

OPEN

Integrated Transcriptomics, Proteomics, and Glycomics Reveals the Association between Up-regulation of Sialylated N-glycans/Integrin and Breast Cancer Brain Metastasis

Wenjing Peng, Rui zhu, Shiyue Zhou, Parvin Mirzaei & Yehia Mechref*

Breast cancer brain metastasis has been recognized as one of the central issues in breast cancer research. The elucidation of the processes and pathways that mediate this step will provide important clues for a better understanding of breast cancer metastasis. Increasing evidence suggests that aberrant glycosylation patterns greatly contribute to cell invasion and cancer metastasis. Herein, we combined next-generation RNA sequencing with liquid chromatography-tandem mass spectrometry-based proteomic and N-glycomic analysis from five breast cancer cell lines and one brain cancer cell line to investigate the possible mechanisms of breast cancer brain metastasis. The genes/proteins associated with cell movement were highlighted in breast cancer brain metastasis. The integrin signaling pathway and the up-regulation of α -integrin (ITGA2, ITGA3) were associated with the brain metastatic process. 12 glycogenes showed unique expression in 231BR, which could result in an increase of sialylation during brain metastasis. In agreement with the changes of glycogenes, 60 out of 63 N-glycans that were identified exhibited differential expression among cell lines. The correlation between glycogenes and glycans revealed the importance of sialylation and sialylated glycans in breast cancer brain metastasis. Highly sialylated N-glycans, which were up-regulated in brain-seeking cell line 231BR, likely play a role in brain metastasis.

Breast cancer, the leading cause of cancer mortality and morbidity among females, represents the second highest frequency of brain metastasis^{1,2}. The incidence of brain metastasis among breast cancer patients is continuously increasing due to the development of detection technologies for better prognosis and the improvement of therapeutic methods leading to a longer survival of primary breast cancer patients³. Thus, breast cancer brain metastasis is considered an essential issue in breast cancer studies^{4–8}.

The key event of breast cancer brain metastasis is the penetration of breast cancer cells through the blood-brain barrier⁹. The blood-brain barrier acts as a selective “physical barrier” around the brain, due to the tight junction and adherens junction between adjacent brain microvascular endothelial cells (BMECs)^{4,10}. BMECs provide low permeability, high electrical resistance, and a series of transport channels that supply the brain with nutrients and mediate the waste products of brain metabolism, so that the blood-brain barrier can protect the brain from toxins, unnecessary proteins, and cells^{4,11}. However, breast cancer cells can penetrate this defensive layer by attaching to and invading the blood-brain barrier. In addition to protecting the brain, the blood-brain barrier can also prevent drugs from entering the brain; consequently, brain metastasis treatments pose a great challenge. Current systematic therapy, including chemotherapy, hormone therapy, and targeted-drug therapy, is not effective when treating breast cancer brain metastatic sites¹². Thus, in recent years breast cancer brain metastasis has attracted much research interest, and many efforts have been made to elucidate the mechanisms of this process^{5,7,12–14}. Overexpression of HER2 is considered to be associated with breast cancer brain metastasis and can be used as a positive prognostic

Department of Chemistry and Biochemistry, Texas Tech University, Lubbock, TX, USA. *email: yehia.mechref@ttu.edu

Cell Line Name	subtype	ER	PR	HER2	tumor source	Pathology
MDA-MB-231	Basal	–	–	–	Metastasis; pleural effusion	Breast cancer
MDA-MB-231BR	Basal	–	–	–	Metastasis; brain	Breast cancer
MDA-MB-361	Luminal B	+	+	+	Metastasis; brain	Breast cancer
HTB131	Basal	–	–	–	Metastasis; pericardial effusion	Breast cancer
HTB22	Luminal A	+	+	–	Metastasis; pleural effusion	Breast cancer
CRL-1620					Brain	Brain cancer

Table 1. Immunophenotypes of six cancer cell lines.

factor^{15–17}. Parul Gupta *et al.* reported that the suppression of HER2, EGFR, and VEGF may suppress breast cancer brain metastasis *in vivo*⁶. The methylation of GALNT9, CCDC8, and BNC1 genes have also been investigated in brain metastasis¹⁸. In 2009, Massague and coworkers reported that COX2 and HBEGF contributed to breast cancer brain metastasis⁵. However, the comprehensive mechanisms of this process remain unclear.

Glycosylation is one of the most important post-translational modifications (PTM) and has been successfully utilized as a biomarker of many cancers^{19–26}. Recently, the sialylation of Lewis X epitopes was detected and glycan levels were found to be higher in patients with high circulating tumor cells (CTCs)²⁷. In addition, CD44, HPA, and Galectin were found likely to be involved in breast cancer metastasis^{28–31}. However, few current studies are focusing on breast cancer brain metastasis. The Massague group also found that the ST6GALNAC5 glycogene contributes to breast cancer brain metastasis by increasing the adhesion of breast cancer cells to the blood-brain barrier⁵. To our best knowledge, this is the only glycogene which is directly confirmed to be associated with breast cancer brain metastasis.

Herein we have integrated transcriptomics, proteomics, and glycomics among different breast cancer cell lines. Correlations between these different “-omics” were studied to achieve a deeper insight of the biological molecular mechanism of breast cancer brain metastasis. In addition, we focused on the glycosylation and correlations between glycogene expression and N-glycan patterns to further investigate the relationship between glycosylation and breast cancer brain metastasis and potential glyco-biomarkers.

Results and Discussion

Cell line clinicopathological features. Five breast cancer cell lines (MDA-MB-231, MDA-MB-231BR, MDA-MB-361, HTB131, and HTB22) and one brain cancer cell line (CRL-1620) were employed in this study. Each cell line has a different clinicopathological feature, including phenotypes, metastatic properties, and sites (Table 1). Among them, MDA-MB-231BR (231BR) has a 100% trend to metastasize to the brain, and it cannot migrate to any other organs; the brain is not the first target for other cell lines, which can migrate to areas such as bone and lung^{6,32}. 231BR is the subline of MDA-MB-231 (231). 231 was initially injected into a nude mouse, and then tumor cells from its brain metastatic site were isolated and injected into another mouse. After 6 cycles, 231BR was generated and had a specific capacity of brain metastasis. Therefore, 231BR, as a brain-seeking cell line, was special in this study and used as the reference cell line in comparison with five other cell lines. Due to its specific capacity of brain metastasis, any unique expressions in 231BR when compared to other cell lines were expected. These differences probably contributed to the penetration of cells through the blood-brain barrier, thus facilitating breast cancer brain metastasis.

