July 27th, Saturday	
9:00-10:00	**Registration and morning tea**
10:00-11:40	**Transcriptomics, *Auditorium M-1***
10:00	Entropic Hourglass Patterns and the Emergence of Biodiversity *Ivo Grosse*
10:40	Single-cell Systems Biology Approaches to Stem Cell Research and Regenerative Medicine *Antonio del Sol*
11:20	Factor decomposition of structured single-cell omics data *Danila Bredikhin*
11:40-12:10	**Tea break**
12:10-13:30	**Functional genomics, *Auditorium M-1***
12:10	Quantifying the progress in life sciences: Decline in gene and protein function discovery rate despite big omics data *Frank Eisenhaber*
12:50	The 160k Natural Organism Library at the Bioinformatics Institute Singapore: A Treasure Trove for Mining Using Biological, Chemical, Genomic and in silico High-Throughput Screenings *Birgit Eisenhaber*
13:10	Analysis of CRISPR/Cas non-canonical functions *Marko Djordjevic*
13:30-15:00	**Lunch break**

15:00-16:40	**Protein structure and function,** ***Auditorium M-1***
15:00	Resistance to novel antibiotics studied with a combination of protein structure modelling and phylogenetic approach *Olga Kalinina*
15:40	Parallel computing in structural bioinformatics: efficient 3D-alignment of protein superfamilies on a supercomputer *Maksim Shegay*
16:00	Computational method generate highly stable D-amino acid peptides and proteins using a mirror image of the entire PDB *Michael Garton*
16:40-19:00	**Tea break and Poster session**
19:00 + max 10min	**Departure to the welcome dinner (buses)**

July 28th, Sunday	**Day of parallel sessions**
9:30-10:00	**Morning tea**
10:00-11:40	**mRNA - regulation and splicing, *Auditorium M-1***
10:00	Integrative transcriptomic analysis suggests novel autoregulatory splicing events coupled with nonsense-mediated mRNA decay *Dmitri Pervouchine*
10:40	Structural annotation of protein indels associated with splicing aberrations *Aleksei Mironov*
11:00	Complexity and evolution of the mammalian transcriptome: the architecture of alternative transcription and splicing *Svetlana Shabalina*
11:20	Analysis of experimentally derived landscape of microRNA-mRNA human interactome *Olga Plotnikova*
10:00-11:40	**Lifespan and other phenotypes, *Room 557***
10:00	Systems Biology of Aging and Lifespan Control *Vadim Gladyshev*
10:40	Identification and application of gene expression signatures associated with lifespan extension *Alex Tyshkovskiy*

11:00	Bayesian SEM model describes the latent mechanism linking genotypes and complex phenotypes *Anna Igolkina*
11:20	WhoGEM, a genomic admixture-based prediction machine, that accurately predicts quantitative functional traits in plants *Laurent Gentzbittel*
11:40-12:10	**Tea break**
12:10-13:30	**Chromatin and genome structure, *Auditorium M-1***
12:10	Reconstructing haplotype-specific cancer genome karyotypes with multiple sequencing technologies *Sergey Aganezov*
12:30	Integrating molecular modeling of nucleosomes with experimental data from EM, FRET and footprinting experiments *Alexey Shaytan*
12:50	Chromosome conformation analysis of ependymoma tumors identifies putative target genes activated by distal oncogenic enhancers *Konstantin Okonechnikov*
13:10	Modelling Segmental Duplications in the Human Genome *Eldar Abdullaev*
12:10-13:30	**Lifespan and other phenotypes, cont'd, *Room 557***

12:10	Genetic landscape of human healthspan *Yurii Aulchenko*
12:50	Mining a database of billions of genetic associations to understand biology of complex human traits *Tatiana Shashkova*
13:10	Analysis of human height in the UK Biobank *Sergei Slavskii*
13.30-15.00	**Lunch break**
15:00-16:40	**Promoters and transcription, *Auditorium M-1***
15:00	TransPrise – a deep learning approach for prediction of eukaryotic transcription start sites *Khalimat Murtazalieva*
15:20	Modeling SELEX for Promoters Using a Combination of Royal Road and Royal Staircase Fitness Functions *Anton Eremeev*
15:40	Transcriptome and epigenome landscape of human cortical development modeled in organoids *Alexej Abyzov*
16:00	Identifying mechanisms regulating gene expression variation during embryonic development *Olga Sigalova*

16:20	Inference of transcriptional regulatory network driven by desiccation and rehydration in Polypedilum vanderplanki *Yusuke Hiki*
15:00-16:40	**Algorithms, *Room 557***
15:00	metaplasmidSPAdes: Plasmid Detection and Assembly in Genomic and Metagenomic Datasets *Dmitry Antipov*
15:20	De novo transcriptome assembly with spISO-seq data *Andrey Prjibelski*
15:40	Estimating the true number of genome rearrangements between species *Nikita Alexeev*
16:00	Annotation of genome graphs *Sergey Petrov*
16:20	A Soft Alignment of Multiple Omic t-SNEs *Laleh Haghverdi*
16:40-17:10	**Tea break**
17:10-18:50	**Molecular evolution, part 1, *Auditorium M-1***
17:10	Signatures of genetic exchange in a natural population of the bdelloid rotifer Adineta vaga inferred from whole-genome data *Olga A. Vakhrusheva*
17:30	Inference of changes of fitness landscape from sequence data with single-position resolution *Galya Klink*

17:50	Allele-specific non-stationarity in evolution of Influenza A surface proteins *Anfisa Popova*
18:10	Tandem repeats are selfish elements which mark the level of hidden recombination in animal mitochondrial genomes *Alina A. Mikhailova*
18:30	Mitochondrial mutational spectrum in vertebrates is shaped by temperature and generation time *Alina Mikhaylova*
17:10-18:50	**A variety of omics, *Room 557***
17:10	Sense and nonsense in studying the antisense transcription in prokaryotes *Maria N. Tutukina*
17:30	Comparative genomics analysis revealed novel types of bacterial microcompartments in the human gut microbiome *Dmitry Ravcheev*
17:50	Omics approaches to unravel molecular responses to root microbes in legumes *Cécile Ben*
18:10	Using museomics to solve historic puzzles *Tatiana Tatarinova*

July 29th, Monday	
9:30-10:00	**Morning tea**
10:00-11:40	**Memorial M.A.Roytberg session, *Auditorium M-1***
10:00	Fun with clumps and automata *Mireille Regnier*
10:20	Change point identification in long Poisson series: application to DNase I and ChIP-seq data *Vsevolod Makeev*
10:40	Algorithm for finding all combinatorially complete datasets in random mutagenesis data *Dmitry Ivankov*
11:00	Predicting polygenic phenotype from genotype Shamil Sunyaev
11:40-12:10	**Tea break**
12:10-13:30	**Memorial M.A.Roytberg session, cont'd, *Auditorium M-1***
12:10	Red-C – a new method for RNA-chromatin interactome discovery *Andrey Mironov*
12:30	Accounting for RNA secondary structure allows improved classification and prediction of RNA base triples *Eugene Baulin*

12:50	The Intelligent Man's Burden: modern biology and biomedicine as an effective driving force on the boundary of science and education *Petr Vlasov*
13:10	M.A. Roytberg and the first library of disordered patterns in 3D protein structure *Oxana Galzitskaya*
13:30-15:00	**Lunch break**
15:00-16:20	**Translation, *Auditorium M-1***
15:00	Computational modelling of novel gene expression rules and their applications *Tamir Tuller*
15:40	Translation initiation in mammals: Why so many starts? *Pavel Baranov*
16:20-16:50	**Tea break**
16:50-17:50	**Comparative genomics, *Auditorium M-1***
16:50	Comparative genomics analysis of translational frameshifting in aerobic cobalt chelatase genes Ivan Antonov
17:10	Ancient genomic regions of extreme conservation: Insights into evolution, structural organization, and function *Dmitry Korkin*
17:50 + max 10min	**Meeting of the group for the boat cruise**

July 30th, Tuesday

9:30-10:00	**Morning tea**
10:00-11:40	**Single cell transcriptomics, *Auditorium M-1***
10:00	Inferring complex pseudo-time trajectories in single cell data using elastic principal graphs and STREAM pipeline *Andrei Zinovyev*
10:40	Transcriptional Dynamics of Single Cells *Ruslan Soldatov*
11:20	Transcriptome map of the human, chimpanzee, bonobo, and macaque brains at the single-cell resolution *Ekaterina Khrameeva*
11:40-12:10	**Tea break**
12:10-13:30	**Molecular evolution, part 2, *Auditorium M-1***
12:10	Clustered mutations in human germline *Vladimir Seplyarskiy*
12:30	Mutational patterns in E.coli evolution *Sofya Garushyants*
12:50	mRNA editing: a genome-scale pre-adaptation in soft-bodied cephalopods *Mikhail Moldovan*
13:10	*Prevalent epistatic interactions between amino acid sites in Schizophyllum commune* *Anastasia Stolyarova*
13.30-15.00	**Lunch break**
15:00-16:10	**Medical genomics, *Auditorium M-1***

15:00	Assessment of network module identification across complex diseases *Sven Bergmann*
15:40	Systems Medicine - or - What I learned about Arnold Schwarzenegger while studying breast cancer survival *Jan Baumbach*
16:00	Scientific background for personal DNA tests (sponsored talk) *Alexandr Rakitko*
16:10-16:40	**Tea break**
16:40-18:00	**Cancer, *Auditorium M-1***
16:40	Probabilistic approaches to positive and negative selection inference on coding regions in cancer *Donate Weghorn*
17:20	Analysis of proteome-transcriptome correlations with self-contained GSA enhances consensus colon cancer classification *Ancha Baranova*
17:40	Uncovering hidden sources of transcriptional dysregulation arising from inter- and intra-tumor heterogeneity *Alexander V Favorov*
18:00	Prediction of functional effect of short in-frame indels *Vasily Ramensky*
18:20-00:00	**Farewell party**

