Session 1. Genomics and Genomics – Driven Research

Lectures

L1.1

Activity of retroposed genes in tumors

Izabela Makałowska, Joanna Ciomborowska, Patrizia Marciniak, Rafał Grochala, Michał Szcześniak

Adam Mickiewicz University in Poznan, Institute of Antropology, Laboratory of Integrative Genomics, Poznań, Poland
Izabela Makałowska <izabel@amu.edu.pl>

Retrogenes represent cDNA copies of mRNA transcripts that are by-products of LINE-mediated retrotransposition. Such duplicates of genes are in majority considered as pseudogenes. However, recent studies demonstrated that some of them play important regulatory roles and are critical in carcinogenesis. Nevertheless, retrocopies of protein coding-genes have not been systematically studied in cancer. We performed complex analyses of retroposed genes in lung and breast tumors as well as in corresponding healthy tissues. Differential expression analysis revealed retrogenes with significant changes in expression in all studied tumors. Moreover, a number of retrocopies demonstrated expression pattern specific for particular type of cancer (adenocarcinoma, breast ER+, breast TNBC). Differential expression of the most promising candidates for biomarkers was further examined experimentally. In addition, a series of analyses was performed in order to pinpoint potential functions of retrocopies. For example, transcriptome wide miRNA target analysis, jointly with expression correlation study, was carried out to identify retrocopies that could play a role as miRNA sponges.

L1.2

Liquid biopsy: challanges and opportunities

Tomasz Zemojtel

Berlin Institute of Health, Genomics Platform, Berlin, Germany
Tomasz Zemojtel <tomasz.zemojtel@bihealth.de>

Liquid biopsy, due to the minimally invasive sampling method, allows for serial monitoring of cancer treatment and can reveal treatment resistance. The wide implementation of next generation sequencing enables swift adoption of liquid biopsy by clinical labs.
In this talk challenges and opportunities associated with further development of liquid biopsy-based testing will be discussed.

L1.3

AthCNV – a map of DNA copy number variations in *Arabidopsis thaliana* genome based on the 1001 Genomes Project data

Agnieszka Zmienko[1,2], Małgorzata Marszałek-Zenczak[1], Pawel Wojciechowski[1,2], Anna Samelak-Czajka[1], Marek Figlerowicz[1,2]

[1]Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznań, Poland; [2]Institute of Computing Science, Poznan University of Technology, Poznań, Poland
Agnieszka Żmieńko <akisiel@ibch.poznan.pl>

Increasing evidence confirms the pivotal role of copy number variations (CNV) in shaping intra-species genetic variation. Although CNV affects multiple genes important for agronomic traits in plants, the current information on CNV in *Arabidopsis thaliana* has been limited, restricting the number of CNV-focused studies in this model plant. Recently, the 1001 Genomes Project provided whole-genome sequence data of > 1 000 *A. thaliana* accessions. With these data we set up a pipeline combining 6 well validated CNV detection tools and applied it for large (0.5 kb to 1 Mb) CNV discovery in a worldwide *A. thaliana* population. The resulting AthCNV map consists of 19 000 nonredundant variants that together cover on average 35% of *A. thaliana* nuclear genome (27% outside the centromeres). Unlike previous predictions, this coverage resembles the high CNV genomic load observed in other plant species. The identified CNVs overlap 17.6% of 27445 protein coding genes by at least 80%. We computed the copy numbers of CNV-affected genes in 1060 accessions using whole-genome sequencing data. Experimental validation with MLPA performed for about 40 genes in ~240 accessions proved high accuracy of our predictions. The AthCNV is a powerful resource that will support functional analyses and genome-wide association studies in *A. thaliana*.

Acknowledgements:
This study has been supported by National Science Centre grants 13/B/NZ2/03837 and 2017/01/X/NZ2/00144. Computations were performed on PLGrid Infrastructure.

Oral presentations

O1.1

Half the costs keep the quality. Targeted enrichment for the Next Generation Sequencing of ancient genomes.