NGS transcriptome mapping and differential expression analysis. Total RNAs from six cell lines were applied in this study. Using the next generation sequencing (NGS) technique, overall 1.25 billion paired end reads were generated with 66X sequencing depth by sequencing the pooled cDNA libraries from the six cell lines. Function annotation against a reference human genome (Homo sapiens-build 36.1.hg18) revealed 24763 transcripts from the cell lines.

Differential expression analysis was achieved by the use of QSeq software. 14551 genes were found differentially expressed in all six cell lines when they were compared with 231BR, with a fold change of more than 2 and a p-value less than 0.01. Figure 1 depicts the quantitative transcriptomic visualization of differentially expressed genes. The color codes (color bar on the right) denotes the up-regulation or down-regulation of the genes in five cell lines relative to 231BR. Each comparison has a different expression pattern distinct from the others which denotes that there are differences among all cell lines at the transcriptomic level. 231 shows the fewest difference with 231BR (least red or blue color denotes least up or downregulated genes). Interestingly, as a brain cancer cell line, gene expression difference between CRL and 231BR were less than those between other breast cancer cell lines and 231BR. It may be because both CRL and 231BR are from brain tumor sites. In addition, the number of differentially expressed genes varied among the different cell lines. There were 4922 (19.9%), 8022 (32.4%), 6585 (26.6%), 7348 (29.7%), and 8038 (32.5%) genes that demonstrated up/down-regulation in 231, 361, CRL, HTB22, and HTB131 comparisons, respectively. Compared to 231BR, 231 showed less difference than other cell lines, which was expected because 231BR is a subline of 231.

Gene ontology (GO) analysis. The differentially expressed genes of the six cell lines were enriched by GO analysis. Fig. S1 depicts the top five biofunction enrichments among five comparisons (231BR was utilized as the baseline). From the inner circle to the outer circle are 231, 361, HTB22, HTB131, and CRL, respectively, compared with 231BR. The most significant function in all comparisons was cellular movement, suggesting that 231BR had the most significant modifications associated with cell movement. These gene expression changes of

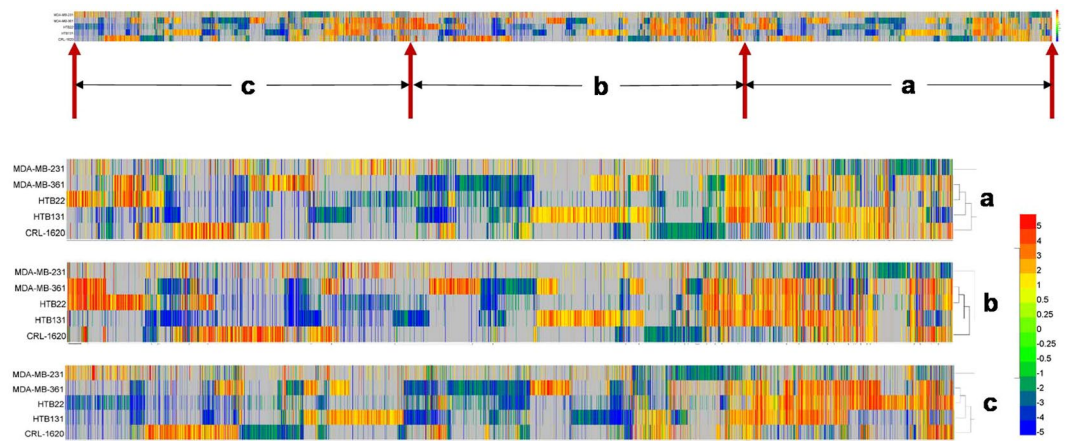


Figure 1. Gene differential expression patterns of five cell lines relative to MDA-MB-231BR. Distinct gene expression patterns of five cell lines were observed when compared to 231BR. (a–c) Denote the corresponding zoomed in areas on original heatmap. The color codes represent gene expression fold changes (RPKM) of each cell line relative to 231BR. Red color denotes the up-regulation of genes in other cell lines when compared to 231BR, and down-regulation in 231BR. Blue color denotes the down-regulation in other cell lines, and the up-regulation in 231BR. Gray color denotes the gene expressions had no significant difference when compared to 231BR.

cell movement may enhance the brain metastatic ability of 231BR. In addition, CRL showed the largest difference due to its different cancer type.

Pathway analysis. Pathway analysis of differentially expressed transcriptome was performed by Ingenuity Pathway Analysis (IPA, QIAGEN Inc., <https://www.qiagenbioinformatics.com/products/ingenuitypathway-analysis>). Again, 231BR was utilized as the baseline. The transcriptomic data from other cell lines were compared with 231BR and then uploaded to IPA. IPA searched its database and provided the pathway and biofunction enrichment, then predicted the upstream regulator or the influence on the downstream bioprocesses based on a z-score and p-value. The p-value was calculated by using Fisher's Exact Test to estimate the significance of the clustering, and the z-score was applied to infer the activation states of predicted transcriptional regulators (upstream) and functions/pathways (downstream). Fig. S2 depicts the diseases and disorders activation status of the five cell lines compared with 231BR. The invasion of tumor, growth of tumor, and metastasis abilities of all cell lines were predicted to be inhibited when compared to 231BR, indicating that 231BR had a higher invasive and metastatic capacity over the other cell lines. Figure 2 depicts the functional activation status of the cell lines when compared with 231BR. The darker blue color denotes that the corresponding function is less active compared to 231BR. The migration of cells was predicted to be more active in 231BR when compared with other breast cancer cell lines (blue color denotes the inhibition in other cell lines, and activation in 231BR). These results suggest that the differentially expressed genes that were associated with cell migration may contribute to the high brain metastatic ability of 231BR.

