the condition of patients with Iia-IIb-III stages of CYLE according to selected key performance criteria. Clinical and laboratory parameters, which were being studing during the entire period of treatment and monitoring of all patients, showed that there were no significant deviations from the original values. There weren't noted any adverse or serious adverse events associated with the using of the drug for the entire period of observation, indicating a good tolerability.

ARTIFICIAL POLY-CTL-EPITOPES DNA VACCINES

Karpenko L.I.*, ReguzovaA.Yu., Starostina E.V., Borobova E.A., Antonets D.V., Ilyichev A.A., Bazhan S.I.
State Research Center of Virology and Biotechnology Vector, Koltsovo, Novosibirsk, Russia
e-mail: karpenko@vector.nsc.ru
** Corresponding author*

Key words: *DNA vaccine, poly-CTL-epitope immunogen, HIV-vaccine, anti-melanoma vaccine*

Motivation and Aim: Polyepitope immunogens based on selected T-cell epitopes effectively induce virus-specific CTL-responses and can be used as components of the effective vaccine against different virus infection. Besides, design antitumor CTL- immunogenes is one the most promising approach to therapeutic anti-cancer vaccine.

Methods and Algorithms: As is known, CTLs recognize the viral protein antigens synthesized in the cell as short peptides (8–12 amino acid residues) associated with specific MHC class I molecules rather than full-sized proteins . These short antigenic epitopes are produced by the proteasome-mediated processing with subsequent transportation to the lumen of endoplasmic reticulum by the TAP proteins to bind to the formed MHC class I molecules. Therefore, in the case of the vaccines inducing CD8+ CTL responses, the most natural way to present the antigens to the immune system is DNA immunization.

Consequently, the following requirements should be provided to elicit a high level of CD8+ CTL responses at various stages of poly-CTL-epitope immunogen presentation to the immune system: (i) a high expression level of the gene encoding the target immunogen; (ii) an efficient proteasome-mediated processing of the expression product of target gene and transport of the formed peptides to the endoplasmic reticulum; (iii) the maximal expression of the complexes [peptide–MHC class I] by the antigen-presenting cells; and (iv) the specificity of MHC class I repertoire for the selected epitopes.

There are several strategies allowing this goal to be achieved, namely:

(1) Targeting proteins or polyepitope constructs for their proteasome degradation by genetic appending of Ub to their N-terminus;

(2) Inclusion of flanking residues between the determinants to ensure their proteasome liberation;

(3) Use of the TAP recognition motifs for flanking the epitopes;

(4) Use of the epitopes with high half-life time values for [peptides–MHC class I] complexes;

(5) The constructs should contain determinants for inducing the CD4+ CTL responses, as they are known to boost the CD8+ CTL responses;

(6) To increase the efficiency of the induction responses of CD4 + T-lymphocyte target poliepitope should contain additional signal sequences (N-terminal leader peptide and C-terminal fragment of protein LAMP-1); and

(7) The codons encoding the determinants should be optimized for the expression of target genes in human cells.

Results: For the designing artificial polyepitope immunogens we have developed software PolyCTL Designer. This software allows (a) to select the minimal set of CD8+ T cell epitopes with the known (or predicted) specificity towards various allelic variants of MHC class I molecules; (b) to predict binding affinity of peptides to TAP and, if necessary, the addition of N-terminal flanking residues to optimize this binding; (c) to provide engineering the optimal spacer sequences for each pair of peptides; and (d) to carry out the design of target polyepitope constructs. Using this software we have developed structure of artificial poly-CTL-epitope DNA vaccines able to induce high levels of the specific CD8+ CTL responses against HIV and melanoma.

CANCER CELLS TARGETING TOXICITY OF CROTAMINE, MILD TOXIN FROM BRAZILIAN RATTLESNAKE

Kerkis I.*[1], Rádis-Baptista G.[2], Pereira A.[1], Oliveira E.B. [3], Kerkis A.[1], Prieto da Silva A.R.B.[1], Yamane T.[4]

[1]Laboratory of Genetics, Butantan Institute, San Paulo, SP; [2]Institute of Marine Sciences, Federal University of Cearб, Fortaleza; [3]University of San Paulo, Ribeirao Preto Medical School, Ribeirao Preto, SP; [4]State University of Amazon, National Institute of Science and Technology, Environment, Energy and Biodiversity, Manaus, AM, Brazil.

e-mail: ikerkis@butantan.gov.br

** Corresponding author*

Key words: *crotamine, therapeutic cell penetrating peptide, drug development*

MOTIVATION AND AIM: Crotamine is a low molecular weight cationic polypeptide found in the venom of the South American rattlesnake (*Crotalus durissus terrificus*). The mature crotamine, the post-translationally modified protein precursor, (YKQCHKKGGHCFPKEKICLPPSSDFGKMDCRWRWKCCKKGSG), is comprised of 42 amino acids and is rich in basic residues (nine lysines and two arginines) with a net positive charge and high cationic character (pI ~ 10). Structurally, crotamine is stabilized by six cysteine residues that participate in three disulfide bonds (Cys4–Cys36, Cys11–Cys30 and Cys18–Cys37). Additionally, interchain disulfide bonds appear to contribute to crotamine self-association in solution and a type of supramolecular structure formation. Since 2004 our group, has accelerated the pace of discovery of crotamine, which was shown to be useful as a template for drug development with distinct biomedical applications.

