



Published in final edited form as:

J Proteomics Bioinform. ; 6(12): 302–312. doi:10.4172/jpb.1000295.

Plasma Anti-Glycan Antibody Profiles Associated with Nickel level in Urine

Marko Vuskovic^{1,*}, Anna-Maria Barbuti², Emma Goldsmith-Rooney², Laura Glassman², Nicolai Bovin³, Harvey Pass², Kam-Meng Tchou-Wong⁵, Meichi Chen⁴, Bing Yan⁴, Jingping Niu⁴, Qingshan Qu⁵, Max Costa⁵, and Margaret Huflejt²

¹Department of Computer Science, San Diego State University, San Diego, 92182 CA, USA

²Department of Cardiothoracic Surgery, NYU School of Medicine, New York, 10016 NY, USA

³Shemyakin Institute of Bioorganic Chemistry, Russian Academy of Sciences, 117997 Moscow, Russia

⁴Lanzhou University School of Public Health, Lanzhou, Gansu 730000, China

⁵Department of Environmental Medicine, NYU School of Medicine, New York, 10016 NY, USA

Abstract

Nickel (Ni) compounds are widely used in industrial and commercial products including household and cooking utensils, jewelry, dental appliances and implants. Occupational exposure to nickel is associated with an increased risk for lung and nasal cancers, is the most common cause of contact dermatitis and has an extensive effect on the immune system. The purpose of this study was two-fold: (i) to evaluate immune response to the occupational exposure to nickel measured by the presence of anti-glycan antibodies (AGA) using a new biomarker-discovery platform based on printed glycan arrays (PGA), and (ii) to evaluate and compile a sequence of bioinformatics and statistical methods which are specifically relevant to PGA-derived information and to identification of putative “Ni toxicity signature”. The PGAs are similar to DNA microarrays, but contain deposits of various carbohydrates (glycans) instead of spotted DNAs.

The study uses data derived from a set of 89 plasma specimens and their corresponding demographic information. The study population includes three subgroups: subjects directly exposed to Nickel that work in a refinery, subjects environmentally exposed to Nickel that live in a city where the refinery is located and subjects that live in a remote location. The paper describes the following sequence of nine data processing and analysis steps: (1) Analysis of inter-array reproducibility based on benchmark sera; (2) Analysis of intra-array reproducibility; (3) Screening of data - rejecting glycans which result in low intra-class correlation coefficient (ICC), high coefficient of variation and low fluorescent intensity; (4) Analysis of inter-slide bias and choice of data normalization technique; (5) Determination of discriminatory subsamples based on multiple bootstrap tests; (6) Determination of the optimal signature size (cardinality of selected feature set) based on multiple cross-validation tests; (7) Identification of the top discriminatory glycans and their individual performance based on nonparametric univariate feature selection; (8) Determination of multivariate performance of combined glycans; (9) Establishing the statistical significance of multivariate performance of combined glycan signature.

Copyright: © 2013 Vuskovic M, et al.

This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

*Corresponding author: Department of Computer Science, San Diego State University, San Diego 92182 CA, USA, Tel: 858-344-7857; mvuskovic@mail.sdsu.edu.

The above analysis steps have delivered the following results: inter-array reproducibility $\rho=0.920 \pm 0.030$; intra-array reproducibility $\rho=0.929 \pm 0.025$; 249 out of 380 glycans passed the screening at $ICC>80\%$, glycans in selected signature have $ICC = 88.7\%$; optimal signature size (after quantile normalization)=3; individual significance for the signature glycans $p=0.00015$ to 0.00164 , individual AUC values 0.870 to 0.815; observed combined performance for three glycans $AUC=0.966$, $p=0.005$, $CI=[0.757, 0.947]$; $specificity=94.4\%$, $sensitivity=88.9\%$; predictive (cross-validated) AUC value 0.836.

Keywords

Printed glycan arrays; Immune response to occupational exposure to Nickel; Plasma anti-glycan antibodies; Quality analysis of printed glycan arrays; Immunoprofiling; Discriminatory signatures; Immunoruler

Introduction

Nickel is used in large number of every-day applications including coins, jewelry, household and cooking utensils, orthodontic and orthopedic implants, and batteries [1]. Exposure to Ni as a result of daily dermal contact or through occupational or environmental sources is known to produce a variety of pathologies including allergic contact dermatitis [2] and cardiovascular diseases [3]. Long-term occupational exposure is associated with lung and nasal cancers [4], and an increased risk for acute respiratory syndromes such as mild irritation and inflammation, bronchitis, pulmonary fibrosis, asthma, and pulmonary edema [5]. The general population is also exposed to Nickel via ingestion, since Nickel is a contaminant in drinking water and is present as a component or contaminant of foods [6,7]. Nickel can also be secreted in human breast milk leading to early dietary exposure of infants [8].

Due to the growing evidence about their toxic effects, in 1990, certain Ni compounds have been classified by the International Agency for Research on Cancer (IARC) as being carcinogenic to humans [9]. Excellent reviews of carcinogenic activities of Nickel are given [10-13].

Understanding the mechanisms of Nickel carcinogenicity and identification of biomarkers of susceptibility to Nickel toxicity is needed to allow development of tests for earlier identification and protection of individuals with high risk for the development of Nickel-related diseases.

This paper presents early results of linking Nickel exposure and presence of Nickel in human body to immune response measured by expressions of anti-glycan antibodies (AGA) using a new platform which is based on printed glycan arrays (PGA).

PGA is a new biomarker-discovery platform, which has been developed and utilized for biomarker discovery in the last several years [14]. PGA has useful advantages over nucleic acid-based testing and other platforms including: minimal invasiveness of blood sampling, stability of antibodies, low cost and short turnaround time. The printed glycan arrays (PGA) are similar to DNA microarrays, but contain depositions of various carbohydrate structures (glycans) instead of spotted DNAs. Most of these glycans can be found on the surfaces of human normal and cancer cells, as well as on the surfaces of human infectious agents such as bacteria and other pathogenic microorganisms. Pathologies including infection, inflammation and malignant transformation are associated with the appearance of abnormal

glycosylation of proteins and lipids present on surfaces of altered cells in tissues and in circulation.

The malignancy-related abnormal glycans are called tumor-associated carbohydrate antigens (TACA) [15]. There is evidence that numerous TACAs are immunogenic [16], and that the human immune system can generate antibodies against them. We have also demonstrated that the dynamics of anti-glycan antibodies detected by PGA can indicate the status of immune response to malignancies [17-19]. A prototype of PGA with a library of 200 glycan structures was built at Scripps Research Institute, La Jolla, California, under the auspices of the Consortium for Functional Glycomics (CFG), [20]. Further development and standardization of a PGA with 211 glycans was conducted at Cellexicon, Inc., La Jolla. The latest PGA version with a total of 392 probes, containing 380 glycans of pharmacological purity grade and 12 control probes was developed in the Tumor Glycome Laboratory at NYU SoM. The 211 and 380 PGA versions were developed in collaboration with Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, of the Russian Academy of Sciences, Moscow, Russia.

This paper describes a sequence of data processing and analysis steps, starting with quality analysis of raw PGA data and ending with a putative glycan signature which provides a basis for identification of individuals with high concentration of Nickel in urine. To the best of our knowledge, this is the first study in which the systematic immune-profiling of AGAs using PGA in plasma of individuals with different levels of occupational exposure to Nickel identifies the putative immune signature of Ni toxicity. The presented results demonstrate a potential of PGA-based analyses of serum/plasma anti-glycan antibodies coupled with the described bioinformatics approach in search of the disease biomarkers.

Methods and Materials

Demographic and clinical data

The study population included 3 groups of subjects: (1) Nickel refinery workers in Jinchang (30 cases), (2) Residents of Jinchang (30 cases) and (3) residents of adjacent city of Zhangye (29 cases). The latter two groups of subjects had only environmental exposures to nickel. The human subject protocol for this study was approved by the Institutional Review Boards of both the New York University School of Medicine and the Lanzhou University School of Public Health (IRB # 09-0726). Written informed consent was obtained from all participating subjects.

The demographic information included: age, smoking status, urinary creatinine [$\mu\text{g/g}$] and Urinary Ni [$\mu\text{g/L}$]. A characteristic of the study subjects is presented in Table 1 (Additional information about the study population can also be found in [21].)

Urinary Ni, used to index the individual's personal exposure to Ni, was analyzed for all study subjects by inductively coupled plasma mass spectroscopy (Elan DRCII; PerkinElmer, Norwalk, CT USA, [22]). Urinary cotinine, a major metabolite of nicotine and a valid biomarker of environmental tobacco smoke, was measured in each subject to confirm smoking status and control its potential confounding effects. Urinary cotinine measurements were measured using a Cotinine Direct ELISA kit (Immunoanalysis, Pomona, CA, USA [23]).

A simple preliminary analysis of demographic data shows that there is a significant difference in the level of Nickel in urine among directly exposed subjects in refinery and the subjects who are not working at refinery. For example Figure 1 shows the distributions of urinary Nickel concentrations among three subgroups of subjects; the Kolmogorov-Smirnov test of unequal distributions gives the following p-values: 0.0017 between refinery workers

and Jinchang residents, and 0.0073 between refinery workers and Zhanghye residents. However, the distributions for subjects at Jingchang and Zhanghye are significantly equal. In addition, Figure 2 illustrates the fact that there are a greater number of subjects within the refinery workers with high content of Nickel in urine [$>9.98 \mu\text{g/L}$] as compared with the number of subjects in the two other study groups. It is important to notice that some individuals not exposed occupationally to Nickel may also have high [$>9.98 \mu\text{g/L}$] content of Nickel in urine.