231BR was derived from 231 and showed the highest similarity with 231. The expression changes between 231BR and 231 were considered to be of highest importance and probably play a role in brain metastasis. Thus, the pathway and upstream regulator analyses were focused on 231 and 231BR in this study. The integrin signaling pathway (Fig. 3) and VEGF signaling pathway (Fig. S3) were shown to be inhibited in 231 by IPA, and thus were enhanced in 231BR (green color denotes the down-regulation of gene in 231, and up-regulation in 231BR). The integrin family, as a transmembrane receptor, is one of the key elements in integrin signaling pathway that regulates many bioprocesses, such as cell cycles and cytoskeleton organization³³. It is made of α -integrin and β -integrin that mediate cell proliferation, migration, metastasis, and survival. It has recently become clear that ligated or unligated integrin can positively or negatively influence tumor cell survival, which are often contradicted in some cases, thus facilitating tumor growth and metastasis³⁴. Figure 3 depicts the integrin signaling pathway in the 231 vs. 231BR comparison. The up-regulation of α -integrin (including ITGA2 and ITGA3) was observed in 231BR (green color denotes the down-regulation of integrin in 231 and up-regulation in 231BR). According to IPA, the integrin signaling pathway was inhibited in 231 when compared to 231BR while it was activated in 231BR. The higher activity of the integrin signaling pathway could positively affect cell adhesion and mobility, which are important in brain metastasis. Therefore, the up-regulation of integrin may facilitate the brain metastatic ability of 231BR. As shown in Fig. S3, the VEGF signaling pathway was also predicted to be inhibited in 231, which meant more activity in 231BR. VEGF can regulate vascular permeability by activating VEGF receptors³⁵. Therefore, the activation of the VEGF pathway may help 231BR cells to pass through the blood-brain barrier.

In functional annotation, genes that are associated with specific biological functions can be grouped and the activation status of their relative functions can be predicted by the up/down-regulation of these genes. Figure 4 depicts the metastasis function annotation in 231 vs. 231BR. Most of the relative genes were down-regulated in 231, resulting in the inhibition of the two metastasis functions. Therefore, the metastasis of breast cancer cells and metastasis of cells were predicted to be more active in 231BR (these two functions were inhibited in 231 and

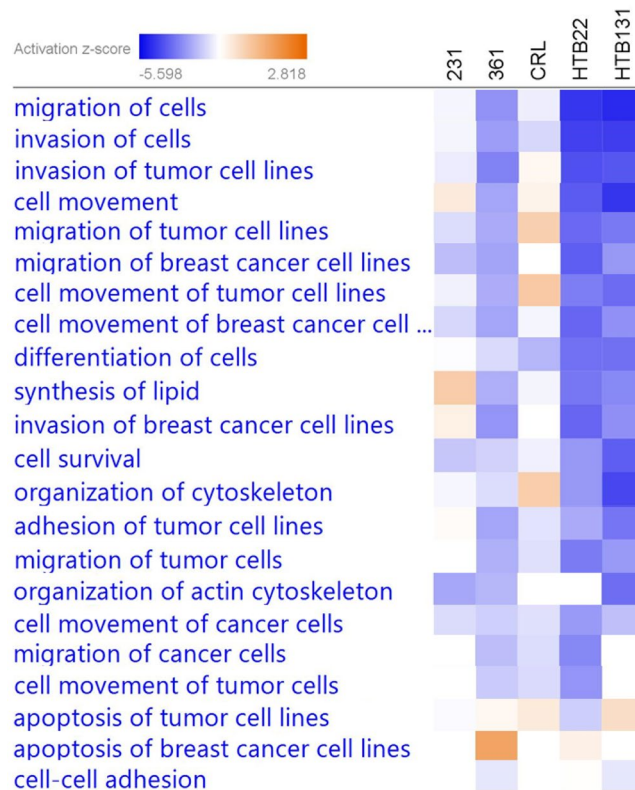


Figure 2. Activation/inhibition of molecular functions of different cell lines when compared to 231BR. Activation z-scores were generated by IPA based on the differential gene expressions in each cell line relative to 231BR. The [networks, functional analyses, etc.] were generated through the use of IPA (QIAGEN Inc., <https://www.qiagenbioinformatics.com/products/ingenuity-pathway-analysis>). Red color denotes the activation of corresponding functions in each cell line, and the inhibition in 231BR. Blue color denotes the inhibition in each cell line, and the activation in 231BR.

activated in 231BR), meaning that 231BR had a better metastasis capacity than 231. Moreover, Fig. S4 depicts the most significant upstream regulator as MTF1, which was predicted by IPA based on the changes of observed genes. The down-regulation of MTF1 could result in the down-regulation of a series of genes, thus decreasing the activity of cell movement.

Glycosylation gene enrichment. Aberrant glycosylation has been proven to be associated with breast cancer metastasis⁷. However, only one glycosylation gene has been demonstrated to be associated with breast cancer brain metastasis⁵. In this study, we manually selected the glycosylation genes from differentially expressed genes and investigated their expression change to find the possible glycozymes which may positively facilitate breast cancer brain metastasis. Overall, 91 glycosylation genes were selected for further study. Fig. S5 depicts the differential expressions of these glycozymes. Among them, 12 genes (ST6GALNAC2, ST6GALNAC5, FUT5, POFUT1, B4GALNT3, B4GALNT4, B4GALT6, MGAT3, MGAT4A, XYLT2, GALNT13, and PIGK) showed the same up- or down-regulation in at least four comparisons, making the expression of these genes unique in 231BR. These 12 genes can influence sialylation (ST6GALNAC2, ST6GALNAC5), fucosylation (FUT5), glycan anchor (PIGK), and complex glycan structure synthesis (B4GALNT3, B4GALNT4, B4GALT6, MGAT3 and MGAT4A), which make the unique glycan expression pattern of 231BR possible. Thus, the alteration of the expression of these genes may play important roles in breast cancer brain metastasis. However, the function of these genes need to be verified in future studies. Figure S6A depicts the functional annotation of the glycosylation genes in IPA. Most of the glycosylation genes cannot be predicted by IPA due to insufficient studies of these genes. However, based on the known correlation between glycosylation genes and their functions, sialylation and glycosylation were enhanced in 231BR and inhibited in 231, which was consistent with the cell line glycomics results (see section “Cell Line Glycomics Analysis”). Figure S6B shows the interactions of glycosylation genes/proteins and other related genes/proteins. The glycosylation alteration caused by the expression change of glycozymes may influence the integrin pathway and the cancer mechanism pathway *via* ITGAs.