July 27th, 10:00

Entropic Hourglass Patterns and the Emergence of Biodiversity

Ivo Grosse

One surprising observation going back to pioneering works of Karl Ernst von Baer in 1828 and Ernst Haeckel in 1866 is that embryos of different animal species express on average evolutionarily young genes at the beginning of embryogenesis, evolutionarily old genes in mid-embryogenesis, and again evolutionarily young genes at the end of embryogenesis. This phylotranscriptomic hourglass pattern as well as the resulting morphological hourglass pattern that animals of different phyla look morphologically different at the beginning of embryogenesis, morphologically similar in mid-embryogenesis, and again morphologically different at the end of embryogenesis show that trancriptomic and morphologic biodiversity emerges in a nonlinear and even non-monotonic manner. Focusing our attention on plants, which represent the second major kingdom in the tree of life that evolved embryogenesis, we have found that the phylotranscriptomic hourglass pattern also exists in plant embryogenesis. This observation is surprising as multicellularity and embryogenesis evolved independently in animals and plants and suggests the convergent evolution of phylotranscriptomic hourglass patterns in animal and plant embryogenesis. Moreover, we have found that phylotranscriptomic hourglass patterns also exist in the two main transitions of post-embryonic plant development, germination and floral transition, suggesting the convergent evolution of phylotranscriptomic hourglass patterns in embryonic and post-embryonic plant development.
The origin of these phylotranscriptomic hourglass patterns has remained concealed, but here we find that not only the mean age of expressed genes changes in an hourglass-like manner, but the whole age distribution of expressed genes changes. When studying the entropy of these age distributions as functions of time, we find hourglass patterns that surprisingly are orders of magnitude more significant than the original phylotranscriptomic hourglass patterns of the mean, which might indicate that the phylotranscriptomic hourglass patterns of the entropy are more fundamental than, and possibly even the origin of, the original phylotranscriptomic hourglass patterns of animal and plant development.

July 27th, 10:40

Single-cell Systems Biology Approaches to Stem Cell Research and Regenerative Medicine

Antonio del Sol

The application of computational biology approaches to stem cell research and regenerative medicine is becoming more and more necessary in order to tackle a variety of fundamental questions in these fields. In particular, advances in single cell sequencing technologies have allowed the implementation of novel computational methods for addressing challenges, such as the in-vitro reprogramming of specific cell subtypes for cell transplantation, and the identification of signals from the tissue microenvironment impairing tissue regeneration in ageing or disease conditions, that would allow the development of novel strategies for stem cell rejuvenation. In this regard, I will give an overview of computational methods developed in my lab, whose predictions have been validated by experimentalists in various cellular systems. In particular, one of the methods, which employ a multivariate generalisation of mutual information, aims at identifying synergistic transcriptional identity cores characterising cell subpopulations. Perturbations of these core transcription factors are shown to induce transitions between cellular subpopulations, allowing the in-vitro generation of novel cell subpopulations, such as specific types of dopaminergic neurons. Further, I will discuss a recently implemented computational method that integrates cellular signalling and gene regulatory networks to identify key signalling molecules controlled by the niche to maintain specific cellular phenotypes, and whose perturbations could trigger cellular transitions in-vivo. Application of this method to the mouse ageing brain identified inflammatory signals that maintain stem cells in a quiescent state, and proposes strategies to activate neural stem cells to repair the ageing brain. In summary, these methods have been shown to be useful in guiding experimental research, and are currently being used in designing strategies for cell therapy treatment in patients with Parkinson’s disease, as well as in patients with partial vision loss due to depletion of corneal limbus stem cells.

July 27th, 11:20

Factor decomposition of structured single-cell omics data

Danila Bredikhin

Single-cell RNA sequencing (scRNA-seq) has become an ubiquitous method for studying gene expression regulation with unprecedented resolution. It provides an opportunity to discover novel cell types and characterise differentiation trajectories in an unsupervised fashion. However the high dimensionality of scRNA-seq datasets and inherent amount of technical noise make the analysis of such data one of the most challenging problems in computational biology. An important computational strategy for analysing scRNA-seq data is dimensionality reduction. By exploiting the correlation structure between genes one can learn a latent representation of the cells which can be used for data interpretation, visualisation, feature extraction, or imputation. For that, a variety of approaches are commonly applied, including widely used PCA and t-SNE. However these approaches assume cells being independent given a set of latent factor and ignore the rich structure of single-cell experiments, which may include multiple groups of cells or multiple omics profiled, and it still remains a challenge to infer biologically relevant and interpretable factors from the datasets in an unsupervised manner. We propose a statistical framework for learning the latent sources of cell-to-cell variability in structured data sets with multiple sample groups. The model builds upon group factor analysis, a Bayesian framework that includes hierarchical sparsity priors on factor loadings to efficiently integrate multi-view data. Here we re-define the sparsity priors to include a group-specific regularisation in order to disentangle the activity of factors across multiple groups of cells. Effectively, this allows the quantification, for every latent factor, of how much variability is shared between the different groups of cells, e.g. different cell types, tissues, or donor cohorts. Importantly, we employ stochastic variational inference and graphics processing unit (GPU)- accelerated computations in order to accommodate large volumes of single-cell sequencing data. It is also designed to be easily incorporated into common scRNA-seq analysis workflows, including Bioconductor, Seurat, or scanpy. We demonstrate how our framework can be used to interpret gene expression variation in single-cell RNA sequencing experiments. For instance, we show that it’s capable of learning cell type-specific and tissue-specific factors across large collections of cells such as Tabula Muris and can also be applied to perturbation experiments in order to highlight gene sets that capture the response to the perturbation at the transcriptome level.

July 27th, 12:10

Quantifying the progress in life sciences: Decline in gene and protein function discovery rate despite big omics data

Frank Eisenhaber

It is generally believed that full human genome sequence published in 2001 was a watershed event in human history that boosted biomedical research, biomolecular mechanism discovery, and life science applications. If the mentioning of gene names in the scientific literature is used as a proxy to assess the novelty of biological discovery, we find that the gene function discovery rate per year has grown until 2000 but is in drastic decline thereafter. A group of elite genes (4817 protein-coding genes and 119 non-coding RNAs) attracts the overwhelming majority of the scientific literature about biomolecular mechanisms.

July 27th, 12:50

The 160k Natural Organism Library at the Bioinformatics Institute Singapore: A Treasure Trove for Mining Using Biological, Chemical, Genomic and in silico High-Throughput Screenings

Birgit Eisenhaber

Natural organism libraries (NOLs) experience a renaissance, the demand from industries such as pharmaceutical, chemical, food and nutrition, consumer care, biotechnology or agriculture for new active compounds, colorants, enzymes, etc. remains unsatisfied. The Bioinformatics Institute (BII) Singapore does not only house one of the largest NOLs in the world but also a Natural Product Discovery Platform that provides expertise in high-throughput screening, genomics, analytical chemistry, synthetic biology, and bioinformatics. As an example, we present the discovery of a new compound, BII-Rafflesfungin, and its biosynthetic gene cluster found in a Phoma species.

July 27th, 13:10

Analysis of CRISPR/Cas non-canonical functions

Marko Djordjevic

It is widely accepted that CRISPR/Cas is an advanced bacterial immune system, but only few relevant physiological systems have been demonstrated to exhibit the defense function. Experimental evidence of non-canonical CRISPR/Cas functions starts to accumulate, but, at least for now, they may appear more as isolated incidents. We bioinformatically address how widespread are these functions, by systematically analyzing some of Type I and Type II systems, and propose a method which can make analysis of non-cannonical functions more straightforward.

July 27th, 15:00

Resistance to novel antibiotics studied with a combination of protein structure modelling and phylogenetic approach

Olga Kalinina

Antimicrobial resistance is a major health hazard when it arises in pathogenic bacteria. Nature is an inexhaustible resource of both novel compounds with antibacterial activity and mechanisms of resistance against them. In this talk I will give a few examples of how the mechanisms of action and resistance towards these compounds can be studied with a combination of comparative and structural bioinformatics.

July 27th, 15:40

Parallel computing in structural bioinformatics: efficient 3D-alignment of protein superfamilies on a supercomputer

Maksim Shegay

A 3D-alignment of multiple protein structures is fundamentally important for a variety of tasks in modern biology. Today, non-redundant collections of protein superfamilies are represented by hundreds of 3D-records, making it problematic to use the available single-CPU software to perform such a superimposition. We have developed the parMATT – the first program intended for large-scale 3D-alignment of protein families/superfamilies running in a parallel environment of a classical cluster/supercomputer.

July 27th, 16:00

Computational method generate highly stable D-amino acid peptides and proteins using a mirror image of the entire PDB

Michael Garton

Biologics are a rapidly growing class of therapeutics with many advantages over traditional small molecule drugs. A major obstacle to their development is that proteins and peptides are easily destroyed by proteases and, thus, typically have prohibitively short half-lives in human gut, plasma, and cells. One of the most effective ways to prevent degradation is to engineer analogs from dextrorotary (D)-amino acids, with up to 105-fold improvements in potency reported. We here propose a general peptide-engineering platform that overcomes limitations of previous methods. By creating a mirror image of every structure in the Protein Data Bank (PDB), we generate a database of ~2.8 million D-peptides. To obtain a D-analog of a given peptide, we search the (D)-PDB for similar configurations of its critical—"hotspot"—residues. As a proof of concept, we apply our method to two peptides that are Food and Drug Administration approved as therapeutics for diabetes and osteoporosis, respectively. We obtain D-analogs that activate the GLP1 and PTH1 receptors with the same efficacy as their natural counterparts and show greatly increased half-life.

