

## Development of a Data-mining System for Differential Profiling of Cell Glycoproteins Based on Lectin Microarray

Atsushi Kuno<sup>1#</sup>, Yoko Itakura<sup>1#</sup>, Masashi Toyoda<sup>2</sup>, Yoriko Takahashi<sup>3</sup>, Masao Yamada<sup>4</sup>, Akihiro Umezawa<sup>2</sup>, and Jun Hirabayashi<sup>1\*</sup>

<sup>1</sup>Research Center for Medical Glycoscience (RCMG), National Institute of Advanced Industrial Science and Technology (AIST), 1-1-1, Umezono, Tsukuba, Ibaraki 305-8568, Japan

<sup>2</sup>Department of Reproductive Biology and Pathology, National Research Institute for Child Health and Development, 2-10-1, Okura, Setagaya, Tokyo, 157-8535, Japan

<sup>3</sup>Bioscience Group, Mitsui Knowledge Industry Co., Ltd., Hitotsubashi SI bldg., 3-26, Kandanishikicho, Chiyoda-ku, Tokyo 101-0054, Japan

<sup>4</sup>Glycomics Research Laboratory, Moritex Corporation, 1-3-3, Azamino-Minami, Aoba-ku, Yokohama City, Kanagawa 225-0012, Japan

# These authors contributed equally to this study

\*Corresponding author: Jun Hirabayashi, Research Center for Medical Glycoscience, National Institute of Advanced Industrial Science and Technology, AIST Tsukuba Central 2, 1-1-1, Umezono, Tsukuba, Ibaraki 305-8568, Japan, Tel: +81-29-861-3124; Fax: +81-29-861-3125; E-mail: jun-hirabayashi@aist.go.jp

Received April 15, 2008; Accepted May 14, 2008; Published May 20, 2008

**Citation:** Atsushi K, Yoko I, Masashi T, Yoriko T, Masao Y, et al. (2008) Development of a data-mining system for differential profiling of cell glycoproteins based on lectin microarray. *J Proteomics Bioinform* 1: 068-072. doi:10.4172/jpb.1000011

**Copyright:** © 2008 Atsushi K, 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.

### Abstract

Lectin microarray is an emerging technique enabling multiplex glycan profiling in a direct, rapid and sensitive manner. So far, there has been no robust system available for efficient data-mining to realize differential profiling, which is an effective approach to biomarker investigation. In the present paper, we describe a practical strategy for proteomics-based glycan-related biomarker discovery, with an example of mice embryonal carcinoma and embryonic stem cells and their differentiated forms with retinoic acid. Data were processed by the microarray system using a max-normalization procedure after a gain-merging process, followed by principal component analysis.

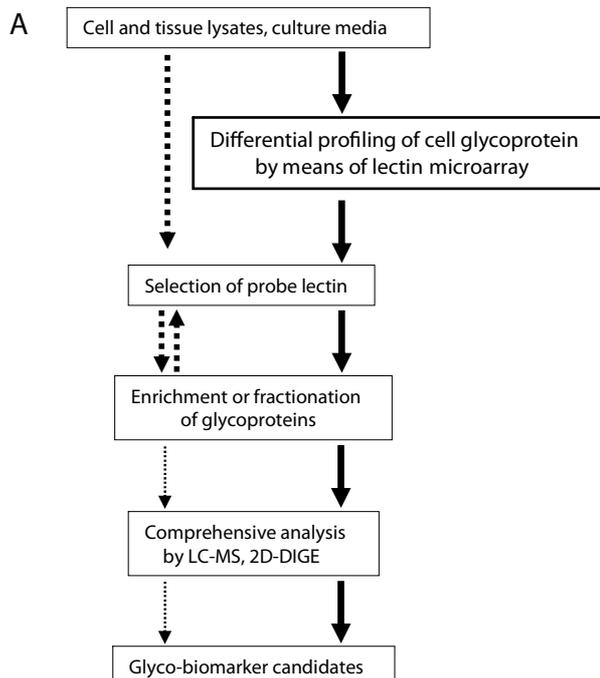
**Keywords:** Differential glycan profiling; Biomarker discovery; Lectin microarray; Principal component analysis

### Abbreviations

EC cells: Embryonal Carcinoma cells; ES cells: Embryonic Stem cells; MS: Mass Spectrometry; PCA: Principal Component Analysis; TBSTx: Tris-buffered Saline containing 0.1% Triton X-100.