Printed glycan arrays

A printed glycan array (PGA) is used as a biomarker-discovery platform in a form of a “glycochip”. This glycochip is generated using standard robotic nano-printing technology that allows printing of a large range of amine-functionalized glycans/probes on amine-reactive N-hydroxysuccinimide (NHS)-activated glass slides (Schott GmbH, Germany) with surface modified for rapid covalent coupling. Glycans are printed in concentration of $50 \mu\text{M}$ generating spot sizes of ~ 70 microns. In addition to 380 glycans, 12 control probes including print buffer samples as reagents for background quality control and a spot-reference location are included in each print set of 392 probes. All these 392 probes are distributed within two sub-arrays, each containing replicate subarrays of $14 \times 14 = 196$ probes/spots. Each glycochip divided into two subarrays, accommodates 24 sub-grids, with 12 replicates for each printed glycan/probe.

The measurement of binding of human anti-glycan antibodies (AGA) to arrayed glycans, also called “immunoprofiling”, is achieved as described in [24]. Briefly, the glycochip is first incubated at 37°C with the subject’s plasma diluted 1:15 in a Carrier Buffer, allowing the binding of plasma antibodies to arrayed glycans. Plasma IgG, IgM and IgA immunoglobulins bound to glycans are visualized simultaneously with the “combo” biotinylated secondary goat anti human IgG, IgM and IgA antibodies (Pierce Biotechnology, Inc., Rockford, IL), and streptavidin-Alexa⁵⁵⁵ (Invitrogen/Molecular Probes, Carlsbad, CA). Fluorescence signal intensities that correspond to the binding of antibodies to glycans are scanned using Perkin Elmer ScanArrayG at 90% laser power, and quantified with ImaGene software (BioDiscovery, Inc., El Segundo, CA). The total relative fluorescence signal intensity values (appx. range: 1,000 – 12,000,000 Relative Fluorescence Units) are used for further data processing and analyses.

The quantified images are automatically analyzed for missing, or for overflowed spots (rare events) which are excluded from the final summarization of replicates performed by median. This process has enabled robust and accurate measurements.

Figure 3 shows an excerpt from a scanned image of a glycochip developed with plasma of one of the study subjects. The four subgrids represent two replicates (columns) and two complementary subarrays (rows) which contain all 380 glycans in our current library and 12 control probes.

Quality analysis and data pre-processing

In order to begin statistical analysis of PGA data with the goal of discovering potential putative signatures associated with the immune-responses to Nickel exposure, it is vital to first address the accuracy and the reproducibility of measurements, specifically the intra- and inter-array reproducibility of PGA signals. The former relates to precision of measurement of AGA bindings within a single array, while the latter relates to between-array biases induced by the platform, or even by the quality of plasma.

The first step in this quality analysis is to test the platform with the benchmark sera. The process of printing, development and scanning of glycochips at NYU Tumor Glycome

Laboratory is interleaved with immunoprofiling of the test sera called here “pooled arrays”. The test sera is prepared by pooling sera from several healthy subjects, which are then stored in larger quantities for usage in all similar experiments. The interleaving is performed for each new glycochip print batch, and in each of the immunoprofiling experiments/days. In this project 8 pooled arrays have been used. The scatter plots in Figure 4 show concordance between four typical pairs of pooled arrays out of total of 28 combinations (others were omitted to fit the page). The labels at the diagram axes indicate the instances of pooled sera arrays. The numbers above diagrams are Lin’s concordance correlation coefficient [25] CCC (left) and the Pearson correlation coefficient PCC (right). The fact that PCC is consistently higher than CCC, suggests a linear inter-slide bias in location and scale. This bias can be mostly removed by various normalization techniques. The results for all 28 combinations of pooled arrays are summarized in Figure 5.

The four diagrams in Figure 5 show Lin’s Concordance Correlation Coefficient (CCC), Pearson Correlation Coefficient, Overall Lin’s Concordance Correlation Coefficient [26,27] (OCCC), and Coefficient of Variation (CV). The first diagram (CCC) suggests relatively large bias between arrays, while the Pearson CC shows that the inter-array concordance is satisfyingly good. The OCCC and CV are computed independently for each array: the former is obtained by applying CCC to 66 combination pairs of 12 replicated subgrids on the same glycochip. These two diagrams show excellent intra-array concordance, i.e. small measurement error. All diagrams are obtained after screening out the glycans with small intensities, small Inter Class Correlation coefficient (ICC, discussed later) and the highly correlated glycans ($\rho > 0.95$). The number of glycans that have survived the screening is 249 out of total 380 glycans in our current PGA library.

To demonstrate the effect of inter-array bias among pooled sera, distributions of signal intensities are shown in Figure 6. The distributions are presented as sorted signal intensities for each array, across all glycans which have survived the screening. For the purpose of this diagram the highly correlated glycans were not screened out. In order to emphasize the differences, the diagram shows only 35 glycans associated with highest signal intensities of bound AGAs, however similar differences can be observed with all other glycans. The first diagram confirms the finding in Figures 4 and 5 about the inter-array bias. The second diagram shows the distributions after intra-array normalization (A-normalization). The dotted line in both diagrams shows mean distribution taken across all arrays, which is equivalent to distributions after quantile normalization [28,29] (Q-normalization). As shown, the A-normalization is not entirely effective as the Q-normalization. Other normalization techniques, such as between-array normalization, which is essentially equivalent to IQR-normalization, and sequential A-Q normalization were also considered, but they performed unfavorably.

Figure 7 shows similar intensity distributions for the study data used in discriminatory analyses.

The next step addresses the ICC of the measurements, i.e. the ratio between the biological variability across all subjects and the total variability which includes the controlled measurement error. The results are shown in Figure 8. The common x-axis of all diagrams in the figure represents glycans sorted according to decreasing values of ICC (black line). The ICC is estimated as in [30]. The upper diagram in blue color shows the robust version of ICC obtained by appropriate replacements of means and standard deviations by medians and median absolute deviations (MAD) respectively. The lower two diagrams (green and magenta) represent coefficient of variation and its robust version. The relatively high ICC values for most of the glycans, as well as the low CV values assure that the data in Nickel Exposure Study qualify for further discriminatory analyses.

Discriminatory samples and signature length

After having established that the data obtained from AGA immunoprofiling qualify for further investigation it is necessary to define the subsamples which can be used in discriminatory analysis. Our original goal was to identify glycan-based immuno-signature that would allow identification of individuals exposed to potentially dangerous Nickel sources, such as occupational exposure to Nickel in Nickel refinery. We have therefore first compared three study subpopulations where one subpopulation included 30 workers of Nickel refinery, and two other subpopulations with subjects exposed to only environmental Nickel sources. Interestingly, this analysis did not deliver plausible “signature of hazardous Nickel exposure”, most likely due to yet unknown factors, such as lack of adequate glycan probes.

We have then started a search for differences in immunoprofiles of individuals based on their Urinary Nickel content regardless of their location and assignment to a study subgroup. A straightforward approach would be to define a simple cutoff value which separates individuals with low from high concentration of Nickel in urine. However, this approach did not yield satisfactory results, seemingly due to the adverse impact of cases with medium concentrations of urinary Nickel, which produced undesired clutter. The approach that has offered better results was to exclude the cases with medium concentrations. For example, the cutoff value can be the number of cases in balanced subsamples with low and high concentrations. Determination of this cutoff value is a matter of compromise: large value increases the clutter, while the small value causes loss of discriminatory power, both resulting in diminished statistical significance.

For the purpose of finding the optimal solution we have run a series of bootstrap tests for various subsample sizes and for various signature lengths. The replicate statistic has been chosen to be area under the ROC curve (AUC), which has a number of desirable properties over discrimination accuracy, including independence from discriminant bias, good resolution and ranking ability [31]. The AUC is computed for a combination of normalized signals associated with a chosen glycan signature, i.e. set of features. The combination is performed through projection determined with multivariate logistic regression (MLR). Each bootstrap run contained 500 replications based on permutations instead of usual resampling with replacement, the former being more conservative. The results for quantile-normalized data are presented in Figure 9, which shows the achieved significance level (ASL) [32]. The figure suggests the optimal subsample sizes of 18 subjects. It should be noted that the same repeated bootstrap tests were run for other normalization approaches and for sample sizes smaller than 16 and larger than 20, all giving inferior results. This cutoff value implies low concentrations less than or equal to 4.44 $\mu\text{g/L}$ of Urinary Nickel and high concentrations greater than or equal to 9.98 $\mu\text{g/L}$ of Urinary Nickel.

So determined subsamples can be now used to perform a simple univariate feature selection, for example the non-parametric Wilcoxon-Mann-Whitney rank-sum test, the results are presented in the next section.

Once the samples of two discriminatory groups are determined, it remains to decide the signature length, i.e. the number of top selected glycans, which minimize the likelihood of over fitting. This can be done by multiple cross-validations performed for various sizes of signatures, then by choosing the signature size which yields the optimal cross-validated performance. This is shown in Figure 10, which presents the 100 times repeated 10-fold cross-validation. For the performance measure is again used the combined AUC value. The diagram shows that the optimal cross-validated AUC value can be achieved with three-glycan signatures, which yields the predictive AUC value 0.836, while the training (observed) AUC value is 0.966. In addition, the diagram in Figure 10 presents the Kuncheva

stability index (SI) [33], which reaches the maximum at two glycans. The reason SI drops after two glycans is that the third and fourth glycan ($GID=133$ and 136 , Figure 11) alternate in various cross-validation folds.