July 28th, 10:00 (Auditorium M-1)

Integrative transcriptomic analysis suggests novel autoregulatory splicing events coupled with nonsense-mediated mRNA decay

Dmitri Pervouchine

Nonsense-mediated decay (NMD) is a eukaryotic mRNA surveillance system that selectively degrades transcripts with premature termination codons (PTC). Many RNA-binding proteins (RBP) regulate their expression levels by a negative feedback loop, in which RBP binds its own pre-mRNA and causes alternative splicing to introduce a PTC. We present a bioinformatic analysis integrating three data sources, eCLIP assays for a large RBP panel, shRNA inactivation of NMD pathway, and shRNA-depletion of RBPs followed by RNA-seq, to identify novel such autoregulatory feedback loops. We show that RBPs frequently bind their own pre-mRNAs, their exons respond prominently to NMD pathway disruption, and that the responding exons are enriched with nearby eCLIP peaks. We confirm previously proposed models of autoregulation in SRSF7 and U2AF1 genes and present two novel models, in which (1) SFPQ binds its mRNA and promotes switching to an alternative distal 3'-UTR that is targeted by NMD, and (2) RPS3 binding activates a poison 5’-splice site in its pre-mRNA that leads to a frame shift and degradation by NMD. We also suggest specific splicing events that could be implicated in autoregulatory feedback loops in RBM39, HNRNPM, and U2AF2 genes. Taken together, these findings indicate that autoregulatory negative feedback loop of alternative splicing and NMD is a ubiquitous form of post-transcriptional control of gene expression among splicing factors.

July 28th, 10:40 (Auditorium M-1)

Structural annotation of protein indels associated with splicing aberrations

Aleksei Mironov

It is well known that more than 90% of human genes are alternatively spliced, but the function of alternative transcripts is still a matter of debate. On the one hand, many alternative isoforms are associated with diseases, which indicates their functionality. On the other hand, proteomic studies revealed only a small proportion of the mRNA isoforms, which are translated into proteins. Among the repertoire of non-annotated RNA isoforms found in NGS data, there are specific splicing events that are characterized by a close (~ 30 nt) tandem arrangement of alternative splicing sites (TASS). We focus on TASS in protein-coding genes that preserve the reading frame, i.e., usage of alternative sites cause short insertions and deletions in the protein sequence (protein indels) rather than lead to nonsense-mediated degradation of the transcript. We investigate such sites using RNA-seq data from GTEx (Genotype Tissue Expression) and TCGA (The Cancer Genome Atlas) projects.

July 28th, 11:00 (Auditorium M-1)

Complexity and evolution of the mammalian transcriptome: the architecture of alternative transcription and splicing

Svetlana Shabalina

Alternative splicing (AS) and alternative transcription (AT) create the extraordinary complexity of transcriptomes and lay the basis for the structural and functional diversity of mammalian proteomes. We present evidence that the acquisition of new exons in spliced genes can occur by mosaic extension of gene functional domains, where new alternative coding exons can be incorporated in the course of evolution, preferentially at the ends of CDSs. In this study, it is shown how gene architecture and evolutionary rates of human receptor genes (nuclear receptors, opioid receptors, etc.) influence their expression patterns and translation. Notably, the acquisition of novel exons at the boundaries of CDSs and UTRs is mainly mediated by alternative transcription events - initiation (ATI) and termination (ATT). Extended 5' and 3' ends, associated with AT events, make major contribution to the diversity of the protein isoforms and harbor approximately five times more alternative nucleotides in the coding exons than in the core protein-coding regions which are subject to AS. Thus, alternative transcription makes an even larger contribution to transcriptome and proteome diversity than alternative splicing, specifically for tissue- and condition- specific expression. Our results suggest that differential processing of the 5' and 3' ends reflect two different regulatory strategies employed by large gene groups: regulation by ATI at the level of transcription initiation, and regulation by ATT that alters post-transcriptional stability of mRNA, enriches C-terminal variability of protein isoforms and provides options for differential posttranslational regulation. We also showed that during evolution, compact protein domains are typically encoded by highly structured mRNAs suggesting that alternative mRNA structures might control protein folding of alternative isoforms. The role of alternative isoforms as biomarkers will be discussed for the disease states.

July 28th, 11:20 (Auditorium M-1)

Analysis of experimentally derived landscape of microRNA-mRNA human interactome

Olga Plotnikova

We present the landscape of microRNA-mRNA human interactome, which we derived from a combination of microRNA and mRNA expression data with direct microRNA-mRNA interactions experimentally defined in HEK293 and Huh7.5 cell lines. We have also collected and examined the data from 79 CLIP datasets of microRNA binding sites to report and systematically describe 46,800 experimentally confirmed interacting duplex regions that are available as web tool (http://score.generesearch.ru/services/mirna).

July 28th, 10:00 (Room 557)
Systems Biology of Aging and Lifespan Control
Vadim Gladyshev

July 28th, 10:40 (Room 557)

Identification and application of gene expression signatures associated with lifespan extension

Alex Tyshkovskiy

Dozens of pharmacological, genetic and dietary interventions leading to lifespan extension are known for a variety of organisms ranging from yeasts to mammals. Although some key molecular players behind lifespan extension have been uncovered, common mechanisms of different interventions remain unclear. Here we fill this gap by performing systemic analyses of gene expression across longevity interventions in mouse using our RNA-seq data and integrating publicly available data into a single study. By carrying hepatic RNA-seq across 8 longevity interventions in mice, we showed that interventions tend to regulate number of common pathways. We also discovered a feminizing effect associated with growth hormone regulation and diminution of sex-related differences in response to the longevity interventions. Using public data, we expanded this analysis to 17 interventions and observed that many interventions induced similar gene expression changes. We identified hepatic gene signatures associated with lifespan extension across interventions, including upregulation of oxidative phosphorylation and drug metabolism, and showed that perturbed pathways may be shared across tissues. We further found that genes related to oxidative phosphorylation and hepatic regulation of immune response could serve as both qualitative and quantitative predictors of lifespan extension. We applied the discovered longevity signatures to identify new lifespan-extending candidates, such as chronic hypoxia, KU-0063794 and ascorbyl-palmitate. Finally, we developed GENtervention, an app that visualizes associations between gene expression changes and longevity. Overall, our work investigates gene expression of longevity interventions, describes general and specific transcriptomic programs of lifespan extension in mice and provides tools to discover new interventions.

July 28th, 11:00 (Room 557)

Bayesian SEM model describes the latent mechanism linking genotypes and complex phenotypes

Anna Igolkina

SEM-based genotype-phenotype model for estimating complex phenotypes (tens traits) was developed. The model describes the latent mechanism which maps genotype information to phenotypes; estimation of parameters is handled by Gibbs sampling after Bayesian Inference. The model not only accurately predicts phenotypic trait but also extracts SNPs supplementing GWAS results

July 28th, 11:20 (Room 557)

WhoGEM, a genomic admixture-based prediction machine, that accurately predicts quantitative functional traits in plants

Laurent Gentzbittel

The explosive growth of genomic data provides an opportunity to make increased use of sequence variations for phenotype prediction. We have developed a prediction machine for quantitative phenotypes (WhoGEM) that, in plants, equals or outperforms all existing algorithms for prediction of quantitative trait prediction. WhoGEM analysis is using geographical locations as covariates for admixture analysis, thank to the newly developed ProvenancePedictor algorithm.

July 28th, 12:10 (Auditorium M-1)

Reconstructing haplotype-specific cancer genome karyotypes with multiple sequencing technologies

Sergey Aganezov

In the presented research a problem of reconstructing haplotype-specific rearranged cancer genomes with both short, linked, and long read sequencing technologies is undertaken. On several breast-cancer genomes an ensemble of methods for structural and copy number variations is deployed with subsequent integration of them into inferred karyotypes with the RCK method. Presented results demonstrate that long reads outperform short/linked reads in structural variations detection, and a comprehensive approach is demonstrated for recovering haplotype-specific karytoypes for rearranged cancer genomes.

July 28th, 12:30 (Auditorium M-1)

Integrating molecular modeling of nucleosomes with experimental data from EM, FRET and footprinting experiments

Alexey Shaytan

Nucleosomes are highly involved in chromatin functioning by providing epigenetic markup to the genome and participating dynamically in all DNA processing pathways. In this report we will dwell on our efforts to use molecular modeling techniques combined with the analysis of various experimental data sources to elucidate the structures and dynamical properties of nucleosomes and their complexes.

July 28th, 12:50 (Auditorium M-1)

Chromosome conformation analysis of ependymoma tumors identifies putative target genes activated by distal oncogenic enhancers

Konstantin Okonechnikov

In order to strengthen the identification of enhancer target genes in ependymoma tumors, we applied HiC technique to map the 3-dimensional organization of tumor chromatin in the two most common and aggressive subgroups: posterior fossa group A and supratentorial group with gene fusions involving the NF-κB subunit gene RELA. By an integrative analysis of enhancer and gene expression in the context of the newly derived HiC data, we identify many new putative tumor-dependency genes activated by long-range promoter-enhancer interactions. In addition we use the novelty of HiC to detect complex structural variants that also could be potential molecular targets for the future medical application.

July 28th, 13:10 (Auditorium M-1)

Modelling Segmental Duplications in the Human Genome

Eldar Abdullaev

Segmental duplications (SDs) are long (> 1kbp) DNA sequences that are repeated several times in a genome and have high sequence identity. There are multiple well-studied mechanisms that are responsible for propagation of segmental duplications: non-homologous end joining, DNA polymerase slippage, non-allelic homologous recombination etc. However, we do not have a general understanding that would explain the distribution of SDs in the genome, for example: why do SDs appear more often in some regions than in others, why do SDs overlap with each other so often, how does selection affect their distribution in the genome etc. Up to now there is no mathematical model for the propagation of SDs proposed that would explain simple statistical features of SDs. The goal of our project is to find such a model. We use a graph representation of all SDs in the human and try to approach this system from a complex network point of view. Nodes and edges of the graph represent genomic regions and homology between two regions (alignment), respectively. The SD graph appears similar to those of scale-free complex networks and some graph features shed light on the dynamics of the propagation process. By simulating the growth of such a network we predicted basic “rules” of SDs evolution and propagation in the genome. This model also agrees with biological data, such as recent SDs or structural variants from the human population that behave as predicted based on our model.