Cell line proteomics analysis. Unlabeled bottom-up proteomics of six cell lines were achieved using nano-C18-LC-MS/MS with a 2-hour LC gradient and 60 K MS resolution. The optimized LC condition and high MS resolution permitted a total of 1569 proteins to be identified and quantified across the six cell lines (Supplementary Information Table S1A). Brain-seeking cell line 231BR was applied as a baseline and compared

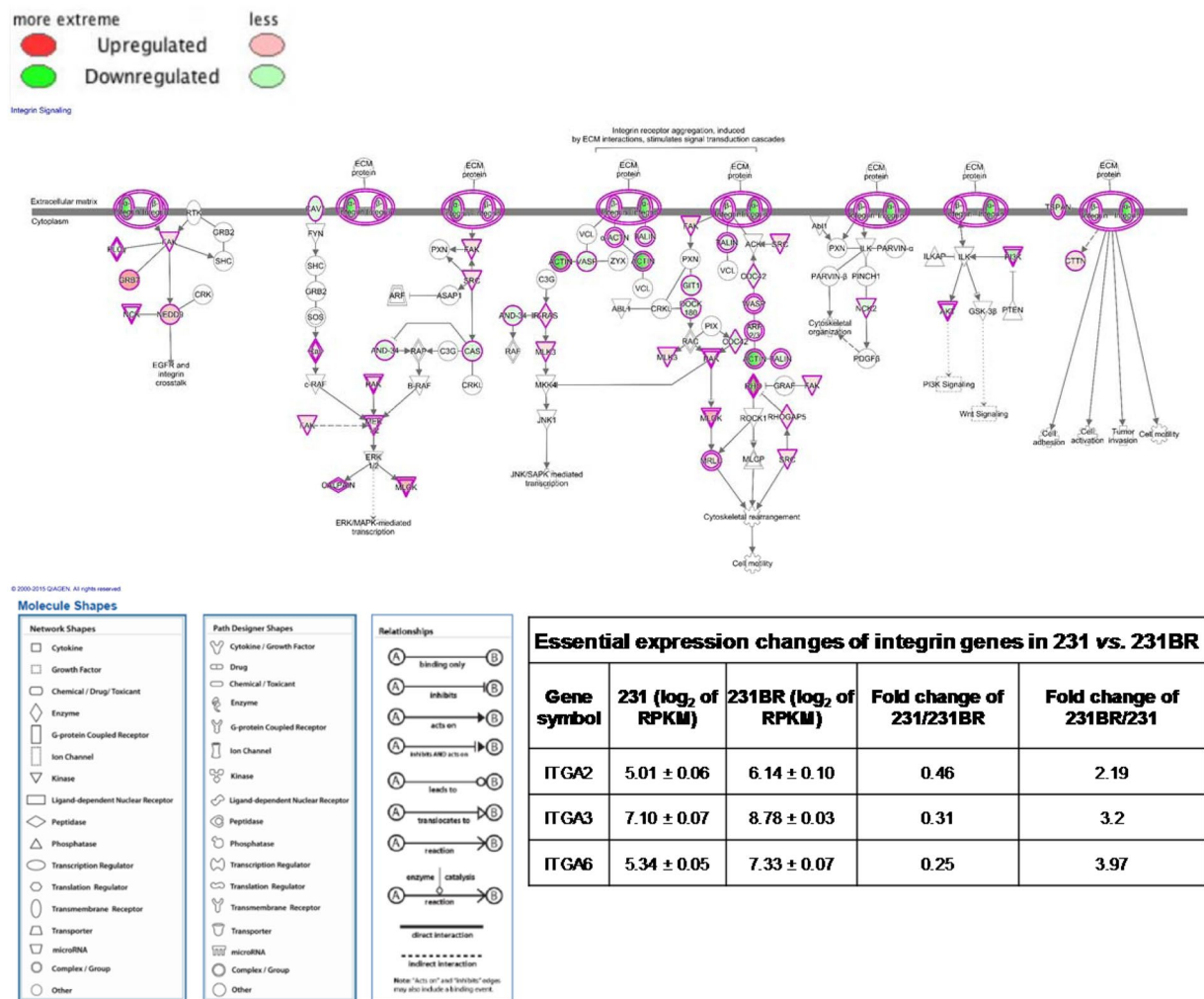


Figure 3. Integrin signaling pathway of 231 vs. 231BR. Green color denotes the down-regulation of genes in 231, and up-regulation in 231BR. The down-regulation of integrin in 231 (up-regulation in 231BR) may affect the cell adhesion, activation, invasion and mobility functions. The inserted table depicts the expression changes of essential integrin genes in 231 vs. 231BR.

with the other cell lines. In total, 1096 proteins exhibited significantly differential expressions ($p < 0.05$) when compared with 231BR (Supplementary Information Table S1B). Among the different cell lines, 231/231BR had the fewest differentially expressed proteins (151 proteins) due to the high similarity between 231 and 231BR. There were 581, 508, 515, and 454 significantly altered proteins ($p < 0.05$) in 361, HTB22, HTB131, and CRL, respectively. Figure 5A depicts the Venn diagram of the significant proteins in five comparisons. 50 proteins exhibited significant up/down-regulation in all comparisons, which made their expressions unique in 231BR (Fig. 5B). Among these, 37 proteins showed the same up- or down-regulation in all five comparisons; 30 proteins were up-regulated in 231BR, while 7 of them were down-regulated. Lymphocyte cytosolic protein 1 (LCP1) was the most significant up-regulated protein in 231BR, while clathrin heavy chain (CLTC) was the most significant down-regulated protein (Fig. S7). LCP1 belongs to the class of actin-bundling proteins. It is usually ectopically expressed in cancer cells³⁶ and has been reported to promote tumor metastasis with phosphorylation³⁷. In this study, we observed the obvious up-regulation of LCP1 in the brain-seeking cell line 231BR, suggesting that the up-regulation of LCP1 may play a role in breast cancer brain metastasis. CLTC is a clathrin heavy chain that participates in endocytosis, membrane trafficking, and tumorigenesis^{38,39}. However, few reports have associated CLTC with cancer metastasis. Although we are unable to conclude that CLTC contributes to breast cancer brain metastasis, a significant down-regulation in brain specific cell line 231BR was observed, which may provide useful information for further studies. In addition, the up-regulation of adhesion G protein-coupled receptor E5 (CD97) was also observed in 231BR.