### Introduction

Cell surface dynamics are characterized by altered glycosylation in the development and differentiation stages. Drastic glycosylation change has also been proposed for tumor progression and metastasis. For instance, cell surface sialylation and 1-6 branching of N-linked oligosaccharides are strongly correlated with differentiation of embryonal carcinoma cells and metastatic potential of cancer cells (Dennis et al., 1982; 1987; Heffernan et al., 1993). Therefore, it is highly likely that finding of novel cell differentiation-related or tumor-specific glycoproteins with significant structural changes will become reliable biomarkers. From these points of view, proteomics-based biomarker discoveries have now been complemented by extensive glyco-technologies, such as chemical capturing targeting N-linked glycoproteins (Zhang et al., 2003; Nishimura et al., 2005) and affinity capturing with the use of various glycan-binding proteins, i.e., lectins (dashed arrows in Fig. 1A) (Kaji et al., 2003).



**Figure 1:** (A) A proposed strategy for an alternative proteomics-based glyco-biomarker discovery with differential glycan profiling (*bold arrows*). An optimal set of lectins was systematically determined following a lectin microarray-based data-mining procedure. In the conventional strategy (*dashed arrows*), such a lectin set must be selected based on previous knowledge or repeated trial-and-error experiments.

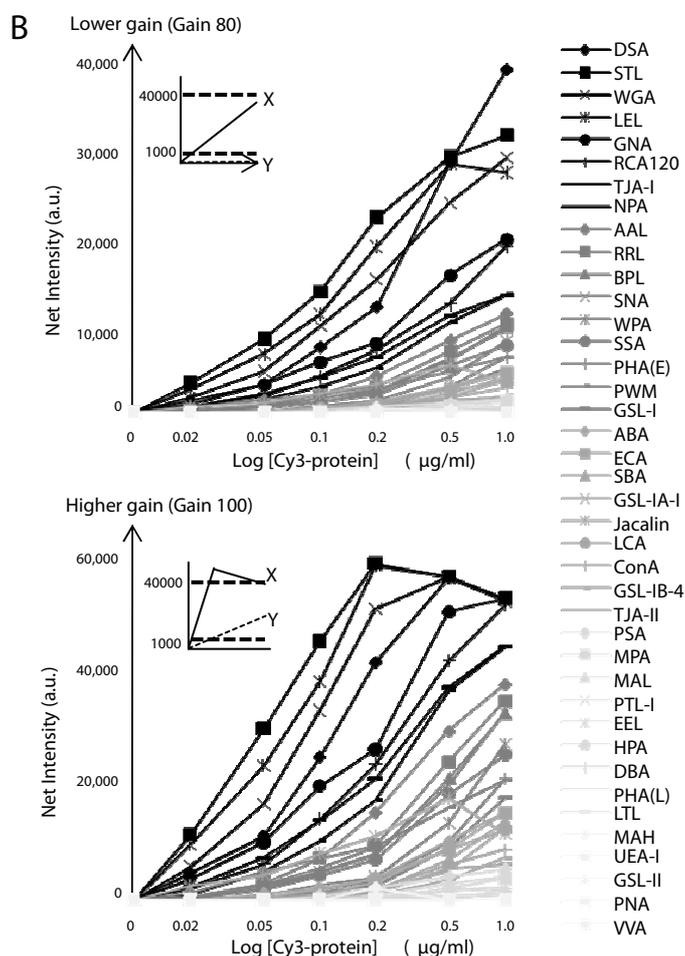
One of the successful reports involving the concept of glycoproteomics includes the discovery of GP73, a novel glycoprotein discovered as a serological biomarker candidate for liver cancer (Block et al., 2005; Drake et al., 2006). Traditionally, serial lectin affinity chromatography (Cummings et al., 1982) has been a procedure for enrichment of particular glycoproteins with a target glycan structure of either N- or O-glycosylation (Madera et al., 2005; Qiu et al., 2005). In this case, selection of a highly-effective set of lectins is essential for success in the biomarker discovery (*dashed arrows* in Fig. 1A). If a systematic data-mining procedure which follows differential glycan analysis were to be available, it would facilitate the design of an optimal set of lectins (*bold arrows* in Fig. 1A).

