Data life cycle in plasma-medicine - what can we learn from the proteomics community

K. Wende[1], J.-W. Lackmann[1], S. Bekeschus[1], T. von Woedtke[2], K-D. Weltmann[2]

[1] Leibniz-Institute for Plasma Science and Technology, ZIK *plasmatis*, Greifswald, Germany

[2] Leibniz-Institute for Plasma Science and Technology, Greifswald, Germany

E-mail: kristian.wende@inp-greifswald.de

In plasma medicine researchers seek for answers how a gaseous system can have impact in biological systems and search for further clinical applications. Important avenues to tackle these kind of questions are the various omics-approaches using molecular biology and chemical analytics on the molecular level. Approaches as transcriptomics and proteomics have been successfully applied since years, younger members such as lipidomics or glycomics catch up.

Typically, untargeted experiments are performed in shot gun manner, resulting in large datasets easily reaching terabyte sizes. Careful analysis is necessary to unveil the information content in general and the specific evidence to allow the drawing of a valid conclusion. This comprises a bioinformatics workflow: raw data conversion, data curation (normalization etc.) and annotation, and multivariate analysis using statistical approaches such as cluster analysis or principal component analysis. Ultimately, analysis of the biological network, e.g. gene ontology (GO) term enrichment follows. Increasingly more popular but not mandatory for publication in all journals is storing the data in public repositories, allowing a re-evaluation by independent researchers. For most of this process, several software solutions are marketed or freely distributed, each coming with weaknesses and strength. Such, -omics data handling is a seemingly streamlined process, but many issues are unsolved and are approached differently by each researcher. Among these, sequence coverage, false discovery rates, non-existing databases, and missing values can be found. Ideally, the analysis of the converted data should be unbiased by the researcher, using machine-learning algorithms. Final challenge is to formulate the biological or clinical relevance. Integrating software solutions can help to do so, but suffer from high demand on programming skills, or are not flexible enough for research purposes.

In the talk, various approaches are discussed in the light of proteomics, an approach which has been used by our group since several years to unlock the fundamentals of plasma – cell or plasma – tissue interaction (1-3).

1. Schmidt, A., Dietrich, S., Steuer, A., Weltmann, K. D., von Woedtke, T., Masur, K., and Wende, K. (2015) Non-thermal plasma activates human keratinocytes by stimulation of antioxidant and phase II pathways. *J Biol Chem* **290**, 6731-6750
2. Bekeschus, S., Lackmann, J.-W., Gümbel, D., Napp, M., Schmidt, A., and Wende, K. (2018) A Neutrophil Proteomic Signature in Surgical Trauma Wounds. *International journal of molecular sciences* **19**, 761
3. Wende, K., Barton, A., Bekeschus, S., Bundscherer, L., Schmidt, A., Weltmann, K.-D., and Masur, K. (2013) Proteomic tools to characterize non-thermal plasma effects in eukaryotic cells. *Plasma Medicine* **3**