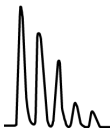

GPF Symposium 2017
Integrative Omics - From Data to Biology

Abstracts

Asaph Aharoni

Unravelling Solanaceae Secondary Metabolism through the Integration of Heterogeneous and Spatial Data from Metabolomics, Genetics and Informatics

Department of Plant & Environmental Sciences, Faculty of Biochemistry, Weizmann Institute of Science, Rehovot, Israel

The regulation of metabolic pathways in plants is constantly tuned in order to suit the needs of development and fitness. Our main research objective is to unravel networks of genes and proteins which coordinate the activity of metabolic pathways, predominantly secondary metabolism, during plant development and stress response. An integrated investigation of several members of the Solanacea family (mainly tomato, potato and eggplant), rather than studying a single plant, provided us with unprecedented insights to metabolic biology in these species. Most if not all processes characterized, impact to a certain degree key quality, nutritional and post-harvest traits of these crop plants. Integrating cutting-edge transcriptomics, proteomics and metabolomics tools together with genes co-expression assays were of great value in making several key discoveries. In a recent example, combined co-expression analysis and metabolic profiling in tomato and potato led to the discovery of the multi-step, core pathway leading to the formation of the renowned Solanum alkaloids including the biosynthesis of their precursor, cholesterol. This class of molecules represents important anti-nutritional compounds in these crop plants. In the presentation, I will highlight several advanced technologies and genetic research tools and the invaluable knowledge on core metabolic traits obtained through combining them in a single study. Most if not all could be applied in the coming years to the study of key traits in other, less studied plant species.

Douglas Armstrong

Network topology analysis reveals substructures within the synaptic proteome that have close association with complex traits

School of Informatics, University of Edinburgh, Scotland

The synaptic proteome is widely believed to be the molecular machine that underpins the core functions of neurons – the integration and transfer of information from one cell to another. Perhaps unsurprisingly it is enriched for proteins whose genes are linked to a wide range of human neurological conditions. However GWAS datasets from these conditions map less clearly onto the synaptic proteome often with weak or no significant enrichment. The synaptic proteome can be subdivided, on the basis of network topology into cluster that each have enriched functional associations. We hypothesised that these topological communities form natural groups for gene set analysis and contain information not only about protein encoding genes with a mechanistic association with the phenotype but also with interaction partners whose role is important but less obvious. We combined 30 published synaptic proteomic studies from 2000 to date to obtain a list of 6500 molecules. We retrieved protein-protein interactions (PPIs) for combined list and built the most complete up-to-date PPI networks for presynaptic and postsynaptic compartments. We then divided this proteome up into sub-communities on the basis of network topology. We analysed three large Genome Wide Associations Studies of genetic associations with human cognitive ability or educational attainment. A single community was found to have a significantly enriched association with educational attainment (n>100,000) and this was then replicated across the other cohorts. The community contained half of the synaptic glutamate receptors. However, the glutamate receptors (as a group) are under strong evolutionary selection pressure and showed no significant association with the phenotype on Genome Wide Association analysis (GWAS). The majority of the signal in the community arose from other proteins in a dense subgraph closely linked to the glutamate receptors.

Tim Beißbarth

Integrating paired proteomics and transcriptomics data using upstream and downstream analysis

Statistische Bioinformatik, Institut für Medizinische Statistik, Universitätsmedizin Göttingen

Integrating time-course information from different data types has emerged as essential element of systems biology approaches to improve our understanding of dynamic cellular responses. Driven by the growing interest in generation of matched high-throughput datasets, we developed a methodology to systematically identify individual signaling axes that are triggered by activated receptors and to link them to their transcriptional response. We implemented this methodology in the R-package *pwOmics*. In this talk we will introduce the methodology and demonstrate it on an example time-resolved phosphoproteome dataset and transcriptome data-set to analyze activated B cell receptor signaling dynamics. We integrated these datasets by a cellular layer-specific pathway-based approach, using public knowledge from biological databases. Thus, integration of matched high-throughput datasets from different cellular layers is a promising approach to broaden our view on complex cellular signaling processes and thereby refine our mechanistic understanding of the cell.