July 28th, 12:10 (Room 557)

Genetic landscape of human healthspan

Yurii Aulchenko

Aging populations face diminishing quality of life due to increased disease and morbidity. These challenges call for research focusing on understanding the pathways controlling healthspan. Here, we aimed to understand determinants of human healthspan using genetics and genomics. We used the data from the UK Biobank mega-cohort and observed that the risks of major chronic diseases increased exponentially and double every eight years, at a rate compatible with the Gompertz mortality law. Assuming that aging drives the morbidity rates acceleration, we built a risk model to predict the age corresponding to the end of healthspan depending on their age, gender, and the genetic background. Using a large database including tens of billions of genetic associations, we performed functional genomic investigation into molecular pathways underlying healthspan, and explored which (modifiable) risk factors are likely to be causatively related to the healthspan. A number of genome-wide significant healthspan loci acted in sex-specific manner. Phenome-wide association scan demonstrated that action of majority of loci was restricted to specific disease domain (e.g. cancer, dementia, cardio-metabolic). The strongest genetic correlations were observed between healthspan and all-cause mortality (as derived from parental survival, with genetic correlation equal to -0.76). Other strongly (absolute genetic correlation > 0.3) genetically correlated traits included life-history (metrics of obesity, age at first birth), and lifestyle traits (e.g. smoking behaviour). A large number of candidate healthspan-modifying risk factors were discovered in Mendelian randomisation analysis. We conclude that healthspan is a very complex trait integrating genetic predisposition to major ageing diseases and responses to environmental exposures. Although genetic correlation between health- and life-span is high, the overlap between loci that are genome- wide significantly associated to these traits is limited, underlining their differences.

July 28th, 12:50 (Room 557)

Mining a database of billions of genetic associations to understand biology of complex human traits

Tatiana Shashkova

Genome-wide association studies (GWAS) investigating the relationship between genotype and phenotype across the human genome is a research methodology that has become very popular over the past decade. We developed the GWAS-MAP platform for storing, processing, and analysis of GWAS results and a web-service PheLiGe (https://phelige.com) for their display and analysis. In this topic, we describe our platform and demonstrate its utility by analysis of coronary artery disease, varicose vein and back pain.

July 28th, 13:10 (Room 557)

Analysis of human height in the UK Biobank

Sergei Slavskii

Height is one of the most studied complex human traits in different scientific fields. However, there is a controversion between additive and multiplicative models that are used for height analysis in genetic and socio-economic studies respectively. Using the data from the UK Biobank, we show that human height most likely results from a multiplicative process and should be analyzed on a logarithmic scale in epidemiological and genetic studies.

July 28th, 15:00 (Auditorium M-1)

TransPrise – a deep learning approach for prediction of eukaryotic transcription start sites

Khalimat Murtazalieva

TransPrise is an effective and efficient deep learning approach to precise prediction of eukaryotic transcription start sites. Our approach offers significant improvement over existing promoter-prediction methods. We provide the full basis for the comparison and encourage users to freely access our computational method to facilitate and streamline their own analyses.

July 28th, 15:20 (Auditorium M-1)

Modeling SELEX for Promoters Using a Combination of Royal Road and Royal Staircase Fitness Functions

Anton Eremeev

The well-known biotechnological procedure SELEX (Systematic Evolution of Ligands by EXponential enrichment) is considered as an experimental implementation of an evolutionary algorithm (EA). Theoretical bounds on EA runtime are applied to model SELEX search for a promoter sequence, including an enhancer. A comparison of theoretical bounds to the results of computational simulation indicates some cases where the theoretical bounds give favorable prediction, while simulation requires prohibitive computational resource.

July 28th, 15:40 (Auditorium M-1)

Transcriptome and epigenome landscape of human cortical development modeled in organoids

Alexej Abyzov

Genes implicated in neuropsychiatric disorders are active in fetal brain, yet hardly possible to study in a longitudinal fashion using human tissue. By comparing organoids derived from human pluripotent cells and isogenic cortex we demonstrate that organoids model cerebral cortical development before 16 weeks postconception. Analyses of histone marks, transcriptome and chromatin conformation revealed genes and enhancers (bound by activators or repressors) differentially active in step-wise progression from neural stem cells to neurons. Networks of correlated gene and enhancer modules could be assembled into six and four global patterns of expression/activity across time. A pattern with progressive downregulation was enriched with evolutionary gained enhancers, suggesting their importance for very early human brain development. A few convergent gene and enhancer modules were enriched in autism-associated genes, genomic variants in autistic children, and credible SNPs associated with Schizophrenia. Combined, the organoid model points to genes and regulatory elements driving disease onset.

July 28th, 16:00 (Auditorium M-1)

Identifying mechanisms regulating gene expression variation during embryonic development

Olga Sigalova

Gene expression occurs in an intrinsically noisy environment. Yet, living systems have a remarkable capacity to give rise to robust and highly reproducible phenotypes. Here, we investigate the mechanisms underlying robustness and variation in gene expression in a population of developing embryos of Drosophila melanogaster.

July 28th, 16:20 (Auditorium M-1)

Inference of transcriptional regulatory network driven by desiccation and rehydration in Polypedilum vanderplanki

Yusuke Hiki

Water is essential for active life of living organisms. The organisms living on the ground were steadily exposed to desiccation stress, which causes the death of its life. Therefore, these organisms have evolved to survive, tolerating the desiccation stress. One of these strategies is anhydrobiosis, an ametabolic state to tolerate extreme desiccation stress. Sleeping chironomid, Polypedilum vanderplanki, is the insect uniquely having the ability to withstand complete desiccation at a larval stage by entering anhydrobiosis. In this research, we inferred the regulatory relationships of transcription factors, transcriptional regulatory network, from time-series RNA-seq data during desiccation and rehydration for P. vanderplanki, and tried to raise the candidates of transcription factors to critically regulate desiccation tolerance system.

July 28th, 15:00 (Room 557)

metaplasmidSPAdes: Plasmid Detection and Assembly in Genomic and Metagenomic Datasets

Dmitry Antipov

Although there are about 10,000 plasmids listed in the RefSeq database, many plasmids remain undetected since it is not trivial to assemble plasmids from genomic and metagenomic datasets and to detect plasmids hidden among chromosomal contigs. Here we present a metaplasmidSPAdes algorithm for plasmid assembly and detection in metagenomes. Additionally we show that same iterative pipeline can be modified to be used for detection and assembly of DNA viruses.

July 28th, 15:20 (Room 557)

De novo transcriptome assembly with spISO-seq data

Andrey Prjibelski

This work presents a novel algorithm for de novo transcriptome assembly from barcoded mRNA sequencing data obtained with recently emerged spISO-Seq protocol (Tilgner et al., 2018). Using sequencing data collected from several H.sapiens individuals, we show that the developed assembly method allows to accurately reconstruct full-length transcript sequences even for complex alternatively spliced isoforms -- something that is not possible using conventional short-read RNA-Seq datasets.

July 28th, 15:40 (Room 557)

Estimating the true number of genome rearrangements between species

Nikita Alexeev

Genome rearrangements are evolutionary events that shuffle genomic architectures. In the current study, we propose a new method for estimating the true evolutionary distance between two genomes under the INFER model. We estimated the parameters of the model for both the general case (when there are no assumptions about the distribution of the regions' fragilities) and the case when fragilities are distributed according to the Dirichlet distribution. We obtained surprisingly nice mathematical formulas, and showed that the estimation results are very accurate.

July 28th, 16:00 (Room 557)

Annotation of genome graphs

Sergey Petrov

Genome graph is a well known structure in bioinformatics, widely applied in genome assembly, proceeding of gene regulatory networks, alignment, etc. At present, genome graphs are also used as final result of the genome assembly that contains all possible haplotypes that can be assembled from the certain dataset. This work is dedicated to the problem of the gene prediction in this kind of data structure.

July 28th, 16:20 (Room 557)

A Soft Alignment of Multiple Omic t-SNEs

Laleh Haghverdi

New data integration methods for understanding the relation between several molecular levels (e.g. genomics, epigenomics, transcriptomics) that regulate and define cells identity are in demand. We present a new multi-omics data integration method, which uses the similarity between the shapes of data sets (i.e. structure of cell populations with respect to each other) in the cell samples collected from a same biological system to map them in a common latent space provided by the t-distributed stochastic neighbours embeddings (t-SNE). Our method optionally can use available information of correspondence for a number of co-asseyed cells or any shared features collected for both data modes to leverage the alignment quality.

July 28th, 17:10 (Auditorium M-1)

Signatures of genetic exchange in a natural population of the bdelloid rotifer Adineta vaga inferred from whole-genome data

Olga A. Vakhrusheva

Sexual reproduction is near-ubiquitous among living eukaryotes. Although transitions to asexual reproduction are frequent, they usually lead to a quick extinction of the involved species. Therefore, obligate asexuality is commonly thought to be an evolutionary dead end.

However, the existence of ancient asexual lineages poses a challenge to the indispensability of sexual reproduction for the long-term evolutionary success. On these grounds, anciently asexual lineages are frequently referred to as ‘evolutionary scandals’. Bdelloid rotifers are one of the most striking examples of such a lineage. Presumably, bdelloid rotifers abandoned conventional sexual reproduction tens of millions of years ago. However, this does not exclude the possibility of some other mode of interindividual genetic exchange in their populations. Recent analyses based on several genomic regions suggested genetic exchanges in this group, although one of these studies has been controversial. To address the possibility of genetic exchange in bdelloid rotifers on a whole-genome scale, we sequenced genomes of 11 individuals from a natural population of the bdelloid rotifer Adineta vaga. Using several types of analysis, we show that the observed patterns of population structure in A. vaga are incompatible with strict clonality. We detected several signatures suggestive of genetic exchange and recombination in A. vaga. First, we observed that linkage disequilibrium in A. vaga rapidly declines with the distance between polymorphic loci and showed that this decline cannot be attributed solely to the action of gene conversion. Next, we detected a substantial number of triallelic sites harboring all three possible heterozygous genotypes, with the observed number significantly exceeding the one expected due to recurrent or back mutations. Finally, we reconstructed haplotype-based phylogenies at many genomic loci and identified multiple cases when the two haplotypes of a single individual clustered with haplotypes from different individuals. As the above- mentioned patterns are not to be expected under obligate asexuality, our findings provide compelling evidence for genetic exchange in A. vaga, which was previously thought to be an ancient asexual species. Although the nature of evolutionary advantage of sex remains unclear, genetic exchange in A. vaga, irrespective of its mechanism, is likely to protect it from the long-term costs of abandoning sex such as Muller’s ratchet. Thus, the results of our study underscore an important role of genetic exchange in counteracting forces leading to the extinction of populations.