METHODS AND RESULTS: The studies revealed that crotamine is a versatile molecule that is amenable to *in vitro* peptide engineering, protein evolution, and synthesis of structural derivates. The biological activities of crotamine include its ability (1) to cross the biological membrane and penetrate selectively into different types of actively proliferating cells; (2) to interact with several cell organelles, particularly the nucleus and sub-nuclear structures, (3) to deliver nucleic acids into eukaryotic cells *in vitro* and *in vivo*, (4) to act as anti-microbial compounds against lower eukaryotes and prokaryotes , as well as (5) crotamine has specific and selective antineoplastic cancer cell cytotoxicity both *in vitro* and *in vivo*. This peptide sensitivity against tumor cells is at least fivefold higher than against normal cells, when used at the same concentration. Therefore, *in vivo* crotamine delayed melanoma's implantation, inhibited tumor growth and prolonged the lifespan of the mice. Additionally, crotamine accumulated within tumor mass and traced metastatic invasion of B16-F10 cells, suggesting its possible applications as an imaging agent and metastasis marker in living organisms.

CONCLUSION: Crotamine is unique among several classes of venom peptides. Authentic features of crotamine seem to originate in the evolution and diversification of the ancestral structural scaffold β- defensin, which is shared among peptides of the innate immunity system of vertebrates and venom peptides.

TARGETING OF FUNCTIONALLY IMPORTANT SITES OF BACTERIAL RNA POLYMERASE BY SINGLE-STRANDED DNA APTAMERS

Kulbachinskiy A.V.*[1], Pupov D.[1], Miropolskaya N.A.[1], Esyunina D.M.[1,2]
[1]*Institute of Molecular Genetics, Russian Academy of Sciences, Moscow, Russia;*
[2] *Molecular Biology Department, Biological Faculty, Moscow State University, Moscow, Russia*
e-mail: akulb@img.ras.ru
** Corresponding author*

Key words: *RNA polymerase, aptamers, transcription inhibition, promoter recognition*

Motivation and Aim: RNA polymerase (RNAP) is the central enzyme in gene expression and a major target for transcription regulation. Bacterial RNAP is one of the most attractive targets for antibacterial therapy. Several RNAP-specific drugs, including antibiotic rifampicin, have been used for treatment of infectious diseases (such as tuberculosis) for many years. However, bacteria often develop antibiotic resistance due to RNAP mutations that make it insensitive to antibiotic action and decrease the therapeutic efficiencies of available antibiotics. Thus, characterization of novel inhibitors of bacterial RNAP is essential for development of novel antibacterial compounds.

Methods and Algorithms: To obtain novel RNAP inhibitors we selected single-stranded DNA aptamers against RNAPs from *Escherichia coli* and *Thermus aquaticus* using the SELEX procedure. The starting oligonucleotide library contained 32 nt long randomized region with about 10^{14} different sequences used for aptamer selection in each experiment. After 10-15 rounds of selection, the enriched libraries were cloned and individual aptamers were tested for their functional properties and RNAP inhibition.

Results: Several classes of high affinity ssDNA aptamers were obtained that specifically bind to the sigma subunits, core enzymes and holoenzymes of *E. coli* and *T. aquaticus* RNAPs. The apparent K_ds for different RNAP-aptamer complexes lie in the range of 10^{-11}-10^{-8} M and the complexes are highly stable at a wide range of conditions. Using various mutant RNAP variants, we demonstrated that the aptamers target different functionally important regions of RNAP, including parts of the DNA and RNA binding channels and the rifampicin-binding site. The aptamers were shown to prevent RNAP interactions with promoters and to inhibit RNAP activity with high efficiency. Finally, we developed aptamer-based assays that allow to perform screening of various RNAP inhibitors and to characterize their functional properties.

Conclusion: The RNAP-specific aptamers represent a novel type of nucleic-acid based transcription inhibitors that prevent RNAP-DNA interactions with high efficiency and specificity. The aptamers can be used for performing high-throughput screenings and characterization of novel antibacterial compounds and antibiotics.

Acknowledgements: This work was supported by the Russian Foundation for Basic Research grant 10-04-00925, by Federal Targeted Program "Scientific and scientific-pedagogical personnel of innovative Russia 2009-2013" (state contract 02.740.11.0771), and by the grant of the President of Russian Federation MD-618.2011.4.

BIENZYME NANOCONJUGATE FOR VASCULAR PROTECTION

Maksimenko A.V.*, Vavaev A.V., Turashev A.D.
Institute of Experimental Cardiology, Russian Cardiology Research-and Production Complex, Moscow, Russia
e-mail address: alexmak@cardio.ru
** Corresponding author*

Key words: *antioxidant therapy, superoxide dismutase, catalase, chondroitin sulfate, vascular wall, oxidative stress, hydrogen peroxide, bienzyme conjugate, vasoprotective effect*

Motivation and Aim Oxidative stress is known to play an important role in the pathogenesis of most cardiovascular disorders. We have suggested to use the covalently connected (via chondroitin sulfate /CHS/) system of antioxidant enzymes (superoxide dismutase /SOD/ and catalase /CAT/) for prevention and decrease of oxidative stress injurious action on vascular state. The aim of investigation was the determination of vasoprotective properties of obtained bienzyme SOD-CHS-CAT conjugate.

Methods and Algorithms The methods of platelet aggregation study, tone changes of rat ring arterial fragment, hemodynamics parameter registration at oxidative stress condition after intravenous infusion of hydrogen peroxide in rabbits and rats were used for evaluation of antioxidant efficacy of SOD-CHS-CAT action.

Results Antiplatelet potential of bienzymic conjugate was elicited due to its antiaggregation effect manifested through the connection of enzyme activities and an acquired supramolecular structure. The influence on arterial fragment tonus was the same as for SOD and CAT in native and conjugated form. Blood pressure and heart rate were significant and effectively normalized with SOD-CHS-CAT conjugate in rabbits and rats after hydrogen peroxide administration as perturbance agent of model oxidative stress.

