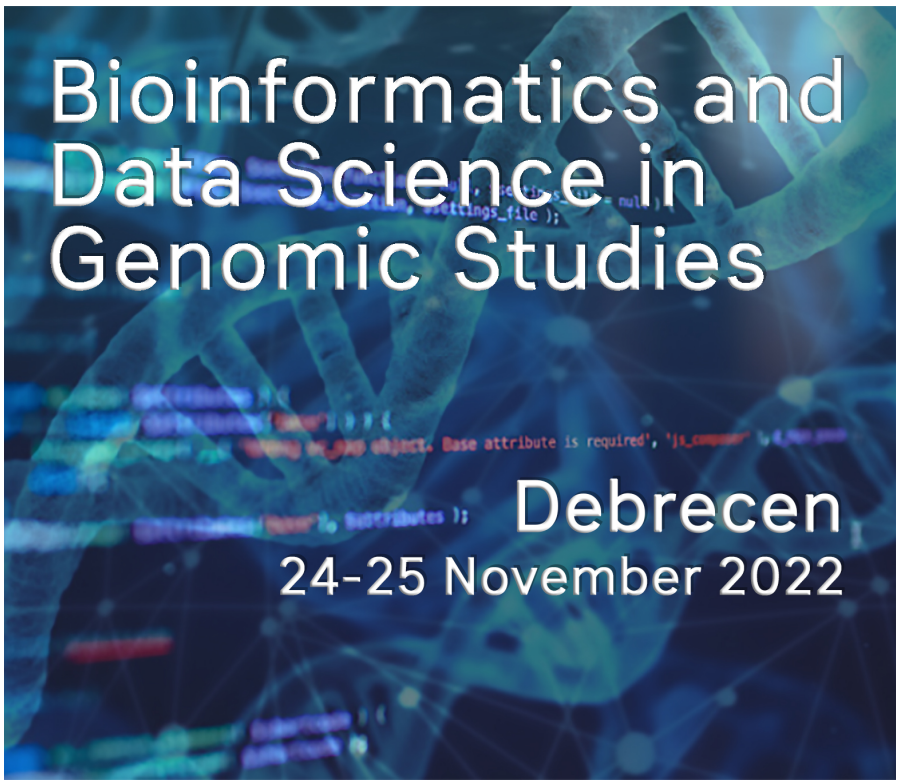
BDG 2022

Book of Abstracts



ORGANIZING & SCIENTIFIC COMMITTEE

Tímea Bárányné Pap
Zsuzsa Novák
Krisztián Bányai
Gábor Kardos
Géza Hegedűs
Alex Váradi
Krisztina Szarka
Eszter Virág
Eszter Kaszab
Eszter Fidrus
Péter Takács
Nikoletta Nagy
Levente Laczkó
Zoltán Rádai

Institute of Metagenomics
University of Debrecen

The work is supported by the GINOP-2.3.4-15-2020-00008 project.
The project is co-financed by the European Union and the European Regional Development Fund.

PROGRAMME

24th November, Thursday (forenoon)

10:00-10:10 Welcome speech
10:10-10:20 *Gábor Kardos*: Opening speech

MODELING, STATISTICAL ANALYSES, MACHINE LEARNING, BIG DATA I.

Chairperson: *Gábor Kardos*

Keynote presentation (30 minutes + 15 minutes discussion)

10:25-11:10 *Katalin Gombos:* Synthesis of molecular genetics and bioinformatics to develop clinical diagnostics and surveillance

Short presentations (15 minutes + 5 minutes discussion)

11:10-11:30 *Zoltán Rádai:* Complex effects of Next Generation Sequencing error sources on the quality of *de novo* genome assembly
11:30-11:50 *Alex Váradi*: Effects of genome size and composition and duplicate removal on the results of *de novo* transcriptome assembly and DGE analysis: an *in silico* study
11:50-12:10 *Otília Menyhart*: Preserved correlations identify extracellular matrix organization as a critical factor in pancreatic ductal adenocarcinoma

24th November, Thursday (afternoon)

MODELING, STATISTICAL ANALYSES, MACHINE LEARNING, BIG DATA II.

