IM FOR FUTURE

https://imforfuture.eu

H2020-MSCA-ITN 721815

AGEING

WORKSHOP

FEBRUARY 8-10, 2021

Introduction

It is my pleasure to present the abstracts of the first workshop organized by the *IMforFUTURE* (Innovative Training in Methods for Future Data) network to showcase our work in the past three years. A team of early-stage researchers within the network has put forward a workshop programme on the topics of glycomics, ageing, omics, and their interface. I am pleased to announce that some keynote and invited speakers from within and outside our network will present their research on those topics.

IMforFUTURE is an innovative multidisciplinary and intersectoral research training programme which addresses current shortcomings in omics research. We develop innovative methods for high throughput omics and for integrative analysis of omics data. We focus on ageing, which is the biggest single risk factor for many diseases. By applying our novel methods to emerging datasets representing inflammation and immunology, *IMforFUTURE* will contribute to the understanding of underlying biological processes involved in diseases and ageing.

I would like to thank the scientific committee for selecting poster and contributed talk abstracts, the organizing committee for planning and managing this online workshop and for designing the evening programme. I would specially thank Jessica Brennan for always supporting us and making our work so much easier.

The next workshop will be held on 28 June 2021 and we hope to offer a hybrid version of talks at the school of Mathematics in Leeds, which can be followed online as well. Finally, I hope you will enjoy this workshop and I am looking forward to attending the presentations and discussions, and meeting you online.

Jeanine Houwing-Duistermaat

Workshop Committees

Organising committee:	Jeanine Houwing-Duistermaat (chair), Jessica Brennan, Gastone Castellani, Azra Frkatović, Zhujie Gu, Arief Gusnanto, Tamás Pongrácz, Claudia Sala, Frania Zuñiga Bañuelos
Scientific committee:	Arief Gusnanto, Jeanine Houwing-Duistermaat, Claudia Sala, Jim Wilson, Manfred Wuhrer
Poster committee:	Ivo Ugrina (chair), Maria Giulia Bacalini, Angga Fuady

Index

		Page number
Introduction		2
Workshop committees		2
Workshop programme		4 – 6
Abstracts:		
Maria Giulia	Bacalini	7
Iva	Budimir	8
Jacob	Cancino-Romero	9
Juan	Castillo-Fernandez	10
Andrew	Clegg	11
Lorenzo	Dall'Olio	12
Noortje	de Haan	13
Sonia	Dembowska	14
Azra	Frkatović	15
Angga	Fuady	16
Zhujie	Gu	17
Julija	Jurić	18
Aida	Karray	19
Arianna	Landini	20
Gordan	Lauc	21
Carlito	Lebrilla	22
Annah	Muli	23
Tamás	Pongrácz	24
Md Shafiqur	Rahman	25
Sylvie	Ricard-Blum	26
Eline	Slagboom	27
Samira	Smajlović	28
Maarten	van Schaik	29
Minzhen	Xie	30
Frania	Zúñiga-Bañuelos	31

Workshop programme

Monday 8 February: Glycomics **[Session chairs: Samira Smajlović and Tamás Pongrácz]**	
13.45 - 14.15	Workshop opening
14.00 – 14.15	Introduction, **Jeanine Houwing-Duistermaat**
14.15 - 14.45	**Noortje de Haan** (invited speaker) "High-Throughput Workflows Targeting the Human Glycome in the Search for Markers of Aging and Disease"
14.45 - 15.00	Contributed talk 1 – **Azra Frkatović** "Epigenetic regulation of the human immunoglobulin G N-glycosylation"
15.00 - 15.15	Contributed talk 2 – **Arianna Landini** "Same role but different actors in genetic regulation of transferrin and immunoglobulin G N-glycosylation"
15.15 - 15.30	Contributed talk 3 – **Frania Zúñiga-Bañuelos** "Optimized Workflow for the In-Depth N-Glycoproteomic Analysis of Human Blood Plasma"
15.30 - 15.45	Contributed talk discussion, moderator **Erdmann Rapp**
15.45 - 16.00	Break
16.00 – 16.55	Poster session, with presenters arranged by presentation topics:
Developing molecular tools for studying ageing related diseases	- **Aida Karray** "Hydrolysis of three different head groups phospholipids by chicken group V phospholipase A2 using the monomolecular film technique" - **Samira Smajlović** "Modulation of the HNF1A and FOXA2 genes, using CRISPR/dCas9 molecular tools, in studies of their role in glucose stimulated insulin secretion"
Statistical models for ageing	- **Annah Muli** "Use of semiparametric frailty model in analysis of survival data in twins" - **Jacob Cancino-Romero** "Prediction accuracy and variable selection for joint models of longitudinal and time-to-event data"
Statistical methods for complex omics datasets (longitudinal)	- **Maarten van Schaik** "Modelling high dimensional count data using random effects" - **Sonia Dembowska** "Multivariate Functional Partial Least Squares models for longitudinal OMICS datasets" - **Minzhen Xie** "Non-parametric clustering of longitudinal functional data with application to NMR spectra of 18 kidney transplant patients"
Glycosylation involved in inflammation	- **Tamás Pongrácz** "Immunoglobulin G1 Fc glycosylation as a potential early inflammation marker in COVID-19" - **Md Shafiqur Rahman** "Bisecting N-acetylglucosamine (GlcNAc) enriched IgG N-glycans are associated with chronic widespread musculoskeletal pain"
16.55 - 17.00	Group photograph, **Azra Frkatović**
17.00 - 17.45	**Carlito Lebrilla** (keynote speaker) "Towards linking diet, ageing, and Alzheimer's disease through glycosylation"
17.45 - 18.00	Closing