The stability of feature selection can also be illustrated by the frequency of occurrences of each feature in total of $100 \times 10 = 1000$ cross-validation folds, presented in Figure 11. As seen the glycan $GID=191$ has been selected 100% of times, while the glycans 264 and 133 were selected 97% and 93% of times respectively. After the third glycan, the frequencies drop significantly.

Results and Discussion

In the previous section we have determined the subsamples associated with low and high level of Nickel in urine which can be now used in discriminatory analysis and in identification of putative glycan signature. In addition, we have determined the optimal signature size, which will least likely cause over fitting.

Discriminatory analysis

A first step in discriminatory analysis is to perform some univariate test for all glycans of interest. Since the PGA signals depart significantly from normal distribution (they even for the most glycans have multinomial distributions) we prefer to use some non-parametric test, such as the Wilcoxon-Mann-Whitney two-sample rank sum test. An additional benefit of this test is that the AUC values are directly linked with the p-values of the test. The same test was employed in the previous section, where the statistic used for sample selection and cross-validation was the AUC value.

The test was applied to quantile-normalized PGA signals obtained by median summarized replicates. The result for 10 glycans with lowest p-value, or highest AUC value is shown in Table 2.

The first column of the table represents the glycan identification numbers (GID). The corresponding glycan structures are shown in Table 3. The signs of the z-statistic indicate whether the PGA signals decrease (negative Z), or increase (positive Z) with the increase of urinary Nickel levels. The relatively high AUC values suggest high discriminatory power of the samples. Low values of the false discovery rate (FDR) imply a good confidence in the results, especially for the first three glycans, which is in compliance with the finding in cross-validation test.

The sixth column of the table, AUC_c , shows the cumulative AUC values obtained for combination of all glycans above each respective glycan. For example the cumulative value for combination of three top glycans, $GID=191, 264, 133$, is $AUC_c=0.966$. The combination of glycans is performed by multivariate logistic regression.

A convenient way to visualize the performance of the training set for the selected glycan signature can be achieved with the Immunoruler [34]. The immunoruler is a bar graph which presents the subjects with low (left bars in blue) and the subjects with high (right bars in magenta) urinary Nickel. The bars indicate the risk scores, which are quantified as probability of membership to the group of high urinary Nickel. The bars are sorted according to these probabilities for each group separately. The two shades of each color indicate the quartile regions. The risk scores are computed by combining the quintile-normalized signals for the signature determined above. The combination of the intensities associated with glycans from the signature is based on MLR. This kind of visualization makes apparent the number of false positives $FP=1$, false negatives $FN=2$, true positives

$TP=16$ and true negatives $TN=17$, as well as specificity $S_p=94.4\%$ and sensitivity $S_n=88.9\%$. The training precision is 91.7% and the observed AUC value is 0.966 (Figure 12).

The individual statistical significance of each glycan from the selected signature is evident from p-values and FDR shown in Table 2.

Now we need to establish the statistical significance of the observed $AUC=0.966$ obtained for the combined three-glycan signature. This can be achieved with a bootstrap test. In order to keep a conservative approach we have performed the unbiased permutation bootstrap with 1000 replications, rather than the bootstrap based on resampling with replacement. The empirical probability density function under null hypothesis is presented in Figure 13.

As shown, the two-sample confidence interval is $CI=[0.757, 0.947]$, or one-sided upper bound $CU=0.935$. Consequently the achieved significant level is $ASL=0.005$, which gives a very strong evidence against the null hypothesis.

Finally, the comparison of normalized, but not transformed intensities of low (diagrams at left) and high (diagrams at the right) urinary Nickel for selected signature glycans is presented in Figure 14. As shown, the intensities are decreasing in subjects with high level of Urinary Ni for glycan $GID=191$, while the effect is opposite for glycans $GID=264$ and 133, which complies with the Z-values in Table 2. The bottom diagrams show the combined intensities.

Association with other demographic factors

The focus of this study was to find association of Urinary Ni with the immune response measured by anti-glycan antibodies (AGA). A natural question after the analysis presented above is whether there is a possibility of association of AGA with other demographic factors, such as age, smoking and creatinine, listed in Table 1.

Since the age of subjects is of interest we are showing the result of linear regression of the Urinary Ni and the AGA for the most expressed glycan in signature, $GID=191$, with the age (Figure 15).

In order to decrease the influence of outliers we have used in the second diagram the log-transformed quantile normalized intensities. As shown the age does not correlate with the Urinary Nickel and with the AGA, at least for the data at hand ($\rho=0.076$ and $\rho=-0.093$, respectively).

Another variable of interest would be the smoking status of subjects. Unfortunately, the number of non-smoking subjects is only 15 versus 74 smokers, which makes a proportion of smokers 5:1. The small nonsmoking group and high sample imbalance makes it difficult to draw any plausible conclusion.

Yet another variable from the list is creatinine. The bottom diagram of Figure 15 indicates that the Urinary Ni and creatinine are relatively highly correlated considering the variation in measured data ($\rho=0.761$), suggesting that these factors are potentially interchangeable in our analysis. Therefore we have used Urinary Ni instead of creatinine which has offered slightly better performance in terms of stability of feature selection and statistical significance of combined AUC value. This however needs to be further investigated in a future study with larger samples.

Nickel exposure, glycosylation and AGA

The direct link between the nickel exposure, aberrant glycosylation and anti-glycan antibodies has not yet been established. Cellular glycosylation is a highly dynamic process carried out by a concerted action of hundreds of glycosyltransferases, glycosidases and other proteins, and it is most likely that any nickel-related molecular damage on the nucleic acid level that results in cellular malignant transformation will also result in aberrant glycosylation. Salnikow and Kasprzak [35] discuss the direct effects of the chronic exposure to nickel on altering glycosylation and assembly of the extracellular matrix components as well as assembly and function of surfactants and complement by depleting intracellular ascorbate. One of the mechanisms of nickel on the human innate immunity has recently been elucidated by the elegant demonstration that nickel directly triggers human Toll-like receptor 4 (TLR4) and Pattern Recognition Receptor (PRR) signaling which results in the expression of multiple proinflammatory genes [36]. It is therefore likely that the expected modifications of glycosylation on the cell surface and in the extracellular matrix participate in the immune response that then targets molecules with the aberrant glycosylation pattern.

The putative Ni-toxicity AGA-based signature identified here brings novel, and potentially very important, findings which reveal a significance of glycans some of which are yet known to the field of the disease biomarkers. For example, phosphorylated LacNAc, 6-P-Gal β 1-4GlcNAc β (GID 191), a main discriminatory glycan in the putative Nitoxicity signature, has not been reported as a component of the glycome. This glycan has been synthesized in the laboratory of Prof. Nicolai Bovin as an analog of 6-O-Su-Gal β 1-4GlcNAc(6'-O-Sulfate-LacNAc) to study the specificity and effect of charge versus moiety structure in antibody recognition of Gal β 1-4GlcNAc β and Gal β 1-4GlcNAc β -containing negatively charged carbohydrates. To our surprise, while 6-P-Gal β 1-4GlcNAc β has shown major decrease in antibody bindings in subjects with high versus low urinary Ni, its sulfated analog 6-O-Su-Gal β 1-4GlcNAc did not show any significant differences in antibody binding intensities between these two groups. We have also not found any significant correlation of antibody signal binding between 6-P-Gal β 1-4GlcNAc β and other glycans present on our PGA. Because of the well-known toxic metal-chelating properties of phosphate, it is tempting to hypothesize that the antibodies binding on the PGA to this phosphorylated glycan, are in fact involved in the clearance of Nickelating phosphorylated and glycosylated macromolecules.

While N-acetyllactosamines, in particular poly-N-acetyllactosamines consisting of repeated units of Gal β 1-4GlcNAc (GID 384), have been well-recognized as Tumor Associated Carbohydrate Antigens [37-39], association of Gal β 1-3GlcNAc or Lewis^C (a disaccharide in GID 379) with malignant transformation is much less known. We have recently found antibodies differentially binding several glycans containing Lewis^C disaccharide in sera of patients presenting with Non-Small Cell Lung Carcinoma and Malignant Pleural Mesothelioma as compared with sera of control subjects. The Lewis^C glycans appeared in the putative signatures of these tumors, and we are currently investigating the significance of this glycan in malignant transformation.

Highly correlated, differential antibody binding to several lactose (Gal β 1-4Glc) glycans containing glycine or two amino acids as a spacer is a novel finding, and the true antibody target remains unknown. These molecules probably mimic aberrant molecular patterns on malignant cell surface where glycosphingolipids (inner Lac) play a significant role [39,40].

Based on the presented results, we hypothesize that the elevated level of nickel in urine signals a certain type of cellular and possibly systemic damage which is reflected by the differential expression of specific anti-glycan antibodies. The biological targets that include the glycans identified here by the antibody recognition remain unknown. Altogether, the

findings reported here could lead to the further investigations of the mechanisms and biomarkers of an individual's susceptibility to the nickel toxicity.