Chairperson: *Péter Takács*

Keynote presentation (30 minutes + 15 minutes discussion)

13:15-14:00 *Serhii Vakal*: Why protein folding is not yet solved, and you should be careful when using AlphaFold as a "black box"

Short presentations (15 minutes + 5 minutes discussion)

14:00-14:20 *Szonja Anna Kovács*: Investigation of predictive biomarkers in cancer patients treated with immune-checkpoint inhibitors
14:20-14:40 *János Tibor Fekete*: Predictive biomarkers in cancer cell lines
14:40-15:00 *Zoltán Pethő*: Protein homology modeling as a tool to study intra- and interprotein interactions

PROGRAMME

25th November, Friday (forenoon)

GENOME, PANGENOME, METAGENOME, POPULATION GENETICS, PHYLOGENETICS, PHYLOGENOMICS

Chairperson: *Eszter Virág*

Keynote presentation (30 minutes + 15 minutes discussion)

10:00-10:45 *Péter Oláh*: Analysis of whole-genome and marker gene-based cutaneous microbiome signatures in autoimmunity and allergy

Short presentations (15 minutes + 5 minutes discussion)

10:45-11:05 *Levente Laczkó*: Assessment of the accuracy of plasmid prediction tools using draft genome sequences

11:05-11:25 *Valter Péter Pfliegler*: Phylogenomics of *Saccharomyces* yeasts infecting and colonizing humans

11:25-11:45 *Eszter Ari*: Global map of evolutionary dependencies between antibiotic resistance and virulence genes in *E. coli*

25th November, Friday (afternoon)

TRANSCRIPTOME, PROTEOME, METABOLOME, FUNCTIONAL GENOMICS

Chairperson: *Nikoletta Andrea Nagy*

Keynote presentation (30 minutes + 15 minutes discussion)

13:00-13:45 *Eszter Virág*: Time-course gene expression analysis of the effect of SC-CO2 garlic extract encapsulated in nanoscale liposomes

Short presentations (15 minutes + 5 minutes discussion)

13:45-14:05 *William Jayasekara Kothalawala*: Transcriptomic and cell type enrichment analysis of colorectal cancer by combining multiple independent cohorts

14:05-14:25 *Eszter Kaszab*: Genetic characterization of probiotic candidate microbial strains

14:25-14:45 *Gergely Nagy*: Dissecting the cistromes determining bone marrow-derived macrophage identity

14:45-15:05 *Gyula Hoffka*: Structural analysis of the binding of nirmatrelvir to the SARS-CoV-2 main protease

15:05-15:15 Concluding remarks

ABSTRACTS

Keynote presentations

SYNTHESIS OF MOLECULAR GENETICS AND BIOINFORMATICS TO DEVELOP CLINICAL DIAGNOSTICS AND SURVEILLANCE

Katalin Gombos, Ágoston Hamar, Alex Váradi

The molecular diagnostics laboratory unit of the Department of Laboratory Medicine at Pécs University is involved in different areas of molecular genetic diagnostics and research. We identify molecular biomarkers from different human sample types including conventional blood samples as well as urine, liquor, saliva, body fluids and mucosal swabs. The applied workflow follows PCR amplification or hybridization of the target DNA or RNA molecules and corresponding fluorescent colorimetric detection in qPCR or droplet digital PCR detection formats. Our focused target analyses are often applied in infectious disease diagnostics and surveillance, eg. the SARS-CoV-2 pandemics or in tumor identification and follow up. In these applications molecular genetic data is integrated with other clinical or epidemiological data and analyzed by various bioinformatical tools. We demonstrate a dynamic map of SARS-CoV-2 that was constructed based on 270 000 genetic tests, clinical, geographic and demographic data to follow the spatial and timely spread of the virus within the South-Transdanubian population. We also introduce a 3D visualization of cancer organization field in a combined genomic and radiology based diagnostical approach we achieved by integrating different types of medical data by bioinformatics.

Keynote presentations

WHY PROTEIN FOLDING IS NOT YET SOLVED, AND YOU SHOULD BE CAREFUL WHEN USING ALPHAFOLD AS A "BLACK BOX"