Lectin microarray is an emerging technology enabling an ultra-sensitive measuring of multiplex lectin-glycan interaction analysis (Angeloni et al., 2005; Pilobello et al., 2005; Kuno et al., 2005). Taking advantage of the merits of this technology, i.e., sensitive detection and simple manipulation, an increasing number of studies using lectin microarray report that cell-surface glycans are closely associated with the functions, states and relation to diseases of individual cells (Ebe et al., 2006; Pilobello et al., 2007; Tatenno et al., 2007). Among biological interests in glycans, a current trend is the focus on glycan-related biomarkers. However, there is no established strategy and optimized protocols for cell glycoprotein profiling, in particular regarding data-mining procedures. In this study, we describe logistic processes for differential cell glycoprotein profiling including data-mining as an alternative approach to conventional proteomics-based biomarker discovery. Key points of the strategy for cell glycoprotein profiling include: (1) fitting the protein concentrations in the appropriate range between 0.2 and 0.5  $\mu\text{g/ml}$  to obtain robust and reproducible signal patterns, (2) a gain-merging technique to expand the dynamic range of the lectin-glycan interaction signals, and (3) the max-normalization procedure using the merged data for normalization. The data thus processed were found to be useful for systematic determination of the best set of lectins among more than 40 probe candidate lectins immobilized on the microarray (*bold arrows* in Fig. 1A). A model study focused on regenerative medicine is described for mice embryonal carcinoma and embryonic stem cells as well as their differentiated forms with retinoic acid.

## Results and Discussion

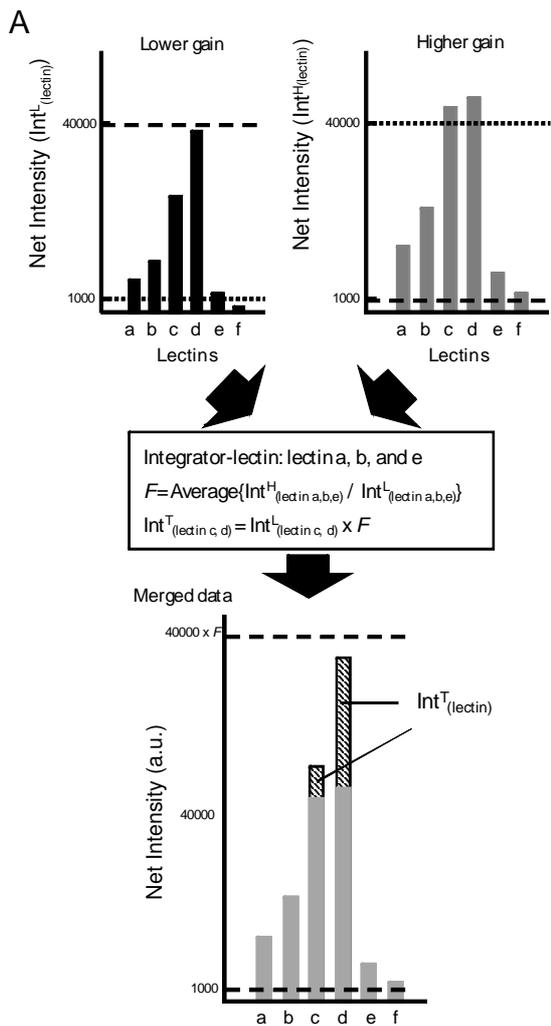
**Optimization of lectin microarray manipulations.** For improved proteomics-based biomarker discovery, cell glycoproteins are proposed as targets. Glycosylation change is analyzed by a high-sensitivity, robust, and reproducible method using lectin microarray, if the evanescent-field fluorescence-assisted detection method is adopted. However, the previous protocol for cell glycoprotein analysis has not fulfilled the recent requirements for detailed cell profiling and biomarker discovery (Ebe et al., 2006). To address these issues, we first established a strict protocol for differential analysis of cell glycoproteins using mouse embryonal carcinoma cells (mouse teratocarcinoma cell line F9) as a model. The analyte (i.e., glycoprotein) was focused on hydrophobic, raft-associated membrane-bound proteins isolated using a CellLytic MEM Protein Extraction kit (Sigma, St. Louis, MO), because we found the proteins to be analyzed showed the highest signal-to-noise ratio. A small aliquot of the obtained protein (200 ng from approximately  $1 \times 10^3$  cells) was labeled with Cy3-succinidyl ester (designated as Cy3-labeled glycoprotein). Various concentrations of the Cy3-