Conclusion

The goal of this study was to investigate the evidence of immune-response in workers of a Nickel refinery to hazardous levels of Nickel in their work place and the immune-responses reflecting elevated levels of Urinary Ni, as measured by anti-glycan antibodies (AGA). For this purpose we have evaluated immunoprofiles of plasma specimens from 89 subjects, some directly exposed to high levels of airborne Nickel in a flash-smelting workshop of nickel refinery, and some exposed to only environmental sources of Nickel. The plasma specimens were processed using the new high-throughput platform based on printed glycan arrays (PGA) in the form of a glycochip developed in the Tumor Glycome Laboratory of NYU School of Medicine. The extensive quality analysis has shown that data obtained from PGAs qualify for subsequent discriminatory analysis: the mean Pearson inter-array concordance coefficient for test sera was $\rho=0.921 \pm 0.03$, proving a good platform inter-array concordance, and the mean coefficient of variation across PGA replicates $CV=19.8 \pm 2\%$, proving good intra-array reproducibility. Similar analysis performed on actual Nickel data (89 arrays) has resulted in $CV=21.6 \pm 4.9\%$, and inter class correlation coefficient (ICC) larger than 85% for 260 glycans on the chip (88.7% for glycans found in the signature). The successful feature selection has been achieved after the entire collection of Nickel specimens were divided into two groups, one containing subjects with low (4.44 $\mu\text{g/L}$) and the other containing subjects with high (9.98 $\mu\text{g/L}$) level of urinary Nickel. The screened and the quantile-normalized PGA signals were then used in univariate feature selection based on non-parametric Wilcoxon-Mann-Whitney rank-sum test, which has suggested the signature $GID=191, 264$ and 133. The number of glycans in the signature is limited to three glycans to avoid over-fitting, as determined by an unbiased 100 times 10-fold cross-validation test applied to various sizes of signatures. The signals associated with the signature can be combined using projection based on multivariate logistic regression, thus forming a single discriminative marker. The observed AUC value for this marker is 0.966. The statistical significance of this result has been confirmed with the permutation bootstrap test with 1000 repetitions which has provided a strong evidence against the null hypothesis at achieved significance level $ASL=0.005$.

The work presented in this paper entails additional investigation, such as:

- Correlation (regression) of Urinary Ni concentration with the combined discriminatory marker and other confounding factors; for example, length of exposure and subjects' other demographic and clinical information such as age, smoking, inflammation, presence of other diseases, etc.
- Investigation of the impact of direct exposure to airborne Nickel on the level of antibodies against glycans as opposed to the environmental exposure.
- Linking the urinary Nickel levels with the potential clinical presentation of disease symptoms. A wide range of urinary Nickel levels in refinery workers with a long-term, direct occupational exposure to Nickel suggests significant differences in individual biological responses to this occupational carcinogen.

All these investigations require stratification of the already small cohort, which would significantly reduce the operative sample sizes, thus lowering the statistical significance and the plausibility of inference. Therefore much larger study cohorts are required. In addition, more extensive clinical and demographic information is also needed, including the follow-up health status information of the study subjects.

Acknowledgments

This project has been supported by NIEHS 5P30 ES000260 grant through a Pilot Project awarded to Margaret E. Huflejt, Ph.D. The subject recruitment, demographic information and sample collections, as well as measurements of nickel and creatine in urine have been performed at Lanzhou University School of Public Health, Lanzhou, Gansu 730000, China. The glycans used in PGA were synthesized and purified in the Carbohydrate Chemistry Laboratory at Shemyakin Institute of Bioorganic Chemistry, Moscow, Russia, under the support of the grant from RAS Presidium Program "Molecular and Cell Biology". Printing of glycan arrays and their development with plasma, scanning and signal quantification of bound antibodies has been performed in the Tumor Glycome Laboratory, NYU SoM. This study has been carried out upon human subject protocol approved by the Institutional Review Boards (IRBs) of both the New York University School of Medicine and the Lanzhou University School of Public Health, China. Authors wish to acknowledge very valuable discussions with Professor Richard Levine from Department of Statistics, SDSU. Finally, authors acknowledge excellent and very helpful comments from the reviewers.

References

1. Norseth, T. Handbook on the Toxicology of Metals. Elsevier/North-Holland Biomedical Press; Amsterdam: 1979.
2. Basketter DA, Angelini G, Ingber A, Kern PS, Menné T. Nickel, chromium and cobalt in consumer products: revisiting safe levels in the new millennium. *Contact Dermatitis*. 2003; 49:1–7. [PubMed: 14641113]
3. Lippmann M, Ito K, Hwang JS, Maciejczyk P, Chen LC. Cardiovascular effects of nickel in ambient air. *Environ Health Perspect*. 2006; 114:1662–1669. [PubMed: 17107850]
4. Grimsrud TK, Berge SR, Haldorsen T, Andersen A. Exposure to different forms of nickel and risk of lung cancer. *Am J Epidemiol*. 2002; 156:1123–1132. [PubMed: 12480657]
5. Grimsrud TK, Berge SR, Resmann F, Norseth T, Andersen A. Assessment of historical exposures in a nickel refinery in Norway. *Scand J Work Environ Health*. 2000; 26:338–345. [PubMed: 10994800]
6. Agency USEP. Drinking water regulations and health advisories. Water, Oo, editor. Washington, D.C.: 1993.
7. Rojas E, Herrera LA, Poirier LA, Ostrosky-Wegman P. Are metals dietary carcinogens? *Mutat Res*. 1999; 443:157–181. [PubMed: 10415439]
8. Grandjean, P. Human exposure to nickel. In: Sunderman, FW., Jr; Aitio, A., editors. Nickel in the Human Environment: Proceedings of a Joint Symposium; Lyon, France. March 1983; Lyon, France: International Agency for Research on Cancer; 1983.
9. Chromium, nickel and welding. *IARC MonogrEvalCarcinog Risks Hum*. 1990; 49:1–648.
10. Arita A, Munos A, Chervona A, et al. Gene Expression Profiles in Peripheral Blood Mononuclear Cells of Chinese Nickel Refinery Workers with High Exposures to Nickel and Control Subjects. *Cancer Epidemiol Biomarkers Prev*. 2013; 22:261–269. [PubMed: 23195993]
11. Chervona Y, Arita A, Costa M. Carcinogenic metals and the epigenome: understanding the effect of nickel, arsenic, and chromium. *Metallomics*. 2012; 4:619–627. [PubMed: 22473328]
12. Muñoz A, Costa M. Elucidating the mechanisms of nickel compound uptake: a review of particulate and nano-nickel endocytosis and toxicity. *ToxicolApplPharmacol*. 2012; 260:1–16.
13. Costa M. Molecular mechanisms of nickel carcinogenesis. *BiolChem*. 2002; 383:961–967.
14. Bovin NV, Huflejt ME. Unlimited Glycochip. *Trends in Glycoscience and Glycotechnology*. 2008; 20:245–258.
15. Hakomori S. Glycosylation defining cancer malignancy: new wine in an old bottle. *ProcNatlAcadSci U S A*. 2002; 99:10231–10233.
16. Aarnoudse CA, Garcia Vallejo JJ, Saeland E, van Kooyk Y. Recognition of tumor glycans by antigen-presenting cells. *CurrOpinImmunol*. 2006; 18:105–111.
17. Huflejt ME, Cristofanilli M, Shaw LE, Reuben JM, Fritsche HA, et al. Detection of neoplasia-specific clusters of anti-glycan antibodies in sera of breast cancer patients using a novel glycan array. *ProcAmerAssoc Cancer Res*. 2005a; 46
18. Huflejt, ME.; Vuskovic, M.; Blixt, O.; Xu, H.; Shaw, LE., et al. Glycan array identifies specific signatures of anti-glycan auto antibodies in sera of breast cancer Patients: diagnostic, prognostic

- and therapeutic opportunities. *Breast Cancer Res Treat*; 28th Annual San Antonio Breast Cancer Symposium; San Antonio, TX. 2005b. p. S85
19. Hufleit, ME.; Blixt, O.; Vuskovic, M.; Xu, H.; Shaw, LE., et al. Anti-glycan autoantibodies as markers of malignancy status. 5th Internat. Symposium on Minimal Residual Cancer; San Francisco. September 11-14; 2005c.
 20. Blixt O, Head S, Mondala T, Scanlan C, Huflejt ME, et al. Printed covalent glycan array for ligand profiling of diverse glycan binding proteins. *ProcNatlAcadSci U S A*. 2004; 101:17033–17038.
 21. Arita A, Niu J, Qu Q, Zhao N, Ruan Y, et al. Global Levels of Histone Modifications in Peripheral Blood Mononuclear Cells of Subjects with Exposure to Nickel. *Environ Health Perspect*. 2012; 120:198–203. [PubMed: 22024396]
 22. Oliveira JP, de Siqueira ME, da Silva CS. Urinary nickel as bioindicator of workers' Ni exposure in a galvanizing plant in Brazil. *Int Arch Occup Environ Health*. 2000; 73:65–68. [PubMed: 10672494]
 23. Hu Y, Li G, Xue X, Zhou Z, Li X, et al. PAH-DNA adducts in a Chinese population: relationship to PAH exposure, smoking and polymorphisms of metabolic and DNA repair genes. *Biomarkers*. 2008; 13:27–40. [PubMed: 17896209]
 24. Huflejt ME, Vuskovic M, Vasiliu D, Xu H, Obukhova P, et al. Anti-carbohydrate antibodies of normal sera: findings, surprises and challenges. *MolImmunol*. 2009; 46:3037–3049.
 25. Lin LI. A concordance correlation coefficient to evaluate reproducibility. *Biometrics*. 1989; 45:255–268. [PubMed: 2720055]
 26. Barnhart HX, Haber M, Song J. Overall concordance correlation coefficient for evaluating agreement among multiple observers. *Biometrics*. 2002; 58:1020–1027. [PubMed: 12495158]
 27. King TS, Chinchilli VM. A generalized concordance correlation coefficient for continuous and categorical data. *Stat Med*. 2001; 20:2131–2147. [PubMed: 11439426]
 28. Amaratunga D, Cabrera J. Analysis of Data from Viral DNA Microchips. *Journal of the American Statistical Association*. 2001; 96:1161–1170.
 29. Bolstad BM, Irizarry RA, Astrand M, Speed TP. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics*. 2003; 19:185–193. [PubMed: 12538238]
 30. Kilem, LG. *Wiley Encyclopedia of Clinical Trials*. John Wiley & Sons, Inc.; 2008. Intrarater Reliability.
 31. Ling, CX.; Huang, J.; Zhang, H. AUC: a statistically consistent and more discriminating measure than accuracy. *Proceedings of the Eighteenth International Joint Conference of Artificial Intelligence (IJCAI)*; 2003.
 32. Efron, B.; Tibshirani, RJ. *Monographs on Statistics and Applied Probability*. Vol. 57. CRC Press; 1993. *An Introduction to the Bootstrap*.
 33. Kuncheva, LI. A Stability index for feature selection. *Artificial Intelligence and Applications; Proceedings of the 25th IASTED International Multi-Conference*; Innsbruck, Austria. 2007.
 34. Vuskovic MI, Xu H, Bovin NV, Pass HI, Huflejt ME. Processing and analysis of serum antibody binding signals from Printed Glycan Arrays for diagnostic and prognostic applications. *Int J Bioinform Res Appl*. 2011; 7:402–426. [PubMed: 22112531]
 35. Salnikow K, Kasprzak KS. Ascorbate depletion: a critical step in nickel carcinogenesis? *Environ Health Perspect*. 2005; 113:577–584. [PubMed: 15866766]
 36. Schmidt M, Raghavan B, Müller V, Vogl T, Fejer G, et al. Crucial role for human Toll-like receptor 4 in the development of contact allergy to nickel. *Nat Immunol*. 2010; 11:814–819. [PubMed: 20711192]
 37. Henou C, Didier E, De Latour M, Kwiatkowski F, Communal Y, et al. Association of N-acetyllactosamine with tumor progression in human breast cancer: a study using a 16 kDa chick embryo lectin. *Int J Mol Med*. 1998; 1:771–775. [PubMed: 9852296]
 38. Dennis JW, Carver JP, Schachter H. Asparagine-linked oligosaccharides in murine tumor cells: comparison of a WGA-resistant (WGA^r) nonmetastatic mutant and a related WGA-sensitive (WGA^s) metastatic line. *J Cell Biol*. 1984; 99:1034–1044. [PubMed: 6547960]
 39. Dennis JW, Laferté S, Vanderelst I. Asparagine-linked oligosaccharides in malignant tumour growth. *BiochemSoc Trans*. 1989; 17:29–31.

