JCSE Journal of Computing Science and Engineering

Special Issue
Journal of Computing Science and Engineering,
Vol. 11, No. 4, December 2017, pp. 111-120

Identifying Post-translational Modification Crosstalks for Breast Cancer

Chi-Hua Tung
Department of Bioinformatics, Chung-Hua University, Hsinchu City, Taiwan **chihua.tung@chu.edu.tw**

Pei-Wei Shueng
Division of Radiation Oncology, Far Eastern Memorial Hospital, New Taipei City, Taiwan;
Faculty of Medicine, School of Medicine, National Yang-Ming University, Taipei, Taiwan;
Department of Radiation Oncology, Tri-Service General Hospital, National Defense Medical Center, Taipei, Taiwan
shuengsir@gmail.com

Yen-Wei Chu*
Institute of Genomics and Bioinformatics, Biotechnology Center, and Agricultural Biotechnology Center, Taichung City, Taiwan;
Institute of Molecular Biology and Graduate Institute of Biotechnology, National Chung Hsing University, Taichung City, Taiwan
ywchu@nchu.edu.tw

Chi-Wei Chen
Institute of Genomics and Bioinformatics, Taichung City, Taiwan;
Department of Computer Science and Engineering, National Chung Hsing University, Taichung City, Taiwan
d103056006@mail.nchu.edu.tw

Chian-Ying Chen
Institute of Genomics and Bioinformatics, Taichung City, Taiwan **csichocolate@gmail.com**

Abstract
Post-translational modifications (PTMs) of proteins play substantial roles in the gene regulation of cell physiological functions and in the generation of major diseases. However, the majority of existing studies only explored a certain PTM of proteins, while very few have investigated the PTMs of two or more domains and the effects of their interactions. In this study, after collecting data regarding a large number of breast cancer-related and validated PTMs, a sequence and domain analysis of breast cancer proteins was carried out using bioinformatics methods. Then, protein-protein interaction network-related tools were applied in order to determine the crosstalks between the PTMs of the proteins. Finally, statistical and functional analyses were conducted to identify more modification sites of domains and proteins that may interact with at least two or more PTMs. In addition to exploring the associations between the interactive effects of PTMs, the present study also provides important information that would allow biologists to further explore the regulatory pathways of biological functions and related diseases.

Category: Bioinformatics

Keywords: Post-translational modification; Crosstalk; Sequence analysis; Breast cancer

Open Access **http://dx.doi.org/10.5626/JCSE.2017.11.4.111** **http://jcse.kiise.org**

Received 25 May 2017; **Accepted** 25 November 2017
*Corresponding Author

 pISSN: 1976-4677 eISSN: 2093-8020

I. INTRODUCTION

Breast cancer is a malignant disease, and the metastasis of breast cancer to other organs occurs with relative ease. The occurrence of the cancer itself is likely due to random genetic mutations or genetic defects [1, 2]. In Taiwan, breast cancer is the type of cancer with the peak age range of initial diagnosis occurring from 45–69 years old. More specifically, approximately 9,600 women are diagnosed with breast cancer each year, with 1,900 dying from the disease per year. That equates to roughly 26 women being diagnosed with and five women dying of breast cancer per day. In short, breast cancer is among the most common cancers among women in Taiwan, while also being quite common elsewhere. For example, an average of one in every nine women in the United States and Europe will suffer from breast cancer at some point in their lives. In light of these impacts, researchers and clinicians have, in recent years, sought increasingly better understanding of the internal mechanisms and biological regulatory pathways of breast cancer and other major diseases, in addition to researching and applying effective prevention and treatment methods.

A post-translational modification (PTM) is a significant step in the production of proteins. The significance of PTMs lies in their capacity to alter the physical and chemical properties of proteins and their functions, including their folding, stability, and activity. Furthermore, most PTMs are reversible, which is why PTMs can regulate the physiological state of a cell. Examples of PTMS include G protein phosphorylation, the modification of auxiliary enzymes factors, and histone acetylation or methylation. At the same time, there are also some irreversible PTMs, such as cleavage and ubiquitination. Meanwhile, the molecular mechanisms of PTMs are extremely complex in terms of their effects, and such effects constitute one of the important research topics in the field of proteomics at present. The PTMs of proteins play significant roles in the physiological and biological characteristics and functions of those proteins, including the transmission of biological signals, gene expression-related effects, protein and cellular stability, and genetic transcription, among others.

Phosphorylation [3-5] is not only the most common reversible PTM in cells but also plays a role in a variety of biological signaling pathways. Fundamentally, protein phosphorylation consists of the addition of a phosphoric acid molecule (PO4) group or groups to a protein. The sites for such phosphorylation usually consist of tyrosine, threonine, and serine. Some protein phosphorylation will generate structural changes and increased chemical activity. In terms of biological signaling pathways, protein phosphorylation and dephosphorylation supply the regulatory mechanism for eukaryotic cells. In addition, phosphorylation controls the activation and inhibition of many enzymes and receptors [4, 6].

Glycosylation [3, 4, 7] can be categorized under various subtypes, including N-glycosylation, which primarily involves aspartic acid, and O-glycosylation, which primarily involves serine and threonine. Glycosylation is widely found in the extracellular matrix and affects a variety of immune-related proteins, such as selectin and lectins, among others. Moreover, many of the receptors of cells and microorganisms are composed of glycoproteins, such as the antigens of red corpuscles, which consist of glycoproteins composed of fucose and galactose. Glycosylation can affect protein folding, thereby affecting whether the affected protein is able to serve its function or not. According to previous research, some proteins will degrade faster than other normal proteins if they experience a relative lack of glycosylation [4, 6, 8].