Malgorzata Marcinkowska-Swojak[1], Ireneusz Stolarek[1], Michal Zenczak[1], Luiza Handschuh[1,2], Anna Philips[1], Anna Juras[3], Piotr Kozlowski[1,4], Marek Figlerowicz[1,4]

[1]Institute of Bioorganic Chemistry PAS, Department of Molecular and Systems Biology, Poznan, Poland; [2]Poznan University of Medical Sciences, Department of Hematology and Bone Marrow Transplantation, Poznan, Poland; [3]Adam Mickiewicz University in Poznan, Department of Human Evolutionary Biology, Poznan, Poland; [4]Poznan University of Technology, Institute of Computing Science, Poznan, Poland
Małgorzata Marcinkowska-Swojak <m-marcinkowska@o2.pl>

Next-generation sequencing (NGS) technologies are routinely used to sequence DNA of many ancient specimens, including humans. In practice, the genetic material isolated from fossil remains is a mixture of endogenous and exogenous DNA, in varying proportions, thus, direct NGS is costly and ineffective. Therefore, strategies like targeted enrichment are developed to increase the amount of endogenous human ancient DNA (aDNA) available for the NGS. Targeted enrichment uses biotinylated RNA probes (baits) to hybridize with the complementary DNA sequences and pull out either the entire genome or its selected fragments. We examined the enrichment efficiencies of commercially available MyBaits (Arbor Biosciences) kit with 3 types of baits for: whole genome enrichment (WGE), mitochondrial genome (mtDNA) enrichment (MGE) and Y chromosome (Ychr) enrichment (YCE). We showed that WGE was most efficient with only 1 round of capture, whereas 2 rounds of capture produced better efficiency for mtDNA and Ychr. Moreover, in samples rich in human DNA (≥10%) WGE alone was sufficient to enrich for both whole genome and mtDNA. For the Ychr analyses the YCE was required for male samples containing 1–50% human DNA. We have also found that using only half of the baits provides equally good enrichment results for the YCE and MGE. Hence, we put forward the protocol where half-portions of mtDNA and Ychr baits are combined reducing time and cost of the enrichment step with no decrease in the efficiency.

O1.2

Catechins as potential epigenetic modulators

Patrycja Jakubek[1], Monika Baranowska[1], Jovana Rajić[2], Melita Vidaković[2], Jacek Namieśnik[3], Agnieszka Bartoszek[1]

[1]Department of Food Chemistry, Technology and Biotechnology, Faculty of Chemistry, Gdańsk University of Technology, Gdańsk, Poland; [2]Institute for Biological Research "Siniša Stanković", Molecular Biology Department, University of Belgrade, Belgrade, Serbia; [3]Department of Analytical Chemistry, Faculty of Chemistry, Gdańsk University of Technology Gdańsk, Poland

Patrycja Jakubek <patrycja.jakubek93@gmail.com>

Year by year, the public interest in relationship between nutrition and health is growing. Many dietary components that are known to provide health benefits may also act at the epigenetic level, which makes them attractive as new potential agents in prevention and treatment of chronic diseases. Antioxidants constitute one of the most thoroughly studied groups of food compounds that are expected to promote health and delay ageing. This group includes catechins known to play a role in modulation of cellular redox homeostasis and various signalling pathways. Recently, catechins have been also reported to affect the methylation profile of DNA.

Our previous study showed that catechins, depending on the concentrations used, can modulate the expression of genes involved in redox homeostasis. Following this observation, we used MS-PCR and MS-HRM to analyse DNA methylation profiles of promoter areas of genes whose expression was affected by studied catechins. The objective of the research was to find out whether the observed dose dependence is caused by epigenetic modulation of gene expression.

The current literature reports that catechins act as inhibitors of methylation, however, the results we obtained indicate that they also can increase the methylation level. Seeing that the same result was observed for glutathione, taken as a reference antioxidant, this change in methylation levels seems rather upstream of cellular redox state than related to chemical structure.

Reference:

Baranowska M. *et al.* (2018) *Redox Biology* **17**: 355-366.

Acknowledgements:

This work was supported by the National Science Centre in the framework of programme "Maestro 6" (application no: 2014/14/A/ST4/00640). Part of this work was supported by the COST Action CA16112 "NutRedOx: Personalized Nutrition in aging society: redox control of major age-related diseases".

O1.3

Investigating the pangenome of the plant pathogenic *Pectobacterium parmentieri* provides insight into its outstanding adaptation abilities

Sabina Zoledowska[1], Agata Motyka[1], Wojciech Sledz[1], Alessio Mengoni[2], Ewa Lojkowska[1]

[1]Department of Biotechnology, Intercollegiate Faculty of Biotechnology UG and MUG, Gdańsk, Poland; [2]Department of Biology, University of Florence, Sesto Fiorentino, Florence, Italy