Serhii Vakal

The recent introduction of AI-based protein modeling tools such as AlphaFold (AF) and RosettaFold, showing impressive results in the CASP14 benchmark, is considered one of the major scientific breakthroughs of the last few years. Indeed, the accuracy of the predictions offered by these tools is much higher than in previous approaches, almost comparable to experimental methods. Soon, the hype around AlphaFold emerged, with strong claims like "50-year-old problem finally resolved", "protein folding problem solved," or "AlphaFold will revolutionize drug discovery." However, despite the irrefutable achievements and advantages of AlphaFold, such claims are somewhat premature. Many limitations and modeling challenges should be kept in mind when using AI-driven tools like AF. For example, AF struggles with modeling disordered regions, especially in the case of intrinsically unstructured proteins. It cannot predict the oligomeric state of a protein. Moreover, it is not able to handle unnatural or modified amino acids. For some proteins, its predictions are heavily biased towards ligand-bound or apo-conformation. In the case of complex-forming proteins, AF usually predicts only one conformation – either free or complex-bound. Furthermore, it cannot predict conformation-changing events, like pathological misfolding of prion-like proteins. It can model only proteins but not co-enzymes and bound ligands. In the case of large proteins, it struggles to predict the mutual arrangement of domains and connecting loops. AlphaFold sometimes shows low confidence for correct models, which disallows researchers to rely on the prediction, and so on. Regarding drug discovery, although reliable modeling of drug targets is essential, especially in structure-based rational design, most of the time and costs are still spent not on initial discovery stages but on multphase clinical trials. This report will expand on current limitations and details that must be considered by any researcher using AlphaFold or RosettaFold as a "black box" without solid previous experience in protein modeling. In addition, useful hints and recommendations on how to get the best of these tools will also be provided.

Keynote presentations

ANALYSIS OF WHOLE-GENOME AND MARKER GENE-BASED CUTANEOUS MICROBIOME SIGNATURES IN AUTOIMMUNITY AND ALLERGY

Péter Oláh

The human skin harbors a diverse and dynamic microbiota, with high geographical, interpersonal and intrapersonal (topological and temporal) variability. While during the past decade, high-throughput sequencing efforts in combination with improved culture-based screening methods have provided insights into cutaneous species diversity with unprecedented detail, our understanding of the functional relationships and factors of homeostatic balance in the skin ecosystem is still limited. Thus, metagenomic studies of patient cohorts and *in vitro* models of diseases in which cutaneous homeostasis is upset are of key importance. Here, we present large-scale functional microbiome studies in two such model diseases, psoriasis and atopic dermatitis, with special regard to pathology-driven shifts observed in skin microbial communities.

Keynote presentations

TIME-COURSE GENE EXPRESSION ANALYSIS OF THE EFFECT OF SC-CO2 GARLIC EXTRACT ENCAPSULATED IN NANOSCALE LIPOSOMES

Eszter Virág, Barbara Kutasy, Kincső Decsi, Géza Hegedűs

The biostimulant phytochemicals as alternatives to synthetic chemicals are gaining ground in sustainable agricultural production nowadays. The medicinal herb, garlic (*Allium sativum*) has a spectacular therapeutic reputation due to its antimicrobial properties. The effectiveness of supercritical carbon dioxide (SC-CO2) extraction of *A. sativum* could help preserve bioactive compounds and be used as a biostimulant agent. The SC-CO2 garlic was formulated in liposomes and used as a nanoscale drug delivery system to reach better efficiency of penetration and translocation. Time course gene expression profiling was analysed as response to this agent in wheat. A comprehensive bioinformatics work involving combined pathway analysis showed a strong trigger of defence pathways of abscisic acid (ABA) and pathogenesis-related (PR) genes.

Short presentations

COMPLEX EFFECTS OF NEXT GENERATION SEQUENCING ERROR SOURCES ON THE QUALITY OF *DE NOVO* GENOME ASSEMBLY

Zoltán Rádai

Illumina next-generation sequencing is one of the most widely used technologies to acquire *de novo* genome sequences for bacteria. The quality of *de novo* assemblies is known to be shaped by a number of factors, such as the intrinsic error rate of the sequencing technology, sequencing depth, or the presence of PCR duplicates and optical duplicates. While the effects of these parameters on assembly quality are generally well described, it is still largely overlooked how they might interact with one another, potentially modifying their effects on assembly quality. In our study, we used 13 bacterial reference genomes to simulate sequencing data with varying error loads, in order to investigate the joint effects of error rate, sequencing depth, PCR duplicate ratio, and optical duplicate ratio, and we formally tested whether or not these sample parameters have significant interactions. We found significant interactions between the tested parameters, hinting at a complex picture on how they shape the quality of *de novo* assemblies.