Unsupervised principal component analysis (PCA). Principal component analysis is a statistical approach to simplify high dimensional complex data sets. It uses the orthogonal transformation to convert possibly correlated variables into principal components which are linearly uncorrelated, thus displaying the similarities and differences of data groups by plots on a map⁴⁰. Overall, 1569 proteins, including 1096 differentially expressed

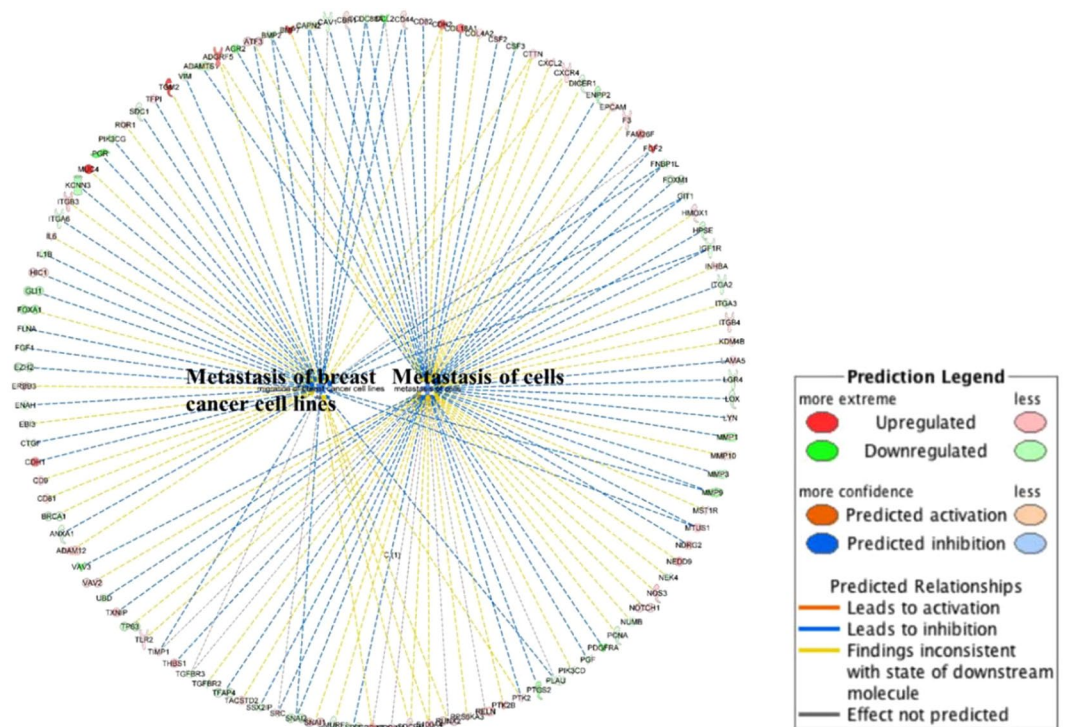


Figure 4. Functional annotation. Genes that exhibited different expression changes in 231 vs. 231BR were annotated to metastasis of breast cancer cell lines function and metastasis of cells function, respectively, using IPA. The networks and functional analyses were generated through the use of IPA (QIAGEN Inc., <https://www.qiagenbioinformatics.com/products/ingenuity-pathway-analysis>). Both functions were inhibited in 231 when compared to 231BR. Lines between genes and functions represent how these changes of genes affect the corresponding functions. Blue line denotes that the corresponding gene change cause an inhibition to the function in 231, and the activation in 231BR. Red line denotes that the corresponding gene change cause an activation to the function in 231, and the inhibition in 231BR, while yellow line denotes the inconsistent changes of corresponding genes and their downstream molecules (this prediction was performed by matching the gene changes observed in this experiment to those in IPA's database).

proteins, were submitted to PCA. Figure S8 depicts the unsupervised PCA plots of quantitative proteomic data from six cell lines. Each cell line had three biological replicates and was well-grouped. The differences between the six cell lines can be observed through primary principal component (PC1) and secondary principal component (PC2). The three biological replicates of each cell line were closely located, suggesting the satisfactory reproducibility of the cell culturing and sample preparation procedures used in this study. All cell line groups were separated, denoting the proteome differences among the different cell lines. As expected, 231BR showed the highest similarity with 231. Meanwhile, 231 and 231BR, considered to be the most invasive breast cancer cell lines, showed significant distinction from other cell lines. The brain cancer cell line CRL exhibited the largest difference from other cell lines due to its different cancer type. The PCA result suggests that the proteomic data from all cell lines are reliable and obviously distinct, which may be crucial for a better understanding of breast cancer brain metastasis.