Tuesday 9 February: Ageing and Glycomics **[Session chairs: Frania Zúñiga-Bañuelos and Azra Frkatović]**	
13.45 - 14.00	Welcome
14.00 - 14.45	**Gordan Lauc** (keynote speaker) "Protein glycosylation as a biomarker and functional effector of ageing and age-related diseases"
14.45 - 15.00	Contributed talk 4 – **Julija Jurić** "Effects of estradiol on biological age measured using the glycan age index"
15.00 - 15.15	Contributed talk 5 – **Angga Fuady** "Association analysis of menopausal status and biological age measured using GlycanAge index from different platforms"
15.15 - 15.30	Contributed talk 6 – **Juan Castillo-Fernandez** "Increased transcriptome variation and localised DNA methylation changes in oocytes from aged mice revealed by parallel single-cell analysis"
15.30 - 15.45	Contributed talk discussion, moderator **Frédérique Lisacek**
15.45 - 16.00	Break
16.00 - 16.45	**Sylvie Ricard-Blum** (keynote speaker) "Extracellular matrix aging: a focus on the role of the lysyl oxidase family"
16.45 - 17.00	Contributed talk 7 – **Lorenzo Dall'Olio** "Prediction of vascular aging based on smartphone acquired PPG signals"
17.00 - 17.15	Contributed talk 8 – **Iva Budimir** "Characterization of DNA methylation correlation structure in Down Syndrome"
17.15 - 17.30	Contributed talk 9 – **Zhujie Gu** "Investigating accelerated aging in Down syndrome by integrating methylation and glycomics"
17.30 - 17.45	Contributed talk discussion – moderator **Hae-Won Uh**
17.45 - 18.30	Break
18.30 - 19.15	**Yoga session, Ana Rangel** at the Sattva Institute Ana will guide you through this Yoga session and help you to relax your mind and body after a long day of presentations. To join this session you need a Yoga mat, and a little bit of space. If a yoga mat isn't available there is also the option of practicing on a blanket, carpet or any other anti-slippery surface. Participants can practice with their partners, colleagues and children. Pregnant participants should check with their doctors if they haven't been exercising or if they suffer from any particular condition. Ana will recommend alternative postures during the class.
19.15 - 19.30	Break
19.30 - 20.30	**Anthony Newman**, Elsevier Author workshop on writing journal articles

Wednesday 10 February: Ageing and Omics **[Session chairs: Arianna Landini and Zhujie Gu]**	
13.45 - 14.00	Welcome
14.00 - 14.45	**Claudio Franceschi** (keynote speaker) "INFLAMMAGING: centenarians and the complex basis of human longevity"
14.45 - 15.15	**Maria Giulia Bacalini** (invited speaker) "Epigenetic signature of Down Syndrome: atypical aging and beyond"
15.15 - 15.30	Break
15.30 - 16.00	**Andrew Clegg** (invited speaker) "The electronic Frailty Index (eFI): development, national implementation and next steps"
16.00 - 16.45	**Eline Slagboom** (keynote speaker) "Metabolomics markers of biological age, mortality and exceptional longevity"
16.45 - 17.00	Break
17.00 - 18.30	Panel discussion
18.30 – 19.00	Summary of posters and announcement of poster prize

Abstracts

Maria Giulia Bacalini
IRCCS Istituto delle Scienze Neurologiche di Bologna

Invited speaker

"Epigenetic signature of Down Syndrome: atypical aging and beyond"

Down Syndrome (DS) is caused by trisomy of chromosome 21 and is characterized by a wide spectrum of multi-systemic manifestations, including physical abnormalities, congenital heart disease and intellectual disabilities. DS is considered a segmental progeroid disease, in which premature/accelerated aging particularly affects the immune and the nervous system. Persons with DS tend to exhibit an accelerated cognitive decline after 40 years and are at higher to develop Alzheimer's disease (AD). However, a wide heterogeneity exists among persons with DS, and while some of them show signs of AD relative early, others avoid or postpone the disease.

Intensive research efforts are being made to unravel the molecular basis at the basis of DS phenotype. Epigenetic mechanisms are likely to play an important role, as large DNA methylation alterations occur in DS. Furthermore, persons with DS show signs of premature/accelerated aging according to the epigenetic clock. Finally, the study of DNA methylation could also provide some hints on the heterogeneity among persons with DS and provide useful markers to identify individuals at higher risk to develop AD.

Iva Budimir
University of Bologna

Contributed talk

"Characterization of DNA methylation correlation structure in Down Syndrome"

Cytosine methylation in the humane genome is an important and well-studied epigenetic mark which has the potential to regulate gene expression. Since the process of methylation predominantly occurs at CG dinucleotide sequences, the target of studies are these so-called CpG sites. One of the available tools, Infinium 450K assay measures the level of methylation of ~480,000 CpGs spread across the genome.

Studies of methylation patterns suggest the bimodal behaviour where different clusters of CpGs tend to be either hypermethylated or hypomethylated, rarely existing in intermediate states. This group behaviour suggests that CpGs are not independent, but rather that the methylation profile is guided by the complicated network structure of CpGs. This study aimed to reconstruct and characterize the correlation network of DNA methylation in blood in the context of Down syndrome. To reconstruct the network, we considered a publicly available DNA methylation data set of 728 healthy control blood samples. For computational reasons, we focused on a smaller subset of CpGs. Specifically, we studied only ~4000 CpGs located on chromosome 21. We demonstrated that partial correlation approach is not an appropriate choice in the presence of highly correlated groups of CpGs. Thus we estimated the relationships between CpGs with the Pearson correlation matrix calculated between the residuals obtained after age, gender and cell count correction. Lastly, we compared the control network obtained on the 728 control samples with the networks obtained from the Down syndrome data set consisting of methylation measurements for Down syndrome patients (29), their unaffected siblings (29) and their mothers (29).

We observed that the CpGs work in small highly correlated groups which can be viewed as unique entities in the methylation network. Interestingly, the nodes of this network, that is, the groups of highly correlated CpGs were often scattered across the chromosome. We identified the problem of the small sample size of Down syndrome data set which resulted in the instability of correlations. Thus only the small differences between Down syndrome data sets were observed. Specifically, our results suggest that in Down syndrome patients, with respect to controls, new correlations between CpGs were created.

Jacob Cancino-Romero
University of Leeds

Poster

"Prediction accuracy and variable selection for joint models of longitudinal and time-to-event data"

Frailty is dynamical process of a reduction in the physical, psychological and social function associated with aging. It predicts mortality and hospital admissions in the elderly. We are interested in building a model that jointly predicts frailty and mortality using data from the Community Ageing Research 75+ (CARE75+) study. Joint modelling frailty and mortality with the CARE75+ data is challenging because it involves specifying a model for each outcome with several highly correlated risk factors and covariates in a relatively small data set. Variable selection for a joint model of longitudinal and time-to-event outcomes has been explored before with the aim of maximizing goodness-of-fit. We propose a variable selection strategy to optimize prediction of both outcomes. This strategy combines penalized likelihood with the LASSO penalty and cross-validation methods to select the fixed effects that optimize simultaneously the mean-squared error (MSE) and the Integrated Brier Score (IBS). Our simulation studies suggest that it is not always possible simultaneous optimization of MSE and IBS, but there seems to be a region defined by the constraints close to an optimal solution. In such a case a small compromise between MSE and IBS is required, depending on which outcome is the priority. As secondary criterion of performance, we assessed our strategy by its accuracy in variable selection with respect to the true model. The final joint model that optimizes prediction of frailty and mortality for the CARE75+ data should include sex, ethnicity, marital status, education, smoking, alcohol consumption, number of falls, and time (since recruitment) as covariate in the frailty submodel and no covariates in the mortality submodel.