Conclusion The SOD-CHS-CAT conjugate was effective for application in chronic prophylaxis regime and eligible for preventive therapeutic using. The development of enzyme conjugates destined for therapy can be medically significant, as a promising approach for the creation and release of new generation drugs. Bienzyme SOD-CHS-CAT conjugate proves interest object of computational biology at present.

Acknowledgements This work was supported partially by the State Contract 8/3-280H-10, the RFBR grants 09-04-00023 and 09-08-00023, Ministry of Health Care and Social Development of Russian Federation.

The authors are grateful to adjacent department coworkers of CardioCenter, Moscow for enlivening work cooperation.

Reference:

1. A.V. Maksimenko, A.V. Vavaev, L.I. Bouryachkovskaya, V.P. Mokh, I.A. Uchitel, V.L. Lakomkin, V.I. Kapelko, E.G. Tischenko. (2010) Biopharmacology of enzyme conjugates: vasoprotective activity of supramolecular superoxide dismutase-chondroitin sulfate-catalase derivative, Acta Naturae, 2(4): 82-94 (English edition), 90-103 (Russian edition).

REVERSE GENETICS AND ANTI-CYTOKINE THERAPY

Nedospasov S.A.
Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, and Lomonosov Moscow State University, Moscow, Russia;
German Rheumatism Research Center (DRFZ), a Leibniz Institute, Berlin 10117, Germany

Methods of "forward genetics" allow to identify genetic locus and eventually DNA sequence variation responsible for a mutant phenotype. Methods of reverse genetics uncover phenotypic features resulted from inactivation or mutation of a known gene or genomic sequence. Genetic knockouts, knockdowns, knocked-in mutations, as well as transgenosis, - all fall into the latter category. Mice are attractive mammalian model organisms which currently are best suited for reverse genetics. Additionally, human diseases can be modeled in mice. We are interested in physiological functions of TNF and in the consequences of genetic or pharmacological TNF ablation. TNF is the founding member of a large family of cytokines, therefore, the issue of non-redundant functions of a single gene is of significant interest. In particular, TNF possesses many beneficial physiological functions, related to host defense and homeostasis of the immune system. However, when overexpressed TNF may become pathogenic and is implicated in several autoimmune diseases in humans. Panels of mice with conditional ablation of TNF helped to define non-redundant functions of a particular cellular source of this cytokine in health and disease. We now want to extend these findings to novel modalities of anti-cytokine therapy.

.

CHARACTERIZATION OF CYTOTOXIC FACTOR OF HUMAN MILK, LACTAPTINE, AND DESIGN OF ITS RECOMBINANT ANALOGUES

Potapenko M.O.*, Savelyeva A.V., Fomin A.S., Semenov D.V., Koval O.A., Kuligina E.V., Richter V.A.
Institute of Chemical Biology and Fundamental Medicine SB RAS, Novosibirsk, Russia
e-mail: miroslava_potapenko@yahoo.com
** Corresponding author*

Key words: *human milk, bioactive peptides, apoptosis*

Motivation and Aim. Human milk is an abundant source of biologically active peptides. Lactaptin, a fragment of κ –casein, is one of biologically active peptides of human milk that induces apoptotic death of cultured human breast adenocarcinoma MCF-7 cells. Recently we have characterized lactaptin as a 66 – 123 fragment of κ –casein with molecular weight 8.6 kDa. Unfortunately the precise amino acid sequence of investigated peptide remains uncertain.

The objective of the present work was the precise identification of N-ternimal sequence of lactaptin and development of recombinant analogues, corresponding to primarily structure of natural peptide.

Methods and Algorithms. For that purpose native peptide extracted by chromatography was analyzed by degradation of amino acids by Edman and by MALDI-TOF.

Results and Conclusion. Results of lactaptin sequencing by Edman and MALDI-TOF analysis suggest two possible N-terminal sequences, corresponding to 31 – 102 and 49 – 122 fragments of human κ-casein.

Based on our finding, cDNAs corresponding to 31 – 102 and 49 – 122 fragments of κ-casein are cloned in *pTwin1* and *pQE30* expression systems. Expression of these vectors in *E. coli* cells allows receiving recombinant proteins corresponding to determined amino acid sequences of natural lactaptin for further investigation of their bioactive properties.

APTAMER-BASED METHODS FOR SCREENING AND CHARACTERIZATION OF NOVEL INHIBITORS OF BACTERIAL RNA POLYMERASE

Pupov D.*[1], Esyunina D.M.[1,2], Kulbachinskiy A.V.[1]
[1] *Institute of Molecular Genetics, Russian Academy of Sciences, Moscow, Russia;*
[2] *Molecular Biology Department, Biological Faculty, Moscow State University, Moscow, Russia*
e-mail: danila@pupov.ru
** Corresponding author*

Key words: *RNA polymerase, aptamers, transcription inhibitors, in vitro transcription assay, high throughput screening, antibiotics*

Motivation and Aim: Transcription is an essential step in gene expression in all living organisms. Bacterial RNA polymerase (RNAP) is the principal enzyme of transcription and one of the most attractive targets for antibacterial compounds that may be used for treatment of infectious diseases. However, only a few available antibiotics affect RNAP. Development of novel high-performance and low-cost approaches to the screening of RNAP inhibitors is important for creation of new antibiotics. The aim of this work was to obtain specific aptamers to RNAP and to develop new aptamer-based approaches to the screening of transcription inhibitors.

Methods and Algorithms: For selection of aptamers, we used the SELEX protocol, followed by design of optimized ligands, testing of their interactions with RNAP and functional properties. Analysis of the effects of aptamers on transcription and development of aptamer-based methods of analysis of RNAP-inhibitor interactions were performed using fluorescently and radioactively-labeled aptamers and *in vitro* transcription assays.