40. Bovin NV. Natural antibodies to glycans. *Biochemistry (Mosc)*. 2013; 78:786–797. [PubMed: 24010841]

Submit your next manuscript and get advantages of OMICS Group submissions

Unique features

- User friendly/feasible website-translation of your paper to 50 world's leading languages
- Audio Version of published paper
- Digital articles to share and explore

Special features

- 300 Open Access Journals
- 25,000 editorial team
- 21 days rapid review process
- Quality and quick editorial, review and publication processing
- Indexing at PubMed (partial), Scopus, EBSCO, Index Copernicus and Google Scholar etc
- Sharing Option: Social Networking Enabled
- Authors, Reviewers and Editors rewarded with online Scientific Credits
- Better discount for your subsequent articles

Submit your manuscript at: <http://www.editorialmanager.com/proteomics>

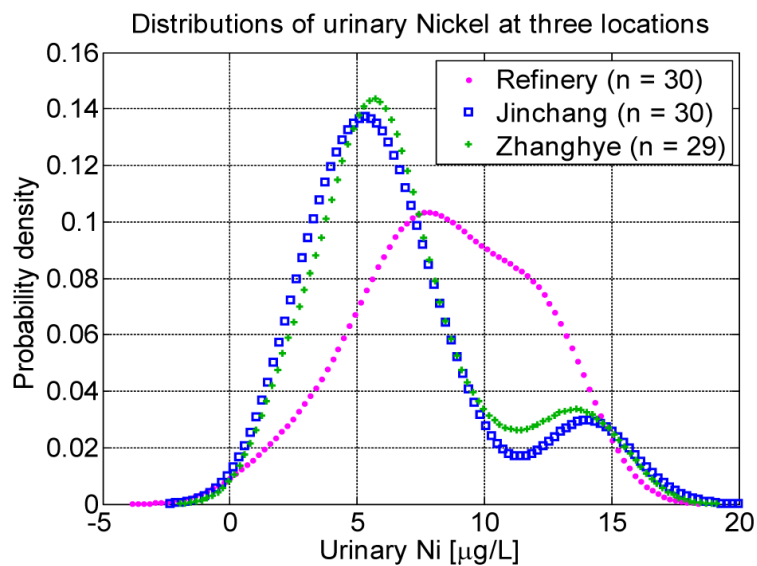


Figure 1. Probability density functions of Urinary Ni in subjects from three locations. The distributions show that a significantly larger number of workers of the refinery have a significantly higher content of Nickel in urine as compared with the subjects at the two other locations, with no occupational exposure to Nickel.

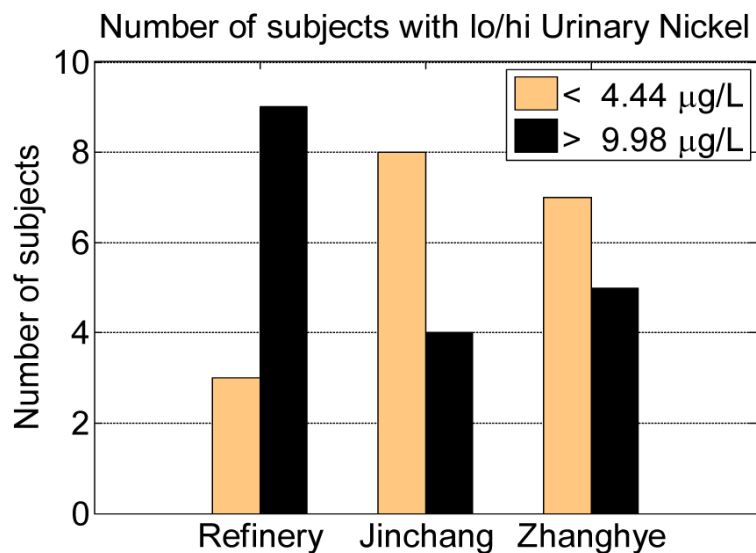


Figure 2. Number of subjects with low and high content of Urinary Ni within three groups: workers of the refinery located in Jinchang, residents of Jinchang who are not working in refinery and residents of a remote city of Zhanghye. The subjects of the former group are directly exposed to Nickel, while the subjects of the latter two groups are only environmentally exposed to Nickel.

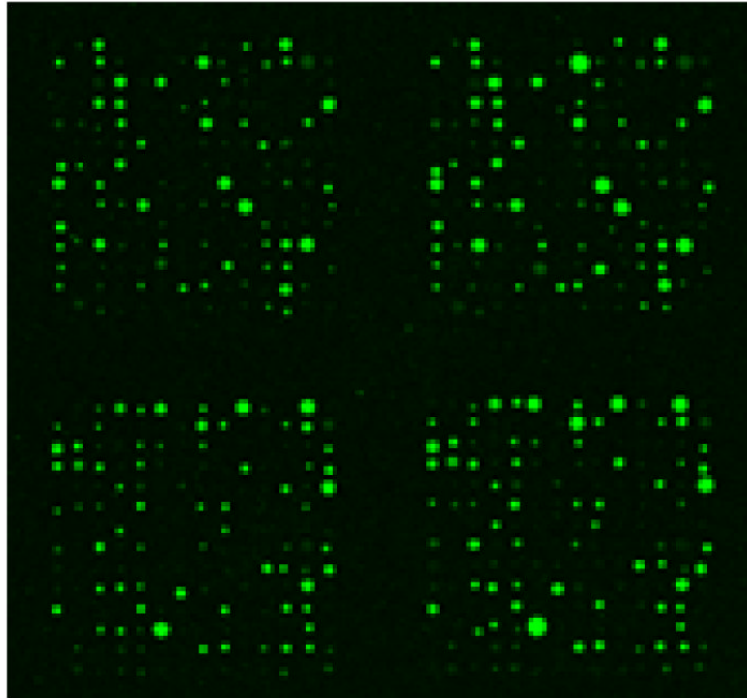


Figure 3. Excerpt from a scanned image of a glycochip developed with plasma of one of the study subjects. The microarray contains 392 probes replicated 12 times and arranged into 24 subarrays. The 392 probes contain a library of 380 glycans printed in concentration of 50 μ M, and 12 control probes. The glycans are distributed into two subarrays, each containing 196 probes. Each sub array contains 14 by 14 spots. The figure shows two replicated sub grid pairs.

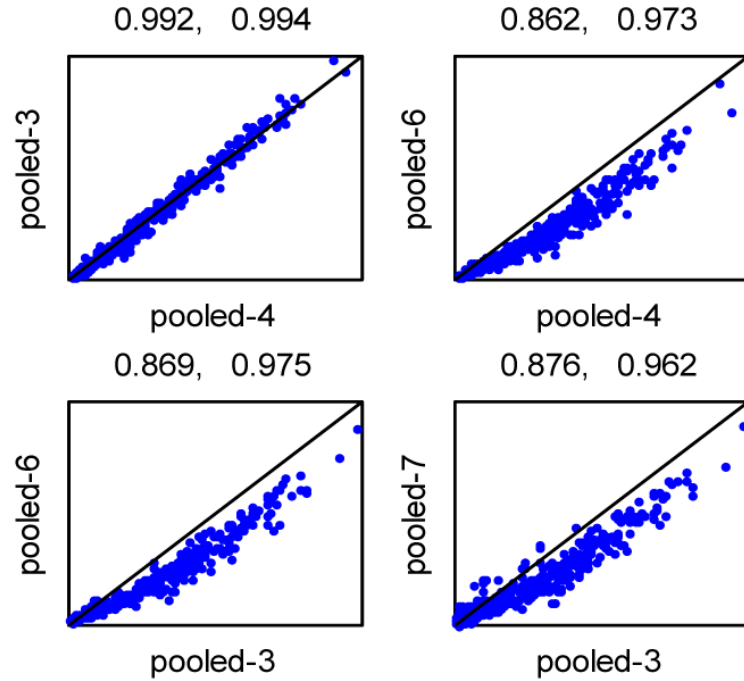


Figure 4.