July 28th, 17:30 (Auditorium M-1)

Inference of changes of fitness landscape from sequence data with single-position resolution

Galya Klink

Sequeces of homologous proteins from many organisms along with their reconstructed phylogenetic relationships can be used to obtain inference about the evolutionary role of shifts in single-position fitness landscapes. Using phylogenetic approach, we have previously shown that the fitness landscapes of metazoan mitochodrial proteins change with time.Now we modified our method in a way that it finds amino acids with variable fitness between two evolutionary lineages. We applied this approach to viral proteins and our results were in agreement with the results of Deep Mutational Scanning experiments.

July 28th, 17:50 (Auditorium M-1)

Allele-specific non-stationarity in evolution of Influenza A surface proteins

Anfisa Popova

Influenza A virus is a major public health problem and a pandemic threat. Prediction of its evolution remains a major challenge, and one of the obstacles is the ever-changing environment which determines strain fitness. Here, we show that the fitness conferred by individual amino acids at influenza A surface proteins, hemagglutinin (HA) and neuraminidase (NA), changes systematically with time, and the direction of this change depends on the function of the amino acid.

July 28th, 18:10 (Auditorium M-1)

Tandem repeats are selfish elements which mark the level of hidden recombination in animal mitochondrial genomes

Alina A. Mikhailova

It is commonly accepted that uniparental inheritance of mitochondrial genomes (mtDNA) helps to resist expansion of selfish-like elements. However, majority of Chordata species still harbor some selfish-like tandem repeats with unknown function in control region of mtDNA. Analyzing mtDNA structure of thousands Chordata species we propose that these tandem repeats can be sensitive to very low level of mtDNA recombination and thus might be used as a marker of mtDNA recombination level in different species.

July 28th, 18:30 (Auditorium M-1)

Mitochondrial mutational spectrum in vertebrates is shaped by temperature and generation time

Alina Mikhaylova

It has been shown, that various exogenic mutagens (such as UV light, tobacco smoke, etc) affect nuclear but not mitochondrial DNA, leading to a hypothesis, that mtDNA is a subject of very strong internal mutagen of unknown nature. To shed light on mtDNA mutagenesis we reconstructed mtDNA mutational spectrum for hundreds of mammalian species and dozens of human tissues. We observed, that main factor, shaping mtDNA mutational specrtrum is cell's turnover rate and proposed a chemical damage hypothesis, which explains our results.

July 28th, 17:10 (Room 557)

Sense and nonsense in studying the antisense transcription in prokaryotes

Maria N. Tutukina

Antisense transcription is widespread among both eukaryotic and prokaryotic genomes. However, we are now only in the beginning on the way to understand its functions. Specific features of bacterial antisense promoters, and possible functions for their products detected from within and upstream of the genes coding for regulatory proteins in Escherichia coli will be discussed.

July 28th, 17:30 (Room 557)

Comparative genomics analysis revealed novel types of bacterial microcompartments in the human gut microbiome

Dmitry Ravcheev

A comparative genomics approach was analyze the bacterial microcompartments in 646 individual genomes of organisms commonly found in the human gut microbiome. This analysis revealed the presence in the human gut microbiome all four known metabolosomes, for utilization ethanolamine, 1,2-propanediol, choline, and fucose/rhamnose. Moreover, two novel microcompartment associated pathways were reconstructed, for catabolism of 1-amino-2-propanol/1-amino-2-propanone and xanthine.

July 28th, 17:50 (Room 557)

Omics approaches to unravel molecular responses to root microbes in legumes

Cécile Ben

-Omics' high-throughput analyses of microRNA and gene expression were conducted to shed light on the molecular root plasticity in response to symbiotic and pathogenic interactions in legumes. The combination of rigorous differential expression statistical analyses and gene co-expression networks has revealed key biological pathways and regulatory factors as promising breeding targets.

July 28th, 18:10 (Room 557)

Using museomics to solve historic puzzles

Tatiana Tatarinova

The identity of the Khazar rulers has been debated for centuries, and one controversial hypothesis posits that they are progenitors of the present-day Ashkenazi Jews. We used whole genome NGS data from reliably identified archaeological sites associated with the Khazars to put this controversy to rest. We find that the Khazar elite had genomic and matrilineal affinities to East Siberian and Caucasian tribes, but not to present-days Jewish populations.

July 29th, 10:00

Fun with clumps and automata

Mireille Regnier

Clumps, e.g. sequences of overlapping words, play a key role in word enumeration. We exhibit two unexpected combinatorial bijections, that lead to define a canonic decomposition for clumps and an equivalence relation. A reduced automata follows, that will be run offline on genomic sequences.

July 29th, 10:20

Change point identification in long Poisson series: application to DNase I and ChIP-seq data

Vsevolod Makeev

Change points in a series of independent random trials are the points when the parameters of random distribution change their value step-wise. A modification of dynamic programming suggested by M.A Roytberg allowed construction of an efficient precise algorithm for segmentation of Bernoulli and polynomial sequences into uniform segments, which allowed identification of compositionally homogeneous domain in nucleotide sequences. The algorithm allowed identification of the true optimal configuration of segment boundaries in O(Seq_length2). New types of high throughput data allow application of modifications of this algorithm for new biological problems. Particularly, segmentation of read densities modeled as a Poisson series allows identifying open chromatin regions and ChIP-Seq peaks. A new approach based on the folded Bernoulli distribution allows identifying CNVs in the samples with low sequence coverage.

July 29th, 10:40

Algorithm for finding all combinatorially complete datasets in random mutagenesis data

Dmitry Ivankov

To approach genotype-to-phenotype prediction it is important to investigate all cases of experimentally measured connections of genotype and phenotype. Particularly, we want to compare effect of multiple mutations with the net effect of single constituent mutations. For this, we should be able to identify all combinatorially complete datasets in experimental genotype-phenotype maps. Here, we propose an effective algorithm to solve this task.

July 29th, 11:00

Predicting polygenic phenotype from genotype

Shamil Sunyaev

Most human complex traits appear to be highly polygenic. We have developed a new non-parametric method for phenotype prediction. We show that an extreme phenotypic presentation could be either monogenic or polygenic, and the segregation data usually lacks statistical power to discriminate between these two possibilities.

July 29th, 12:10

Red-C – a new method for RNA-chromatin interactome discovery

Andrey Mironov

We developed a new high thruput method for RNA-chromatin interactions discover. Our protocol is similar to the HiC protocol but designed for RNA-chromatin interactome. We found a number of new unannotated RNAs that have contacts with chromatin but are invisible in RNA-seq.

July 29th, 12:30

Accounting for RNA secondary structure allows improved classification and prediction of RNA base triples

Eugene Baulin

A classification is proposed for RNA tertiary motifs regarding the secondary structure environment of their nucleotides. The classification was applied to RNA base triples, A-minor motifs and non-canonical base-pairs, and allowed distinguishing particular motifs with distinct properties

July 29th, 12:50

The Intelligent Man's Burden: modern biology and biomedicine as an effective driving force on the boundary of science and education

Petr Vlasov

My late friend and teacher, Mikhail Roytberg, was not just an excellent scientist - but also a man of outstanding passionarity. He realized it in many educational projects that involved hundreds of children and adults. I would like to devote my story to a discussion of how exactly scientists can “project” their understanding of world and their professional (research) activities in the educational field. I want to reveal this topic with some examples of particular and very different educational projects created thanks to Mikhail Roytberg - from simple online courses to serious research tasks - where young people, primarily schoolchildren, took part.

July 29th, 13:10

M.A. Roytberg and the first library of disordered patterns in 3D protein structure

Oxana Galzitskaya

M.A. Roytberg was invited in our paper “Library of disordered patterns in 3D protein structures” after defending his doctoral dissertation. His mathematician talent was very important for formulation of the rules for the construction of the first library of disordered patterns. This paper is my single paper with Proposition and Proof. In my talk I will describe my first meeting with M.A. Roytberg.

July 29th, 15:00

Computational modelling of novel gene expression rules and their applications

Tamir Tuller

In recent years we gained a better understanding that various codes related to the regulation of all gene expression stage are interleaved, in a non-modular manner, in the coding regions, and populate the untranslated regions (UTRs). In the talk I will describe our multidisciplinary strategies for deciphering novel gene expression codes and modelling such unknown codes. It will specifically report some recent discoveries of novel fundamental universal translation codes in prokaryotes. Moreover, I will demonstrate how novel gene expression models can be very useful for solving various problems in the biotechnology and pharmaceutical industries.

July 29th, 15:40

Translation initiation in mammals: Why so many starts?

Pavel Baranov

In mammals, translation initiation of most mRNAs is cap-dependent and involves scanning of mRNA molecules in 5’ to 3’ direction by preinitiation complexes (PIC) consisting of a small ribosomal subunit in a complex with several initiation factors. Start codon recognition is a stochastic process. AUG codons in strong Kozak contexts are recognized very efficiently with a probability approaching 1. AUG codons in a weak context, as well as many non-AUG codons, are recognized less efficiently with probabilities varying in a wide range between 0 and 1 leading to low, yet productive levels of protein synthesis. PICs that fail to recognize such codons as starts continue to scan. This leads to translation initiation at multiple start codons on the same mRNA. The complexity of translation initiation in mammals is further exacerbated by the ability of ribosome complexes to re-initiate downstream of short ORFs. Thus, translation initiation is possible even downstream of AUGs in strong Kozak contexts. Recently we and others began to reveal functional and phenotypical consequences of start codons plurality. Many starts function to translate short uORFs that regulate translation of downstream ORFs encoding proteins. They often serve as sensors of intracellular stress conditions and of specific metabolites. Some uORFs encode functional proteins enabling polycistronic mRNAs. Multiple starts also enable the production of proteoforms differing at their N-termini. Such proteoforms may have different properties, for example, different localization. Evidently, the scanning nature of translation initiation shapes the evolution of mRNA sequences including that of protein coding regions. We have demonstrated that the occurrence of AUG codons at the beginning of protein coding regions depends on the strength of the starting AUG.