Acetylation [9] is a PTM which occurs primarily on histone and lysine. Acetylation and deacetylation of histones are regulated by two proteins in the cell, names HAT and HDAC. Acetylation has significant associations with cell stability, gene transcription, gene expression, and a variety of diseases such as cancers, neurodegenerative diseases, and cardiovascular diseases.

Methylation [3, 10] is most widely known due to the phenomenon of DNA methylation. DNA histone methylation makes the expression of individual genes less likely, in addition to inducing structural changes to chromosomes. Methylation plays a role in cell plasticity, cell growth, and gene expression.

Over the past 10 years, the rapid development of research into PTMs has produced a huge amount of related data. In contrast, however, with the majority of past research studies that explored only a certain PTM, in this study, we sought to investigate the associations of four PTMs, namely, phosphorylation, glycosylation, acetylation, and methylation, with breast cancer, as well as their effects on the organism. This included comprehensive investigations of relevant sequences and domains in order to further integrate the analysis and discussion. Apart from identifying more potential protein PTM sites, looking at the biological functions and processes associated with particular domains can help us to better understand the relationships between PTMs and breast cancer. At the same time, it also allows researchers to have a greater number of biological references as a result of preliminary validations.

II. MATERIALS AND METHODS

A. Data Collection

The sequence data for human breast cancer proteins investigated in this study were all downloaded from the dbPTM, which is an informative resource regarding PTMs. Specifically, the investigated data consisted of all the protein sequence data available for proteins associated

with human breast cancer collected through February 28, 2014. All of the human breast cancer associated protein sequences downloaded were complete in terms of sequence length (that is, they were full length), and the sequences are all distinct from each other.

The total number of breast cancer associated protein sequences obtained from the dbPTM was 67, which includes proteins from humans and others species. However, this study was focused on humans; therefore, screening was conducted to exclude non-human species. The resulting total number of human breast cancer associated genes was 42, with those genes being, namely, *ABCG2, AKP13, AN30A, ANR17, ARI4B, BCAR1, BCAR3, BCAS1, BCD1, BIN2, BRCA1, BRCA2, BRM1L, BRMS1, CCAR2, CE85L, ERGI3, FA84B, GREB1, HEAT6, K0100, KIF15, MAGD2, NCOA3, NCOA6, NFIP1, NRG1, PKHA8, PP14C, PRP31, PSMD6, SEPT1, SFXN4, SGOL1, SMG8, STRAA, SVEP1, SYTL2, SYUG, TFF1, TRFF1,* and *VMA5A*.

B. Research Framework

The experimental procedure for this study was as follows (Fig. 1).

1) Data Collection: Human breast cancer associated protein sequences were downloaded from the dbPTM.

2) Sequences Analysis: Human breast cancer associated protein sequences were analyzed using a two-pronged approach looking at the relevant sequence, and domain.

3) PTM Prediction and PROTTER: First, all the human breast cancer associated protein sequences were analyzed to predict their own N-glycosylation sites, O-glycosylation sites, phosphorylation sites, acetylation sites, methylation sites, and sites of interaction between phosphorylation and glycosylation. Second, a membrane protein topology-related tool called PROTTER [11] was used to draw the structure and location map of each individual human breast cancer protein. Next, each of the PTM sites was annotated with its own color.

4) Domain Prediction: Domain sections were obtained through the InterPro tool [12, 13] by using the domain information for the 42 human breast cancer associated protein sequences. The final results were received from relative position maps of PROTTER, protein-protein interaction information with PTMcode, and domain-domain interaction information with the database of three-dimensional interacting domains (3did).

In order to avoid false positive results, each of the human breast cancer associated protein structure and location maps drawn with the PROTTER tool had to be verified using information from previous studies. In addition, a check was made to determine whether the tyrosine sites of phosphorylation were located near the serine and threonine sites of O-glycosylation.

C. PTMs Sites Prediction and PROTTER

This study sought to identify and explore potential but not yet validated PTMs and explore the interactions between them. Therefore, the study made use of prediction tools and various previously identified PTM sites in order to comprehensively search for other potential PTM sites.

PROTTER is a membrane protein topology-related visualization tool that performs the opposite localizations between protein sequences and PTM sites [11]. After finishing all the PTM sites of the prediction tools, PROTTER was used to illustrate the structure and location maps of each individual human breast cancer protein, and to annotate all the PTMs sites with their own colors (Fig. 2).

First, the prediction of the N-glycosylation sites, O-glycosylation sites, phosphorylation sites, acetylation sites, methylation sites, and sites of interaction between phosphorylation and glycosylation of the 42 human breast cancer associated protein sequences was accomplished using PTM prediction tools, namely, NetNGlyc, NetOGlyc, NetPhos, NetAcet, PMes, and YinOYang. As C-glycosylation is associated with relatively rare PTMs and as information about it is lacking, this study only explored N-glycosylation and O-glycosylation sites. YinOYang sites are interaction sites which generate the interactions between phosphorylation and glycosylation.

The membrane protein topology tool PROTTER was used to draw the structure and location maps of every

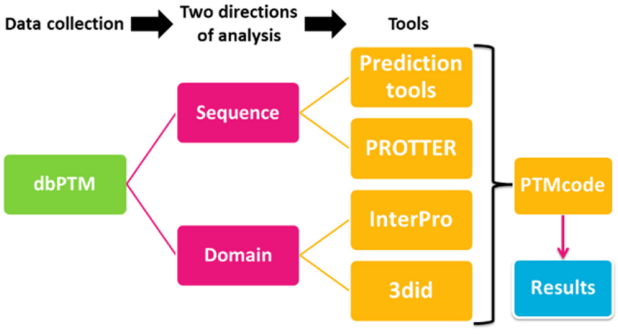

Fig. 1. The experimental method flow chart.