Concordance log-log scatter plots for Nickel Exposure Study. The plots are part of inter-slide quality control which uses slides developed on different dates and on different PGA print batches with benchmark sera obtained from a pool of four healthy subjects. In this study there were 8 such slides whose development was interleaved with that of study subjects, resulting in 28 combination pairs. The scatter plots show logarithms of raw intensity signals. Numbers above plots are Lin's concordance correlation coefficient (CCC, left) and Pearson correlation coefficient (right). The latter coefficient ignores linear bias in scale and location, while the former shows the true differences. The figure presents only 4 combinations out of 28, with patterns typical for this study; correlation coefficients for all 28 combinations are shown in Figure 5.

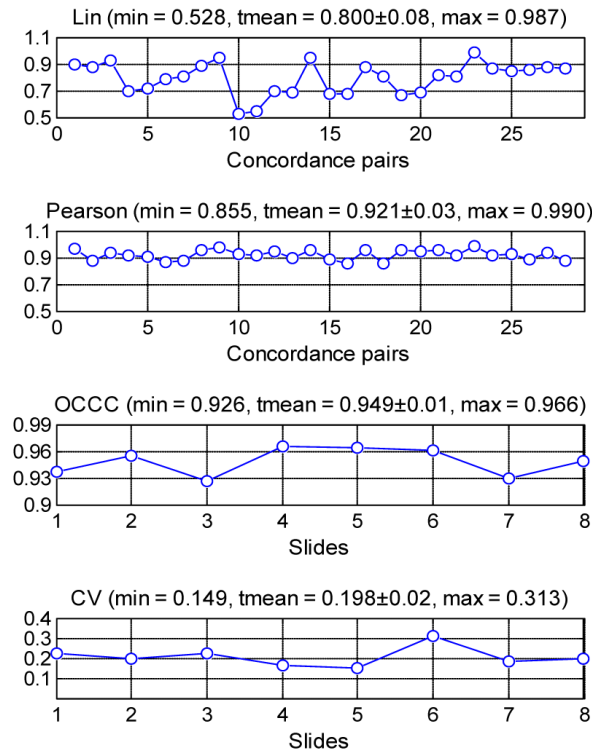


Figure 5.

Platform reproducibility for the Nickel Exposure Study. The diagrams show Lin's concordance correlation coefficient (CCC), Pearson's correlation coefficient (PCC), overall Lin's concordance correlation coefficient (OCCC), and coefficient of variation of replicates averaged across all glycans in array (CV). The CCC and PCC are computed for 28 combination pairs of 8 slides developed with benchmark sera. The OCCC and CV are computed independently for each array: the former is obtained by applying CCC to 66 combination pairs of 12 replicated subarrays from the same PGA. The first diagram (CCC) suggests marginally large bias between slides (20% trimmed mean across all combinations is 0.8), while the PCC shows that the inter-array concordance is satisfyingly good if we ignore the inter-array bias, which can be achieved with proper normalization. The last two slides show excellent intra-array concordance, i.e. small measurement error.

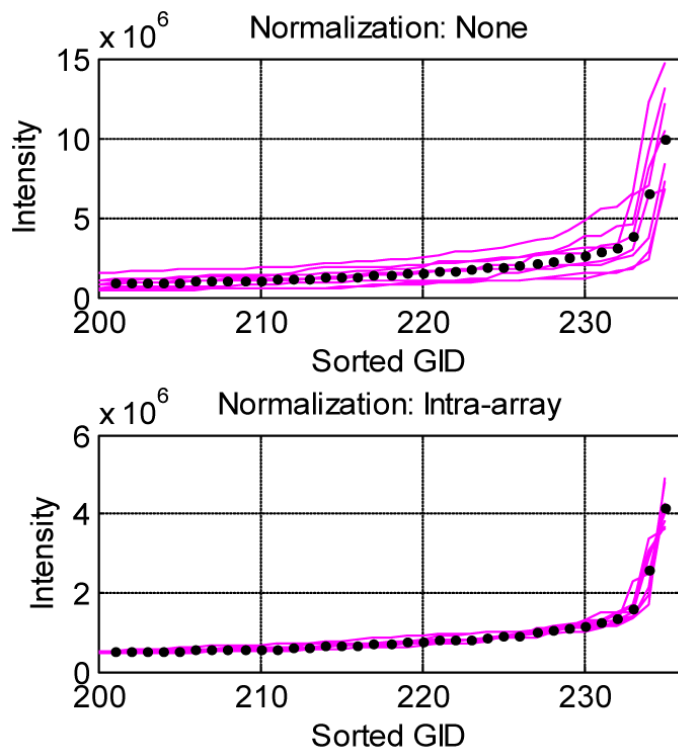


Figure 6.

Distribution of intensities for 8 arrays of benchmark sera before and after normalization. The distributions are presented as sorted medium summarized intensities across all glycans which have survived the screening of low intensity and low ICC glycans. In order to emphasize the differences, the diagram shows only 35 glycans which correspond to highest intensities, however similar differences can be observed with all other glycans. The upper diagram confirms the finding in figures 4 and 5 about the inter-array bias. The lower diagram shows the distributions after intra-array normalization. The quantile normalization as much stronger than inter-array normalization (not shown here) would make distributions of all subjects the same and equal to the mean value (represented by the black dotted line in both diagrams).

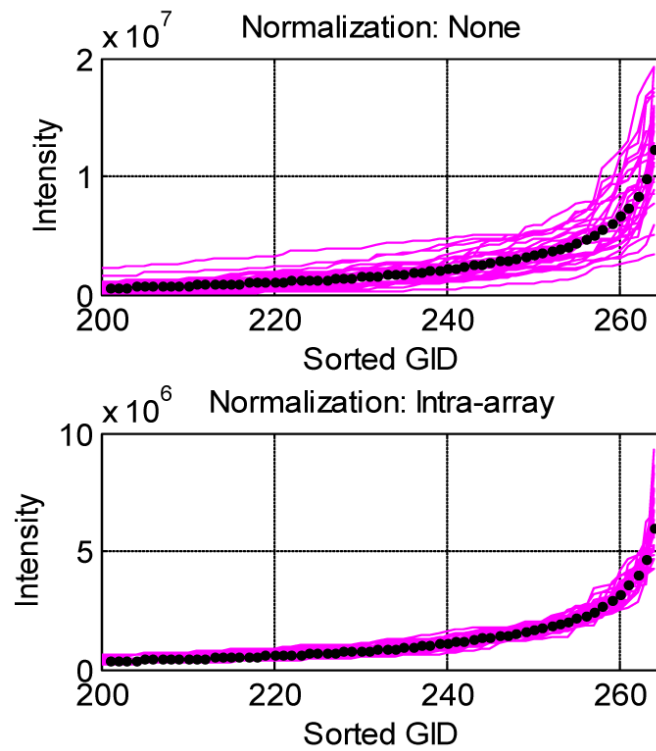


Figure 7. Distribution of intensities for subjects in the Nickel Exposure Study before and after normalization. The diagram is similar to the diagram in Figure 6, only it contains 89 arrays and the range of glycans is doubled. As seen, the intra-array normalization might not be entirely effective; therefore the quantile normalization is used in further discriminatory analysis.

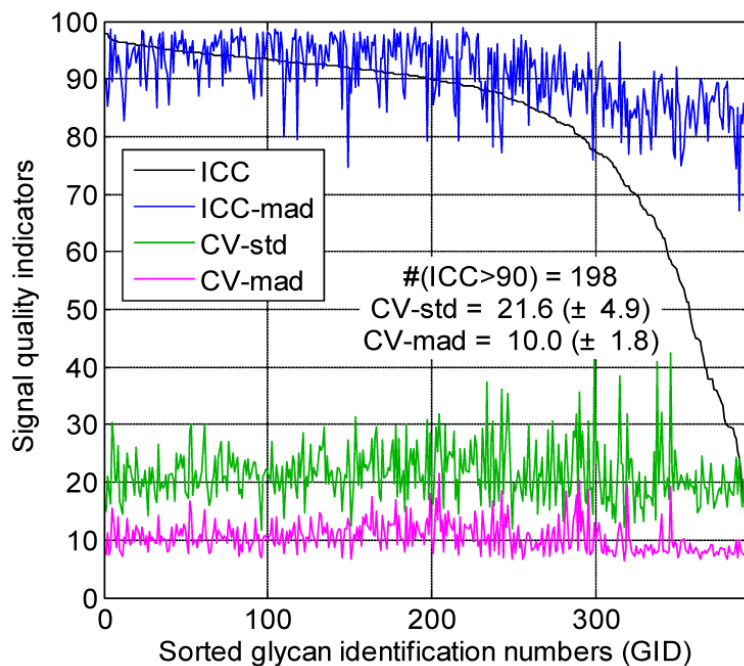


Figure 8.