July 29th, 16:50

Comparative genomics analysis of translational frameshifting in aerobic cobalt chelatase genes

Ivan Antonov

Cobalt chelatase CobNST is one of about 25 enzymes required for aerobic biosynthesis of cobalamin (vitamin B12) in prokaryotes. It has been shown that the large, medium and small subunits of this enzyme are encoded by the cobN, cobT and cobS genes, respectively. A later computational study has revealed a number of prokaryotic genomes where the cobT and cobS genes are missing from the cobalamin biosynthesis pathway. Instead these genomes contain the chlD and chlI genes encoding the medium and small subunits of the magnesium chelatase chlIDH - the enzyme required for chlorophyll biosynthesis. Given the high similarity between the magnesium and cobalt chelatases, the authors have hypothesized that the products of the chlD and chlI genes can replace the missing subunits of the cobNST enzyme. Recently, we have discovered a functional programmed ribosomal frameshifting (PRF) signal located inside the cobT gene from diverse bacteria and archea that efficiently diverts translation to the -1 reading frame. The goal of the present study is to determine the possible biological function of this conserved recoding event. For this purpose we performed a comparative genomics analysis of the PRF-utilization by the cobalt and magnesium chelatase genes from more than 1200 prokaryotic genomes. In total, there were 135 and 36 cobT genes with putative -1 and +1 frameshifting, respectively. Given the high similarity between the CobS protein and the N-terminal part of the CobT protein, we hypothesized that translational frameshifting may allow cobT mRNA to produce two cobaltochelatase subunits. Indeed, the phylogenetic analysis of proteobacteria revealed a strong correlation between the presence of the PRF signal in the cobT gene and the absence of a separate cobS gene in the genome. Particularly, the cobS gene is missing in the genome of the cobalamin synthesizing pathogenic bacteria P.aeruginosa while its cobT gene contains a typical -1 PRF signal. Thus, it may have a potential to produce the missing small chelatase subunit via premature termination of the cobT translation. Therefore, we describe yet another case where translational frameshifting allows mRNA to produce two different proteins. This mechanism is highly conserved and can be found in both archaea and bacteria. Our results may also provide some useful information to the evolution of chlorophyll and/or cobalamin biosynthesis pathways.

July 29th, 17:10

Ancient genomic regions of extreme conservation: Insights into evolution, structural organization, and function

Dmitry Korkin

Genomic regions of extreme conservation are nucleotide sequences shared between multiple distantly related species that are identical in the corresponding genomes, or allow several mutations at most. In our work, we have explored the extreme conservation beyond tetrapods, which includes jawless fish and even invertebrates. By integrating data from multiple high-throughput experiments, we share our recent findings about these ancient extreme elements, their possible origins, structural organization, and functional roles.

July 30th, 10:00

Inferring complex pseudo-time trajectories in single cell data using elastic principal graphs and STREAM pipeline

Andrei Zinovyev

We present STREAM, an interactive pipeline capable of disentangling and visualizing complex branching trajectories from both single-cell transcriptomic and epigenomic data (such as scATAC-seq data). As its core algorithm for complex pseudo-time trajectory inference STREAM uses elastic principal graphs and the framework of topological grammars for graph structure optimization, a general-purpose machine learning approach for creating complex data approximators. STREAM is an important tool to study cellular development and differentiation: it can accurately recover and describe complex developmental trajectories, it provides informative and intuitive visualizations to highlight important genes that define cell-fate decisions and subpopulation composition and it is an accessible to researchers even with limited computational skills to analyse and share their single-cell-based trajectory analyses.

July 30th, 10:40

Transcriptional Dynamics of Single Cells

Ruslan Soldatov

Single-cell transcriptomics captures static snapshot of dynamic biological systems, such as differentiation or cell cycle. Analysis of spliced and unspliced reads reveals signal of transcriptional kinetics and enables accurate modeling of cellular dynamics.

July 30th, 11:20

Transcriptome map of the human, chimpanzee, bonobo, and macaque brains at the single-cell resolution

Ekaterina Khrameeva

We searched for gene expression traits separating humans from other primates by analyzing more than 100,000 cell nuclei and 422 tissue samples representing 33 brain regions of humans, chimpanzees, bonobos, and macaques. We show that gene expression evolves rapidly within cell types, with more than two-thirds of cell-type-specific differences not detected using conventional RNA sequencing of tissue samples. Neurons tend to evolve faster in all hominids, but astrocytes show more differences on the human lineage, including alterations of spatial distribution across neocortical layers.

July 30th, 12:10

Clustered mutations in human germline

Vladimir Seplyarskiy

Detailed investigation of mutational patterns essential to find cancer driver genes or genes associated with severe neurodegenerative disorders, to study natural selection and population history. Moreover, statistical analysis of cancerous mutations or population polymorphisms may provide an insight into underling mechanisms. Clustered mutations occurred as a complex event and manifested as two or more linked mutations in a same locus. This type of mutations especially informative for quantifying contribution of specific mutational mechanisms. We analyzed more than million clustered mutations extracted from the rare human polymorphism. Types of single nucleotide mutation non-randomly co-segregate in clusters and frequently demonstrate specific order. Mutations within the cluster created by single, may be complex mutational process and we applied neural networks to find mutational processes responsible for the observed clusters. First process is intensive only in 5% of the genome and de novo data shows that this process major contributor to the maternal clustered mutations. First mutation in clusters generated by the process frequently oriented by the direction of transcription, suggesting damage-induced nature and following mutations usually are C>G or C>T and likely introduced by error-prone polymerase. Therefore, recently discovered maternal clustered mutations are created by multi-step mutational process: polymerase introduce mismatched nucleotide opposite DNA lesion and recruited low fidelity polymerase to continue replication. Two other processes inferred by neural networks characterized by strand coordinated mutations of the same type and by unknown two-step mechanism. Statistical description of clustered mutations shows that in many cases underling mechanism involves few steps and uncover biology beyond maternal clusters.

July 30th, 12:30

Mutational patterns in E.coli evolution

Sofya Garushyants

In this work we reconstruct the signatures of mutational processes from available data on mutators and mutation accumulation lines and use the obtained signatures to describe the relative contribution of different mutational processes to divergence of E.coli.

July 30th, 12:50

mRNA editing: a genome-scale pre-adaptation in soft-bodied cephalopods

Mikhail Moldovan

In soft-bodied cephalopods, nucleotides which are edited on mRNA demonstrate mutational patterns strikingly different from those of their unedited counterparts. These patterns are consistent with the notion of editing sites being pre-adaptations for subsequent mutations. As mRNA editing is widely spread in soft-bodied cephalopod genomes, pre-adaptation is, at least in this case, one of the major forces driving the genome evolution.

July 30th, 13:10

Prevalent epistatic interactions between amino acid sites in Schizophyllum commune

Anastasia Stolyarova

The basidiomycete fungus S. commune is the most variable eukaryotic species known, providing the first opportunity to use population sequencing data to reveal epistatic interactions between deleterious mutations using patterns of short- and long-range linkage disequilibrium between biallelic sites.

July 30th, 15:00

Assessment of network module identification across complex diseases

Sven Bergmann

Many bioinformatics methods have been proposed for reducing the complexity of large gene or protein networks into relevant subnetworks or modules. Yet, how such methods compare to each other in terms of their ability to identify disease-relevant modules in different types of networks remains poorly understood. We launched the “Disease Module Identification DREAM Challenge”, an open competition to comprehensively assess module identification methods across diverse protein-protein interaction, signaling, gene co-expression, homology, and cancer-gene networks. Predicted network modules were tested for association with complex traits and diseases using a unique collection of 180 genome-wide association studies (GWAS). Our robust assessment of 75 module identification methods reveals top-performing algorithms, which recover complementary trait-associated modules. We find that most of these modules correspond to core disease-relevant pathways, which often comprise therapeutic targets. This community challenge establishes biologically interpretable benchmarks, tools and guidelines for molecular network analysis to study human disease biology (https://synapse.org/modulechallenge, https://doi.org/10.1101/265553).

July 30th, 15:40

Systems Medicine - or - What I learned about Arnold Schwarzenegger while studying breast cancer survival

Jan Baumbach

One major obstacle in current medicine and drug development is inherent in the way we define and approach diseases. We discuss the diagnostic and prognostic value of (multomics panels. We have a closer look at breast cancer survival and treatment outcome, as case example, using gene expression panels - and we will discuss the current "best practice" in the light of critical statistical considerations. In addition, we introduce computational approaches for network-based medicine. We discuss novel developments in graph-based machine learning using examples ranging from Huntington's disease mechanisms via Alzheimer's drug target discovery back to where we started, i.e. breast cancer treatment optimization - but now from a systems medicine point of view. We conclude that multi-scale network medicine and modern artificial intelligence open new avenues to shape future medicine. We will also have a short glimpse on novel approaches for privacy-aware sensitive medical data sharing. We quickly introduce the concept of federated machine learning and blockchain-based consent management to build a Medical AI Store ensuring privacy by design and architecture.

July 30th, 16:00

Scientific background for personal DNA tests (sponsored talk)

Alexandr Rakitko

Introduction. The availability of SNP array genome-wide genotyping technology increases the sample sizes dramatically. Nowadays the meta-analysis of genome-wide association studies (GWAS) could involve more than 1 million samples. The result of such work can be genetic risk scores (GRS) based on an effect of hundreds of thousands of genetic markers.

Aim. Estimate the predictive power of the previously reported GRSs using coronary artery disease (CAD) as an example. Build nonlinear binary classification algorithms taking into account genetic data and biochemical parameters.