Results: We developed several different approaches for screening of transcription inhibitors using ssDNA aptamers obtained to the core and holoenzymes of RNAPs from *Escherichia coli* and *Thermus aquaticus*. Core RNAP-specific aptamers were shown to have hairpin/G-quartet structures and bind at the natural DNA-binding sites in the main RNAP channel. Holoenzyme RNAP-specific aptamers also have hairpin structures, contain specific promoter elements and target the promoter recognition sites of RNAP. These aptamer properties allowed us to design several tests for analysis of transcription inhibitors affecting different steps of RNA synthesis, including promoter recognition, transcription initiation and elongation. The principles of these methods include analysis of aptamer-inhibitor competition, measurements of fluorescence quenching and FRET. The high sensitivity and operability of these methods was confirmed using known RNAP-specific antibiotics.

Conclusion: Aptamers against bacterial RNAP are a promising new class of RNAP-binding molecules due to their high affinity, specificity and the ease of *in vitro* selection. The aptamer-based assays developed in this work can be adapted for performing highly efficient screenings of novel transcription inhibitors.

Acknowledgements: This work was supported by Federal Targeted Program "Scientific and scientific-pedagogical personnel of innovative Russia 2009-2013" (state contract 02.740.11.0771), Russian Foundation for Basic Research (grant 10-04-00925) and by the grant of the President of Russian Federation MD-618.2011.4.

ZYBODIES: MULTI-SPECIFIC, ANTIBODY-BASED THERAPEUTICS

Roschke V.V.
Zyngenia, Inc., Gaithersburg, MD, USA
e-mail: vroschke@zyngenia.com

Zybodies comprise fully functional monoclonal antibodies with short target-binding peptides recombinantly fused to the termini of the mAb heavy and light chains. The production and characterization of trastuzumab-based Zybodies will be presented. These Zybodies simultaneously target multiple signaling pathways, inhibit tumor cell proliferation, and demonstrate superior efficacy in xenograft tumor models. Applicability of the technology to other antibody scaffolds, such as adalimumab, for treatment of immune mediated inflammatory diseases will be also discussed. The data establish Zybodies as stable macromolecules that coordinately target multiple biological pathways, thereby conferring novel functional properties not achievable with conventional mAbs.

TRANSFECTION OF HUMAN CELLS BY ARTIFICIAL RNA-NUCLEOPHOSMIN 1 COMPLEXES

Savelyeva A.V.*[1,2], Semenov D.V.[1], Koval O.A.[1,2], Rabinov I.V.[1], Kuligina E.V.[1], Richter V.A.[1]

[1] *Institute of Chemical Biology and Fundamental Medicine SB RAS, Novosibirsk, Russia;*
[2] *Novosibirsk State University, Novosibirsk, Russia*
e-mail: savelevaav@gmail.com
** Corresponding author*

Key words: *NPM 1, box C/D snoRNA, AluY RNA, MCF-7 cells*

Motivation and Aim: Homeostasis is the property of a system that regulates its internal environment and tends to maintain a stable, constant condition of properties. Homeostasis maintenance plays a key role for existence of multicellular organisms. Cell-to-cell communication – one of the most important homeostatic mechanisms. Findings over the past several years identified a new cell-to-cell communication pathway based on horizontal transfer of secreted noncoding RNA (including micro-RNA that regulate the post-transcriptional silencing of protein-coding genes in eukaryotes). Such noncoding RNA, also called "circulating RNA", is stabilized outside the cells as a component of exosomes, microvesicles, high-density lipoproteins and ribonucleoprotein complexes. Wang *et al.* showed that one of exported proteins nucleophosmin 1 (NPM1) could play a role in protecting secreted miRNAs from degradation.

In the present work the aim to research is the influence of NPM1 on human MCF-7 cells transfection efficacy.

Methods and Algorithms: The plasmid for expression NPM1 was constructed, using pET23a vector (Novagen). *E. coli* strain producing recombinant NPM1 was obtained. The protein was purified from cell lysate via affinity Ni-NTA chromatography. Structure of recombinant analogue NPM1 was verified by MALDI-TOF of NPM1 tryptic fragments. The influence of recombinant NPM1 protein on MCF-7 cells viability was estimated by MTT assay. The ability of recombinant NPM1 to associate with artificial U25 box C/D RNA and AluY RNA was analysed by the electrophoretic mobility shift assay. Human epitheliocytes MCF-7 were transfected with the complex of artificial analogue of U25 box C/D snoRNA or AluY RNA with NPM1.The efficacy of snoRNA delivery was evaluated by real-time RT-PCR.

Results: It was found, that NPM1 didn't influence on MCF-7 cells viability during the transfection. NPM1 interacts with the artificial U25 box C/D snoRNA and AluY RNA. After transfection of MCF-7 cells with artificial complexes of AluY RNA-NPM1 we determined that NPM1 increases the efficacy of AluY RNA delivery into human cells in ≈ 30 times. However, NPM1 didn't significantly affect the penetration of artificial U25 box C/D snoRNA into MCF-7 cells.

Conclusion: Our result show that recombinant NPM1 could effectively bind artificial RNA and promote the penetration of highly structured RNA into MCF-7 cells.

Availability: The direct application of our research is a development of new transfection agent for biological and medical researches.

Acknowledgements: The work is supported by the Russian Foundation of Basic Research grants: № 10-04-01386-a, № 10-04-01442-a, "OPTEC" grant for support young scientists.