labeled glycoprotein solution (60  $\mu\text{l}$ , 0.02~1.0  $\mu\text{g/ml}$ ) were then subjected to the lectin microarray analysis. Due to the specificity of the CCD camera, a gain value should be set so that the observed fluorescence intensities of almost all positive-spots on the glass slide fall within the range 1,000 to 40,000, which provides a dynamic range with sufficient linearity. Each glass slide was successively scanned under different gain conditions. A dose-dependent increment of signal intensity was observed on most of the positive-spots (**Fig. 1B**). However, we could not confirm satisfactory linearity for all of the spots under a single gain condition. For instance, the signals of some positive-spots (e.g., GSL-I, ECA, SBA, LCA, ConA, TJA-II, and PSA) were kept below 1,000 under the lower gain (80) condition as shown in the *top* of **Fig. 1B**. Under the higher gain (100) condition, the intensities of four lectins (DSA, STL, WGA, and LEL) were above the upper limit 40,000, at protein concentrations of 0.2  $\mu\text{g/ml}$  or more (the *bottom* of **Fig. 1B**). Such uneven linearity could cause inappropriate interpretation of the data. A useful data-optimization procedure needed to be introduced to solve this basic problem.



**Figure 1:** (B) Quantitative analysis of lectin-glycoprotein interaction. Various concentrations of Cy3-labeled glycoproteins (0.02~1.0  $\mu\text{g/ml}$ ) were subjected to lectin microarray analysis. After the interaction reaction, each glass slide was successively scanned under different gain conditions (gain 80 and 100). Dose-dependent fluorescent signals are observed except for some saturated signals under the higher gain condition.

**Data-processing by gain-merging and max-normalization:** Provided the intensities of all positive-spots are kept within the acceptable dynamic range (1,000 to 40,000), signal patterns of each analyte should be theoretically the same even under different gain conditions, i.e., higher gain intensity ( $\text{Int}^H$ ) over lower gain



**Figure 2.** (A) Principle of the gain-merging procedure to expand the dynamic range described in the Results and Discussion.

intensity ( $Int^L_i$ ) ratios for lectin  $i$  should be almost the same value. To ensure high-reproducibility, the dynamic range was expanded by a “gain-merging” procedure. An outline of the procedure (Fig. 2A) is as follows: a slide glass is scanned under two different gain conditions; *higher gain* to “rescue” weak signals (e.g., lectin f in Fig. 2A) below 1,000 ( $Int^H_{(lectin f)}$ ) and *lower gain* to “suppress” excessively strong signals (e.g., lectin d) over 40,000 ( $Int^L_{(lectin d)}$ ). At this point, selection of appropriate “merging”-lectins is important (lectins a, b, and e in the case of Fig. 2A)

the signal intensities of which fall within the range 1,000 to 40,000 under both higher and lower gain conditions. With these selected merging lectins, a “Factor ( $F$ )” is determined as the average of higher/lower ratios calculated for individual merging lectins by eq

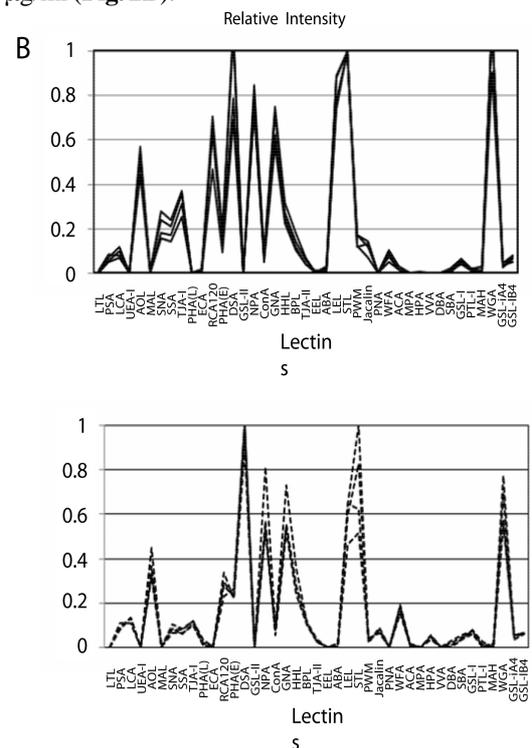
$$(1). \quad F = Average (Int^H_i / Int^L_i) \dots eq (1)$$

The gain-merging procedure is completed by replacement of the over-range intensities ( $>40,000$ ) obtained under the higher gain condition (e.g.,  $Int^H_{(lectin c)}$ ) with theoretical intensities ( $Int^T_{(lectin c)}$ ) by eq (2).  $Int^T_{(lectin c)} = Int^L_{(lectin c)} \times F \dots eq(2)$