Conclusion. Using the CAD example, we demonstrated that GRSs can significantly improve the accuracy of predicting the multifactorial diseases development. Moreover, taking into account non-genetic factors using non-linear models also improves the risk estimation quality. Validation of existing GRSs and the development of new ones seem to be the next step in the development of models for predicting risk of multifactor diseases.

July 30th, 16:40

Probabilistic approaches to positive and negative selection inference on coding regions in cancer

Donate Weghorn

Selection inference in cancer can identify genes that drive tumorigenesis when they are mutated. For this, a proper null model of the expected number and distribution of mutations is fundamental. We have developed a method for selection inference which is based on a probabilistic model that accounts for mutation rate variability as well as DNA sequence context.

July 30th, 17:20

Analysis of proteome-transcriptome correlations with self-contained GSA enhances consensus colon cancer classification

Ancha Baranova

Self-contained GSA may reconciliate of transcriptome and proteome data, while highlighting additional actionable protein level targets. Using Colorectal Cancers “Consensus Molecular Subtypes” as example, we show that CMS1 immunogenicity may be, at least in part, dependent on abundance of MHC-II related proteins, ‘Inflammatory’ and ‘Mesenchymal-like’ phenotype CMS4 showed upregulation of both ECM receptor interactors and components of complement pathway, proteome-level descriptors of CMS2 are reminiscent of ‘garden-variety’ pan-tumor metabolic signature, while CMS3 analysis was indicative that lipid-lowering drugs may aid in its CMS3 treatment.

July 30th, 17:40

Uncovering hidden sources of transcriptional dysregulation arising from inter- and intra-tumor heterogeneity

Alexander V Favorov

This study develops an innovative computational framework, Expression Variation Analysis (EVA), to model transcriptional dysregulation in cancer. Heterogeneity poses a major challenge in translational research. For example, inter-tumor heterogeneity limits the biomarker discovery and intra-tumor heterogeneity enables therapeutic resistance. Moreover, in some cancers driver mutations are insufficient to account for the widespread transcriptional variation responsible for these outcomes. Thus, new computational tools to model transcriptional variation are essential.

July 30th, 18:00

Prediction of functional effect of short in-frame indels

Vasily Ramensky

Small insertion and deletion polymorphisms (indels) are the second most common mutations in the human genome, after single nucleotide variants. Indels that occur in protein coding regions and do not cause frameshifts are of special interest because their potential influence on human traits remains largely unstudied and ranges from severe diseases to the lack of any phenotypic consequences. The aim of this project is to analyze the discriminative features of human in-frame indels and develop a tool to predict the functional effect of such indels.

Poster session

Joint mechanical, thermodynamic and electrostatic machine-learning model for prediction of nucleosome positions *Darya Afenteva*
Analysis of out of reference transcripts from RNA-seq libraries in crops *Dmitry Afonnikov*
Prediction of clusters of miRNA binding sites in mRNA candidate genes of luminal A and B subtypes breast cancer *Dana Aisina*
The interaction of miRNAs with mRNAs of genes in esophageal squamous cell carcinoma *Aigul Akimniyazova*
Variety of proline-specific peptidases in higher fungi *Nikita Alkin*
In Silico Discovery of Novel HIV-1 Entry Inhibitor Scaffolds Based on Broad and Potent Neutralizing Antibody N6 *Alexander Andrianov*
Development of Novel 1,2,4-Triazole-Based Compounds as Potential Aromatase Inhibitors: a Computational Study *Alexander Andrianov*
Identification of Novel Potential HIV-1 Entry Inhibitors Targeting CD4-Binding Site of the Envelope gp120: Click Chemistry In Silico, Structure-based Docking and Molecular Dynamics Studies *Alexander Andrianov*
Metagenomic analysis of Leptinotarsa decemlineata virome *Denis Antonets*

Combining data from TCGA and GEO databases to predict the level of alternative spliced isoforms in cancer cells
Ksenia Anufrieva

Diversity of CRISPR/Cas systems in human microbiomes in health and disease
Irena Artamonova

Genomic encyclopedia of bifidobacterial carbohydrates utilization
Aleksandr Arzamasov

CH-π stacking interactions in antibody-oligosaccharide binding
Vladimir Arzhanik

Genomics-based inference of metabolic capabilities for biosynthesis and degradation of amino acids in the human gut microbiome
German Ashniev

A Rapid Exact Solution for the Guided Genome Aliquoting Problem
Maria Atamanova

Medium spiny neurons genes pathways form four distinct clusters in stress affected murine model
Vladimir Babenko

FTO gene affects obesity specifically in Europeans
Roman Babenko

Proposal of a new alignment free resampling method for phylogeny construction
Yasin Bakis

Characteristics of miRNA interaction with CDS mRNA genes associated with CAG trinucleotide repeat disorders development
Ayaz Belkozhayev

Genome-wide association study of quantitative disease resistance against Verticillium wilt in the model legume Medicago truncatula *Cécile Ben*
Interplay of the Escherichia coli transcription factors involved in control of pathogenicity and translated as multiple protein forms *Tatiana Bessonova*
The dependence of homologous recombination rate on the level of heterozygosity in hypervariable fungus Schizophyllum commune *Aleksandra Bezmenova*
Searching for compensatory frameshifting indels and determining their role in protein evolution *Dmitry Biba*
Evolution of multi-chromosomal bacteria Vibrio spp *Olga Bochkareva*
Mitochondrial phylogeny of Alticola Blanford, 1881 with dN/dS estimation *Semyon Bodrov*
On the use of 12S rRNA marker fragment for resolving phylogenetic relationships *Alexander Bolbat*
Signatures of natural selection in mitochondrial genome of underground rodents *Olga Bondareva*
The search for bird mitochondrial genome adaptations to high altitude, migration, diving, wintering and flight *Valentina Burskaia*

Identification of the human ribosomal RNA contacts with mRNA-binding proteins using eCLIP data from ENCODE *Andrey Buyan*
Phylogenetic-based phenotype-related genetic marker selection method in Sechium edule *Ernesto Borrayo Carbajal*
Desiccation mediates stabilization of 3D chromatin structure in the cells of Polypedilum vanderplanki *Alexander Cherkasov*
A Review of Current Bioinformatic Tools to Assess Mitochondrial Heteroplasmy *Samantha Chill*
Batch effect visualization and correction with R package “proBatch” *Jelena Čuklina*
Novel cross-regulatory events via AS-NMD among RNA-binding proteins *Anastasia Danchurova*
Differentially Methylated Super-Enhancers Regulate Target Gene Expression in Human Cancer *Ludmila Danilova*
Tandem alternative splice sites may represent noise of splicing machinery *Stepan Denisov*
Structural modeling of MOAP-1-guided proapoptotic mitochondrial recruitment of tBid *Hanna Dudko*
svist4get: a simple visualization tool for genomic tracks from RNA-Seq, Ribo-Seq and other high-throughput sequencing data *Artyom Egorov*

Genome-wide association study of genetic principal components of four chronic musculoskeletal human pain types *Elizaveta E. Elgaeva*
Effect of mutations associated with the development of paraganglioma on the structure of succinate dehydrogenase *Evgenia Elizarova*
miRNA-directed activation of genes in macrophages during Mycobacterium tuberculosis infection *Anna Elizarova*
Diversity of Restriction-Modification Systems in Surface and Deep Layers of World Ocean *Anna Ershova*
Horizontal gene transfer of defense systems in microbial communities *Anna Ershova*
Gene coevolutionary network for drug resistance of M. tuberculosis *Gennady Fedonin*
Indomethacin enhances anti-tumor efficacy of a MUC1 peptide vaccine against breast cancer in MUC1 transgenic mice *Jacob Ferrier*
ImmGen Open Source Mononuclear Phagocyte Profiling Analysis *Anastasiia Gainullina*
New approaches for in silico artificial riboswitches design *Marina Galchenkova*
Functional annotation based on proteins alignment and paralogous expression in Polypedilum vanderplanki *Rozalina Galimullina*

Trinucleotide Base Pair Stacking Free Energy for Understanding TF-DNA Recognition and the Functions of SNPs *Jun Gao*
Promoter-level expression atlas of skeletal muscle in hibernating edible dormouse *Guzel Gazizova*
Construction of ensemble model for CNV calling using WES data *Veronika Gordeeva*
Mapping of genetic associations for agronomically important traits in Russian collection of rapeseeds (Brassica napus) *Rim Gubaev*
Seeking for mtDNA structural determinants of organisms longevity *Konstantin V. Gunbin*
Mistranslation optimization of mitochondrial genes drives mammalians longevity. A hypothesis *Konstantin V. Gunbin*
Predicting the percentage of duplicate reads in exome sequencing *Tatiana Gurbich*
Reconstruction and analysis of the of protein-protein interaction network involved in the human body weight regulation *Elena Ignatieva*
Comparison of Cysteine alkylating strategies in proteomics *Irina Ilina*
Mutually exclusive exons evolution *Timofei Ivanov*

Prediction of osa-miRNA binding sites in human mRNA genes *Anatoliy Ivashchenko*
Detection of the protein targets for small molecular ligands with use of the local sequence similarity estimation *Dmitry Karasev*
Computational approach for prediction of autoimmune disease-associated T-cell receptor epitopes *Vadim Karnaukhov*
Are metagenomic data sufficiently informative to robustly predict inflammatory bowel disease status? *Lusine Khachatryan*
Combining GTRD ChIP-Seq datasets with FANTOM5's transcription start sites for prediction of gene expression levels *Semyon Kolmykov*
Human cistrome - genome-wide map of human transcription factor binding sites derived from GTRD database *Semyon Kolmykov*
Role of guanine quadruplexes CpG methylation in epigenetic regulation *Dmitry Konovalov*
WI14 peptide corresponding to the inner domain of H1N1 Influenza hemagglutinin changes its secondary structure in acidic medium *Larisa V. Kordyukova*
Mutagenesis data in the modern age *Ilia Korvigo*
The whole-genome analysis of heat shock proteins system in extremophile nonbiting midges *Olga Kozlova*