Juan Castillo-Fernandez
Babraham Institute, Cambridge

Contributed talk

"Increased transcriptome variation and localised DNA methylation changes in oocytes from aged mice revealed by parallel single-cell analysis"

Advancing maternal age causes a progressive reduction in fertility. The decline in developmental competence of the oocyte with age is likely to be a consequence of multiple contributory factors. Loss of epigenetic quality of the oocyte could impair early developmental events or programme adverse outcomes in offspring that manifest only later in life. Here, we undertake joint profiling of the transcriptome and DNA methylome of individual oocytes from reproductively young and old mice. We find reduced complexity as well as increased variance in the transcriptome of oocytes from aged females. This transcriptome heterogeneity is reflected in the identification of discrete sub-populations. Oocytes with a transcriptome characteristic of immature chromatin configuration (NSN) clustered into two groups: one with reduced developmental competence, as indicated by lower expression of maternal-effect genes; and one with a young-like transcriptome. Oocytes from older females had on average reduced CpG methylation, but the characteristic bimodal methylation landscape of the oocyte was preserved. Germline differentially methylated regions of imprinted genes were appropriately methylated irrespective of age. For the majority of differentially expressed transcripts, the absence of correlated methylation changes suggests a post-transcriptional basis for most age-related effects on the transcriptome. However, we did find differences in gene-body methylation at which there were corresponding changes in gene expression, indicating age-related effects on transcription that translate into methylation differences. Interestingly, oocytes varied in expression and methylation of these genes, which could contribute to variable competence of oocytes or penetrance of maternal age-related phenotypes in offspring.

Andrew Clegg
University of Leeds
Health Data Research UK North (HDRUK North)

Invited speaker

"The electronic Frailty Index (eFI): development, national implementation and next steps"

Andy Clegg is Professor of Geriatric Medicine at the University of Leeds and Associate Director for Health Data Research UK North (HDRUK North). In this talk Andy will outline the development and national implementation of the eFI, enabling the routine identification of frailty using existing primary care electronic health record data, with major impact on health policy. Andy will also outline next steps, including the development of the next iteration of the eFI.

Lorenzo Dall'Olio
University of Bologna

Contributed talk

"Prediction of vascular aging based on smartphone acquired PPG signals"

Lorenzo Dall'Olio[1], Nico Curti[2], Daniel Remondini[1], Yosef Safi Harb[3], Folkert W. Asselbergs[4,5,6], Gastone Castellani[2,+], and Hae-Won Uh[7,+,*]

[1]University of Bologna, Department of Physics and Astronomy, Bologna, 40126 Bologna BO, Italy

[2]University of Bologna, Department of Specialised, Diagnostic and Experimental Medicine, Bologna, 40126 Bologna BO, Italy

[3]Happitech, Amsterdam, 1017 CB, the Netherlands

[4]University Medical Center Utrecht, Department of Cardiology, Utrecht, 3584 CX, Netherlands

[5]University College London, Institute of Cardiovascular Science, Faculty of Population Health Sciences, London, WC1E 6BT, United Kingdom

[6]University College London, Health Data Research UK and Institute of Health Informatics, London, NW1 2BE, United Kingdom

[7]University Medical Center Utrecht, Department of Biostatistics and Research Support, Utrecht, 3584 CX, the Netherlands

[*]h.w.uh@umcutrecht.nl

[+]these authors contributed equally to this work

Biological ageing is a different and more complex phenomenon with respect to chronological ageing; it can vary among organs and in a non-uniform way over time. We aimed to investigate its effects on the vascular system, using photoplethysmography (PPG) and a database of thousands of smartphone measured signals crowd-sourced from the Heart for Heart initiative. In particular, we explored the feasibility of using PPG to predict healthy vascular ageing (HVA) based on two approaches: machine learning (ML) and deep learning (DL). For ML, we performed data preprocessing of the raw PPG signals, including detrending, demodulating and denoising, and used ridge penalized regression methods to perform feature selection. Alternatively, for DL, several convolutional neural networks (CNNs) were applied to the whole PPG signals as input. The prediction performance of ML using two features (AUC of 94.7%) - the ***a*** wave of the second derivative PPG and ***tpr*** (turning point ratio), including four covariates, sex, height, weight, and smoking - was similar to that of the best performing CNN, a 12-layer ResNet (AUC of 95.3%). Therefore, the usage of ML could be of considerable importance in highlighting possible biomarkers for HVA. Moreover, the ***a*** wave appeared to decrease with age, indicating a lower elasticity (higher stiffness) of blood vessels walls. In contrast, ***tpr*** was observed to increase with age, displaying higher randomness in the inter-beat intervals, possibly due to the loss of some biological ageing mechanisms.

Noortje de Haan
University of Copenhagen

Invited speaker

"High-Throughput Workflows Targeting the Human Glycome in the Search for Markers of Aging and Disease"

Healthy aging is one of the big healthcare challenges in the modern society and markers to diagnose and predict diseases in an early stage are warranted. A growing body of evidence suggests that the human glycome, in addition to the human genome and proteome, can play an important role in the search for markers that can support healthy aging. To exploit the human glycome for this purpose, high-throughput methods are needed to study glycan changes in large clinical cohorts. Here, various approaches fit for this purpose will be discussed and examples will be given of promising glycomic markers related to inflammatory bowel diseases and colorectal cancer.

Sonia Dembowska
University of Leeds

Poster

"Multivariate Functional Partial Least Squares model for longitudinal OMICS datasets"

The use of statistical methods to forecast patient outcomes using high dimensional datasets in medicine is becoming increasingly popular. The availability of longitudinal datasets allows for forecasting and monitoring patient health over time. Our work is motivated by a longitudinal dataset containing H1 NMR spectra and outcomes of 18 patients undergoing a kidney transplant. The data is functional in two dimensions, time and the ppm spectra themselves. The patients were separated into three groups determined by renal graft outcomes: acute rejection, delayed graft function and primary function. Current functional partial least squares (FPLS) methods account for univariate scenarios (one time point), which for this dataset limits the amount of information used to forecast patient outcomes.

We propose a functional model that can account for multiple time points which means it is multivariate. Our model is based on the work of Preda and Saporta (2007) on functional data analysis and the work of Said El. Bouhaddani (2020), which uses PLS methods for the analysis of longitudinally measured non-functional omics datasets. The combined models are then applied to the dataset and their performance is compared using ROC curves to the already existing univariate method.