For other lectins with no over-range under the higher gain condition, signal intensities obtained under the higher gain condition are used with no modification. During this process, all the resultant intensities of positive-spots were within the expanded dy-

namic range, from 1,000 to 40,000  $\times F$ . When 1.0  $\mu\text{g/ml}$  of F9 cell proteins were subjected to analysis (Fig. 1B), all 34 positive lectins fell within the merged dynamic range (1,000~132,000) after the gain-merging procedure ( $F=3.3$ ), whereas 85% (29 lectins under the lower gain (80) condition) and 76% (26 lectins under the higher gain (100) condition) of positive lectins were within the original dynamic range (1,000~40,000), respectively.

Using the merged data, a normalization procedure was developed to simplify and stabilize the subsequent differential glycoprotein analysis. Considering the difficulty in selecting a universal lectin, to assure the same level of signal intensities, we selected a practical procedure to calculate the relative intensity in comparison with the strongest intensity among the positive-spots under the given conditions, i.e., max-normalization. The max-normalized data of F9 cells thus processed gave similar signal patterns provided that protein concentrations were maintained within the range 0.2 to 0.5  $\mu\text{g/ml}$  (Fig. 2B).

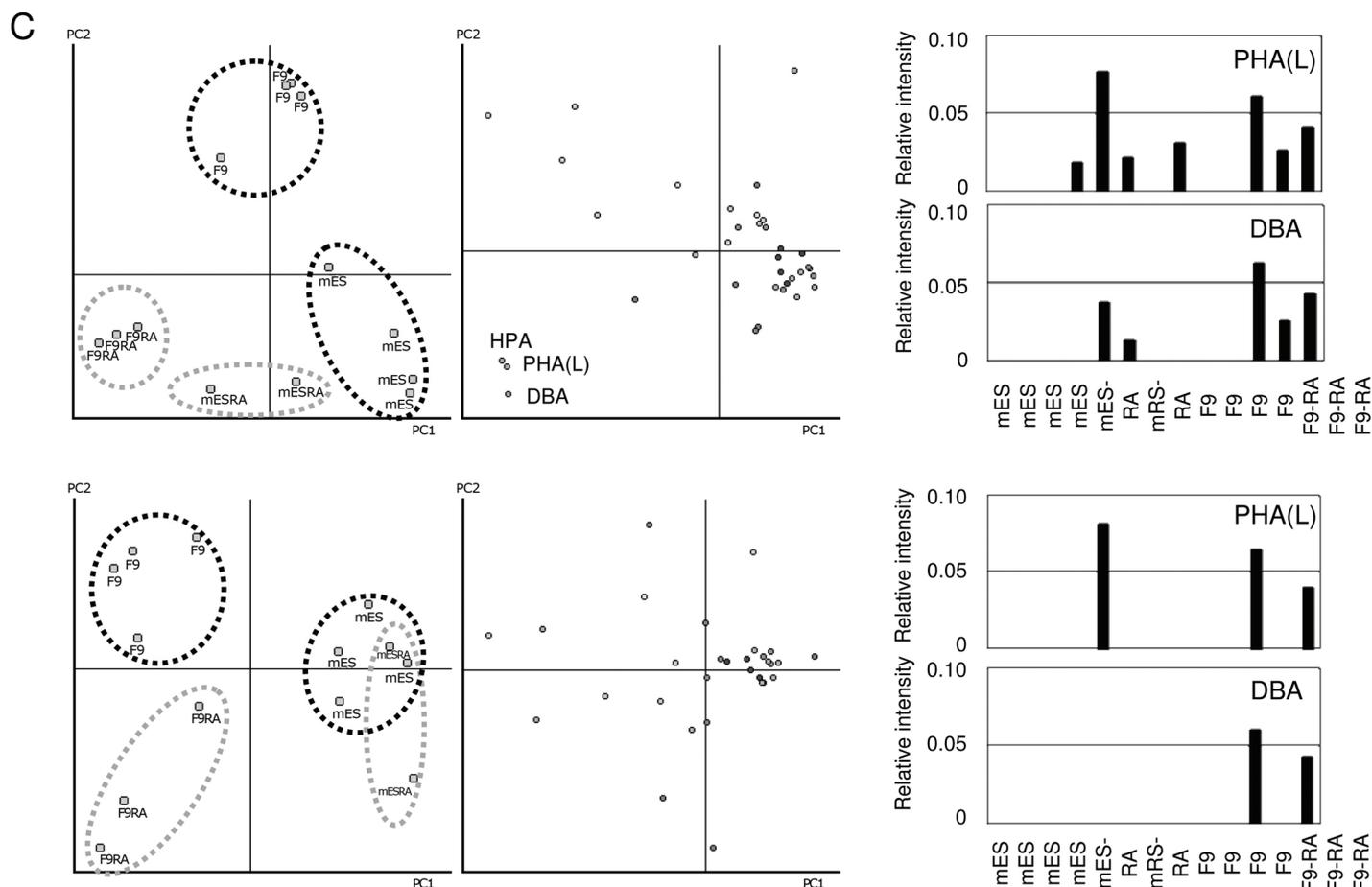


(B) Relative fluorescence intensities of 41 lectins with various concentrations (0.2, 0.3, 0.4, and 0.5  $\mu\text{g/ml}$ ) of proteins extracted from F9 (*plane lines*) and F9-RA (*dashed lines*). Relative intensities were calculated in comparison with the strongest intensity among the positive-spots under the given conditions, i.e., max-normalization.

A similar observation has also been made for the differentiated forms with retinoic acid (F9-RA) (Fig. 2B). These results suggest that the procedure of max-normalization following gain-merging contribute to the establishment of high-reproducible cell glycoprotein profiling with extremely simple and systematic manipulations.

**Principal component analysis:** We next examined whether or not a statistical analysis of the data could actually determine the best set of lectins, which should be useful for an efficient enrichment of relevant glycoproteins associated with glycosylation change induced by retinoic acid treatment. For this purpose, principal component analysis (PCA) using a web-based NIA array analysis tool (<http://lgsun.grc.nia.nih.gov/ANOVA/>; Chapman et al., 2001; Sharov et al., 2005), was chosen and applied to the above processed lectin microarray data of F9 cells (four different preparations) as well as F9-RA (three different preparations). For the sake of comparison, we also analyzed mouse embryonic stem cells (mES)

(four different preparations) and their differentiated forms (mES-RA) (two different preparations). The lectin microarray data processed according to the developed procedures gave two principal components (PCs). The 2D-biplot format thus obtained clearly divided the above 13 preparations into four independent clusters; i.e., F9, F9-RA, mES and mES-RA (the upper left of Fig. 2C). The result also revealed