Correlation between proteome and transcriptome. A biosystem is a complicated and self-adjustable system. After transcription, mRNAs experience translation and post-translational modification to generate functional and structural proteins and enzymes. Many regulatory processes occur between mRNAs and proteins. Thus, identifying the correlation between transcriptome and proteome is necessary to achieve a better understanding of bioprocesses, herein emphasizing brain metastasis. Proteins which had significantly differential expression ($p < 0.01$) were selected to correlate with their transcripts. Most proteins exhibited the same regulation trend as their transcription changes (Supplementary Information Table S2). On average, 83.8% of protein expressions matched the gene expressions. 79.3%, 85.4%, 87.1%, 84.2%, and 83.2% of the differential expressed proteins showed the same expression changes as their transcripts in 231, 361, HTB131, HTB22, and CRL compared with 231BR, respectively. Although the fold change of protein expression and gene expression had a slight difference, it could be expected due to the complicated regulations between genes and proteins and the existence of isoenzymes. Figure S9 depicts the correlation between significant protein expression changes ($p < 0.01$) and transcript expression changes in 231 vs. 231BR. 73 out of 92 proteins showed the same up- or down-regulation as their transcripts (Fig. S9). 40 proteins were observed to be up-regulated in 231BR, while 52 were down-regulated. Among the up-regulated proteins, 2',3'-cyclic nucleotide 3' phosphodiesterase (CNP) exhibited the largest expression fold change, which was 14 times increased, although the increase of its gene

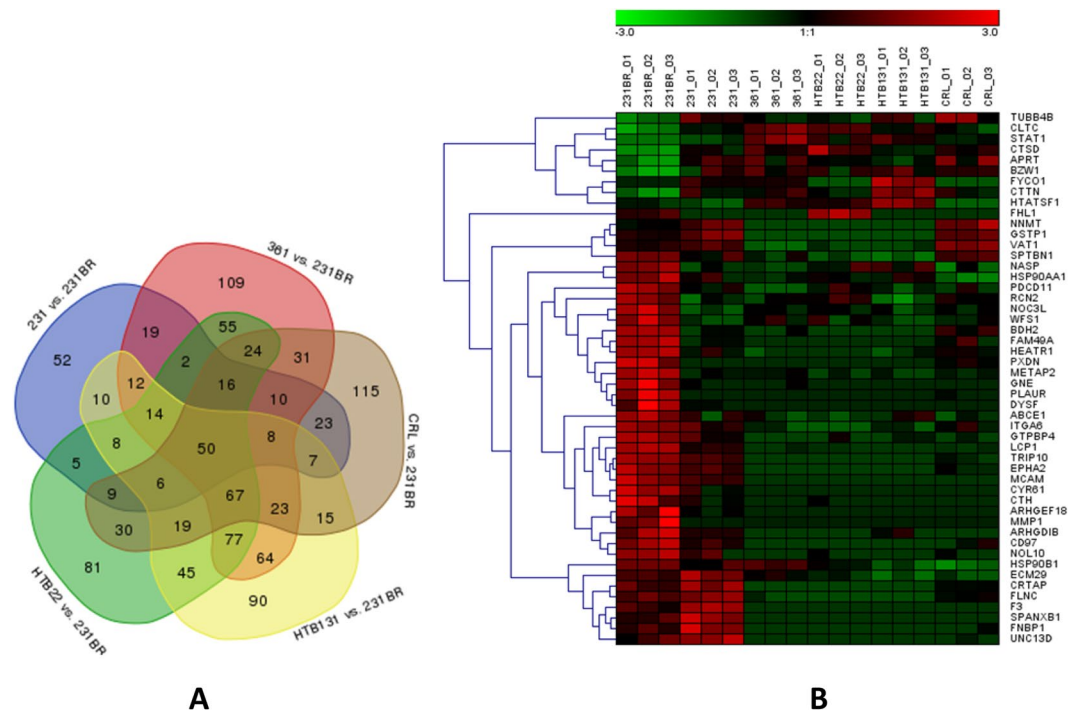


Figure 5. Proteins that had the significant expression alterations in five cell lines when compared to 231BR (N = 3 biological triplicates). **(A)** Venn diagram of differentially expressed proteins in cell lines compared with 231BR, with the p-value less than 0.05; **(B)** Heatmap of 50 common proteins that exhibited significant expression changes in all five comparisons. Color codes represent the expression levels of proteins across six cell lines. Red color denotes the high expression level while green color denotes the low expression level.

expression was not so large. CNP was reported as overexpressed in the brain when compared with other tissues⁴¹. It was initially reported as phosphodiesterase, but has recently been reported to have an association between its differential expression and glioblastoma patient survival⁴². The increase of both CNP transcript and protein in 231BR can most likely be associated with its specific brain metastatic capacity. Interestingly, when looking into the correlation between protein expressions and corresponding gene expressions in all cell lines relative to 231BR, pearson correlation coefficients between protein and gene expression changes were less than 0.77, as shown in Fig. S10. Silimilar to 231/231BR comparison, most of proteins and genes exhibited the same up- or down- regulation trends. The low pearson correlation coefficient maybe due to the fact that the fold changes of genes did not match the fold changes of proteins (e.g. the fold change of CNP protein was 14.2 times, while the fold change of CNP gene was only 1.2 times).

According to GO clustering analysis, differential expression proteins were enriched into different molecular and cellular functions with a high significance ($p < 1E-07$). Similar to their differential expressed genes (Fig. S1), proteins associated with cell movement, cell growth and proliferation, cell death and survival, and cellular assembly and organization exhibited differential expression compared to 231BR (Fig. S11). The GO results stressed the importance of genes/proteins associated with cell movement in the breast cancer brain metastasis process. When we examined proteins with a significance p-value of less than 0.05, we also observed the up-regulation of α -integrin (ITGA) in 231BR, which affected the cell movement and cell adhesion *via* many pathways based on IPA, such as the integrin signaling pathway (Fig. S12, green color denotes the down-regulation of integrin in 231, and up-regulation in 231BR). When we considered cell-to-cell signaling and interactions which were generated by IPA (Fig. S13), the integrin proteins also exhibited a central role in the network. This suggests that they might be important in relative bioprocesses, including brain metastasis.

Although many important correlations between transcriptome and proteome from six cell lines were observed, we could not detect glycosylation proteins in cell line proteomics. The reason may be because the expression level of glycosylation proteins (enzymes) are relatively low in cells, such that it is difficult to detect them using our LC-MS/MS conditions without specific enrichment. Therefore, we investigated the N-glycomic alterations among different cell lines to achieve a better understanding of the functions of glycans and glycozymes in breast cancer brain metastasis.