Evolution of mitochondrial DNA inheritance patterns *Pavel Kravchenko*
Analysis of plant-specific SBP-box transcription factors during flower and fruit development in plants of diverse families *Pushpendra Kumar*
The relationship between DNA secondary structures, epigenetic markers, transcription factors and structural somatic mutations in cancer genomes *Anna Kurilovich*
Comparative analysis of genome sequences of polyphage Xanthomonas arboricola str. 3004 *Elena Kyrova*
BiGAnts: network-constrained biclustering of patients and multi-omics data *Olga Lazareva*
The comprehensive analysis of motifs co-occurrence in ChIP-seq data with MCOT package537 *Victor Levitsky*
Features of DNA sequences learned by neural networks *Anna Litvin*
Mathematical modeling of Vas Deferens Smooth Muscle Electrophysiology: Role of Ion Channels in Generating Electrical Activity *Chitaranjan Mahapatra*
A novel approach for prediction of immunopeptidomes *Dmitry Malko*
Interaction network between genes involved in defensive response to broomrape *Rodica Martea*

Possible role of upstream open reading frames in NMD-dependent degradation of human mRNAs *Lev Mazaev*
Toward genetic selection of dairy cows in Russia *Pavel Mazin*
Predicting DNA:RNA triplexes based on RNA secondary structure *Yulia Medvedeva*
Eliminating transcriptome-wide batch effects results in higher precision in allele-specific expression measurements *Asia Mendelevich*
Evolution of proteins of the nucleoplasmin family *Elizaveta Minina*
The highly-variable enhancers in the intra-specific morphological diversity *Elena Minkina*
Role of breast milk lipid composition in postnatal brain development *Aleksandra Mitina*
Essential and advantageous genes detection in bacterial genomes *Denis Moshensky*
Molecular modelling of T7 RNA polymerase available to work with DNA templates carrying artificial nucleotides *Sabina Mukba*
Characterization of epigenetic landscape of integrated HPV in oropharyngeal cancer *Vera Mukhina*

Modeling of spatial and temporal expression patterns of the gene ensemble active both in early Drosophila segmentation and neurogenesis *Ekaterina Myasnikova*
Accurate fetal DNA variant calling in the presence of maternal cell contamination *Elena Nabieva*
Systems based deconvolution of Immune-defense Transcriptional Signatures in the Arabidopsis Shoot Apical Meristem *Muhammad Naseem*
RNA sequencing as a diagnostic tool for exome- or gene panel-negative cases of neuromuscular Mendelian diseases *Sergey Naumenko*
Phylogenetic analysis of Chordopoxvirinae genomes with alignment-free approach *Tatiana Nepomnyashchikh*
Is habitat diversity associated with bacterial pangenome structure? *Daria Nikolaeva*
Demographic Inference from the Allele Frequency Spectrum *Ekaterina Noskova*
Patterns of associations of G-quadruplexes and histone modifications in human stem cells and brain tissue *Arina Nostaeva*
The genome of new Echinophthiridae Enderlein species Antarctophthirus nevelskoy *Valeriia Novikova*
Comparison of probability distributions for metagenomic sequence data *Vera Odintsova*

Comparative analysis of the lipid organization of various human brain regions *Maria Osetrova*
Search for cis-elements associated with response to dioxin in human *Evgeniya Oshchepkova*
Affinity of estrogen- and antiestrogen-binding to human alpha-fetoprotein: in silico study *Daria Ostroverkhova*
Selection on physical properties of promoters and TFBS DNA affects protein content *Alexander Osypov*
In Silico Docking of Labdane Diterpenoids Reveals Cytoprotective Activity against Cobra Venom and Novel In Vitro Antiophidian Neutralizing Activity *Glenn Oyong*
A hypothesis on the evolution of Myxosporea - parasites that lack major apoptotic pathways *Alexander Panchin*
Developing new inhibitor to diminish Integrase and LEDGF/p75 relation using an in silico approach *Umesh Panwar*
Switching the Post-Translational Modification of Elongation Factor P *Marina Parr*
Impact of somatic mutations disrupting RNA structure on alternative splicing *Dmitry Pervouchine*
The role of RNA secondary structure in expression of non-collinear transcripts *Dmitri Pervouchine*

Identification of new sulfonated inhibitors of human lactate dehydrogenase A through virtual screening and structural filtration *Aleksandra Polenova*
Reason of attraction between selectively neutral loci and its dependence on Ne *Nadezhda Potapova*
Microinversions on the evolutionary path connecting human and primates *Nadezhda Potapova*
Selection of key reprogramming and differentiation factors in the hemopoiesis by machine learning methods *Michael Raevskiy*
Genetic risk scores in multifactorial diseases *Alexandr Rakitko*
DNA methylation changes in regulatory regions in AML patients with mutations in transcription factors *Ekaterina Romanova*
Metabolic shifts during desiccation-rehydration cycle in the anhydrobiotic insect P. vanderplanki *Alina Ryabova*
A computational approach to identify transcription regulation patterns of functionally related genes in Drosophila melanogaster *Olga Rybina*
Epistatic interactions in protein sites that contribute to H5N1 virulence in mammals *Ksenia Safina*
Comparative analysis of aphids microbiomes *Yuliya Sarana*

Prediction of Dictyostelium discoideum chromatin loops from genomic sequence *Victoria Savinova*
Algorithm for searching multi-SNP combinations associated with variation in binary phenotype *Roman Sergeev*
A genetic handicap approach: how to estimate the genome-wide burden of slightly-deleterious variants in a model population *Viktor Shamanskiy*
Genome-wide association study finds new loci affecting N-glycosylation of human blood plasma proteins *Sodbo Sharapov*
Neuraminidase A from Streptococcus pneumoniae: Three Different Ways to Modulate Enzyme Activity *Yana Sharapova*
Identification of species-specific DNA-regulatory elements in brain *Abusaid Shaymardanov*
Construction of TCR's CDR3 loops conformations using in silico step-by-step single amino acid mutation approach *Dmitrii Shcherbinin*
ngsPSMC: genotype likelihood-based PSMC for analysis of low coverage NGS data *Vladimir Shchur*
Identification, evolution, and function of tryptophan leader peptides TrpL in alpha-proteobacteria *Aleksei Shevkoplias*
Molecular components of nematode intestinal pacemaker *Georgy Slivko-Koltchik*

Genome comparison of a group of bacteria by KEGG-modules and pathways *Mikhail Slizen*
Searching for inhibitors against the open conformation of human lactate dehydrogenase A *Elina Smolkina*
Heat shock proteins buffer environmental and genetic perturbations: from short-term to long-term effects *Anastasia Sokol*
SAPPHIRE: automatic pipeline for data analysis of high throughput amplicon sequencing and phylogeny reconstruction *Evgeniia Sotnikova*
Transfer of Evolutionary Computation Techniques to in vitro Evolution: Tests in Silico *Alexander Spirov*
The interplay of Piwi and heterochromatin proteins in transposable element silencing in the germline of Drosophila melanogaster *Anastasia Stolyarenko*
Rare-event probability estimation in the functional gene set enrichment analysis *Vladimir Sukhov*
TP63 and TRIM29 axes regulate landscape of EMT associated active enhancers in the basal epithelium of the prostate *Rinat Sultanov*
Yosshi: the bioinformatic approach to protein disulfide engineering *Dmitry Suplatov*

Open-access Mustguseal platform to explore protein superfamilies *Dmitry Suplatov*
Homology and Vavilov's law of homologous series I; Homology and Vavilov's law of homologous series II; Urbanization, domestication and destabilization *Valentine Suslov*
Comparative study of the IclR family of transcription factors and their DNA binding motifs: structure and co-evolution *Inna Suvorova*
Evolutionary analysis of cis-regulatory elements in the central nervous system of primates *Dmitry Svetlichnyy*
The mitochondrial genome of the moss Polytrichum commune (Polytrichopsida) *Aleksey Troitsky*
Insight into the genetic architecture of back pain and its risk factors from a large genome-wide association study *Yakov Tsepilov*
Improving the prediction of T-cell epitopes by introducing self/non-self discrimination *Vasily Tsvetkov*
Atomic-level modeling of the complexes of small-molecular inhibitors of PD-1/PD-L1 with PD-L1 dimers *Viktor Urban*
Research of pattern for DNA fragmentation in plants *Leonid Uroshlev*
Context analysis of the core promoter region of mouse genes differently expressed in hypothalamic energy-sensing neurons in response to weight-loss *Oleg Vishnevsky*

A model of regenerating patterns using a random-walk model *Hiroshi Yoshida*
Distribution of Runs Of Homozygosity (ROHs) along the human genome is shaped by recombination and purifying selection *Elizaveta Zezyulya*
Serine peptidases from the S1 family in Tenebrionidae beetles *Nikita Zhiganov*
CYP450 9e2 expression in black garden ant Lasius niger *Svetlana Zhukova*
Identification of Differentially Expressed Gene Modules in Heterogeneous Diseases *Olga Zolotareva*
Alterations of plasma lipid profile in schizophrenia *Dmitry Zubkov*
Dicyemida and Orthonectida: Two stories of body plan simplification *Oleg Zverkov*

Some nearby cafes:

1. Hleb Nasushny, Michurinskiy Prospekt, 5
(Хлеб насущный, Мичуринский проспект, 5)

2. Durdin, Michurinskiy Prospekt, 8/29
(Дурдин. Мичуринский проспект, 8/29)

3. Miratorg Burger-n-Fries, Lomonosovskiy Prospekt, 29/1
(Мираторг Бургер & Фрайз. Ломоносовский проспект 29, корп.1)

4. Readymade, Michurinskiy Prospekt, 3
(Рэдимэйд, Мичуринский проспект, 3)

5. Food Court in shopping mall “Kapitoliy”, Prospekt Vernadskogo, 6
(Ресторанный дворик в ТЦ Капитолий, Проспект Вернадского, 6)

Proceedings of the MCCMB2019 can be found at http://mccmb.belozersky.msu.ru/2019/index.html