Data analysis revealed that when considering the average performance of the multivariate and univariate FPLS, multivariate models had better results. Indeed, the average AUC's over different outcome groups were 0.957 and 0.878 respectively. Hence it was suggested that for clinical purposes, multivariate FPLS provides additional insight. It is recommended that future work build on the approach used in this study by applying similar methods, but to a larger patient cohort to determine if a wide use of similar models is applicable to the wider population.

Azra Frkatović
Genos

Contributed talk

"Epigenetic regulation of the human immunoglobulin G N-glycosylation"

Frkatović A[1], Vučković F[1], Wilson JF[2,3], Lauc G[1,4], Klarić L[2]

[1]Genos Glycoscience Laboratory, Zagreb, Croatia
[2]MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK
[3]Usher Institute, University of Edinburgh, Edinburgh, UK
[4]Faculty of Pharmacy and Biochemistry, University of Zagreb, Zagreb, Croatia"

DNA methylation is a well-studied epigenetic mechanism that regulates gene expression via methylation of cytosine residues in CpG islands.
We performed an epigenome-wide association study (EWAS) in ORCADES cohort (n=967) to explore the effects of differential methylation on the composition of immunoglobulin (IgG) N-glycome. We derived glycan traits representing the percentage of the structures containing specific sugar unit, for example fucose or galactose. Methylation levels of CpG sites across the genome were measured with Infinium MethylationEPIC Array, followed by quality control and correction for technical and biological variation using linear mixed modelling. Linear regression was performed for each glycan trait as outcome and CpG methylation value as independent variable. The basic model included age and sex to account for confounding but still allowing the assessment of potential mediation effects of other covariates. The adjusted model further included smoking, BMI and age and sex interaction terms.
The basic model resulted in three significantly associated CpG sites at epigenome-wide significance level ($p<1x10-8$): cg05575921 (associated with bisection), cg06072257 (galactosylation and monosialylation), and cg08946781 (galactosylation and monosialylation). The adjusted model resulted in a significant association between sialylation trait and cg08178031, methylation site in the 5' UTR of sialyltransferase gene, ST6GAL1. The reported CpG sites in EWAS of IgG N-glycome by Wahl et al. were replicated, noting that previously tested glycan traits differed from the current study. Differential methylation of cg06072257 was previously associated with sex-specific age so we further performed sex-stratified association test and showed that the association of galactosylation and cg06072257 is sex-specific (female-only). It is important to note that this association is only suggestive ($p=3.17x10-7$), likely due to lower power for smaller sample size in the sex-stratified analysis. However, there is an implication that the methylation of the associated CpG sites might be sex-specific, thereby affecting the composition of IgG N-glycome in sex-specific manner.
Our initial results suggest the presence of a link between methylation of additional CpG sites and IgG glycosylation. Further steps in the study include meta-analysis and replication of the findings, as well as investigation of potential functional effects of the discovered CpG sites on the IgG N-glycome composition.

Angga Fuady
Leiden University Medical Center, The Netherlands

Contributed talk

"Association analysis of menopausal status and biological age measured using GlycanAge index from different platforms"

GlycanAge index is a recently developed biomarker that addresses both the biological and chronological age of individuals. It is based on three IgG UPLC glycans namely, one nongalactosylated (GP6) and two digalactosylated glycans (GP14 and GP15). Here, we are interested in the relation between biological age, represented by the GlycanAge index, and menopause status. However, different measurement techniques are used and measurements in different studies can come from different platforms. For instance, in our motivating example, glycomics data were generated using two technologies, namely LCMS and UPLC, in cohorts from Vis and Korcula. Specifically, Vis (n=349) has both LCMS (p=50) and UPLC (p=23), while Korcula (n=532) has only LCMS. To assess the relation between GlycanAge and self-reported menopause status in the Korcula cohort, we need UPLC data to compute GlycanAge. The self-reported menopause status for each cohort is obtained from a questionnaire whether the respondent consecutively stopped having periods for one year or more. The number of menopausal women is 342 and 242, for Korcula and Vis, respectively.

We propose to use the LCMS data to obtain the relevant UPLC glycans for the Korcula cohort. We introduce a measurement error model where biological mapping information is used. Vis cohort is used as a calibration study to relate GlycanAge index and LCMS IgG glycans. Logistic regression is used in the main model. The regression coefficient from the calibration study is then used to adjust the estimates in the main model.

We found that there is an association between the estimated GlycanAge index and the self-reported menopause status in the Korcula cohort. Overall, we conclude that our method provides a parameter by combining estimates of multiple surrogate variables (IgG LCMS glycans) using inverse variance weighting which minimizes the variance of the summary estimator.

Zhujie Gu
UMC Utrecht

Contributed talk

"Investigating accelerated aging in Down syndrome by integrating methylation and glycomics"

Author: Zhujie Gu, Said el Bouhaddani, Jeanine Houwing-Duistermaat, Hae-Won Uh

Down syndrome (DS) is a disease that leads to accelerated aging in affected subjects. They develop conditions that are typically observed at higher age. Studies at the molecular level of DS have reported several alterations in methylation and glycomics. In this study, we investigate the aging process in DS patients by jointly analyzing methylation and glycomics, to identify CpG sites and glycans related to aging. Our motivating datasets consist of methylation and glycomics measured on 23 DS patients and their healthy siblings and mothers.

For dimension reduction of high-dimensional correlated omics data, we consider Two-way Orthogonal Partial Least Squares (O2PLS), which constructs a few joint latent variables that explain the covariance between two omics datasets. We implement a two-step approach where we first jointly analyze methylation and glycomics using O2PLS, construct a few joint latent variables, and then link these latent variables to DS via logistic regression. Relevant CpG sites and glycans from the latent variables are selected for interpretation. We then develop a probabilistic framework, supervised probabilistic O2PLS that combines dimension reduction and outcome regression in one model. The parameters in the model are estimated, taking into account the omics data and the outcome variable simultaneously. Results of the DS data analysis using the two-step approach will be shown and interpreted. A simulation study to evaluate the performance of supervised probabilistic O2PLS will be presented.

To conclude, we study accelerated aging in the form of DS by considering methylation and glycomics data together. Our proposed method that jointly analyzes multiple omics data with outcome variable may provide new insight into the underlying mechanism of accelerated aging at different omics levels.