Cell line glycomics analysis. N-Glycomic Analysis of six cell lines was performed using LC-MS/MS. The most obvious advantages of the LC-ESI-MS method are its high sensitivity and adequate structural information. The sample preparation protocol provided a relatively complete release of N-glycans from cells; the solid-phase permethylation enabled higher ionization efficiency of neutral and sialylated glycans and better separation on reversed phase LC with a minimum sample loss; the on line purification removed salts and impurities, which provided a better ionization of glycans; the LTQ Orbitrap Velos mass spectrometer offered a high resolution and

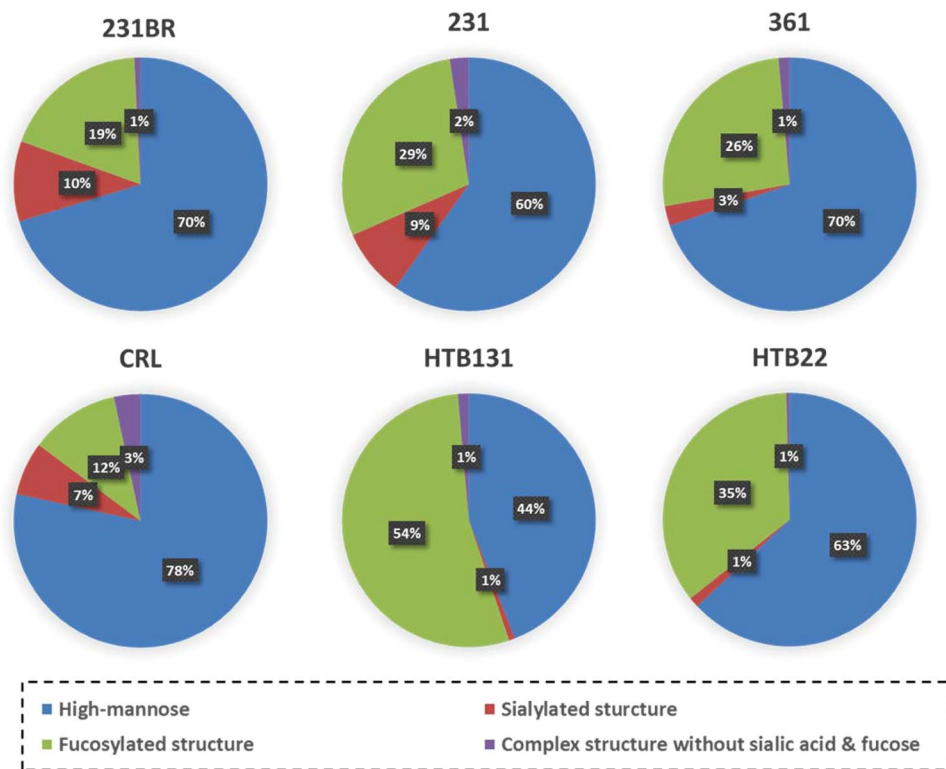


Figure 6. Distribution changes of different glycan types from six cell lines.

mass accuracy, and made the identification of low abundant glycan structures possible. All of the elements of this LC-ESI-MS method contributed to the sensitive and reproducible detection of N-glycans released from cell lines. Overall, 63 structures were identified from cell line samples, with a mass accuracy less than 5 ppm. The number of identified N-glycans was obviously higher than other techniques such as MALDI-TOF⁴³. Among these structures, 60 N-glycans exhibited significantly different expressions in at least one comparison ($p < 0.05$) (Supplementary Information Table S3). The highest number of N-glycans were identified in 231BR (63 N-glycans), followed by CRL and 231 (61 and 56 N-glycans, respectively). However, there were fewer N-glycans detected in 361, HTB131, and HTB22 (48, 25, and 33 N-glycans, respectively). The superior glycan numbers in 231BR and 231 are likely because they are considered to be the most invasive breast cancer cell lines⁴⁴. This result supports the contribution of glycosylation to breast cancer brain metastasis.

Correlation between transcriptome and glycome. Because of the unique brain metastatic capacity of 231BR, it was used as the baseline to compare with the other cell lines. The significant glycosylation alterations in 231BR may contribute to its brain metastasis. The overall intensity of N-glycans identified in 231BR was higher than in other cell lines (Supplementary Information Table S3), which suggests a higher glycosylation level in brain metastasis. Detailed information can be determined when a closer investigation is conducted between these cell lines, especially when identified N-glycans are grouped by their composition characteristics. Figure 6 shows the distribution of different glycan types in different cell lines. 231BR and 231 have more sialylated structures than other cell lines. According to glycogene expressions (Fig. S5), most of the genes associated with sialylation (such as ST6GALNAC5 and ST6GAL1) were down-regulated in other cell lines, which resulted in the higher sialylation of 231BR and 231. Moreover, the sialylated glycans in 231BR were statistically more than 231 ($p < 0.04$). Although some sialylation genes were observed to be up-regulated in 231, two important sialylation genes, ST6GALNAC5 and ST6GAL1, were 13.5 fold and 7.0 fold down-regulated, respectively; only one relative gene had a significant up-regulation (ST6GALNAC2, 14.3 fold up-regulated). Considering that ST6GALNAC5 has been demonstrated to contribute to the penetration of breast cancer cells through the blood-brain barrier, thus facilitating the brain metastasis⁵, ST6GALNAC5 may have a higher catalyzing activity than its isoenzyme ST6GALNAC2. In contrast, ST6GALNAC2 was reported to be a suppressor of breast cancer metastasis⁴⁵. The down-regulation of ST6GALNAC2 may also contribute to breast cancer brain metastasis by helping cancer cells to pass through the blood-brain barrier. The up-regulation of ST6GALNAC5 and ST6GAL1 resulted in a higher sialylation level in 231BR. In addition, sialylated glycan structures appear to increase with the progression of the cancer stage⁴³ and have the capacity to mediate cellular interactions⁴⁶. Therefore, we believe that sialylation and sialylated glycans play an important role in breast cancer brain metastasis.