Jürgen Cox

The MaxQuant and Perseus Computational Platforms for Comprehensive Analysis of Large-scale (Prote)Omics Data

Computational Systems Biochemistry, Max-Planck-Institut für Biochemie, Martinsried

Currently, a main bottleneck in proteomics is the downstream biological analysis of highly multivariate quantitative protein abundance data. The Perseus software supports researchers in interpreting protein quantification, interaction and posttranslational modification data. A comprehensive portfolio of statistical tools for high-dimensional omics data analysis covers normalization, pattern recognition, time series analysis, cross-omics comparisons and multiple hypothesis testing. A machine learning module supports classification and validation of patient groups for diagnosis and prognosis, also detecting predictive protein signatures. Perseus combines a powerful arsenal of algorithms with intuitive usability by biomedical domain experts, making it suitable for interdisciplinary analysis of complex large datasets.

Martin Eisenacher

Bioinformatics and Biostatistics for Mass Spectrometry-based Proteomics

Medizinische Bioinformatik, Medizinisches Proteom-Center, Ruhr-University Bochum

The research unit “Medical Bioinformatics” of the Medizinisches Proteom-Center is coordinated by PD Dr. Martin Eisenacher and links the applied proteomics activities of the other research units of the Medizinisches Proteom-Center and theoretic Bioinformatics / Biostatistics. The unit develops new methods, algorithms, analysis workflows and software for statistically valid analysis of proteomics data.
This presentation presents work pieces of the “Medical Bioinformatics” unit which are linked to the Proteomics performed with mass spectrometry.
One work piece – the PIA software tool – deals with protein inference, which is a necessary step in “bottom-up Proteomics” to report the identified proteins explained by the identified peptides. If no unique peptides for proteins have been identified, these proteins cannot be unequivocally reported and are thus reported in so called “protein ambiguity groups”. This methodology is relevant for all-day protein identification and especially for research questions trying to differentiate sequence isoforms. As a by-product, knowledge about the per-group “unique” peptides also has an impact on the number of quantifiable protein groups. It has been increased in example analyses by up to 50 percent.
Further activities of methodology and tool development focus on the areas of Metaproteomics and Biostatistics.
The presented and all our other tools and expertise are available for the Life Science community within the “German Network for Bioinformatics Infrastructure” (www.denbi.de), including the free download and usage of our software tools, but also Bioinformatics consulting or Biostatistics analysis.

Kirstin Feussner

Metabolomics meets transcriptomics to unravel the wound response in plants

Kirstin Feussner[1,2], Alexander Kaever[3], Manuel Landesfeind[3], Alina Mosblech[1], Ingo Heilmann[1] and Ivo Feussner[1,4,5]
[1]University of Göttingen, Albrecht-von-Haller-Institute of Plant Sciences, Department of Plant Biochemistry, Göttingen, Germany;
[2]University of Göttingen, Albrecht-von-Haller-Institute of Plant Sciences, Göttingen Metabolomics and Lipidomics Platform, Göttingen, Germany;
[3]University of Göttingen, Institute of Microbiology and Genetics, Department of Bioinformatics, Göttingen, Germany;
[4]University of Göttingen, Göttingen Center for Molecular BioSciences (GZMB), Department of Plant Biochemistry, Göttingen, Germany;
[5]University of Göttingen, International Center for Advanced Studies on Energy Conversion (ICASEC), Department of Plant Biochemistry, Göttingen, Germany

Plants as sessile organisms are highly exposed to injury by mechanical stress and pathogen attack, like herbivore assault. A complex signalling network, induced by fatty acid-derived compounds, like jasmonates and other oxygenated fatty acid metabolites allows the plant to adapt and to deal with these challenges.
We used a multi-omics approach to deepen our knowledge on processes induced by wounding and to identify so far unknown genes involved in the wounding response of plants. Therefore, we performed a comprehensive metabolome analyses (non-targeted metabolite fingerprinting, lipidomics and quantitative hormone analysis) as well as transcriptomics by employing a whole genome transcriptome array for a wounding time course of wild type and jasmonate-deficient mutant plants. The multivariant data from both omics techniques were used for integrative data analysis in the extended version of the software package MarVis (http://marvis.gobics.de/). MarVis includes tools for data ranking and filtering as well as for clustering and visualization by means of one-dimensional self-organizing maps (1D-SOM). Pattern comparison of gene expression and metabolite profiles combined with gene and metabolite set enrichment analysis allowed to identify pathways, which are jasmonate-dependent as well as independent for the early wound response. In particular, our approach supports the discovery of so far unknown metabolite and transcript markers involved in the wound response of plants.