Julija Jurić
Genos Ltd

Contributed talk

"Effects of estradiol on biological age measured using the glycan age index"

Julija Jurić[1], Wendy M. Kohrt[2,3], Domagoj Kifer[4], Kathleen M Gavin[2,3], Marija Pezer[1], Peter A. Nigrovic[5,6], Gordan Lauc[1,4]

[1]Genos Glycoscience Research Laboratory, Zagreb, Croatia
[2]Division of Geriatric Medicine, School of Medicine, University of Colorado Anschutz Medical Campus, Aurora, CO 80045, USA
[3]Eastern Colorado VA Geriatric Research, Education and Clinical Center, Aurora, CO 80445, USA
[4]Department of Biochemistry and Molecular Biology, Faculty of Pharmacy and Biochemistry, University of Zagreb, Zagreb, Croatia
[5]Division of Rheumatology, Inflammation, and Immunity, Brigham and Women´s Hospital, Boston, MA 02115, USA
[6]Division of Immunology, Boston Children´s Hospital, Boston, MA 02115, USA

Glycan age is a recently developed biomarker based on glycans attached to immunoglobulin G (IgG). In large population cohorts, glycan age associates well with lifestyle and disease-risk biomarkers, while some studies suggested that glycan changes precede the development of several age-associated diseases. In this study, we evaluated the effects of estrogen on the glycan age. Gonadal hormones were suppressed in 36 healthy young women by gonadotropin-releasing hormone agonist therapy for 6 months. In 15 of them, estradiol was supplemented, while 21 received placebo resulting in very low estrogen levels during the intervention. IgG was isolated from plasma samples before the intervention, after 6 months of intervention, and after subsequent 4-month recovery. At each time point, we performed IgG N-glycoprofiling by hydrophilic interaction ultra-performance liquid chromatography (HILIC-UPLC) and the glycan age index was calculated. Deprivation of gonadal hormones resulted in a median increase of glycan age for 9.1 years which was completely prevented by transdermal estradiol therapy. After the recovery period, the glycan age returned to baseline values in both groups. These results suggest that IgG glycans and consequently also the glycan age are under a strong influence of gonadal hormones and that estradiol therapy can prevent the increase of glycan age that occurs in the perimenopausal period.

Aida Karray
Ecole Nationale d'Ingénieurs de Sfax

Poster

"Hydrolysis of three different head groups phospholipids by chicken group V phospholipase A2 using the monomolecular film technique"

Karray Aida, Madiha Bou Ali, Jallouli Raida and Bezzine Sofiane
Laboratoire de Biochimie et de Genie Enzymatique des Lipases, Ecole Nationale d'Ingénieurs de Sfax, route de Soukra, Sfax 3038, Université de Sfax, Tunisia

Age-related changes at the cellular level include the dysregulation of metabolic and signaling pathways. . Furthermore, the downregulation of genes, such as PLA2, which belong to the arachidonic acid metabolism pathway involved in phosphatidylcholine conversion to anti-inflammatory lipoxins, correlated with increased phosphatidylcholine content in monocytes from older individuals.

The kinetic aspects of lipolysis by pulmonary phospholipase A2 (ChPLA2-V), intestinal phospholipase A2 (ChPLA2-IIA) and pancreatic phospholipase A2 (ChPLA2-IB), from have been compared using the monomolecular films technique, on short-chain phospholipids (with three different head groups) and on long-chain phospholipids. The main conclusions from our experimental data indicate that the maximum catalytic activities of ChPLA2-V on 1,2 phosphatidylcholine and 1,2 phosphatidylethanolamine reached 15.26 and 36.12 moles/cm2.min.mM, respectively, at a pressure of 15 and 35 dynes/cm, respectively. Whereas, those of ChPLA2-IB were 3.58 (at the pressure of 20 dynes/cm) and 4.9 moles/cm2.min.mM. However, hydrolysis of phosphatidylglycerol monolayers (C12PG), were very much higher compared with all the substrates tested with 122 moles/cm2.min. Surprisingly, the hydrolysis rate of ChPLA2-V on long-chain phosphatidylglycerol (C18PG) was very low (1.45 moles/cm2.min) compared with all tested substrates, even with the use of p-cyclodextrin. And thus, the fatty acid preference of ChPLA2-V was 2-decanoyl > 2-oleoyl with a PG head group. In order to gain significant correlations between enzyme's structures and their relative functions, we tried to examine the surface electrostatic potentials of the various secreted phospholipase 2 (sPLA2) from chicken. In the present study, we detailed that the substrate affinity, specificity and the hydrolysis rates of sPLA2 at each interface is governed by the surface electrostatic potentials and hydrophobic interactions operative at this surface.

Arianna Landini
University of Edinburgh

Contributed talk

"Same role but different actors in genetic regulation of transferrin and immunoglobulin G N-glycosylation"

Arianna Landini[1], Irena Trbojevic-Akmacic[2], Pau Navarro[3], Frano Vuckovic[2], Caroline Hayward[3], Tea Petrovic[2], Marija Vilaj[2], James F Wilson[1,3], Gordan Lauc[2,4], Lucija Klaric[2,3]

[1]Centre for Global Health Research, Usher Institute, University of Edinburgh, Edinburgh, United Kingdom
[2]Genos Glycoscience Research Laboratory, Zagreb, Croatia
[3]MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, United Kingdom
[4]Faculty of Pharmacy and Biochemistry, University of Zagreb, Zagreb, Croatia

Post-translational modifications (PTMs) are essential mechanisms used by cells to diversify their protein functions and dynamically coordinate their signaling networks. While PTMs are known to be involved in regulating almost all cellular events, genetic regulation of PTMs themselves has not been extensively investigated. Protein N-glycosylation, acknowledged as one of the major PTMs, has been linked to the ageing process and a wide variety of diseases (including Parkinson's disease, rheumatoid arthritis, Crohn's disease, type 2 diabetes and cancer). Nevertheless, genetic regulation of N-glycosylation is yet not fully understood. To explore whether genes regulating N-glycosylation are protein-specific or rather shared among different proteins, we performed genome-wide association meta-analysis of 35 transferrin N-glycan traits (N=1890) and 24 immunoglobulin G (IgG) N-glycan traits (N=2020) in the European-heritage CROATIA-Korcula and VIKING cohorts. We identified 10 loci significantly associated ($P < 1.43 \times 10^{-9}$) with transferrin N-glycosylation. Three of these genomic regions were never previously associated with the N-glycome of any proteins, while six of them, discovered in CROATIA-Korcula cohort, replicated in VIKING cohort. Using colocalisation methods at genomic regions associated with N-glycosylation of both transferrin and IgG proteins (FUT8 and FUT6), we found strong support for the presence of multiple causal genetic variants, each exhibiting an effect on only one of the two proteins. Uncovering for the first time genes responsible for transferrin N-glycosylation, our results contribute to expanding the current knowledge of the genetic regulation of glycosylation PTMs. Most importantly, they also show that, while the same enzymes are involved in N-glycosylation of both transferrin and IgG, multiple causal variants independently affect the PTM of each protein.