Short presentations

EFFECTS OF GENOME SIZE AND COMPOSITION AND DUPLICATE REMOVAL ON THE RESULTS OF DE NOVO TRANSCRIPTOME ASSEMBLY AND DGE ANALYSIS: AN *IN SILICO* STUDY

Alex Váradi

During the past decade, RNA sequencing methods became a standard tool in transcriptomics, including differential gene expression (DGE) studies. It enables both the sequence reconstruction of all RNA molecules expressed in a sample and the quantification of the absolute expression levels of the transcripts. One of the mainly used transcriptome assembly approaches is the *de novo* assembly, which can be used to study microbial communities or uncultured microorganisms when the study organism's genome is unknown.
However, *de novo* assembly techniques also pose challenges, such as the appropriate reconstruction of repetitive regions, artificial (optical, PCR) duplicates, uneven expression levels, paralogs, and multigene families. Numerous pipelines and computational tools have been developed to address these challenges. However, in most cases, they focused on comparing workflows, preset parameters and assemblers but have not examined how the properties of the studied organism's genome affect the *de novo* assembly outcome and the identification of differentially expressed genes. Furthermore, previous studies have only examined the outcome of duplicate reads and PCR artefact removal on the assembly quality but not on transcript quantification.
In this study, we used coding sequences of nine bacterial species to simulate RNA-seq datasets with different coverages, and varying levels of differentially expressed genes. The simulated reads were the subject of *de novo* transcriptome assembly using rnaSPAdes. We then investigated the effects of genome properties , such as genome size and GC content, on the quality and content of *de novo* assemblies and the results of DGE analysis.

Short presentations

PRESERVED CORRELATIONS IDENTIFY EXTRACELLULAR MATRIX ORGANIZATION AS A CRITICAL FACTOR IN PANCREATIC DUCTAL ADENOCARCINOMA

Otília Menyhart, Áron Bartha, Balázs Győrffy

Correlated gene expression frequently signals shared biological functions with coordinated regulation. We hypothesized that maintained correlations might be essential for cellular survival, representing potential vulnerabilities of cancer cells. We aimed to reveal correlations preserved in pancreatic ductal adenocarcinomas (PDAC) across normal and tumor tissues.

We searched the NCBI GEO for raw microarray data and the TCGA project for RNA-seq data. The microarray dataset consisted of 248 tumors and 108 normal samples, allowing the analysis of 12,210 genes. The RNA-seq dataset incorporated 177 tumors, four normal samples from TCGA, and 248 normal samples from GTEx, enabling the analysis of 21,479 genes. We identified genes with an altered expression with a Mann-Whitney U test at $p<0.01$ and performed a Pearson correlation to identify preserved correlations.

Altogether 371 significant correlations involving 262 genes were preserved across normal samples and tumors in both RNA-seq and gene chip platforms. The identified close-knit gene network is mainly responsible for extracellular matrix organization. Seven genes (SPARC, COL6A3, MMP2, HTRA1, FN1, PALLD, and COL3A1) were heavily overrepresented in maintained correlations, some of them participating in as many as 58 interactions. The high expression of 28 genes was linked to poor disease outcome at FDR≤10%. The growing expression of two genes, MYL12A and MYL12B, across normal tissues, primary and metastatic tumors may drive the acquisition of motility by cancer cells.

Our results propose novel prognostic biomarkers of PDAC and pinpoint fundamental cellular interactions as potential targets for combined therapies. Furthermore, the presence of significant correlations across different data platforms substantiates the validity of our findings.