Intra-class Correlation Coefficient (ICC, black line) estimated for the entire PGA library. There are two ICC curves: the traditional (black curve) and the robust ICC (blue curve), the latter is based on the medians and median absolute deviations as opposed to means and standard deviations. The traditional ICC values are sorted by descending values of ICC, while all other curves in figure are rearranged to reflect the same glycan ordering. As shown, 198 glycans have an ICC value greater than 90%. The diagram also shows the rearranged coefficient of variations: the traditional CV (CD-std, green) and the robust version of CV (CV-mad, magenta). The relatively high ICC values for most of the glycans, as well as the low CV values assure that data in the Nickel Exposure Study qualify for further discriminatory analysis as far as the technical noise is concerned.

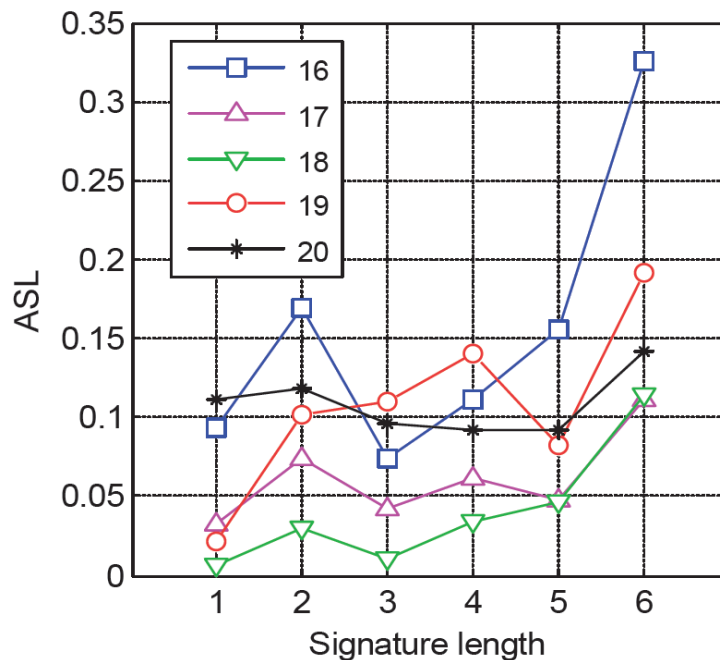


Figure 9.

Unbiased bootstrap tests performed on various sizes of low and high Urinary Ni samples ranging from 16 to 20, and various signature sizes ranging from 1 to 6. The statistic used for bootstrapping is combined AUC. The combined AUC is based on projection of features by multivariate logistic regression. The resampling was based on 500 permutations, as opposed to resampling with replacement, the former being more conservative. The criterion for selection of the best normalization method and the best sample cutoff value was the Achieved Significance Level of the test (ASL). The experiments were performed for several normalization methods but the diagram shows only the quantile normalization, which has delivered the best performance. As shown, the best result was obtained for cutoff value of 18, which yielded $ASL < 0.01$, thus giving very strong evidence against the null hypothesis. This cutoff value is used throughout this report.

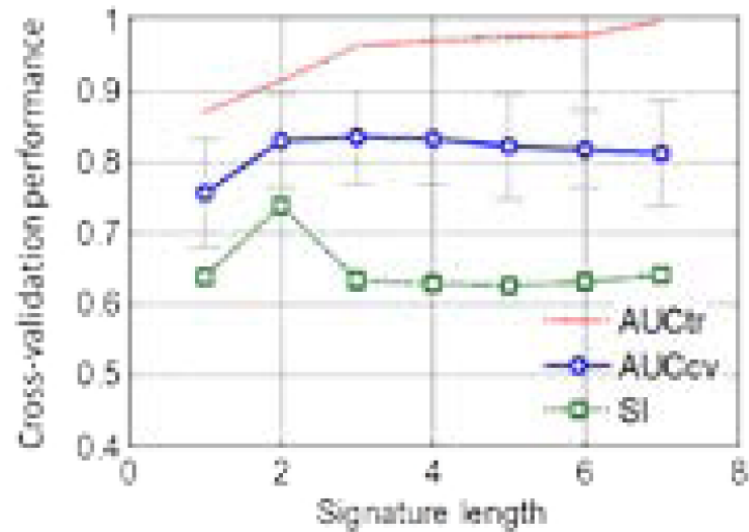


Figure 10.

Cross-validation of low and high urinary Nickel samples each having 18 observations. The diagram shows cross-validated and combined training AUC values for various signature set sizes ranging from 1 to 7. The combined AUC values were obtained by multivariate logistic regression. The algorithm is an unbiased, 100 times repeated, 10-fold cross-validation test. As shown, the optimal result is obtained for three glycans in the signature, resulting in cross-validated AUC = 0.836, while the training (observed) AUC value is 0.966. The diagram also shows the Kuncheva stability index (SI), which reaches the maximum at two glycans. The reason SI drops after two glycans is the fact that the third and fourth glycans (*GID* = 133 and 136, see figure 11) alternate in various cross-validation folds.

GID = 191 264 133 136 135 384 134 379 93 498 137 651 480 138
 Frq = 100 97 93 77 60 51 45 41 37 25 23 12 8 6

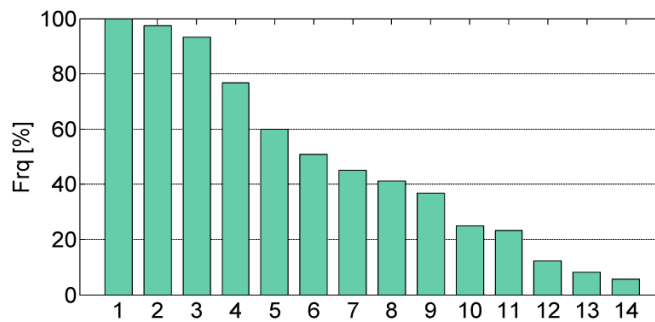


Figure 11. Feature count in 1000 feature selections (100 repetitions, 10 folds each repetition). The diagram shows that the glycan *GID* = 191 has been selected 100% of times, while the glycans 264 and 133 were selected 97% and 93% times, respectively. This diagram is another manifestation of feature selection stability.

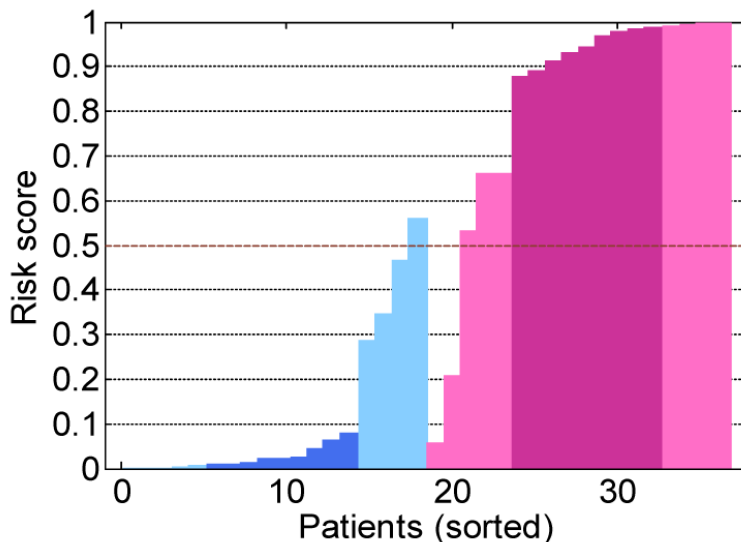


Figure 12.

Immunoruler diagram showing the training risk scores for 18 subjects with low ($4.44 \mu\text{g/L}$) and 18 subjects with high ($9.98 \mu\text{g/L}$) urinary Nickel. The risk scores are obtained by projecting the quantile normalized, median-summarized intensities which correspond to glycan signature $GID = 191, 264, 133$, using multivariate logistic regression. The projection bias is determined under the assumption of equal cost of false positive and false negative rates. In order to facilitate the interpretation of data, the scores are sorted in ascending order for each sample and colored accordingly: low Urinary Ni in blue (left bars), the high Urinary Ni in magenta (right bars). Bars with different color shades represent quartile regions. The bar intensities correspond to the probability of belonging to the high urinary Nickel group, given the training data. This kind of visualization explicitly shows the number of false positives $FP = 1$, false negatives $FN = 2$, true positives $TP = 16$, and true negatives $TN = 17$, all obtained using the cutoff value 0.5. Consequently, specificity is $S_p = 94.4\%$ and sensitivity $S_n = 88.9\%$. The training precision is 91.7% and the observed AUC value is 0.966.

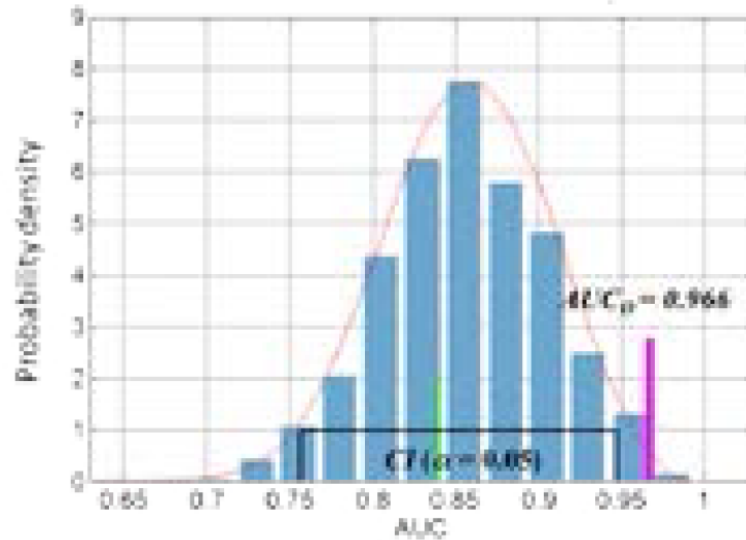


Figure 13.