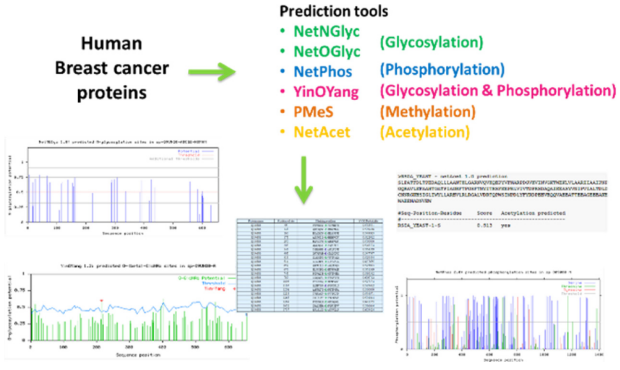

Fig. 2. Data processing flow chart of human breast cancer associated proteins.

human breast cancer protein, and to annotate all the PTM sites with their own colors. In the PROTTER figure, dark green is used to annotate N-glycosylation sites, light green is used for O-glycosylation sites, blue is used for phosphorylation sites, orange is used for methylation sites, yellow is used for acetylation sites, and pink is used for the interaction sites between phosphorylation and glycosylation. This tool provides a better understanding of the relative positions of various PTMs sites.

D. Domain Prediction

Domain data from the InterPro database was used to obtain the domain-related information, including the domains of protein sequences, the regions of each domain, and the associated molecular functions of the domains [12, 13]. Finally, PROTTER was used to integrate the structure and location maps of the PTM sites of every human breast cancer protein. Fig. 3 demonstrates BCAR3_HUMAN as an example, and the results indicate that there are an SH2 domain and PTMs sites, the region of which is 144–256, on the human breast cancer associated protein BCAR3_HUMAN.

E. PTMcode Model

PTMcode is a query interface that utilizes data from a eukaryotic protein-based database [14]. The PTMcode data consists of 25,764 protein sequences which in turn contain 137,525 validated PTM sites and 401,688 functional associations. By typing full protein names or amino acid sequences (in the FASTA format) to get a domain's regions, the PTM sites on protein sequences and their interactions, as well as the associations between PTM sites, can be validated.

The experimental data and results of this study were thus verified using PTMcode. Taking human breast cancer associated protein ABCG2_HUMAN as an example, PTMcode was applied to retrieve data indicating that there are AAA and ABC2_Membrane domains on the protein sequence. In addition to that, the data verified that there are interactions between the phosphorylation site Ser100, which is on the AAA domain, and other sites, including Ser2 and Ser65, as well as the N-glycosylation site Asn596.

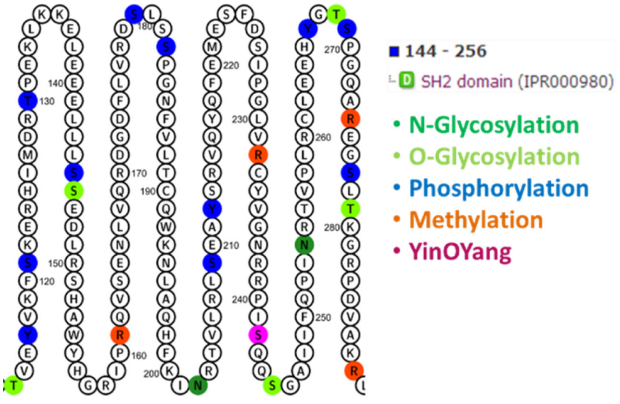

Fig. 3. PTMs sites of SH2 domain with PROTTER in BCAR3_HUMAN protein.

In addition to investigating PTM sites that were previously validated, this study also sought to identify other potential PTM sites. For example, the PTMcode results for ABCG2_HUMAN, which indicate that there are other phosphorylation sites, methylation sites, O-glycosylation sites, and YinOYang sites on the AAA domain. Furthermore, there are other phosphorylation sites, N-glycosylation sites, and methylation sites on the ABC2_Membrane domain.

The AAA domain is related to molecular functions like ATP binding and DNA replication, among others, while the ABC2_Membrane domain is related with molecular functions such as ATP binding and transporter activity. Based on these results, it could be determined that there are interactions between phosphorylation, methylation, O-Glycosylation and N-Glycosylation in AAA and ABC2_Membrane domain.

F. Domain and Function

In addition to using PTMcode results to validate our experimental data, this study used the domain-domain interaction-related tool 3did [15] to retrieve all domain data, and then integrated the retrieved results with PROTTER to connect the PTM sites to their relative locations. The classification and analysis of the results of domain molecular functions and biological processes was thus completed with 3did. For example, the ABC_transporter-like domain on the ABCG2_HUMAN protein sequence can be considered to have something to do with ATP binding function and ATPase activity. These results will be integrated with PROTTER to observe their PTMs sites.

III. RESULTS AND DISCUSSIONS

A. Statistics of PTM Sites and Their Relationships with Human Breast Cancer Proteins

In this study, we obtained the PTM sites for all 42 human breast cancer associated protein sequences. According to the results shown in Table 1, there were 1,064 human breast cancer associated protein sequence O-glycosylation sites, 898 phosphorylation sites, and 828 YinOYang sites. These preliminary statistical results indicated that human breast cancer associated proteins are more significantly associated with O-glycosylation and phosphorylation than other PTMs.

Previous research indicated that protein phosphorylation had a significant association with the signaling pathways of organisms. Most cancer cells will exhibit overreliance

Table 1. The number of PTM sites of human breast cancer associated proteins

PTMs	Quantity
O-Glycosylation	1,064
Phosphorylation	898
YinOYang	828
N-Glycosylation	173
Methylation	104
Acetylation	2

Table 2. The number of PTM sites of human breast cancer associated protein domains

PTMs	Quantity
Phosphorylation	126
Methylation	79
O-Glycosylation	73
N-Glycosylation	51
YinOYang	41
Acetylation	0

on a specific pathway, allowing the cancer cells to survive or be excessive in number [16]. Also, protein glycosylation has been shown to be related to immunity with a lot of the proteins [17]. Protein glycosylation may affect whether the protein folding is normal or not. If a protein cannot fold well then that will affect the immunity-related function of the protein and affect the generation of diseases [18].