Sabina Elzbieta Zoledowska <sabina.zoledowska@biotech.ug.edu.pl>

P. parmentieri is a newly established species within plant pathogenic *Pectobacteriaceae* (Adelou *et al.*, 2016). Microorganisms belonging to this species are causative agents of diseases in economically important crops (e.g. potato) in a wide range of environmental conditions. Severe disease symptoms are caused mainly by the activity of *P. parmentieri* virulence factors, like Plant Cell Wall Degrading Enzymes (PCWDEs; Perombelon 2002). Interestingly, we observe significant phenotypic differences among *P. parmentieri* isolates in terms of PCWDE activities, other virulence factors production or/and ability to macerate potato tissue (Zoledowska *et a*l., 2018). In order to establish the possible genomic basis of these differences, we sequenced 12 genomes of *P. parmentieri* strains (10 isolated in Poland, 2 in Belgium) with the combined use of Illumina and PacBio approaches. *De novo* genome assembly was performed with the use of SPAdes software and gene annotation with NCBI PGAP. The pangenome study was performed on 15 genomes (12 *de novo* assembled, 3 references: *P. parmentieri* CFBP8475TS, *P. parmentieri* SCC3193, *P. parmentieri* WPP163). *P. parmentieri* pangenome includes 3706 core genes, and a high number of accessory (1468) genes, and unique (1847) genes. We were able to determine the presence of all well-known virulence factors in the core genome fraction, but one – pnl gene encoding pectin lyase which was absent in *P. parmentieri* IFB5486, isolated in Belgium. However, a large fraction of horizontally transferred genes, virulence-related gene duplications, as well different as CRISPR arrays were found, which can likely explain the observed phenotypic differences. Finally, we found also for the first time, the presence of a plasmid in *P. parmentieri* (strain IFB5427, isolated in Poland). These data enabled us to speculate on the importance of those processes for *P. parmentieri* adaptation to different environments and led to the current structure and gene content of *P. parmentieri* strains.

References:

Adeolu M, Alnajar S, Naushad S, Gupta RS (2016) Genome-based phylogeny and taxonomy of the '*Enterobacteriales*': proposal for Enterobacterales ord. nov. divided into the families *Enterobacteriaceae*, *Erwiniaceae* fam. nov., *Pectobacteriaceae* fam. nov., *Yersiniaceae* fam. nov., *Hafniaceae* fam. nov., *Morganellaceae* fam. nov., and *Budviciaceae* fam. nov. *International Journal of Systematic and Evolutionary Microbiology* **66**: 5575-5599.

Pérombelon MCM (2002) Potato diseases caused by soft rot erwinias: an overview of pathogenesis. *Plant Pathology* **51**: 1-12.

Zoledowska S, Motyka A, Zukowska D, Sledz W, Lojkowska E (2018) Population structure and biodiversity of *Pectobacterium parmentieri* isolated from potato fields in temperate climate. *Plant Disease* **102**: 154-164.

O1.4

Closure of the *Dickeya solani* pangenome

Agata Motyka[1], Sabina Żołędowska[1], Agnieszka Misztak[1], Wojciech Śledź[1], Alessio Mengoni[2], Ewa Łojkowska[1]

[1]Intercollegiate Faculty of Biotechnology University of Gdańsk and Medical University of Gdańsk, Department of Biotechnology, Gdańsk, Poland; [2]University of Florence, Department of Biology, Florence, Italy
Agata Motyka <agata.motyka@biotech.ug.edu.pl>

Dickeya solani is a recently established species of pectinolytic plant pathogenic bacteria classified to the *Pectobacteriaceae* family. Numerous plants are being affected by the symptoms of soft rot or blackleg caused by *D. solani*. As a result, high economic losses follow, reaching even 25% of the total yield. Previous research pointed to phenotypic diversity between distinct strains of *D. solani*, contrarily to the extremely low genomic variability observed in terms of sequencing of housekeeping genes in addition to REP, BOX and ERIC-profiling. We decided to investigate the pangenome of *D. solani* on the basis of 22 whole genome sequences in search for genetic foundations of the virulence-related variation. It was established that *D. solani* shows enormous core (3726 genes), relatively small accessory (113-271 genes) and exceptionally tiny unique (0-20 genes; except for 286 of RNS 512A strain) pangenome fractions. Categorization of orthologs was performed by function-based classification to Clusters of Orthologous Groups (COGs). Interestingly, the highest percentage of unique COGs after General Function and Function Unknown was assigned to Transcription-related COGs. Therefore, we hypothesize that diverse virulence of *D. solani* strains results from variations in the transcription regulation. Last but not least, by addition of the newly sequenced genomes the closure ($b \approx 0.026$) of *D. solani* pangenome according to Bacterial Pangenome Analysis (BPGA) software was documented.