ARTIFICIAL BOX C/D RNAs AS GENE EXPRESSION MODULATORS

Stepanov G.A.*, Semenov D.V., Koval O.A., Kuligina E.V., Rabinov I.V., Richter V.A.
Institute of Chemical Biology and Fundamental Medicine SB RAS, Novosibirsk, Russia
e-mail: stepanovga@niboch.nsc.ru
** Corresponding author*

Key words: *box C/D small nucleolar RNAs, heat shock cognate protein 70, gene expression modulation*

Motivation and Aim: Box C/D small nucleolar RNAs (snoRNA) are commonly known to guide the site-specific 2'-O-ribose methylation of nucleotides in eukaryotic RNA, such as rRNAs and snRNAs. Eukaryotic RNA to be a natural target of snoRNA-dependent modification is selected by complementary interaction. Hundreds of mammalian snoRNAs haven't guide sequence complementary to rRNA as well as to snRNA so they have unknown function and are referred to as orphan snoRNAs. Some of those orphan RNAs are complementary to pre-mRNAs or mRNAs. Recently some box C/D snoRNAs were reported to regulate post-transcription modifications and alternative splicing of pre-mRNA. The aim of this study is to investigate artificial box C/D RNAs influence on gene expression of human cells.

Methods and Algorithms: In this study artificial U24 box C/D snoRNA analogs directed to nucleotides of pre-mRNA, and mature mRNA HSPA8 (encoding human heat shock cognate protein 70, hsc70) are designed and synthesized. Artificial RNAs contained target recognition regions directed to following pre-mRNA nucleotides: the branch point adenosine; the first and the last nucleotides of the second intron; the donor and the acceptor nucleotides of the spliced exons. Level of mRNA HSPA8 was determined with real time RT-PCR method. Furthermore we analyzed transcriptome of MCF-7 cells transfected by artificial snoRNAs with Illumina HT microarrays.

Results: It was found that transfection of cultured human cells by artificial snoRNA targeted to pre-mRNA induced suppression of mRNA-target in snoRNA-concentration-dependent manner. With increasing transfection time from 6 to 72 h, mRNA HSPA8 amount was determined to be reduced to 10%-level. U24 box C/D RNA analog directed to the branch point adenosine of the pre-mRNA HSPA8 second intron prompted the strongest degree mRNA-target downregulation, compared with other snoRNAs analyzed. The microarray analysis followed by RT-PCR analysis showed reliable upregulation of interferon-induced genes indicating that artificial box C/D RNAs caused the innate immune response of human cells, with STAT1-dependent genes level increasing. Moreover transfection by artificial snoRNAs was found to decrease MCF-7 human adenocarcinoma cell viability.

Conclusion: So the transfection of cultured human cells by artificial snoRNAs directed to pre-mRNA are shown to lower mRNA-target level.

Availability: The data indicate that the snoRNA structure is a promising model to create artificial gene expression regulators.

Acknowledgements: The work was supported by RFBR grants № 10-04-01386-a and № 10-04-01442-a, OPTEC grant for support of young scientists.

EXPRESSION AND ISOLATION OF THE PARKIN PROTEIN CONSTRUCT TV7 THAT DEFINES THE VIOLATION OF E3 LIGASE ACTIVITY IN HEREDITARY PARKINSON DISEASE

Tamarkin M.A.*, Kachalova G.S., Lipkin A.V., Popov V.O.
A.N. Bach Institute of Biochemistry of RAS, Moscow, Russia
e-mail: mtamarkin@mail.ru
** Corresponding author*

Key words: *PARK2; TV7 splice variant; parkin; Parkinson's disease; ARJP*

Motivation and Aim: Mutations in parkin, protein encoding by PARK2 gene, resulted in the loss of parkin function and accumulation of parkin substrates in neurons what lead to ARJP (1). Parkin functions as an E3 ligase - integral component of the cytoplasmic ubiquitin-proteasomal protein degradation pathway. The large size of the human parkin gene - 1380 kb is due to the presence of long length 11 introns (1). Alternative splicing variants of parkin might display different functions and regional distribution and could be associated to the ARJP phenotype (2).

The aim of this study is to isolate and characterize TV7 splice-variant including the unique parkin domain (UPD) flanked by ubiquitin-like domain (Ubl) and two first of four zinc-binding domains (RING0, RING1, IBR, RING2). The characterization of TV7 transcript of PARK2 could contribute to the elucidation of the mechanism of the functioning of parkin on a molecular level.

Methods and Algorithms: Parkin gene was synthesized (GeneArt). Using PCR techniques the TV7 splice-variant was amplified and cloned into pGEX-6P1 vector (GE Healthcare) between BamHI and XhoI sites. The described TV7 construct is IPTG inducible and was overexpressed in E. coli strain *BL21DE3pLysS* and then purified using affinity GST-chromatography (GE Healthcare) followed by size exclusion chromatography. The protein purity evaluated by SDS-PAGE.

Results: TV7 transcript was cloned and expressed in E.coli and purified with the yield of 4mg soluble protein from 1g of cells biomass.

Conclusion: The isolation of soluble and stable folded purified transcript variant of parkin protein provide an excellent starting point for understanding the role of UPD and RING0 domains in hereditary Parkinson disease.

Acknowledgements: We gratefully acknowledge Evgeniy Cherkashin, Anna Cherkashina and Aleksey Lipkin for their support.

References:

1. Velia D'Agata1 and Sebastiano Cavallaro, (2004), Parkin Transcript Variants in Rat and Human Brain, Neurochemical Research, Vol. 29, No. 9. 1715–1724
2. Ditte Bjerre, Lone Bruhn Madsen, Christian Bendixen, Knud Larsen, (2006) Porcine Parkin: Molecular cloning of PARK2 cDNA, expression analysis, and identification of a splicing variant, Biochemical and Biophysical Research Communications 347: 803–813.