As recent research has pointed out, cancer cells will recruit immune suppression cells in the tumor growth region. Cancer cells may also let myeloid-derived suppressor cells (MDSC) and regulatory T (TReg) cells come within and generate the immunosuppressive tumor microenvironment [19]. Glycosylation not only affects the growth and activity of bone marrow-derived cells, but also their survival. An immunosuppressive tumor microenvironment promotes tumor proliferation, in addition to being associated with resistance and cancer metastasis.

These findings preliminarily determine that our experiment direction and methods are keeping with previous research [20, 21]. The amount of glycosylation sites, phosphorylation sites and YinOYang sites are much more than other PTMs sites. Moreover, YinOYang sites could verify that there is a significant association between protein phosphorylation and glycosylation. The material and methods used in this study could also be used to identify many more potential PTM sites, allowing for deeper discussions and explorations [22].

According to the data in Table 1, this study confirmed the finding of a previous study regarding PTM sites of human breast cancer associated protein sequences [23] that protein acetylation has a close association with tumor generation. In view of this, protein acetylation can be supposed to be closely related to breast cancer. However, there were just 2 acetylation sites among the 42 human breast cancer associated proteins in our study. As such, experimental results and those of previous studies could be taken to imply that protein acetylation may act on breast cancer-related pathways to affect the degree of deterioration rather than inducing the generation of breast cancer. We hope that a future study using biologically relevant laboratory data will be undertaken to confirm this inference.

B. Statistics of the Domains of PTMs Sites and Their Relationships with Human Breast Cancer Proteins

In this study, we obtained the domains for all 42 human breast cancer associated protein sequences. According to the results shown in Table 2, there were 126 human breast cancer associated protein domains with phosphorylation sites, 79 domains with methylation sites, and 73 domains with O-glycosylation sites. These preliminary statistical results indicated that human breast cancer associated protein domains are more significantly associated with phosphorylation, methylation, and O-glycosylation than other PTMs.

Phosphorylation [24] is not only the most common reversible PTM in cells but also participates in a variety of biological signaling pathways and controls the activation and inhibition of many enzymes and receptors [25, 26]. Protein methylation [24, 27] makes the expression of genes less likely, induces structural changes to chromosomes, and is related to cell plasticity, cell growth, and gene expression [28]. Glycosylation can affect the folding of proteins, thereby affecting whether the given protein plays its normal role or not [29-31]. Through our experiment and methods, it can be understood that human breast cancer associated proteins have significant associations with phosphorylation, methylation, and glycosylation, as well as their own molecular functions. Among these PTMs, methylation plays the role of affecting gene expression.

Genome instability is one cause of cancer [32]. Genome instability is most likely to cause gene mutations, in addition to causing oncogenes to have increased functionality or causing tumor suppressor genes to lose their function. These effects can in turn induce normal cells to gradually transform into cancer cells, and this transformation process is known as tumorigenesis. The experimental results of this study regarding the domains associated with phosphorylation and O-glycosylation were consistent with those of previous studies. Our experimental results also suggest that methylation on human breast cancer associated protein domains has a significant relationship with the occurrence of breast cancer.

C. Verification of PTMcode model

After comparing the results of this study with the validated study from previous studies, it could be concluded that our study found more potential PTM sites. This study used validated data from PTMcode and dbPTM to verify our experimental data and results. By comprehensive searching from sequence, molecular biology function and biology process which on sequence and domain are investigated the interactions between PTMs. The methods used in this study yielded results consistent with those of previous studies, and also allowed for the identification of more potential PTM sites and interactions. These experimental results may provide valuable data to biological laboratories for use in further research.

1) K0100_HUMAN: Fig. 4 shows the results of the human breast cancer associated protein K0100_HUMAN [33] and other potential PTM sites which were obtained via the experimental methods used in this study. The experimental methods indicated that there may be phosphorylation sites, methylation sites, and YinOYang sites on the Fmp27 domain. Another domain on K0100_HUMAN is the Fmp27_GFWDK domain. There are also other potential phosphorylation sites, methylation sites, and N-glycosylation sites. The other domain on K0100_HUMAN is the Apt1 domain. In addition to the verified phos-

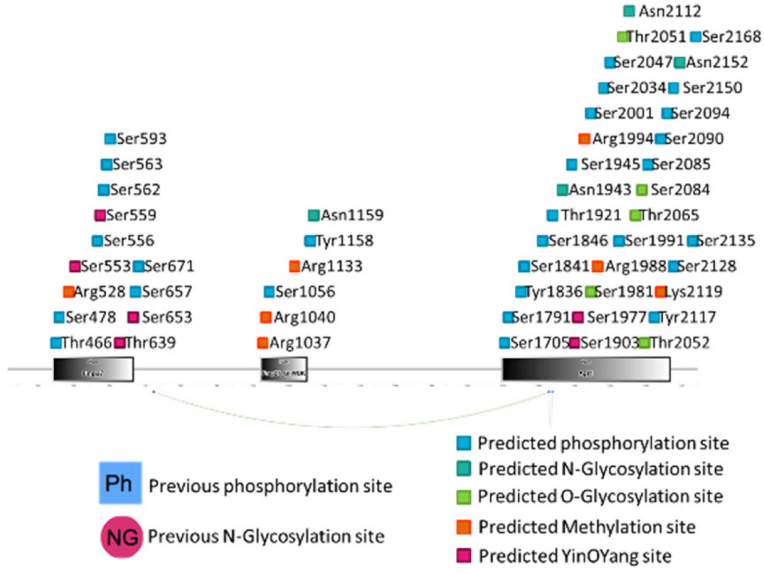

Fig. 4. PTMcode model of protein K0100_HUMAN.