Posters

P1.1

The genetic landscape of pre-state Iron Ages societies of the East Central Europe

I. Stolarek[1], L. Handschuh[1], A. Juras[2], M. Marcinkowska-Swojak[1], P. Kozlowski[1], M. Figlerowicz[1]

[1]Department of Molecular and Systems Biology, Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznań, Poland; [2]Department of Human Evolutionary Biology, Adam Mickiewicz University in Poznan, Poznań, Poland
Ireneusz Stolarek <istolarek@man.poznan.pl>

Despite the rapid progress made in ancient DNA (aDNA) research the processes that shaped the genetic structure of European population during the Iron Age (IA) remain poorly understood. To learn more on the genetic landscape of the pre-state East Central Europe and document the first migrations in the IA we analyzed the aDNA isolated from 87 individuals from two cemeteries of the IA Wielbark culture located east (Kowalewko - Kow-OVIA) and west (Masłomęcz - Mas-VBIA) of the Vistula River. We found high genetic diversity in both groups, undermining the common belief that they were small isolated populations. Kow-OVIA and Mas-VBIA were closely related and showed small genetic distances to the Jutland Iron Age (JIA) population. Our analyses revealed also sex-biased genetic structure of Kow-OVIA population. Males were closely linked with Mas-VBIA and JIA, whereas women shared most similarities with Early-Middle Neolithic farmers. The same phenomenon was not observed for Mas-VBIA. However, in contrast to Kow-OVIA, Mas-VBIA had close genetic links with the ancient Pontic-Caspian steppe populations. In general, our findings depict the first migratory movements of the Goths and disclose the mechanisms that could underlie the formation of the local genetic substructures in the pre-state South Baltic region during the IA.

P1.2

Beta Defensin 1 gene variability as a risk factor for oral pathology: association analysis between –20 G>A and –44 C>G polymorphisms within the 5'UTR region and recurrent aphthous stomatitis

Zuzanna Celmer[1,3], Zuzanna Ślebioda[2], Tomasz Woźniak[1], Robert Zasadziński[1,3], Marta Daszkowska[1,3], Barbara Dorocka-Bobkowska[2], Maria A. Bobowicz[3], Anna Kowalska[1]

[1]Polish Academy of Sciences, Institute of Human Genetics, Poland; [2]University of Medical Sciences, Department of Oral Mucosa Diseases, Poland; [3]Adam Mickiewicz University, Department of Genetics, Faculty of Biology, Poznań, Poland
Zuzanna Celmer <zu.celmer@gmail.com>

Antimicrobial proteins and peptides (AMPs) play a crucial role in maintenance of oral cavity homeostasis as the important element of innate non-specific immune defense. Human beta defensins are AMPs primarily expressed in oral epithelial cells. The human β-defensin-1 (hBD-1) is a small (29-47 amino acids) cationic peptide encoded by *DEFB1* gene located at chromosome 8p22-23. Several functional polymorphisms at 5'-untranslated region in *DEFB1* gene have been reported to promote alteration of hBD-1 expression increasing a risk for a development of many oral lesions, such as: periodontal diseases, caries or oral cancer. Recurrent aphthous stomatitis (RAS) is a common oral mucosa disorder affecting at least 10 to 20% of the general population and characterized by recurrent ulcerations on mucosal surfaces. Multifactoral complex etiology of RAS with a strong impact of genetic and immunological factors remains still unclear.
The aim of the study was to determine whether two coding SNPs (–20 G>A rs11362 and –44 C>G rs1799946) located in 5'UTR of *DEFB1* gene were related to susceptibility to RAS. The study group consisted of 106 unrelated patients with RAS and 96 healthy control subjects. The genotyping of the two SNPs was performed with the use of PCR-RFLP approach. Genotypes and allele frequencies of the –20 G>A rs11362 and –44 C>G rs1799946 single nucleotide polymorphisms were not statistically different between patients and controls ($p<0.05$). In this study, we found that *DEFB1* gene –20 G>A and –44 C>G SNPs were not associated with RAS. Studies, which analyse other polymorphisms of the *DEFB1* gene in large case series, should be conducted to explain fully the role of the *DEFB1* gene variability in RAS etiology.

P1.3

Do retroposed genes make an important contribution to the human proteome?