Oliver Kohlbacher

Going multi-omics – Many issues and (perhaps) a few solutions

Zentrum für Bioinformatik, Eberhard Karls Universität Tübingen und Max-Planck-Institut für Entwicklungsbiologie, Tübingen

While analyzing omics data of a single type is often already challenging enough, more and more we are facing the problem of analyzing large multi-omics data sets in a comprehensive and meaningful matter. Not surprisingly, making sense of more complex data is more complex and comes with many issues attached. Data management, data integration, and smart algorithms for data analysis performing a true integration across omics levels are needed – and the state of the art is often underwhelming. Another big issue that is often overlooked is reproducibility of multi-omics data analysis. Most published multi-omics studies are impossible to reproduce even if the original data is available. We will suggest some generic as well as very concrete software solutions to address these issues.

Bernhard Küster

Chemical proteomics reveals the target landscape of clinical kinase inhibitors

Lehrstuhl für Proteomic und Bioanalytik, Technische Universität München

Kinase inhibitors have developed into important cancer drugs because de-regulated protein kinases are often driving the disease. Efforts in biotech and pharma have resulted in more than 30 such molecules being approved for use in humans and several hundred are undergoing clinical trials. As most kinase inhibitors target the ATP binding pocket, selectivity among the 500 human kinases is a recurring question. Polypharmacology can be beneficial as well as detrimental in clinical practice; hence, knowing the full target profile of a drug is important but rarely available. We have used a chemical proteomics approach termed kinobeads to profile 240 clinical kinase inhibitors in a dose dependent fashion against a total of 320 protein kinases and some 2,000 other kinobead binding proteins. In this presentation, I will outline how this information can be used to identify molecular targets of toxicity, re-purposing existing drugs or combinations for new indications or provide starting points for new drug discovery campaigns.

Alexander Karabatsiakis

Combined metabolite and lipid fingerprinting in women with childhood maltreatment reveals biomarkers linked to inflammation and oxidative stress

Alexander Karabatsiakis[a], Alexandra Maria Koenig[a], Thomas Stoll[b], Sarah Wilker[a], Thomas Hennessy[b,c], Michelle M Hill[b], Iris-Tatjana Kolassa[a]

[a] Clinical & Biological Psychology, Institute of Psychology and Education, Ulm University, Albert-Einstein-Allee 47, 89081 Ulm, Germany

[b] The University of Queensland Diamantina Institute, The University of Queensland, Translational Research Institute, Woolloongabba QLD 4102, Australia

[c] Agilent Technologies, Mulgrave, VIC, Australia

Childhood maltreatment (CM) can increase the risk of adverse physical and mental health consequences in adulthood. However, our understanding of the underlying biomolecular mechanisms is relatively sparse and a deeper insight in associated biological pathways is of high clinical relevance to ideally enable early detection and/or to identify biological targets for intervention. The untargeted investigation of all detectable metabolites and lipids in a biological sample represents a promising new avenue to identify so far unknown biological pathways associated with CM. Using an untargeted approach, liquid chromatography-mass spectrometry (LC-MS) was performed on peripheral blood serum samples collected from 105 healthy adult women with varying degrees of CM exposure. Accurate mass matching with the METLIN Metabolomics Database was used to select putatively identified metabolites and lipids. Comprehensive univariate and multivariate statistical analyses (including group comparisons with multiple comparison corrections, Partial Least Square Discriminant Analysis, and random forests) consistently identified eight putative biomarker candidates belonging to antioxidant-, lipid-, and endocannabinoid-associated pathways, which differentiated between women with and without CM. Classification algorithms allowed for a clear prediction of the CM status with high accuracy scores (~80-90%). In an effort to confirm the identities of these promising biomarker candidates, LC-MS/MS analysis was applied. One of the metabolites was confirmed by LC-MS/MS as bilirubin IXa, a potent antioxidant with immunomodulatory properties. In summary, our results suggest novel pathways that could explain long-term effects of CM on health and disease by influencing biological patterns associated with energy metabolism, inflammation, and oxidative stress.