(C) 2D-biplot representation as a result of principal component analysis with gain-merging processing (upper). The data obtained for F9 and mouse embryonic stem cells were processed in comparison with those obtained for their retinoic acid-induced forms. Glycan alterations associated with cell line difference and differentiation induced by retinoic acid are depicted by PC1 and PC2, respectively (left). Lectins that showed dynamic enhancement with retinoic acid treatment were systematically selected as those showing strong double negative-correlation with respect to PC1 and PC2. Relative intensities of the two lectins thus selected (DBA and PHA(L)) toward glycoproteins from F9, mES and their differentiated forms with retinoic acid are represented by bar graphs (right). These data are compared with the principle component analysis using the raw data set without gain-merging processing (bottom).

double negative-correlation with the PC1 and PC2, i.e., signal enhancement with retinoic acid, for three probe lectins ( $\alpha$ GalNAc binders, DBA and HPA, and  $\beta$ 1-6 branching binder, PHA(L); the upper left of Fig. 2C). Importantly, the normalized intensities of these lectins were relatively low (i.e., 0~0.03; Fig. 2B), which the method could have failed to detect without the use of the rescue process using the gain-merging procedure (<1,000 under the lower gain condition) (see the PCA of the raw data without gain-merging processing in the bottom of Fig. 2C). This observation clearly indicates a practical merit of such a data-mining procedure for the investigation of novel glycan-related biomarkers, which are expected to be fairly minor components in clinical samples.

## Conclusion

A lectin microarray-based data-mining system for differential profiling of cell glycoproteins has been developed by adopting max-normalization following gain-merging processes. This highly-reproducible analysis with simple and systematic manipulations should provide the basis of a robust and logistic strategy for the discovery of proteomics-based glycan-related biomarkers.

## Acknowledgments

We thank N. Uchiyama, Y. Kubo, and J. Murakami for supplying the lectin microarray. We also thank A. Matsuda for critical discussion concerning the preparation of protein solution. This work was supported in part by a grant for New Energy and Industrial Technology Development Organization (NEDO) in Japan.

