Diagnostics of seronegative rheumatoid arthritis using glycomic tools and machine learning

Rheumatoid arthritis (RA) is an inflammatory disease characterized by synovium swelling and destruction. There are two different types of rheumatoid arthritis, sometimes even believed to be two distinct disorders, namely seropositive (S+) and seronegative (S-) rheumatoid arthritis. For S+ rheumatoid arthritis, different markers occur in bloodstream, namely anti-cyclic citrullinated peptide antibodies (anti-CCP) and rheumatoid factors (RF). In S- rheumatoid arthritis patients, these markers are present in very low amount, or are even absent. It is thus impossible to distinguish the patients suffering from S- rheumatoid arthritis based on blood analysis.

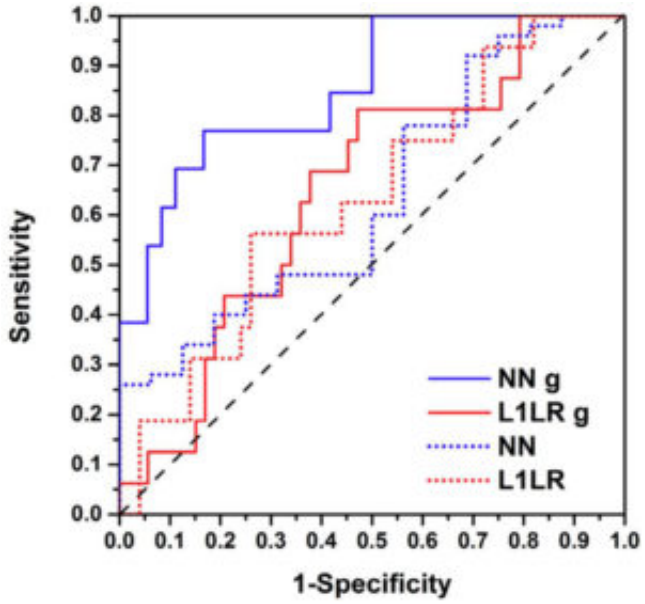

Fig. 1. ROC curve showing an increase for S- RA diagnostics using real human serum samples using glycomic approach (g, full lines) and artificial neural networks (NN, blue). Comparison to L1 regularized logistic regression (red) and immunoassays (dotted lines). The diagonal in the graph means 50% probability of a correct diagnose (basically a guessing). The better a marker, the more should the line be shifted toward upper left corner.

Glycomics is a rapidly evolving field with a huge impact on many different areas, including biomedicine. Glycans and glycoconjugates play an important role in many *in vivo* processes and pathologies. In case of rheumatoid arthritis development, *N*-glycan present on IgG molecules are desialylated and degalactosylated (sialic acid and galactose are the two outermost, terminal saccharide units), leading to a situation, where *N*-acetylglucosamine and mannose residues are more accessible. This makes the glycan able to bind to mannose-binding protein, thus activating the complement *via* lectin pathway. These degalactosylated IgG molecules (*i.e.* G0 or G1 IgG) are

switching from being an anti-inflammatory to pro-inflammatory agents. Glycan analysis is thus a promising tool for an accurate diagnostics of rheumatoid arthritis. Moreover, machine learning algorithms are gaining more attention in healthcare and medical research. In our study, we focused on high throughput, low cost methods for glycan analysis (*i.e.* fluorescent lectin microarray and enzyme-linked lectin binding assay, ELLBA) in a combination with artificial neural networks and L1 regularized logistic regression, as opposed to a “gold standard” mass spectrometry.

Out of 100 human samples (sera from healthy people and patients suffering from S- and S+ RA), 2000 data points were collected and data mined using machine learning algorithms. Healthy cohort and S+ rheumatoid arthritis patients could be distinguished with an accuracy of 83.3%, while adding glycoprofiling to the equation (using two different lectins – SNA-I and RCA) led to much higher accuracy of 92.5%. Moreover, while immunoassay failed completely to identify S- rheumatoid arthritis patients, glycan analysis correctly identified 43.8% of these patients. During these studies, one important aspect of glycan analysis was revealed, precisely the orientation and availability of the binding partners in different formats. The orientation of human IgG can be finely tuned using proteins A and L, adsorbed on an ELLBA plate. Also, using the whole serum samples was the preferred choice in this case instead of using isolated IgG molecules. However, there can be cases, when the differences in glycoprofile of a biomarker won´t be reflected by the overall serum glycoprofiling, most likely in low abundant serum oncomarkers.

Tomas Bertok, Erika Chocholova, Eduard Jane, Jan Tkac
Institute of Chemistry, Slovak Academy of Sciences, Dubravska cesta 9, 84538, Bratislava, Slovak Republic

Publication

Glycomics meets artificial intelligence - Potential of glycan analysis for identification of seropositive and seronegative rheumatoid arthritis patients revealed.
Chocholova E, Bertok T, Jane E, Lorencova L, Holazova A, Belicka L, Belicky S, Mislovicova D, Vikartovska A, Imrich R, Kasak P, Tkac J
Clin Chim Acta. 2018 Jun