Magdalena R. Kubiak, Joanna Ciomborowska-Basheer, Michał W. Szcześniak, Jan G. Kosiński, Izabela Makałowska

Department of Integrative Genomics, Institute of Anthropology, Faculty of Biology, Adam Mickiewicz University in Poznan, Poznań, Poland
Magdalena Regina Kubiak <magdalena.kubiak@amu.edu.pl>

The retroposition of protein coding genes is a significant source of genomic novelties. Here we present our research focused on retroposed genes that contributed to human proteome. Based on previously identified by us retrocopy repertoire stored in the RetrogeneDB2, we have explored protein coding retrogenes in context of their functions and conservation. In total 115 protein coding genes were recognized as retrocopies, including genes such as *GLUD2* or *UBQLN2* linked with the Parkinson disease and amyotrophic lateral sclerosis respectively. Moreover, five of them are resulting from human specific retroposition.
We were also interested in retrocopies that are currently considered as pseudogenes. Hence, we explored mass spectrometry data to identify peptides encoded by them and in a result we pinpointed seven retrocopies encoding unique peptides. Additionally, we analyzed Ribo-seq and ChIP-seq data from previously published experiments to find potentially transcriptionally active and translated retrocopies.
Due to the fact that the number of retrocopies, regardless of their status, was inserted into introns of existing genes, we also looked through known protein coding genes annotated in human genome to determine the retrocopies participating in formation of new transcripts variants and thus alternative proteins. We have found 18 cases, in which a retrocopy is a part of coding DNA sequence. These retrocopies provide alternative start or stop codons, as well as functional domains.

P1.4

Analysis of two novel maize genomes differing in Roundup® resistance

Agata Tyczewska[1], Medhat Helmy[1], Marek Żywicki[2], Joanna Gracz[1], Wojciech Karłowski[2], Tomasz Twardowski[1]

[1]Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznań, Poland; [2]Department of Computational Biology, Institute of Molecular Biology and Biotechnology, Faculty of Biology, Adam Mickiewicz University, Poznań, Poland

Agata Tyczewska <agatat@ibch.poznan.pl>

Glyphosate is an active compound of a systemic, nonselective and most widely used herbicide in the world - Round-Up®. It causes broader range of physiological alterations than previously assumed and some plants gain higher level of resistance without the need to use genetic engineering methods. The holistic understanding of Roundup® mechanism of action is of great importance since it has been shown that glyphosate (an active compound of Round-Up®) affects the growth of plants not only by inhibiting EPSPS (5-enolpyruvylshikimate-3-phosphate synthase) but also through altering several crucial physiological processes (e.g., photosynthesis, carbon metabolism, mineral nutrition, oxidative events).

The 10 chromosomes of the maize genome are structurally diverse and are a result of dynamic changes in chromatin composition (pieces of chromosomes inverted, exchanged, transposed, further duplicated, or lost). The 2.3-billion-base maize genome includes more than 32 000 protein-coding sequences, but sections of DNA called transposable elements, which can move around the genome and cause mutations, are most abundant constituting almost 85% of the genome.

We have analyzed genome structures of two inbred maize lines (S245 and S7957), which substantially differ in resistance to the herbicide Roundup. This has been achieved by using different sequencing technologies (Illumina and SMRT PacBio) to identify the genome structure variations (SV), single nucleotide polymorphism and insertions-deletions (indels), which could be involved in gaining herbicide resistance, without introducing foreign genes into plant system.

Acknowledgements:
The work is supported by a grant no. UMO-2012/06/A/NZ9/00125 from National Science Centre, Poland.

P1.5

Do transcripts of protein-coding genes overlapping at their 5‘ end form RNA:RNA duplexes?

Jarosław K. Sikora[1], Patryk Konieczny[2], Izabela Makałowska[1]

[1]Department of Integrative Genomics, Institute of Anthropology, Faculty of Biology, Adam Mickiewicz University in Poznan, Poznań, Poland; [2]Department of Gene Expression, Institute of Molecular Biology and Biotechnology, Adam Mickiewicz University in Poznan, Poznań, Poland

Jarsosław Krzysztof Sikora <jaroslaw.sikora@amu.edu.pl>

RNA:RNA duplexes is lately of big interest in molecular biology and a considerable number of works demonstrating regulatory potential of coding/non-coding pairs. It was proposed that genes in sense/antisense orientation could regulate each other not only at transcriptional but also post-transcriptional level.

Utilizing TSS-Seq (TSS - transcription start site) data from 73 human tissues and cell lines deposited in the DBTSS database (Yamashita *et al.*, 2010) we identifed pairs of protein-coding genes overlapping at theirs 5’ ends. Analysis revealed 592 gene pairs that overlap in at least one sample (Rosikiewicz & Makałowska, 2016) .

Based on identified gene pairs and RISE database (RNA Interactome from Sequencing Experiments) search we selected most promising candidates for experimental validation. Existing methodologies for RNA:RNA duplexes investigation where modified to meet requirements of this particular study. Pairs of transcripts of protein-coding genes overlapping at their 5‘ end were conducted using HEK923T and HeLa cell lines.