Short presentations

INVESTIGATION OF PREDICTIVE BIOMARKERS IN CANCER PATIENTS TREATED WITH IMMUNE-CHECKPOINT INHIBITORS

Szonja Anna Kovács

The availability of immune-checkpoint inhibitors (ICIs) in the last decade has resulted in a paradigm shift in oncology. Patients can be treated either by anti-CTLA-4, anti-PD-1, or anti-PD-L1. To maximize the benefit of cancer immunotherapy, the prediction of the actual immune response by the identification of clinically useful biomarkers will be required. Our goal was to identify predictive biomarkers of immunotherapy - in particular ICIs - in silico. We searched publicly available datasets for transcriptomic and clinical data gained from patients treated with anti-PD-1, anti-PD-L1, or anti-CTLA-4. Duplicate records, cell lines, immune cells, non-tumor, scRNA-Seq samples were excluded from our analysis. We used unpaired Wilcoxon-test and ROC for statistical analysis. Bonferroni correction was used to screen for the best genes. Our final database contains 20 datasets with 2159 samples from 1688 patients. From the 1,290 melanoma cases (n=720 blood, and n=570 tissue), 501 samples were treated with anti-PD-1. We found that YAP1 (pBonferroni=3.22E-05, FC=2.20), SPIN1 (pBonferroni=1.15E-08, FC=1.70), EIF4H (pBonferroni=2.92E-07, FC=1.69), SLC25A36 (pBonferroni=1.07E-06, FC=1.73), ACER3 (pBonferroni=4.47E-06, FC=1.54), LYPLA1 (pBonferroni=4.46E-06, FC=1.64), KDSR (pBonferroni=2.11E-06, FC=1.53) had the highest potential to predict resistance in patients. YAP1 plays a role in cancer progression and can be inhibited by a clinically available drug. We established a database capable of exploring and validating biomarkers in cancer patients treated with immune-checkpoint inhibitors. *Ex vivo* and *in vivo* studies are currently under progress.

Short presentations

PREDICTIVE BIOMARKERS IN CANCER CELL LINES

János Tibor Fekete, Balázs Győrffy

In vitro cell line models provide a valuable resource to investigate compounds useful in the systemic chemotherapy of cancer. However, due to the dispersal of the data into several different databases, the utilization of these resources is limited. Here, our aim was to establish a platform enabling the validation of chemoresistance-associated genes and the ranking of available cell line models.

We processed four independent databases, DepMap, GDSC1, GDSC2, and CTRP. To evaluate the therapeutic response, IC50 and AUC values from the source databases were categorized into sensitive and resistant cell lines using lower/upper tertile values as cut-off points. Mann-Whitney test, random forest algorithm and ROC analysis were used to analyse the relationship between gene expression and therapeutic response.

In total, data for 1,562 therapeutic agents and 1,250 cell lines are available in the analysis interface. The application can be used for individual and simultaneous analysis of multiple genes. An additional possibility is to analyze the genes of each KEGG pathway together and to identify the most correlated genes with the therapeutic response using a random forest algorithm.

The computational tool is useful to correlate gene expression with resistance, to identify and rank resistant and sensitive cell lines, and to rank resistance-associated genes. The online application is available at www.rocplot.org/cells.

Short presentations

PROTEIN HOMOLOGY MODELING AS A TOOL TO STUDY INTRA- AND INTERPROTEIN INTERACTIONS

Zoltán Pethő

Computational studies have become indispensable in studying the structure and interactions of biomolecules. In parallel with the resolution of numerous novel crystal structures, homology modeling of proteins based on sequence-based homology detection has emerged as one of the most accurate and sensitive approaches to predict protein structure.
In our recent studies, we utilized this homology protein modeling for two reasons. First, we mapped the orientation of linked fluorophores on the human voltage-gated proton channel Hv1. With the help of the model, we interpreted signals obtained with voltage-clamp fluorimetry. In another study, we created homology models in order to study the consequences of mutations occurring in multiple proteins of the human papillomavirus 11. We found that multiple mutations in the E2 and L1 proteins lead to a change of intra- and interprotein interactions that ultimately affect disease progression. Taken together, protein homology modeling helps us understand the structure-function relationship of proteins and is thereby invaluable in computational biomedical research.

Short presentations

ASSESSMENT OF THE ACCURACY OF PLASMID PREDICTION TOOLS USING DRAFT GENOME SEQUENCES

Levente Laczkó, Eszter Kaszab, Gábor Kardos, Zoltán Rádai, Ákos Tóth

Next Generation Sequencing-based genome assemblies often result in draft genomes. In such genomes, plasmids are usually found in multiple pieces. Several methods have been devised for their identification, but it is unclear which of them can be used given certain conditions. Tracking plasmids is only possible after their accurate identification and can be important for assessing the spread of resistance genes, which are often encoded on plasmids. Using the genome sequence of *E. coli* and the sequences of its known plasmids, we benchmarked eight commonly used plasmid prediction tools (plasflow, plasclass, mlplasmids, rfplasmid, plasforest, mob-suite, plasmidfinder, platon). Sequencing libraries were simulated in silico, and the genome contiguity was influenced by setting different sequencing error rates (0, 0.01, 0.001). The performance of tools showed a large variation, with a sensitivity of 0.04 (mob-suite) to 0.99 (plasflow, platon), while the specificity ranged from 0.006 (platon) to 0.99 (plasmidfinder, mob-suite). Based on our preliminary results, we recommend the simultaneous application of multiple tools for the reliable identification of plasmid sequences in draft genomes.