## References

1. Angeloni S, Ridet JL, Kusy N, Gao H, Crevoisier F, et al. (2005) Glycoprofiling with micro-arrays of glycoconjugates and lectins. *Glycobiology* 15: 31-41. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
2. Block TM, Comunale MA, Lowman M, Steel LF, Romano PR, et al. (2005) Use of targeted glycoproteomics to identify serum glycoproteins that correlate with liver cancer in woodchucks and humans. *Proc Natl Acad Sci USA* 102: 779-784. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
3. Chapman S, Schenk P, Kazan K, Manners J (2001) Using biplots to interpret gene expression patterns in plants. *Bioinformatics* 18: 202-204. » [CrossRef](#) » [Pubmed](#)
4. Cummings R, Kornfeld S (1982) Fractionation of asparagine-linked oligosaccharides by serial lectin-agarose affinity chromatography. *J Biol Chem* 257: 11235-11240. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
5. Dennis JW, Waller C, Timpl R, Schirmacher V (1982) Surface sialic acid reduces attachment of metastatic tumour cells to collagen type IV and fibronectin. *Nature* 300: 274-276. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
6. Dennis JW, Laferte S, Waghorne C, Breitman ML, Kerbel RS (1987) Beta1-6 branching of Asn-linked oligosaccharides is directly associated with metastasis. *Science* 236: 582-585. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
7. Drake RR, Schwegler EE, Malik G, Diaz J, Block T, et al. (2006) Lectin capture strategies combined with mass spectrometry for the discovery of serum glycoprotein biomarkers. *Mol Cell Proteomics* 5: 1957-1967. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
8. Ebe Y, Kuno A, Uchiyama N, Koseki-Kuno S, Yamada M, et al. (2006) Application of lectin microarray to crude samples: differential glycan profiling of Lec mutants. *J Biochem* 139: 323-327. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
9. Heffernan M, Lotan R, Amos B, Palcic M, Takano R, et al. (1993) Branching  $\beta$ 1-6N-acetylglucosaminyltransferases and polyactosamine expression in mouse F9 teratocarcinoma cells and differentiated counterparts. *J Biol Chem* 268: 1242-1251. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
10. Kaji H, Saito H, Yamauchi Y, Shinkawa T, Taoka M, et al. (2003) Lectin affinity capture, isotope-coded tagging and mass spectrometry to identify N-linked glycoproteins. *Nat Biotech* 21: 667-672. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
11. Kuno A, Uchiyama N, Koseki-Kuno S, Ebe Y, Takashima S, et al. (2005) Evanescent-field fluorescence-assisted lectin microarray: a new strategy for glycan profiling. *Nat Methods* 2: 851-856. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
12. Madera M, Mechref Y, Novotny MV (2005) Combining lectin microcolumns with high-resolution separation techniques for enrichment of glycoproteins and glycopeptides. *Anal Chem* 77: 4081-4090. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
13. Nishimura SI, Niikura K, Kuroguchi M, Matsushita T, Fumoto M, et al. (2005) High-throughput protein glycomics: Combined use of chemoselective glycoblotting and MALDI-TOF/TOF mass spectrometry. *Angew. Chem Int Ed* 44: 91-96.
14. Pilobello KT, Krishnamoorthy L, Slawek D, Mahal LK (2005) Development of a lectin microarray for the rapid analysis of protein glycopatterns. *Chem Bio Chem* 6: 985-989. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
15. Pilobello KT, Slawek DE, Mahal LK (2007) A ratiometric lectin microarray approach to analysis of the dynamic mammalian glycome. *Proc Natl Acad Sci USA* 104: 11534-11539. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
16. Qiu R, Regnier FE (2005) Use of multidimensional lectin affinity chromatography in differential glycoproteomics. *Anal Chem* 77: 2802-2809. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
17. Sharov AA, Dudekula DB, Ko MSH (2005) A web-based tool for principal component and significance analysis of microarray data. *Bioinformatics* 21: 2548-2549. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
18. Tateno H, Uchiyama N, Kuno A, Togayachi A, Sato T, et al. (2007) A novel strategy for mammalian cell surface glycome profiling using lectin microarray. *Glycobiology* 17: 1138-1146. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
19. Zhang H, Li XJ, Martin DB, Aebersold R (2003) Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry. *Nat Biotech* 21: 660-666. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)