Philip Stegmaier

Multi-omics “upstream analysis”

geneXplain

We present a computational method for the analysis of multiple omics data sets [1]. Our approach is based on the "upstream analysis" [2] which proposes master regulators as potential research targets together with putative downstream signalling pathways and relevant transcriptional regulators. Upstream analysis begins with analyzing DNA regulatory regions in order to infer transcription factors (TFs) that cause observed expression changes. The next step consists of predicting molecular networks connecting upstream signalling regulators, e.g. transmembrane receptors, with the TFs. The underlying computational methods have been designed to take into account additional experimental data if available. When we applied this approach to a data set comprising proteomic, transcriptomic as well as epigenomic data, we identified several potentially causal targets of cancer cell resistance against methotrexate (MTX) therapy for which we could further infer interacting chemical compounds.

1. Kel, A., Stegmaier, P., Valeev, T., Koschmann, J., Poroikov, V., Kel-Margoulis, O. V., Wingender, E. (2016) Multi-omics “upstream analysis” of regulatory genomic regions helps identifying targets against methotrexate resistance of colon cancer. EuPA Open Proteomics 13:1–13.

2. Koschmann, J., Bhar, A., Stegmaier,P., Kel, A. E. and Wingender, E. (2015) “Upstream Analysis”: An integrated promoter-pathway analysis approach to causal interpretation of microarray data. Microarrays 4, 270-286.

Jörg Dojahn

Accelerating Biological Interpretation through the integration of Transcriptomics and Proteomics Results

Sciex

To build a model of a complex system, an understanding of how all the many different types of components of the system respond to a perturbation is needed. Powerful tools exist in each of the omics fields to quantitatively profile changes at the transcript level and the protein level. Often the expertise to perform these in-depth experiments lies in different labs or institutions, creating challenges for the integration of these separate disciplines. There is an urgent need to build tools and infrastructure that will allow integration of these projects and data streams. By bringing all this rich information together, we will have a truly systems approach to biological understanding that will drive translational medicine. The information gained by combination will be more powerful than that obtained from each omics technology in isolation. To do this, we launched the OneOmics™ Project to be able to handle the combination of these massive datasets through more powerful informatics, more scalable computing power and enhanced usability.

In this talk we present how the OneOmics™ Project works and helps to combine multiomics data in one pipeline and get biological answers faster and in a user friendly way.

Johannes PC Vissers

Peak Detecting, De-Multiplexing, and Searching Multidimensional Data Independent Acquisition Omics LC-MS Data

Waters Corporation, Wilmslow, United Kingdom

Data-independent acquisition (DIA) is an emerging quantitative omics profiling technique and is rapidly gaining in popularity due to its comprehensive and unbiased sampling of precursor ions compared to data-dependent acquisition (DDA). A number of variants have been proposed, which all have non-biased precursor selection, or the lack thereof, in common. Moreover, DIA based approaches are aimed at increasing the detectable dynamic range and coverage of the MS analysis, and improve the precision and accuracy of relative or absolute quantification by systematic sampling of the precursor *m/z* space. Improvements in technology, next to unbiased and complete sampling, will contribute to broader acceptance of the methodology and afford more comprehensive analysis and understanding of complex biological samples and processes.

The principles of current and recently introduced DIA techniques will be contrasted using typical omics application examples, as well as an technology outlook provided by describing the concepts and results of integrating DDA and DIA into a single LC-MS experiment. The challenges associated with single and multi-dimensional peak detection will be highlighted and which approaches could be considered to generate product spectra for either targeted or untargeted searches, how this information is currently stored into open-source, community standards. De-multiplexing routines and approaches in order to generate product ion libraries and searchable spectra are discussed as well, including how these affects quantitative accuracy and precision. Lastly, the principles of a data-independent peptide search algorithm and informatics approaches for the identification of small molecules and lipids are discussed.