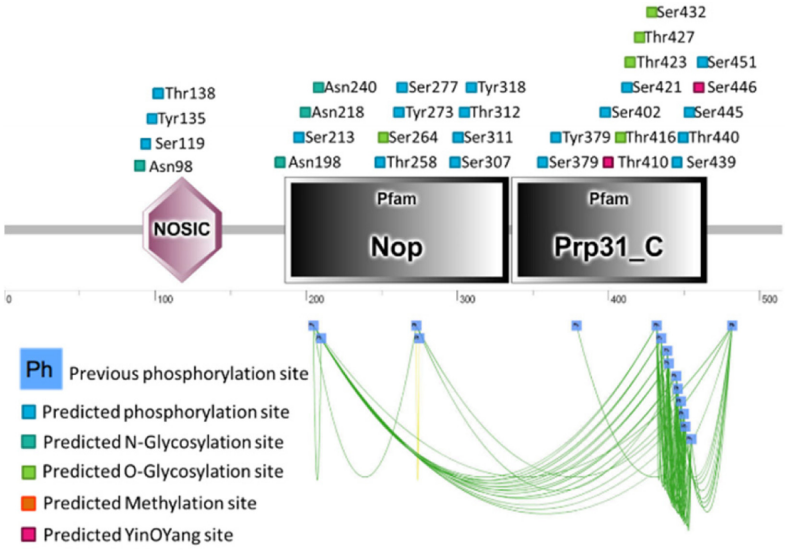

Fig. 5. PTMcode model of protein PRP31_HUMAN.

phorylation sites, there are other potential phosphorylation sites, methylation sites, N-glycosylation sites, O-glycosylation sites, and YinOYang sites. According to the study results, it was possible to determine that there are interactions between phosphorylation, methylation, N-glycosylation and O-glycosylation in these domains on this human breast cancer associated protein.

2) PRP31_HUMAN: Fig. 5 displays the PTMcode model of the human breast cancer associated protein PRP31_HUMAN [34] and other potential PTM sites which were obtained by using the experimental methods in this research. The experimental methods indicated that there may be phosphorylation sites and N-glycosylation sites on the NOSIC domain. Another domain on PRP31_HUMAN is the NOP domain. In addition to the verified phosphorylation sites, there are other potential phosphorylation sites, N-glycosylation sites, and O-glycosylation sites. The other domain on PRP31_HUMAN is the Prp31_C domain. In addition to the verified phosphorylation sites, there are other potential phosphorylation sites, O-glycosylation sites, and YinOYang sites. According to the study results, it was possible to determine that there are interactions between phosphorylation, N-glycosylation and O-glycosylation in three domains on this protein.

3) NFIP1_HUMAN: Fig. 6 shows the results of the NFIP1_HUMAN [35]. The PTMcode model and other potential PTM sites were obtained via the experimental methods used in this study. The experimental methods indicated that, in addition to the verified phosphorylation sites, there may be other phosphorylation sites, O-glycosylation sites, methylation sites, and YinOYang sites on the DUF2370 domain. According to the study results, it was possible to determine that there are interactions between phosphorylation, O-glycosylation, and methylation on the DUF2370 domain of this protein.

4) TREF1_HUMAN: Fig. 7 displays the results of PTMcode model of the TREF1_HUMAN protein [36] and potential PTMs sites which were obtained from this experiment methods. According to the experimental methods of

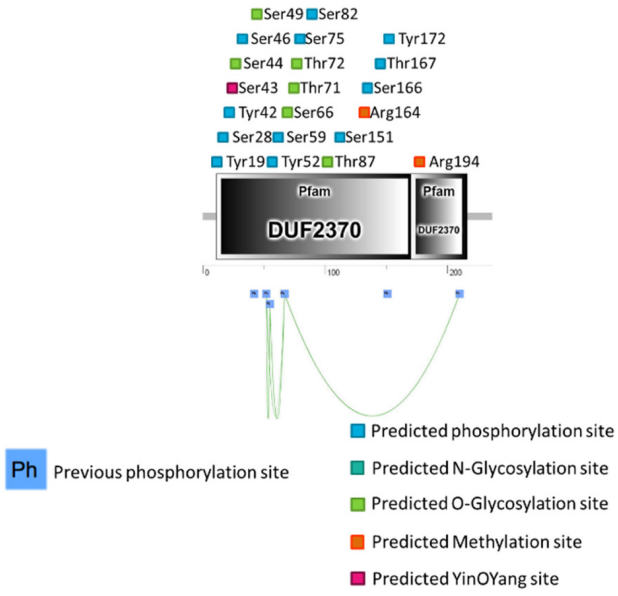

Fig. 6. PTMcode model of protein NFIP1_HUMAN.

this study, it obtains that there are methylation sites and O-glycosylation sites are on the ZnF_C2H2 domain. Another domain on TREF1_HUMAN is the ELM2 domain. There are also other potential phosphorylation sites, O-glycosylation sites, and methylation sites on the ELM2 domain. The third domain on TREF1_HUMAN is the SANT domain, and there are other potential phosphorylation sites. These results show that there are interactions between methylation, O-glycosylation, and phosphorylation in three kinds of domains on TREF1_HUMAN protein.