Gordan Lauc
Genos

Keynote speaker

"Protein glycosylation as a biomarker and functional effector of ageing and age-related diseases"

Gordan Lauc
Director of the Human Glycome Project
Director of the National Centre of Excellence in Personalized Healthcare, Zagreb, Croatia
glauc@pharma.hr

The majority of proteins that evolved the after appearance of multicellular life are glycosylated and glycans significantly affect structure and function of these proteins. However, due to structural complexity of glycans and the absence of a direct genetic template, the analysis of protein glycosylation is much more complicated than the analysis of DNA or proteins. Consequently, the knowledge about the importance of individual variation in glycans for both normal physiological processes and diseases is still limited. In the last few years it is becoming increasingly clear that variations in a DNA sequence are only a beginning of the understanding of complex human diseases. Genetic polymorphisms have to be put in the context of complex biology of life and a more elaborate approach that combines different 'omics phenotypes is needed to understand disease mechanisms and perform patient stratification that transcends genomics. Glycomics, as by far the most complex posttranslational modification, has an immense potential in this respect, which is only beginning to be investigated.

By generating glycomic data for over 150,000 individuals from some of the best characterized clinical and epidemiological cohorts we enabled glycomics to meet other 'omics. The analysis of this rich gold mind is painting a picture of a very complex genetic and epigenetic regulation of glycosylation that fine tunes protein activity in multiple biological systems and also contributes to ageing at the molecular level. In particular, the evidence is accumulating that in cardiometabolic diseases changes in glycosylation are not only biomarkers, but functional effectors that change up to a decade before initial disease diagnosis and in the meantime they actively participate in disease development. Since glycans are under significant environmental influence, lifestyle and pharmacological interventions can be used to preventively correct these disease-causing age-related changes and in this way delay, or even completely avoid disease onset.

Carlito B. Lebrilla
University of California, Davis

Keynote speaker

"Towards linking diet, aging, and Alzheimer's disease through glycosylation"

Protein glycosylation is a posttranslational modification that can guide and alter protein function. While it is a metabolic product, glycosylation is found in 50-80% of all proteins and can be altered even when protein expression is not. As such, glycosylation can be a stronger indication of the physical condition that protein expression. In this presentation, advanced mass spectrometry-based methods for producing glycomic profiles and glycoproteomic analysis are developed and used to understand the role of glycosylation in infant nutrition as well as disease including cancer and Alzheimer. Glycomic profiling has unlocked the role of human milk oligosaccharides in developing the gut microbiota. Glycomic and glycoproteomic analysis are providing new classes of biomarkers that are closer and more sensitive to the diseases. Glycosylation and methods that analyze them are providing new insight into diet and nutrition and disease progression.

Annah Muli
University of Leeds

Poster

"Use of semiparametric frailty model in analysis of survival data in twins

Family based studies help to investigate traits that segregate within families e.g., human longevity which is known to cluster within families. This is attributed to the fact that there is correlation between family members (e.g., twin pairs) as they share genetic and environmental factors. Analysis of such data is challenging due to censoring and correlation among the survival times. The shared frailty model is commonly used for analysis of such correlated right-censored survival data. The frailty represents unobservable effects which are shared between twins and influence survival. A parametric frailty distribution is assumed, for example the gamma distribution which is computationally convenient. Via simulation we have shown that if the frailty distribution is not correctly specified the estimates of the regression coefficient and survival probabilities may be biased.

We shall consider a nonparametric specification of the baseline hazard by making use of splines. Advantages of using splines for baseline hazard include flexibility of the baseline hazard and smooth survival curves. Full maximum likelihood is then used for estimation of parameters. Simulations showed that replacing the parametric baseline hazard by a flexible baseline hazard can adjust for the incorrect frailty distribution and may improve the estimators of the population survival probability.

We therefore propose to use a semiparametric frailty model to estimate individual specific probabilities of fracture in the next time period given covariates using the TwinsUK data.

Tamás Pongrácz
Leiden University Medical Center

Poster

"Immunoglobulin G1 Fc glycosylation as a potential early inflammation marker in COVID-19"

Immunoglobulin G (IgG) plays a major role in humoral immune events. IgGs are N-glycosylated in their Fc portion and part of the induced effector functions are highly influenced by the presence and the structure of the CH2 N-glycans. The presence or absence of individual monosaccharides on these N-glycans subtly modulates the binding activity of the Fc portions to their interaction partners, for example the FcγRIIIa/b receptor. Importantly, one particular structural difference, which is the absence of a core fucose, is known to increase the affinity of IgG1 for the FcγRIIIa/b receptors by 20-100 fold, thereby largely increasing downstream processes such as antibody-dependent cellular cytotoxicity (ADCC) or phagocytosis of bacteria and viruses by immune cells that express the corresponding receptors including NK cells, neutrophils and macrophages.

Antigen-specific IgG fucosylation can potentially be decreased as compared to total IgG, hence leading to enhanced immune activation and clinical inflammatory manifestation. IgG Fc glycosylation changes of antigen-specific IgG have been described to correlate with disease outcome.

In this study, we analyzed anti-Spike protein (anti-S) IgG affinity captured from plasma of patients infected with severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). This approach builds on a recently published study showing low fucosylation on anti-S protein IgG1 in part of the COVID-19 patients, with parallel pro-inflammatory signatures.

Our recently obtained data in the BEAT-COVID study performed at Leiden University Medical Center appear to be largely in line with these findings, and imply skewing of anti-S IgG1 fucosylation in SARS-CoV-2 infection with functional implications. We show that a proinflammatory anti-S IgG1 fucosylation signature characterizes the onset of COVID-19, followed by a consistent increase in fucosylation with time, and may therefore be an early marker and inducer of inflammatory reactions promoting disease severity. Additionally, negative correlations were found between fucosylation and the concentration of numerous inflammatory cytokines and acute phase proteins. We believe that our findings deserve further evaluation at a biomarker and mechanistic level, thus currently additional samples are analyzed to increase the relevance thereof.