CHIMERIC ANTIBODIES AGAINST TICK-BORNE ENCEPHALITIS VIRUS

Tikunova N.V.
Institute of Chemical Biology and Fundamental Medicine SB RAS, Novosibirsk, Russia
e-mail: tikunova@niboch.nsc.ru

Antibodies are used for therapy for a long time due to their exceptional properties – high specificities, availability of effector functions, and involvement of natural mechanisms in pharmacokinetics of administrated antibodies. Currently, serum immunoglobulin is used for tick-borne encephalitis treatment and, like other preparations obtained from donor's blood, it has some disadvantages. In order to produce an alternative preparation, a chimeric antibody against glycoprotein E of tick-borne encephalitis virus was developed in ICBFM SB RAS.

A panel of neutralizing mouse monoclonal antibodies was tested in the model animal protection experiments using lethal doses of tick-borne encephalitis virus. A chimeric antibody binding protein E with nanomolar affinities was constructed based on the variable domains of protective mouse monoclonal antibody and human IgG constant domains. Then, stable CHO cell line producing this antibody was obtained and procedure of its purification from culture medium was worked out. It was shown in the model animal experiments that protectivity of this chimeric antibody was higher than protectivity of commercial preparation of serum immunoglobulin.

MONOCLONAL ANTIBODIES AGAINST THE EXTRACELLULAR FRAGMENT OF CONNEXIN-43: A POTENTIAL ANTITUMOR ACTIVITY IN HIGH-GRADE GLIOMAS

Yusubalieva G.M.[1,2], Baklaushev V.P. [1,2], Gurina O.I. [2], Grinenko N.Ph.[2], Chekhonin V.P. [1,2]

[1] Pirogov Russian State Medical University, Department of Medicinal Nanobiotechnology, Moscow, Russia;
[2] Serbsky National Research Center for Social and Forensic Psychiatry, Ministry of Health and Social Development of the Russian Federation, Department of Fundamental and Applied Neurobiology
e-mail: gaukhar@gaukhar.org

Key word: *tumor, glioma, connexin 43*

Motivation and Aim: The therapeutic effectiveness of monoclonal antibodies against the E2 extracellular loop (MAbE2) of connexin 43 (Cx43) was estimated in experimental rat model of intracranial high-grade glioma. The role of Cx43 in the invasion of glial tumors is being intensely studied. There are some experimental evidences that Cx43 can activate the invasion of human glioblastoma multiforme *(Oliveira et al., 2005). In particular, it has been demonstrated that Cx43-positive C6 glioma* cells have a higher migration capacity than Cx43-negative cells and can be more resistant to oxidative stress and a number of other damaging factors (Bates et al., 2007). We suppose that MAbE2Cx43 could block the formation of heterologous gap junctions between glioma cells and astrocytes in the peritumoral zone of glioma invasion. The purpose of this study was to evaluate the influence of intravenous administration of MAbE2Cx43 on human gliomablastoma model (C6 glioma) in rats.

Seventy adult female Wistar rats (200–220 g) with experimental glioblastoma were involved in this experiment and randomized per 3 groups. Glioblastoma were initiated by intracerebral stereotactic implantation of C6 glioma cells ($5*10^5$ cells per animal) to the striatum region of ketamine-anesthetized (100 mg/kg) rats. The treatments were administered via femoral vein injections 3 times at 7-day intervals. The first group of rats with C6 glioma was received the MAbE2Cx43 treatment. The second and third groups were injected with non-specific IgGm and saline, respectively. The dynamic volume of intracranial glioma was estimated by magnetic resonance imaging (MRI). The survival function was analyzed using Kaplan-Meier estimator.

The results of the dynamic MRI morphometry revealed the significant decrease of volume of glioma during the MAbE2Cx43 therapy. Thus, after treatment of rats using the MAbE2Cx43 the average volume of glioma was twice smaller in comparison with all controls. Moreover, the results of Kaplan-Meier analysis revealed that the survival time of rats was significantly increased as compared with control groups. All rats of control groups died after 25-28 days after implantation of C6 glioma cells. Several rats with C6 glioma after MAbE2Cx43 therapy (approximately 20%) fully recovered. It is important to note, that survived animals were able to produce the offspring which denote the absence of recurrent tumor. *Conclusion:* We demonstrated that MAbE2Cx43 therapy of C6 glioma model can decrease the tumor growth as well as increase the lifespan. The experimental results showed the possibility to use MAbE2Cx43 for treatment of high-grade glioma progression.

CONSTITUTIVE ANDROSTANE RECEPTOR (CAR): XENOSENSOR AND TARGET FOR THERAPY

Yarushkin A.A.*[1], Kazakova Y.A. [1,2], Pustylnyak V.O. [1,2]
[1] *Institute of Molecular Biology and Biophysics SB RAMS, Russia;*
[2] *Novosibirsk State University, Novosibirsk, Russia*
** Corresponding author*

Key words: *constitutive androstane receptor, PEPCK, G6Pase, glucose level, gluconeogenesis*

Motivation and Aim. The constitutive androstane receptor (CAR, NR1I3) has a central role in detoxification processes, regulating the expression of a set of genes involved in metabolism. The dual role of CAR as both a xenosensor and as a regulator of endogenous energy metabolism has recently been accepted. Recent studies suggested that CAR is involved in lipogenesis and gluconeogenesis, showing anti-deabetic and anti-obesity effects. Here, we investigated the mechanism of transcriptional regulation of the glucose metabolising genes phosphoenolpyruvate carboxykinase (*PEPCK*) and glucose-6-phosphatase (*G6Pase*) by a highly effective CAR activator, 2,4,6-triphenyldioxane-1,3 (*cis*TPD) and examined the long-term effect of *cis*TPD on glucose metabolism and homeostasis in normal diet and high-fat diet rats.

Methods and Algorithms. Real-time PCR, Western-blot, ChIP assay, Fasting blood glucose analysis, Glucose tolerance test.