The bootstrap test for low and high urinary Nickel samples (each of size 18). The test claims a strong statistical significance of the observed combined AUC value of 0.966. The bootstrap statistic used is combined AUC value obtained by multivariate logistic regression applied to three top selected glycans. Selection of glycans in each bootstrap iteration is performed by Wilcoxon ranking. The test is performed with 1000 resampling by permutation and the resulting test p-value is $ASL = 0.005$. The diagram shows the empirical distribution under the null hypothesis that the observed AUC value is no larger than any other replicated value. The null distribution has two-sided confidence interval $CI = [0.757, 0.947]$, or one-sided upper bound $CU = 0.935$. As shown, the observed value is beyond both confidence limits. The empirical data is fitted with a Generalized Extreme Value distribution, which has offered the same p-value as the count of replications above the observed value.

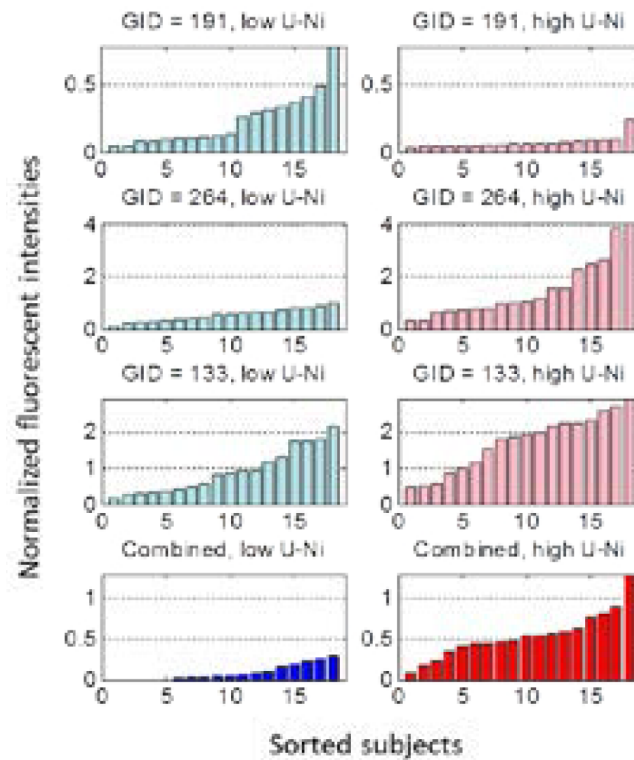


Figure 14.

Sorted intensities for 18 subjects with low Urinary Ni (diagrams at left, blue bars) and for 18 subjects with high Urinary Ni (diagrams at right, pink bars) for three selected discriminatory glycans. The intensities are obtained by quantile normalization of median summarized intensities. As shown, the intensities are decreased in subjects with high level of urinary Nickel for glycan *GID* = 191, while the effect is opposite for glycans *GID* = 264 and 133. The bar graphs at the bottom show intensities combined with logistic regression. The relative intensity scale factor for each diagram is 10^6 .

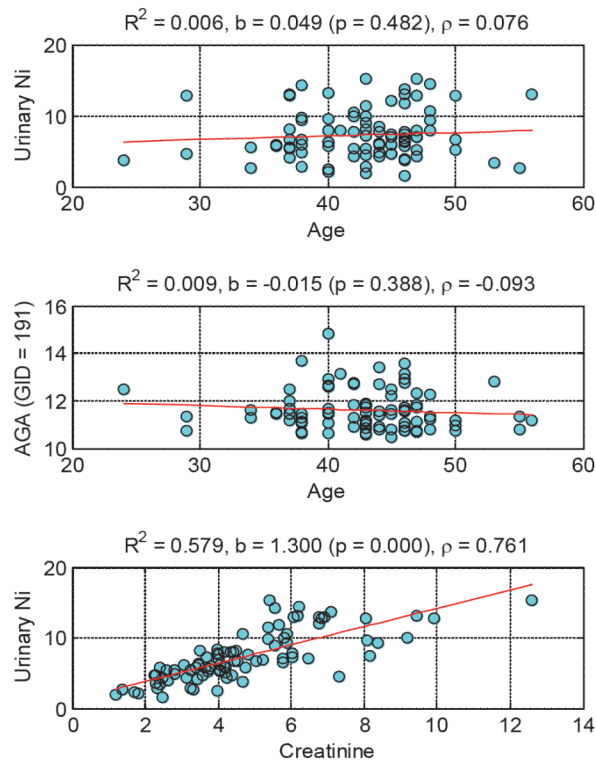


Figure 15.

Linear regression of Urinary Ni with age (top diagram), quantile-normalized and log transformed AGA for GID = 191 with the age (middle diagram), and Urinary Ni with creatinine (bottom diagram). The diagrams show R-squared values, regression coefficients with their p-values and the Pearson correlation coefficient. The low Pearson correlation coefficients in the first two diagrams indicate that there is no association of Urinary Ni or AGA with the age of subjects in this study. The relatively high correlation between Urinary Ni and creatinine suggests that these factors are potentially interchangeable in our analysis.

Table 1

Demographic Characteristics of Participating Subjects*.

	Zhangye Residents	Jinchang Residents	Ni Refinery Workers	P value**
<i>N</i>	29	30	30	
<i>Age (years)</i>	43.4 ± 4.9	41.6 ± 6.3	43.3 ± 4.9	> 0.05
<i>Smokers [n (%)]</i>	25 (86.2)	25 (83.3)	25 (83.3)	> 0.05
<i>Urinary Ni μg/L</i>	6.83 ± 3.53	6.55 ± 3.51	8.43 ± 3.22	0.0246
<i>μg/g creatinine</i>	4.15 ± 1.52	4.13 ± 2.05	5.79 ± 2.08	0.0002

* all participating subjects were male

** comparison between Ni refinery workers and control subjects (residents in Zhangye and Jinchang)

Table 2

Wilcoxon-Mann-Whitney two-sample rank sum test applied to screened, quantile-normalized median summarized data from the Nickel Exposure Study. The samples contain 18 subjects with high (9.98 $\mu\text{g/L}$) and 18 subjects with low (4.44 $\mu\text{g/L}$) level of urinary Nickel. The meaning of columns are as follows: *GID* – glycan identification number, *Z* – z-statistic, *p* – p-value of the test, *FDR* – false discovery rate, *AUC* – area under the ROC curve, *AUC_C* – cumulative AUC value obtained by combining the above glycans through multivariate logistic regression, *ICC* – corresponding Intra-class Correlation Coefficient computed for raw data. The sign of the z-statistic indicates downregulation (negative sign) or upregulation (positive sign) of normalized signals. The low values of *FDR*, at least for top three glycans, imply a good confidence in the results. The glycans in the table are sorted by ascending order of the p-value. The sixth column suggests that combining several glycans can considerably increase the AUC value. For example, combining three top glycans: *GID* = 191, 264, 133, gives *AUC_C* = 0.966. Figure 13 shows a solid statistical significance for this AUC value.

GID	Z	p	FDR	AUC	AUC_C	ICC
191	3.797	0.00015	0.0181	0.8704	0.8704	93.3
264	-3.497	0.00047	0.0353	0.8457	0.9167	89.1
133	-3.149	0.00164	0.0584	0.8148	0.9660	88.7
136	-3.058	0.00223	0.0602	0.8117	0.9722	82.2
135	-2.959	0.00309	0.0756	0.7932	0.9784	93.3
384	-2.927	0.00342	0.1490	0.7901	0.9815	94.5
134	-2.897	0.00377	0.0668	0.7901	1.0000	81.7
93	-2.864	0.00418	0.0629	0.7840	1.0000	91.5
379	-2.864	0.00419	0.1186	0.7840	1.0000	94.5
137	-2.792	0.00524	0.0807	0.7932	1.0000	88.9

Table 3

Structures of glycans from Table 2.

GID	Glycan Structure	Generic Name
191	6-O-P-Gal β 1-4GlcNAc β	6'P-LacNAc
264	Gal β 1-4Gal β 1-4GlcNAc β	Gal β 4LacNAc
133	Gal β 1-4Glc β -NH \leftarrow Gly \leftarrow Ala	Lactose-Gly-Ala
136	Gal β 1-4Glc β -NH \leftarrow Gly \leftarrow Ile	Lactose-Gly-Ile
135	Gal β 1-4Glc β -NH \leftarrow Gly \leftarrow Asn	Lactose-Gly-Asn
384	Gal β 1-4GlcNAc β 1-3Gal β 1-4GlcNAc β	LacNAc β 3LacNAc
134	Gal β 1-4Glc β -NH \leftarrow Gly \leftarrow Arg	Lactose-Gly-Arg
93	Gal β 1-4Glc β -NH \leftarrow Gly	Lactose-Gly
379	Gal β 1-3GlcNAc β 1-3Gal β 1-4GlcNAc β	LewisC β 3LacNAc
137	Gal β 1-4Glc β -NH \leftarrow Gly \leftarrow Nle	Lactose-Gly-Nle

6'P-LacNAc: 6-phosphate N-acetyllactosamine

Gal β 4LacNAc: Galactose β 1-4N-acetyllactosamine

sp4: glycine

aa: aminoacid

LacNAc β 3LacNAc: N-acetyllactosamine β 1-3N-acetyllactosamine

Lewis^C β 3LacNAc: Lewis^C β 1-3N-acetyllactosamine