Md Shafiqur Rahman
King's College London

Poster

"Bisecting N-acetylglucosamine (GlcNAc) enriched IgG N-glycans are linked with Chronic Widespread Pain"

Chronic widespread pain (CWP) is a highly prevalent condition and a diagnostic symptom for fibromyalgia. CWP is heritable. Autoimmune pain causing conditions such as systemic lupus erythematosus have been implicated to immunoglobulin-G (IgG) N-glycosylation. Low back pain, a non-inflammatory condition, has also been associated with N-glycosylation. Transferring IgG from fibromyalgia patients to mouse has shown increased pain sensitivity compared to control IgG (Goebel et al., 2019) and fibromyalgia patients treated with IgG have short-time improvement in pain, tenderness and strength (Caro, Winter, & Dumas, 2008). Taken together, the relevance of immune-system to CWP is clear.

We investigated change in 76 IgG N-glycans (22 directly measured and 54 derived traits) level in 3226 white ancestry participants (CWP cases=836; controls=2390) from TwinsUK. IgG N-glycosylation level was measured using hydrophilic interaction ultra-performance liquid chromatography. Association between N-glycans and CWP was assessed using generalized linear mixed-effects model adjusted for age, sex and smoking status as fixed-effects and family relatedness, batch and run day as random-effects. Multiple testing correction considering 19 independent tests for Sidak's correction provided a two-tailed significant level of p=0.0027. Associated N-glycans were assessed for its causal relationship with CWP using latent causal variable modelling.

Several N-glycans containing bisecting N-acetylglucosamine (GlcNAc) (n=7) and core fucosylated structure (n=2) were associated with CWP. These N-glycans either promote or block antibody-dependent cell-mediated cytotoxicity (ADCC). Sensitivity analysis in females and N-glycans measured five years before CWP assessment supported the internal validity of the primary findings. Using causal inference, we observed partial genetic causality for N-glycans enriched for bisecting GlcNAc (n=3) and core fucosylation (n=1) with CWP. These findings suggest a role of N-glycans in the pathogenesis of CWP but would benefit from being replicated in an independent dataset.

Conflict of interest: None
Acknowledgement: We want to thank all the participants of TwinsUK.
Funding: This work was funded by the European Union's Horizon 2020 research and innovation program IMforFUTURE, under H2020-MSCA-ITN grant agreement number 721815. TwinsUK is funded by the Wellcome Trust, Medical Research Council, European Union, Chronic Disease Research Foundation (CDRF), Zoe Global Ltd and the National Institute for Health Research (NIHR)- funded BioResource, Clinical Research Facility and Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust in partnership with King's College London.

Sylvie Ricard-Blum
University of Lyon

Keynote speaker

"Extracellular matrix aging: a focus on the role of the lysyl oxidase family"

Sylvie Ricard-Blum
Institut de Chimie et Biochimie Moléculaires et Supramoléculaires, UMR5246 CNRS
Université Lyon 1, Villeurbanne, France
sylvie.ricard-blum@univ-lyon1.fr

The extracellular matrix (ECM) undergoes remodeling upon aging, leading to a decrease in collagen biosynthesis, glycosaminoglycans and proteoglycans and to an increase in metalloproteinase expression. These changes in ECM composition are associated with changes in ECM architecture and in its biophysical and mechanical properties, including a decreased solubility and an increased stiffness, which both depend on ECM covalent cross-linking. The lysyl oxidase family contributes to ECM cross- linking by catalyzing the first step of collagen and elastin cross-linking. The five members of this family, lysyl oxidase (LOX) and four lysyl oxidase like proteins (LOXL1-4), are copper amine oxidases, which oxidatively deaminate specific lysyl and hydroxyl residues in these ECM proteins. LOX expression is decreased upon aging, although the anti-aging vitamin C increases its expression, and LOXL2 knockdown protects against age-associated increase in arterial stiffening *in vivo*. In order to integrate all the mechanisms regulating the expression and activity of the LOX family upon ECM aging we have used a systemic approach and built its protein-protein and protein-glycosaminoglycan interaction network. We have also built a 3D model of LOX, which recapitulates its known structural and biochemical features, and will be useful to decipher the molecular mechanisms of LOX interaction with its various substrates, and to design potential anti-aging compounds.

Eline Slagboom
Leiden University Medical Center

Keynote speaker

"Metabolomics markers of biological age, mortality and exceptional longevity"

P. Eline Slagboom
Molecular Epidemiology, Leiden University Medical Center (The Netherlands)
Max Planck Institute for Biology of Ageing (Germany)

When we study phenotypes of biological ageing, we often explore physiological parameters mortality, multimorbidity or longevity as endpoints. Biological age could also be investigated but there is no gold standard composite marker of biological age. Biological age predictors have been generated in the past based on physiological read outs and clinical variables and the last 10 years many studies have added molecular or omics data to this field. Here I will discuss the use of 1H-NMR based metabolomics in studies into biological age prediction. The blood metabolome represents environmental cues as well as the host's genetic background, potentially offering a holistic view of an individual's health status. We have generated metabolomics predictors based on chronological age, on mortality and disease endpoints. We will discuss the exploration of such predictors in the biomedical ageing domain, focussing on physiological decline and clinical endpoints and for intervention studies aimed at health improvement. These studies will have to indicate which set of biological age predictors is most relevant for fundamental studies into ageing.

Samira Smajlović
University of Zagreb

Poster

"Modulation of the HNF1A and FOXA2 genes, using CRISPR/dCas9 molecular tools, in studies of their role in glucose stimulated insulin secretion"

HNF1A (the hepatocyte nuclear factor 1A), together with FOXA2, is a master regulator of a network of genes responsible for proper glucose stimulated insulin secretion (GSIS) in mice pancreatic β cells. Coordinately, they regulate N-acetylglucosaminyltransferase (Gnt)-IV responsible for proper glycosylation and localization of glucose transporter receptors GLUT1 and GLUT2 on the β cells thus enabling proper glucose intake. My aim was to examine if these two genes play the same role in GSIS in human pancreatic β cells since the mutation in HNF1A causes the MODY-type of diabetes and is probably contributing to diabetes type II.

Specific goal was to analyze if inactivation of HNF1A by cytosine methylation would cause the perturbation of GSIS and lead to diabetic phenotype as mutation does. I used CRISPR/dCas9-based molecular tools to manipulate the HNF1A promoter methylation in 1.1B4 cells, a model cell line for human pancreatic β cells. I targeted 4 CpG sites identified to have a role in HNF1A transcriptional regulation. Additionally, HNF1A promoter was simultaneously targeted using dCas9-VPR (for direct gene reactivation) and dSaCas9-TET1 (for gene demethylation). Also, both genes were simultaneously targeted using dSpCas9-KRAB for direct transcriptional repression. Results showed decreased level of FOXA2 transcripts by around 4 times compared to control. Changes in gene expression level were confirmed by changes of FOXA2 protein level by Western blotting and immunofluorescence analysis.