IV. CONCLUSIONS

In this research, the interactions between PTMs of breast cancer-related proteins were comprehensively investigated

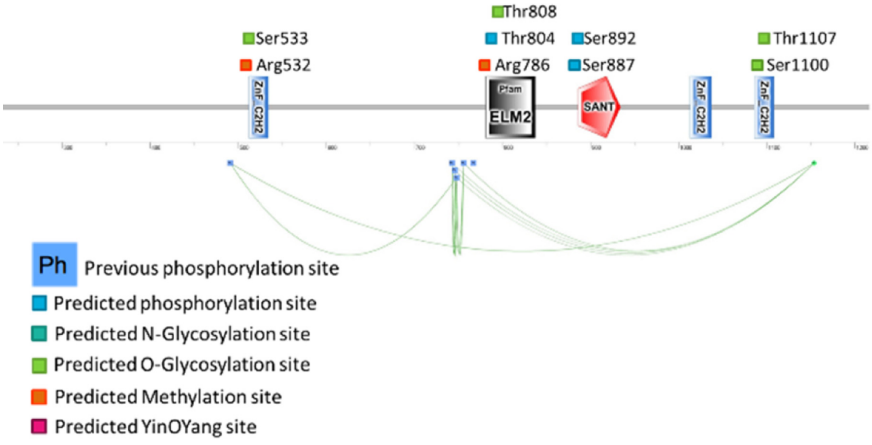

Fig. 7. PTMcode model of protein TREF1_HUMAN.

using sequence and domain information and the application of bioinformatics methods. With these methods, other possible PTMs of proteins were identified. According to the study results, the PTM sites in human breast cancer associated proteins may be more associated with O-glycosylation and phosphorylation than other PTMs. And the third related type of PTM is YinOYang PTMs, which consist of interactions between glycosylation and phosphorylation. The YinOYang results of this study may show that human breast cancer associated proteins have some association with glycosylation and phosphorylation. On the basis of the results regarding the domains of breast cancer associated proteins, it can also be suggested that while human breast cancer associated domains have greater relationships with O-glycosylation and phosphorylation, they also have some association with methylation. This finding would improve the understanding of the associations of human breast cancer with O-glycosylation and phosphorylation, and it also may offer the key that methylation has some relation with human breast cancer.

We first used the membrane protein topology associated tool PROTTER to visualize the opposite locations of the PTM sites. In addition to considering clearly the opposite locations of the PTM sites, the sequence and domain-related results could make more discussion of protein functions and PTMs possible. Moreover, these experimental results may provide biological laboratories with important experimental evidence and validation data.

However, this study only focused on human breast cancer associated proteins in its experiments and discussion. In fact, the amount of data in this study was relatively lacking and weak compared to that of previous studies, and the credibility of related future studies will need to be improved. However, the study did use validated data and information and comprehensive searches for sequences and domains to investigate the interactions of PTMs. The PTM interaction information obtained in the study was thus greater in amount and more carefully obtained than that of previous studies. Because the amount of information and data regarding motifs were less than those regarding domains, most of the experimental results of this study are based on human breast cancer associated protein domains. We hope that more motif-related information and data will be available in the future that could be added to our study to improve the credibility of our experimental model.

In the future, this study hopes to combine metabolic pathways information and extend to other disease applications. Such diseases could include Alzheimer's disease [37] and neurodegenerative diseases that are also major genetic diseases, among others. In addition to expanding upon the scope of this study, future studies could be undertaken in order to better understand and learn about the generation of different diseases and the various factors influencing said diseases. If we could, for example, add the combined data of experimental groups and control groups, then the credibility and reliability of related future research will be significantly improved. It is also hoped that the PTMs found in the experimental results of this study may provide validation and foundational data for use biological laboratories.

AKNOWLEDGMENTS

This research was supported by Ministry of Science and Technology, Taiwan, R.O.C. (Grant No. 106-2221-E-005-077-MY2).

REFERENCES

1. J. Mohandass, S. Ravichandran, K. Srilakshmi, C. P. Rajadurai, S. Sanmugasamy, and G. R. Kumar, "BCDB: a database for breast cancer research and information," *Bioinformation*, vol. 5, no. 1, pp. 1-3, 2010.
2. R. A. Baasiri, S. R. Glasser, D. L. Steffen, and D. A. Wheeler, "The breast cancer gene database: a collaborative information resource," *Oncogene*, vol. 18, no. 56, pp. 7958-7965, 1999.
3. A. M. Butt, D. Feng, M. Idrees, Y. Tong, and J. Lu, "Computational identification and modeling of crosstalk between phosphorylation, O-beta-glycosylation and methylation of FoxO3 and implications for cancer therapeutics," *International Journal of Molecular Sciences*, vol. 13, no. 3, pp. 2918-2938, 2012.
4. A. M. Butt, D. Feng, I. Nasrullah, S. Tahir, M. Idrees, Y. Tong, and J. Lu, "Computational identification of interplay between phosphorylation and O-beta-glycosylation of human occludin as potential mechanism to impair hepatitis C virus entry," *Infection, Genetics and Evolution*, vol. 12, no. 6, pp. 1235-1245, 2012.
5. R. Rao, D. Xu, J. J. Thelen, and J. A. Miernyk, "Circles within circles: crosstalk between protein Ser/Thr/Tyr-phosphorylation and Met oxidation," *BMC Bioinformatics*, vol. 14(Suppl 14), article no. S14, 2013.
6. G. W. Hart, C. Slawson, G. Ramirez-Correa, and O. Lagerlof, "Cross talk between O-GlcNAcylation and phosphorylation: roles in signaling, transcription, and chronic disease," *Annual Review of Biochemistry*, vol. 80, pp. 825-858, 2011.
7. J. S. Chauhan, A. Rao, and G. P. Raghava, "In silico platform for prediction of N-, O- and C-glycosites in eukaryotic protein sequences," *PLoS One*, vol. 8, no. 6, article no. e67008, 2013.
8. J. L. McLarty, S. A. Marsh, and J. C. Chatham, "Post-translational protein modification by O-linked N-acetyl-glucosamine: its role in mediating the adverse effects of diabetes on the heart," *Life Sciences*, vol. 92, no. 11, pp. 621-627, 2013.
9. Z. Lu, Z. Cheng, Y. Zhao, and S. L. Volchenboum, "Bioinformatic analysis and post-translational modification crosstalk prediction of lysine acetylation," *PLoS One*, vol. 6, no. 12, article no. e28228, 2011.
10. G. Alves, A. Tatro, and T. Fanning, "Differential methylation of human LINE-1 retrotransposons in malignant cells," *Gene*, vol. 176, no. 1-2, pp. 39-44, 1996.
11. U. Omasits, C. H. Ahrens, S. Muller and B. Wollscheid, "Protter: interactive protein feature visualization and integration