Short presentations

PHYLOGENOMICS OF *SACCHAROMYCES* YEASTS INFECTING AND COLONIZING HUMANS

Valter Péter Pfliegler

The human mycobiome, including its role in health and disease is increasingly studied, but still lags behind the study of the bacterial microbiome. Several species of yeasts may colonize us, at least temporarily, including *Saccharomyces cerevisiae* yeasts used in food and beverage production, or those used in probiotics. These yeasts may occasionally switch to a pathogenic lifestyle as well. Phylogenomic methods are required to understand the origin and evolution of these microscopic fungi, but many aspects of the yeast genome complicate phylogenomic workflows.
We aimed to analyze a large collection of sequenced human *S. cerevisiae* isolates, along with yeast genomes extracted from stool and vaginal shotgun metagenomic datasets to compare them to the various clades and subgroups of the species. We applied a fine-tuned pipeline for variant calling and phylogenomic analysis that works reliably for polyploid, highly heterozygous yeasts with highly admixed genomes as well. This pipeline helps to understand where the human isolates of the species originated from, and how they adapted to their novel lifestyles.

Short presentations

GLOBAL MAP OF EVOLUTIONARY DEPENDENCIES BETWEEN ANTIBIOTIC RESISTANCE AND VIRULENCE GENES IN *E. COLI*

Eszter Ari, Enikő Kiss, Bálint Vásárhelyi, Gergely Fekete, Ágoston Hunya, Bálint Kintses, Balázs Papp

Genes conferring antibiotic resistance or virulence phenotypes frequently undergo horizontal gene transfer in bacteria, contributing to the emergence of new multidrug resistant pathogenic variants. Mounting evidence indicates that pre-existing genome content variations influence the successful acquisition of such genes. However, the underlying evolutionary dependencies among specific genes, i.e. when one gene facilitates or hinders the acquisition of a second gene, remain poorly understood. Here we chart a high-resolution map of evolutionary dependencies between resistance and virulence genes by phylogenetic analysis of more than 20,000 *Escherichia coli* genomes. Our map reveals that (1) resistance genes generally facilitate each other's gain; (2) key virulence genes lack such a general pattern; and (3) contrary to some previous results, there is no overall negative dependency between the acquisitions of key virulence and resistance genes, indicating largely independent evolution between these two traits in *E. coli*. Strikingly, we found that the presence of efflux pump resistance genes in a genome strongly increases the chance of acquiring various other classes of resistance genes, making these efflux pumps an indicator of potentially emerging new multidrug resistant strains.

Short presentations

TRANSCRIPTOMIC AND CELL TYPE ENRICHMENT ANALYSIS OF COLORECTAL CANCER BY COMBINING MULTIPLE INDEPENDENT COHORTS

William Jayasekara Kothalawala

By linking the cellular content and molecular subtypes of Colorectal Cancer (CRC) we aim to uncover novel features that are useful for targeted therapy. Our first goal was to evaluate gene expression alterations linked to CRC pathogenesis. Then we assessed the cellular composition differences between normal colon mucosa and tumor and between different colon cancer molecular subtypes.
We collected microarray and RNA-seq data of CRC patients from the GEO and TCGA. We combined all cases and performed quantile normalization. Genes with a FC>2 were further investigated. We used xCell for cellular decomposition and cmsCaller for molecular subtyping.
We established an integrated database of normal colon and colorectal cancer using transcriptomic data of 1,082 samples. By using this dataset, we identified genes showing the highest differential expression in colon tumors. The top genes were linked to calcium signaling, matrix metalloproteinases, and transcription factors. When compared to normal, CD4+ memory T-cells, CD8+ naïve T-cells, CD8+ T-cells, Th1-cells, Th2-cells and regulatory T-cells were enriched in tumor tissues. The ImmuneScore decreased in tumor samples compared to normal. The CMS1 and CMS4 molecular subtypes were the most immunogenic, with the highest ImmuneScores but also high infiltration by CD8+ T-cells, Th1-cells and Th2-cells in CMS1 and B-cell subtypes and CD8+ T-cells in CMS4.
Our analysis uncovers features enabling advanced treatment selection and the development of novel therapies in CRC.