Then, I focused on manipulation of HNF1A and/or FOXA2 transcription and potential effect of the induced change on expression of their target downstream genes including glycosyltransferases, fucosyltransferases and fucose biosynthesis genes. When using dSpCas9-VPR and dSaCas9-TET1, increase in HNF1A expression was followed by significant downregulation of most fucosylstranferases, 8th day post transfection. Analysis of methylation level on 4 CpG sites in the 1. exon of HNF1A and gene transcription level revealed an inverse association between the two, confirming the regulatory role of the 4 CpG sites for the transcription of this gene in 1.1B4 cells.

Currently, I am analyzing how manipulations of the HNF1A and FOXA2 transcription affect transcription of their downstream genes and how this influences the whole N-glycome, GLUT receptors' localization and glucose uptake in 1.1B4 cells.

Maarten van Schaik
University of Leeds

Poster

“Modelling high dimensional count data using random effects”

Many studies collect omics data that can unravel biological mechanisms underlying diseases and healthy ageing. Several of these data sets take the form of high-dimensional, multivariate count data. For example, in human microbiome studies, sample counts are observed for several bacterial species. Our work is motivated by count data from a randomized clinical trial where samples were collected from subjects living in a helminth endemic rural area in Indonesia (n=152). The aim of this study is to unravel the relationship between helminth infection and health outcomes. Here we focus on modelling the effects of anthelminthic treatment and infection on the microbiome counts. However, the number of bacterial species (categories) is in hundreds, if not thousands.

Standard models for the effect of covariates on multivariate counts involve many parameters due to the high number of categories. One solution is to pool categories, but this might lead to a loss of information. We propose to model the effect of the covariates on the counts by using mixed effects negative binomial models, where the fixed effect parameters modelling the effects of the covariates on the multivariate counts are replaced by random effects. A simulation study is performed to evaluate the performance of our models in terms of bias and efficiency of the estimators. We consider various models for the effect of the covariates on the multivariate counts.

Preliminary results show that our random effect models result in a lower variance at the cost of a slight bias, and with similar ability to detect covariate effects using global tests. In the application to the Indonesia data, results show the variability at the taxa level in treatment and infection effects, their interaction, and correlations between these covariate effects.

Minzhen Xie
University of Leeds

Poster

"Non-parametric clustering of longitudinal functional data with application to NMR spectra of 18 kidney transplant patients"

The amount of data in the health domain is growing rapidly. These datasets have different forms: omics, imaging and functional. Examples of functional data are NMR spectra (in mass), ECG (time), pollution (time). A relevant question is whether these data can be used to cluster subjects to reveal their underlying health status. The aim of this research is to cluster the 18 kidney transplant patients based on their NMR spectra. An NMR spectrum is a function in mass, and the NMR spectra of each patient are recorded up to nine times (longitudinal design). The health status of the patients at the end of the study is known enabling assessing the accuracy of the proposed cluster methods. A straightforward approach is to compute summaries which are usually scalar of the NMR spectra and apply multivariate clustering methods such as K-means to these obtained summaries to cluster the patients. However this may result in the loss of relevant information available in functional data.

For example in our dataset, we extracted 8 scalar feature points of each patient by performing longitudinal functional principal component analysis (FPCA). Specifically, this research performed FPCA on two dimensions, which are spectra dimension and time dimension. From the spectra dimension, 4 feature curves (function of time) of each patient are extracted. Based on the extracted feature curves, 8 scalar feature points are extracted by performing FPCA on time dimension. By comparing the obtained clusters with the health outcomes, we concluded that the K-means on these 8 scalar feature points didn't provide an accurate clustering.

To use all available information, we propose a non-parametric clustering method for multivariate functional data. Firstly, the distance of multivariate curves is defined and small ball probability is computed based on defined distance. Secondly, we compute the heterogeneity of original sample and partitioned samples by using mean, median and mode. Thirdly, we define a criterion to determine whether the obtained clustering provides homogeneous clusters or require further splits. We compare our method with other (non) functional clustering methods via simulation and apply the method to the kidney data.

Frania Zúñiga-Bañuelos
glyXera GmbH, Max Planck Institute for Dynamics of Complex Technical Systems

Contributed talk

"Optimized Workflow for the In-Depth N-Glycoproteomic Analysis of Human Blood Plasma"

F. Zúñiga-Bañuelos[1,2], M. Hoffmann[2], U. Reichl[2,3], E. Rapp[1,2].
[1]glyXera GmbH, Magdeburg, Germany.
[2]Max Planck Institute for Dynamics of Complex Technical Systems, Magdeburg, Germany.
[3]Otto von Guericke University Magdeburg, Chair of Bioprocess Engineering, Magdeburg, Germany.

The glycoproteomic analysis of human blood plasma proteins poses several challenges: on the one hand, it is the mere number of glycoproteins with their corresponding glycoforms; on the other hand, it is the dynamic protein concentration range of more than ten orders of magnitude. Particularly, the low-abundant plasma proteins, can provide crucial information towards the understanding of diverse human traits, biomarker discovery or therapy development.

As a first step of an in-depth N-glycoproteomic analysis of blood plasma samples, it is necessary to enrich the low-abundant glycoproteins. Therefore, a dedicated workflow was developed that includes two protein fractionation steps. During development of the 1st fractionation step, two high-abundant blood plasma proteins (HAP) depletion columns were compared. For the establishment of the 2nd fractionation step, three methods were compared: two buffer configurations for mixed-mode ion exchange chromatography on native proteins, and gel-free electrophoresis on denatured and reduced proteins. To assess the benefit of the 1st fractionation step, all 2nd fractionation methods were also conducted directly with untreated plasma. In total, nine workflows were evaluated. After conducting a proteolytic digestion, all samples were analyzed by liquid chromatography coupled to high-resolution tandem mass spectrometry. The data processing was done by the search engines Mascot and Sequest HT. The protein identification lists obtained per fraction were merged into one protein list per workflow using visProteomics R package (available through github.com/imforfuture). In addition, VisProteomics retrieved features such as, protein concentration, protein class, biological function, etc.

We conclude that depletion of the top 14 HAP is crucial for increasing the identification-rate of low-abundant proteins. After analyzing untreated plasma and the 14 HAP depleted sample, the concentration of the proteins identified at the lowest level was 106 and 105 pg/mL, respectively. When including the 2nd fractionation step, proteins could be observed down to a concentration of 102 pg/mL. The glycopeptides derived from the best workflows will be enriched and the glycoforms of the low-abundant proteins will be described. Finally, the best workflow will be applied to a cohort of patients with Rheumatoid Arthritis, a disease that has shown a dependency on IgG glycosylation.