with experimental proteomic data," *Bioinformatics*, vol. 30, no. 6, pp. 884-886, 2014.
12. R. D. Finn, T. K. Attwood, P. C. Babbitt, A. Bateman, P. Bork, A. J. Bridge, H. Y. Chang, Z. Dosztanyi, S. El-Gebali, M. Fraser, et al., "InterPro in 2017: beyond protein family and domain annotations," *Nucleic Acids Researchg*, vol. 45(D1), pp. D190-D199, 2017.
13. A. Mitchell, H. Y. Chang, L. Daugherty, M. Fraser, S. Hunter, R. Lopez, C. McAnulla, C. McMenamin, G. Nuka, S. Pesseat, et al., "The InterPro protein families database: the classification resource after 15 years," *Nucleic Acids Research*, vol. 43(D1), pp. D213-D221, 2015.
14. P. Minguez, I. Letunic, L. Parca, and P. Bork, "PTMcode: a database of known and predicted functional associations between post-translational modifications in proteins," *Nucleic Acids Research*, vol. 41(D1), pp. D306-D311, 2013.
15. R. Mosca, A. Ceol, A. Stein, R. Olivella, and P. Aloy, "3did: a catalog of domain-based interactions of known three-dimensional structure," *Nucleic Acids Research*, vol. 42(D1), pp. D374-D379, 2014.
16. P. Minguez, L. Parca, F. Diella, D. R. Mende, R. Kumar, M. Helmer-Citterich, A. C. Gavin, V. van Noort, and P. Bork, "Deciphering a global network of functionally associated post-translational modifications," *Molecular Systems Biology*, vol. 8, article no. 599, 2012.
17. C. Slawson and G. W. Hart, "O-GlcNAc signalling: implications for cancer cell biology," *Nature Reviews Cancer*, vol. 11, no. 9, pp. 678-684, 2011.
18. J. Seo and K. J. Lee, "Post-translational modifications and their biological functions: proteomic analysis and systematic approaches," *Journal of Biochemistry and Molecular Biology*, vol. 37, no. 1, pp. 35-44, 2004.
19. M. A. Daniels, K. A. Hogquist, and S. C. Jameson, "Sweet 'n' sour: the impact of differential glycosylation on T cell responses," *Nature immunology*, vol. 3, no. 10, pp. 903-910, 2002.
20. D. Hanahan and R. A. Weinberg, "The hallmarks of cancer," *Cell*, vol. 100, no. 1, pp. 57-70, 2000.
21. S. I. Hayashi, H. Eguchi, K. Tanimoto, T. Yoshida, Y. Omoto, A. Inoue, N. Yoshida, and Y. Yamaguchi, "The expression and function of estrogen receptor alpha and beta in human breast cancer and its clinical application," *Endocrine-Related Cancer*, vol. 10, no. 2, pp. 193-202, 2003.
22. A. S. Venne, L. Kollipara, and R. P. Zahedi, "The next level of complexity: crosstalk of posttranslational modifications," *Proteomics*, vol. 14, no. 4-5, pp. 513-524, 2014.
23. M. R. Fernandes, D. C. de Carvalho, A. K. dos Santos, S. E. dos Santos, P. P. de Assumpcao, R. M. Burbano, and N. P. dos Santos, "Association of slow acetylation profile of NAT2 with breast and gastric cancer risk in Brazil," *Anticancer Research*, vol. 33, no. 9, pp. 3683-3689, 2013.
24. P. O. Esteve, Y. Chang, M. Samaranayake, A. K. Upadhyay, J. R. Horton, G. R. Feehery, X. Cheng and S. Pradhan, "A methylation and phosphorylation switch between an adjacent lysine and serine determines human DNMT1 stability," *Nature Structural & Molecular Biology*, vol. 18, no. 1, pp. 42-48, 2011.
25. F. Hochgrafe, L. Zhang, S. A. O'Toole, B. C. Browne, M. Pinese, A. Porta Cubas, G. M. Lehrbach, D. R. Croucher, D. Rickwood, A. Boulghourjian, et al., "Tyrosine phosphorylation profiling reveals the signaling network characteristics of Basal breast cancer cells," *Cancer Research*, vol. 70, no. 22, pp. 9391-9401, 2010.
26. N. Blom, S. Gammeltoft, and S. Brunak, "Sequence and structure-based prediction of eukaryotic protein phosphorylation sites," *Journal of Molecular Biology*, vol. 294, no. 5, pp. 1351-1362, 1999.
27. H. Tang and E. Goldberg, "Homo sapiens lactate dehydrogenase c (Ldhc) gene expression in cancer cells is regulated by transcription factor Sp1, CREB, and CpG island methylation," *Journal of Andrology*, vol. 30, no. 2, pp. 157-167, 2009.
28. L. Yan, X. Yang, and N. E. Davidson, "Role of DNA methylation and histone acetylation in steroid receptor expression in breast cancer," *Journal of Mammary Gland Biology and Neoplasia*, vol. 6, no. 2, pp. 183-192, 2001.
29. R. Gupta and S. Brunak, "Prediction of glycosylation across the human proteome and the correlation to protein function," *Pacific Symposium on Biocomputing*, vol. 7, pp. 310-322, 2002.
30. M. V. Dwek, H. A. Ross, and A. J. Leathem, "Proteome and glycosylation mapping identifies post-translational modifications associated with aggressive breast cancer," *Proteomics*, vol. 1, no. 6, pp. 756-762, 2001.
31. C. Steentoft, S. Y. Vakhrushev, H. J. Joshi, Y. Kong, M. B. Vester-Christensen, K. T. Schjoldager, K. Lavrsen, S. Dabelsteen, N. B. Pedersen, L. Marcos-Silva, et al., "Precision mapping of the human O-GalNAc glycoproteome through SimpleCell technology," *The EMBO Journal*, vol. 32, no. 10, pp. 1478-1488, 2013.
32. P. N. Munster, T. Troso-Sandoval, N. Rosen, R. Rifkind, P. A. Marks, and V. M. Richon, "The histone deacetylase inhibitor suberoylanilide hydroxamic acid induces differentiation of human breast cancer cells," *Cancer Research*, vol. 61, no. 23, pp. 8492-8497, 2001.
33. J. Song, W. Yang, M. Shih Ie, Z. Zhang, and J. Bai, "Identification of BCOX1, a novel gene overexpressed in breast cancer," *Biochimica et Biophysica Acta (BBA)-General Subjects*, vol. 1760, no. 1, pp. 62-69, 2006.
34. M. J. Scanlan, I. Gout, C. M. Gordon, B. Williamson, E. Stockert, A. O. Gure, D. Jager, Y. T. Chen, A. Mackay, M. J. O'Hare, et al., "Humoral immunity to human breast cancer: antigen definition and quantitative analysis of mRNA expression," *Cancer Immunity Archive*, vol. 1, article no. 4, 2001.
35. L. H. Low, Y. L. Chow, Y. Li, C. P. Goh, U. Putz, J. Silke, T. Ouchi, J. Howitt, and S. S. Tan, "Nedd4 family interacting protein 1 (Ndfip1) is required for ubiquitination and nuclear trafficking of BRCA1-associated ATM activator 1 (BRAT1) during the DNA damage response," *Journal of Biological Chemistry*, vol. 290, no. 11, pp. 7141-7150, 2015.
36. T. van Agthoven, J. Veldscholte, M. Smid, T. L. van Agthoven, L. Vreede, M. Broertjes, I. de Vries, D. de Jong, R. Sarwari, and L. C. Dorssers, "Functional identification of genes causing estrogen independence of human breast cancer cells," *Breast Cancer Research and Treatment*, vol. 114, no. 1, pp. 23-30, 2009.
37. W. Jiang, Y. Zhang, F. Meng, B. Lian, X. Chen, X. Yu, E. Dai, S. Wang, X. Liu, X. Li, et al., "Identification of active transcription factor and miRNA regulatory pathways in Alzheimer's disease," *Bioinformatics*, vol. 29, no. 20, pp. 2596-2602, 2013.