Short presentations

GENETIC CHARACTERIZATION OF PROBIOTIC CANDIDATE MICROBIAL STRAINS

Eszter Kaszab, Levente Laczko, Eszter Fidrus, Krisztián Bányai, Gábor Kardos

Probiotics are defined as live microorganisms which confer health benefits to the host when administered in adequate amounts. Many lactic acid bacteria (LAB) strains have been classified as probiotics. In this study, novel probiotic candidates were isolated and characterized from various source (fermented products, cervix and stool samples, dairy products) by whole-genome analysis. Analyzing the complete genetic information, including potential virulence genes and antimicrobial resistance genes with a negative impact on health, prediction of metabolic pathways allows us a cost-effective approach for the initial *in silico* microbial risk evaluation. Identification and characterization of strains found in various products and human samples may help us to understand the background of their probiotic properties.

Short presentations

DISSECTING THE CISTROMES DETERMINING BONE MARROW-DERIVED MACROPHAGE IDENTITY

Gergely Nagy

Mouse bone marrow-derived macrophages (BMDMs) are a widely used model to study gene expression regulation; therefore, a large number of various next-generation sequencing (NGS) results are available from them (such as bisulfite-, ATAC-, ChIP-, and GRO-seq data). Using these data, we aimed to study how the interaction of cis- and trans-regulatory elements contributes to transcription regulation in macrophages.
We collected 42 cistromes including those of bZIP, ETS, bHLH, and MEF2 family members, and our first goal was to identify the direct binding elements of the major transcription factors (TFs) participating in the gene regulation of BMDMs. Because of the similar DNA sequence preferences of members of a given TF family, binding sites of related TFs were merged and divided into clusters of genomic sites with similar TF binding pattern to identify the direct binding elements.
Using the DNA sequences enriched in different clusters, it became clear that methylatable but non-methylated sequences and their specific binding TFs are the most potent transcription initiators in the lead with the promoter-specific ETS- and E-boxes. Based on ATAC-seq results, additional promoter-specific motifs (like SP1 and NFY) could be mapped, of which presence rather contributes to chromatin openness. Non-methylatable ETS- and E-boxes are also good “openers” and initiators relative to the bZIP and MEF2 binding sites. Our results provide insight into the initial step of gene regulation, and the exact knowledge of promoter building blocks may help one understand how the expression of individual genes is regulated.

Short presentations

STRUCTURAL ANALYSIS OF THE BINDING OF NIRMATRELVIR TO THE SARS-COV-2 MAIN PROTEASE

Gyula Hoffka, Mohamed Mahdi, József Tőzsér, János András Mótyán

The coronavirus disease 2019 COVID-19 pandemic has led to an unprecedented loss of life in the 21st century, the severe acute respiratory syndrome 2 coronavirus (SARS-CoV-2) had resulted in over 630 million infections and more than 6.5 million deaths worldwide, according to the World Health Organization. Several antivirals were repurposed as experimental treatments against the virus, out of which only remdesivir became widely approved and used. Nirmatrelvir is a novel antiviral medication developed by Pfizer to offer an effective treatment of the COVID-19 disease. Acting as an inhibitor of SARS-CoV-2 main protease (Mpro), the main protease of the virus, it is able to covalently bind to its catalytic Cys145 residue.
Due to the high mutation rate of SARS-CoV-2 proteins, including the Mpro, there is a wide range of possibilities for the emergence of resistance mutations, as a defense against nirmatrelvir. Thus it is important to examine the molecular background of the binding properties of nirmatrelvir to SARS-CoV-2 Mpro and the mechanism behind resistance development.
In order to map the crucial interactions that are responsible for the binding of nirmatrelvir, we have studied several available crystal structures of Mpro complexed with the inhibitor. For comparison, we have investigated enzyme-substrate complexes, as well. To determine the most polymorphic residues, we have also examined the mutation rates of the individual residues, based on the available Mpro sequences of circulating virus variants. In our ongoing studies, we focus on the prediction of possible resistance mutations with the application of multiscale quantum mechanical/molecular mechanical methods.
This project was supported by the grant for high-risk research in the post-COVID field of the Hungarian Academy of Sciences (POST-COVID2021-16).


SZÉCHENYI 2020
HUNGARIAN GOVERNMENT
European Union
European Regional Development Fund
INVESTING IN YOUR FUTURE