Results. It was shown that expression of the gluconeogenic genes *PEPCK* and *G6Pase* was repressed by *cis*TPD treatment under fasting conditions. Western-blot analysis demonstrated a clear reduction in the intensity of PEPCK and G6Pase immunobands from the livers of *cis*TPD-treated animals relative to bands from the livers of control animals. Chromatin immunoprecipitation assays demonstrated that *cis*TPD prevents the binding of FOXO1 to the insulin response sequences in the *PEPCK* and *G6Pase* gene promoters in rat liver. Moreover, *cis*TPD-activated CAR inhibited HNF-4 transactivation by competing with HNF-4 for binding to the HNF-4 -binding element (DR1-site) in the gluconeogenic gene promoters. Long-term activation of CAR by *cis*TPD significantly reduces fasting blood glucose levels in both normal diet and high-fat diet rats. In *cis*TPD-treated normal diet and high-fat diet rats, the hepatic expression of *PEPCK* and *G6Pase* genes and protein levels were significantly inhibited, consistent with the decreased fasting glucose in these animals. Moreover, *cis*TPD treatment improves glucose tolerance in high-fat diet rats.

Conclusion. Thus, our results are consistent with the hypothesis that the *cis*TPD-activated CAR participates in the regulation of the gluconeogenic genes *PEPCK* and *G6Pase*. The metabolic benefits of CAR activation by *cis*TPD may have resulted from the effect of inhibition of gluconeogenesis. It is reasonable to conclude that CAR may be a target to prevent or suppress high level of glucose and improve glucose homeostasis and *cis*TPD may be a novel potential therapeutic tool for the regulation of gluconeogenesis.

Acknowledgements. This work was supported by the Russian Federal Program of the Ministry of Education and Science of the Russian Federation "Scientific and pedagogical manpower of innovational Russia" (GK 14.740.11.1054 and GK 16.740.11.0631) to V.O.P.

THERMOREGULATION OF THE ENZYMATIC DNA HYDROLYSIS

Abrosimova L.A.[1*], Romanova E.A.[2], Volkov E.M.[3], Pingoud A.[4], Kubareva E.A.[2], Oretskaya T.S.[2,3]

[1]Department of Bioengineering and Bioinformatics, [2]A.N. Belozersky Institute of Physico-Chemical Biology and [3]Chemistry Department, M.V. Lomonosov Moscow State University, Leninskie gory 1, Moscow 119991, Russia

[4]Institut fuer Biochemie, Fachbereich Biologie und Chemie, Justus-Liebig-Universitaet, Heinrich-Buff-Ring 58, D-35392 Giessen, Germany

e-mail: abrludmila@rambler.ru

Corresponding author

Key words: *restriction endonucleases, thermoregulation*

Motivation and Aim: In "genome surgery", a number of enzymes are used as molecular tools for cleaving defective genes - restriction and homing endonucleases. However, the enzymes can show undesirable activity during scanning DNA in the cell. Thus, regulation the protein activity is a crucial task. For *in vitro* applications we propose a method of temperature controlling of DNA hydrolysis by homodimeric restriction endonucleases.

Results: The "molecular gate" strategy for regulating activity of DNA-binding proteins is illustrated for the type II restriction endonuclease SsoII (R.SsoII) [1]. Modifying R.SsoII with oligodeoxyribonucleotides at the entrance of the DNA-binding site might make it inaccessible for DNA substrate due to steric hindrance or DNA duplex formation. We attached different 10-mer oligonucleotides to the R.SsoII variant - R.SsoII(2CS/S171C), containing just one cysteine residue in position 171. In the case of $(T)_{10}$ oligonucleotide containing 10 thymidine residues there was no enzymatic activity at 25°C. At 37°C the initial rate of hydrolysis increased 10 times and at 45°C about 15 times comparing to the initial rate at 25°C. This enzymatic change was reversible during 2 cycles of heating-cooling. The effect can be caused by increased conformational mobility of oligonucleotides during temperature increasing and opening the "molecular gate". To create more substantial obstacle for substrate inlet we synthesized 10-mer self-complementary oligonucleotide that was supposed to form DNA-duplex in the dimer's interface. The initial rates at different temperatures were rather similar to those obtained using $(T)_{10}$ oligonucleotide. Probably 10-mer self-complementary oligonucleotides could not form duplex in the dimer's interface or its formation doesn't influence significantly the enzymatic activity.

Conclusion: For further improvement of the "molecular gate" strategy we plan to synthesize longer oligonucleotides that will be able to form DNA duplex in the dimer's interface. For *in vivo* applications it is also possible to incorporate photosensitive azobenzene residues into oligonucleotides. Then the hydrolysis reaction can be regulated via irradiation by light.

This work was supported by program "International Research Training Groups" (RFBR-DFG 11-04-91338 and GRK 1384).

References:

1. L.T.Hien *et al.* (2011) Restriction endonuclease SsoII with photoregulated activity - a «molecular gate» approach, *Bioconjug. Chem.*, **22**: 1366-13734.

Author index

H

I

J

K

L

M

N

O

P

R

S

T

V

W

X

Y

Z

Системы для генетического анализа нового поколения

life technologies™

Секвенатор Ion Proton

Самый быстрый секвенатор
Ion Proton предлагает новый метод секвенировния, основанный на полупроводниках и природной биохимии, что позволяет прочитывать геном человека всего за несколько часов.

Proton I
Чип Ion Proton I
165 миллионов лунок
2 человеческих экзома

Proton II
Чип Ion Proton II
650 миллионов лунок
Полный геном человека

Ion Proton:
Гибкость в выборе производительности прибора
Чип определяет производительность запуска: от 10 до 100 млрд пар оснований высокого качества, возможность использования молекулярных штрихкодов для мультиплексного анализа, длина прочтений от 200 пар оснований, чтение парных концов молекул ДНК.
Непревзойденная однородность покрытия
Природная биохимия секвенирования обеспечивает непревзойденную однородность покрытия, вне зависимости от GC-состава ДНК.
Возможности геномного центра
Быстрое и доступное секвенирование экзомов, транскриптомов и полных геномов, анализ некодирующих и микроРНК, анализ иммунопреципитированного хроматина (ChIP-Seq), изучение метилирования ДНК.