Chi-Hua Tung

Chi-Hua Tung was born in Kaohsiung, Taiwan. He received his Bachelor degree in biology from National Sun Yat-Sen University in 2002. Also, he received the M.S. and Ph.D. degrees in 2005 and 2009, respectively, all from the Institute of Bioinformatics, National Chiao Tung University, Taiwan. Since 2011, he has been with the Chung-Hua University, Taiwan, where he is currently an Assistant Professor with the Department of Bioinformatics. His research is focused on structure genomics and the studying of relationships between protein structures and functions using "Structural Alphabet." He is currently working to apply structural alphabet to protein sequence analysis, structure motif discovery, and network biology.

Pei-Wei Shueng

Pei-Wei Shueng graduated from National Defense Medical Center in Taiwan and used to be a visiting Scientist of MD Anderson Cancer in Houston, USA in 1999–2000. He was one of the pioneers of intensity-modulated radiotherapy in Taiwan. Also, he was the founder of Department of Radiation Oncology, Far Eastern Memorial Hospital (FEMH) in Taiwan. As of July 2017, he and his colleagues of FEMH have published more than 40 peer-reviewed articles outlining its clinical experience with the TomoTherapy System. Furthermore in 2009, Dr. Shueng reported their first experience with total marrow irradiation (TMI). Then in 2013, he announced the first global experience with helical irradiation of the total skin (HITS) for intractable cutaneous lymphoma. With outstanding reputation and innovative approaches for cancer treatment, he currently has treated more than 5,000 patients with the TomoTherapy System.

Yen-Wei Chu

Yen-Wei Chu is an associate professor of Institute of Genomics and Bioinformatics and the joint appointment professor of Institute of Molecular Biology and Graduate Institute of Biotechnology at National Chung Hsing University, Taiwan. He received his Ph.D. degree in Computer Science in 2006 from the National Chiao Tung University, Taiwan. His lab takes the technologies of data mining, machine learning and artificial intelligence as the core algorithms to establish a variety of intelligent decision-making systems for different issues, which is the important part for industry 4.0 and Internet of Things. He mainly focuses on building the learning model of humanoid intelligence, covering the field of bioinformatics, medical science, agriculture, food science, business, astronomy and so on.

Chi-Wei Chen

Chi-Wei Chen graduated from Institute of Genomics and Bioinformatics, National Chung Hsing University, Taiwan in 2012. He is currently a Ph.D. student in the Department of Computer Science and Engineering at National Chung Hsing University. His research interests include bioinformatics and machine learning.

Chian-Ying Chen

Chian-Ying Chen received her bachelor's degree in Department of Healthcare Administration and Medical Informatics from Kaohsiung Medical University, Taiwan in 2012, and a master's degree from the Institute of Genomics and Bioinformatics, National Chung Hsing University, Taiwan in 2014. Her research interests include bioinformatics and statistical analysis.