Система для количественной и цифровой ПЦР QuantStudio 12K Flex

QuantStudio 12K Flex:
Широкий спектр приложений
Анализ экспрессии генов, генотипирование, цифровая ПЦР, HRM-анализ, cast-ПЦР, анализ копийности генов и микроРНК.
Максимальная производительность, минимальные затраты
Система одновременно анализирует до 4 слайдов OpenArray, каждый на 3072 реакций, за 4 часа. Отлаженный рабочий процесс и автоматизированная загрузка плашек сокращают время запуска и работы прибора.
Необычайная функциональность
Система поддерживает 5 взаимозаменяемых термоблоков: OpenArray,TaqMan Array Card, на 384 и 96 лунок для стандартного и быстрого термоциклирования. Замена блока занимает менее минуты и не требует инструментов, обеспечивая максимальную гибкость исследований.
Масштабируемость ваших исследований
5 различных блоков позволяют проводить от 1 до более чем 12000 реакций за запуск в наиболее подходящем для вашего исследования формате.

Технология OpenArray
Система QuantStudio 12K Flex поддерживает технологию OpenArray, позволяющую проводить реакции в нанолитровых объемах. Для этого используются слайды, несущие по 3072 отверстия, которые образуют отдельные реакторы объемом 33 нл. Слайды покрыты гидрофильными и гидрофобными веществами, удерживающими реакционную смесь внутри отверстий за счет поверхностного натяжения.
Эта технология позволяет проводить высокоточное определение числа молекул ДНК без стандартов и контролей методом цифровой ПЦР.

ion torrent by life technologies™

applied biosystems by life technologies™

Московское представительство Life Technologies
117485, г. Москва, ул. Обручева, 30/1
тел.: (495) 651-67-97; факс: (495) 651-67-99
www. lifetechnologies.com

Технология сравнительной геномной гибридизации в биомедицине

BlueGnome

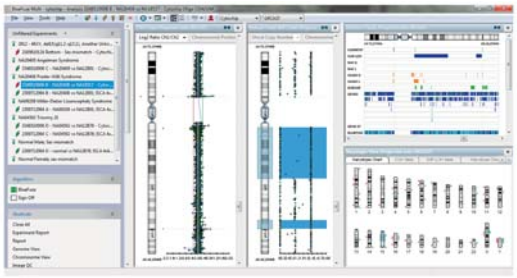
Фирма BlueGnome (UK) предлагает молекулярные цитогенетические тест-системы для предимплантационной, пренатальной, онко- и онкогематологической диагностики

Наборы 24sure™ для скрининга всех хромосом бластоцисты на анеуплоидии за 12 часов

В качестве исходного материала используется ДНК, выделенная из одной клетки

Анализ и управление данными, составление отчетов, формирование баз данных осуществляется с использованием программного обеспечения BlueFuse

- Полностью автоматизированный анализ изображений чипа в формате TIFF
- Хранение отчётов с исходными изображениями для улучшения контроля качества
- Автоматическая идентификация и аннотирование аномалий
- Хранение результатов в централизованной базе данных SQL
- Ограничение доступа к редактированию отчёта после его подписания руководителем



ГРУППА КОМПАНИЙ ВСМ БИОХИММАК

ЗАО БиоХимМак

119991 Москва, Ленинские горы, МГУ им. М.В. Ломоносова. Тел.: (495) 647-2740.
www.biochemmack.ru

BIO-RAD

Качественный прорыв в количественном ПЦР

QX100 Droplet Digital

Представляем технологию ПЦР третьего поколения:

Система **Droplet Digital™ PCR (ddPCR™)** позволяет определить абсолютное количество ДНК мишеней с непревзойденной точностью и достоверностью. Область применения включает определение числа копий гена, детектирование редких последовательностей и мутаций, анализ экспрессии генов.

Система **QX100 ddPCR** обладает разрешением, начиная с одной копии ДНК – мишени для проведения для проведения ПЦР, что позволяет ускорить темпы научных изысканий и разработок новых стратегий исследований наследственных заболеваний, детекции рака и инфекционных заболеваний.

Система **QX100 ddPCR** позволяет:

- Обнаружить сиквенсы редких мутаций;
- Определить даже небольшие отличия в концентрации ДНК-мишеней;
- Определить число копий гена без использования калибровочной кривой.

Научное издание

Тезисы Международной конференции
«Постгеномные технологии для биомедицины»

на английском языке

International Conference
on Postgenomic Technology for Biomedicine

Abstracts

Тезисы конференции печатаются
в авторской редакции

Издание осуществлено при финансовой поддержке
Российского фонда фундаментальных исследований
по проекту № 12-04-06045-г

Подготовлено к печати
в редакционно-издательском отделе
ИЦиГ СО РАН
630090, Новосибирск, пр. акад. М.А. Лаврентьева, 10

Научный редактор *М.Р. Кабилов*
Выпускающий редактор *А.А. Ончукова*
Дизайн и компьютерная верстка *А.В. Харкевич*
Дизайн обложки *С.Ш. Ханинаев*

Подписано к печати 13.06.2012 г.
Формат бумаги 70 × 108 1/16. Печ. л. 17,6. Уч.-изд. л. 17,4
Тираж 170. Заказ 246.

Отпечатано в типографии ФГУП «Издательство СО РАН»
630090, Новосибирск, Морской пр., 2