Regarding individual N-glycans, the glycan expressions (distributions) were altered among the different cancer cell lines (Fig. 7 and Supplementary Information Table S3). Most of the glycan structures showed no significant change between 231 and 231BR ($p > 0.05$). However, highly sialylated structures such as

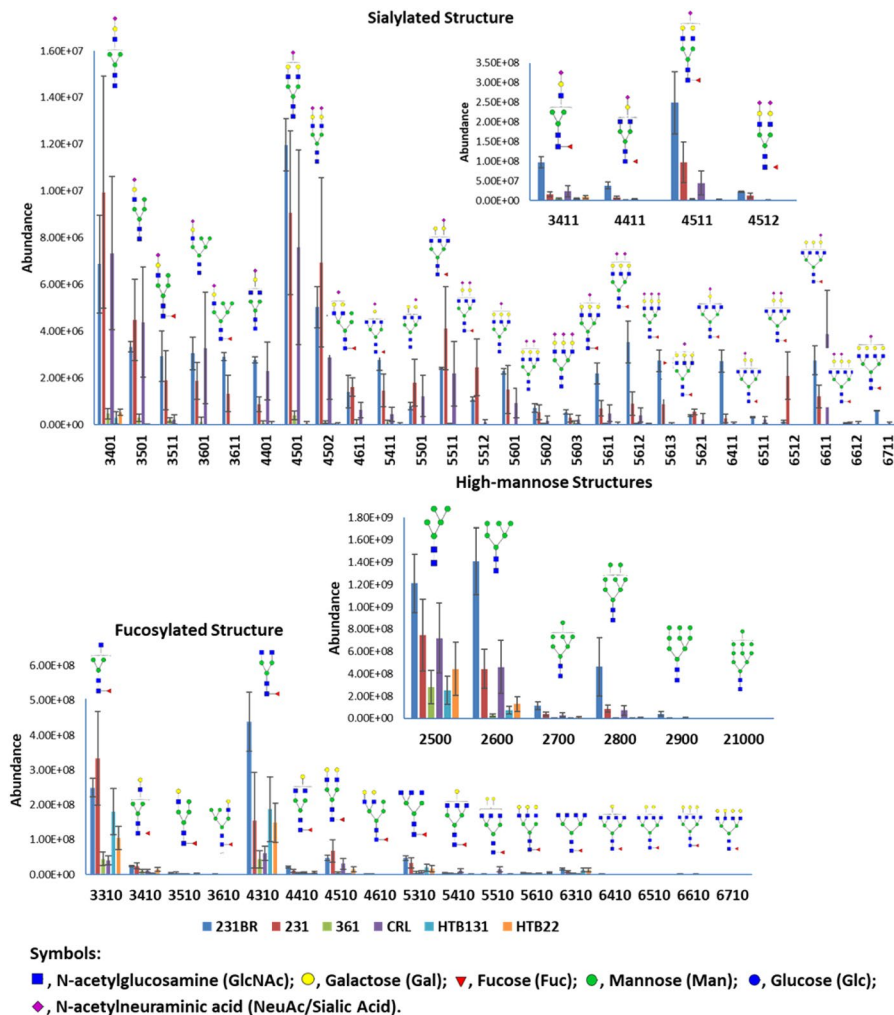


Figure 7. Absolute abundance and putative glycan structures from six cell lines using LC-MS/MS. Error bars denote the standard deviations from 3 biological triplicates. Glycan composition is presented by a 4-digit code. Nomenclature: a 4-digit code is employed to represent glycan structures. The 1st number denotes the number of N-acetylglucosamine (HexNAc); the 2nd number denotes the number of hexose (Hex); the 3rd number denotes the number of fucose (DeoxyHex); the 4th number denotes the number of sialic acid (NeuAc). For example, 4512 represents glycan HexNAc₄Hex₅DeoxyHex₁NeuAc₂.

HexNAc₅Hex₆DeoxyHex₁NeuAc₁, HexNAc₅Hex₆DeoxyHex₁NeuAc₂, and HexNAc₅Hex₆DeoxyHex₁NeuAc₃ were overexpressed in 231BR. For the high abundant sialylated glycans, as shown in the inset of Fig. 7, 231BR also exhibited a significantly higher expression level (p -value < 0.05) in all four glycans (HexNAc₃Hex₄DeoxyHex₁NeuAc₁, HexNAc₄Hex₄DeoxyHex₁NeuAc₁, HexNAc₄Hex₅DeoxyHex₁NeuAc₁, and HexNAc₄Hex₅DeoxyHex₁NeuAc₂). The overexpression of these glycans may be important for breast cancer brain metastasis. In addition, the sialylated structures in 231 and 231BR were predominant compared to other cell lines, which matched the distributions of different glycan types in cell lines (Fig. 6). Similar results can be observed for high-mannose structures. 231BR expressed more high-mannose structures, especially MAN6. According to transcriptomic data, most of the mannosidase genes (such as MANBA or MAN1) were up-regulated in the 361, HTB131 and HTB22 cell lines. Since mannosidase trims a high-mannose structure to form a core structure, the up-regulation of these genes may result in the decrease of high-mannose structures. For the fucosylated structures, most of the fucosylated glycans didn't show any significant difference, especially the low abundant structures. In addition, according to Fig. 6, the distribution of the fucosylated types were relatively low in 231BR. This result may be caused by the down-regulation of the fucosylation genes (FUT) in 231BR (Fig. S5).

Through the correlation of transcriptomics, cell line glycomic analysis has revealed the essential role of sialylation and sialylated N-glycans in breast cancer brain metastasis. Seven structures, both fucosylated and sialylated (HexNAc₅Hex₆DeoxyHex₁NeuAc₁, HexNAc₅Hex₆DeoxyHex₁NeuAc₂, HexNAc₅Hex₆DeoxyHex₁NeuAc₃, HexNAc₃Hex₄DeoxyHex₁NeuAc₁, HexNAc₄Hex₄DeoxyHex₁NeuAc₁, HexNAc₄Hex₅DeoxyHex₁NeuAc₁, and HexNAc₄Hex₅DeoxyHex₁NeuAc₂) may have contributed to the brain metastatic process.

Conclusion

In this study, we investigated and correlated transcriptomics, proteomics, and glycomics of five breast cancer cell lines and one brain cancer cell line. The brain-seeking cell line 231BR was used as a reference to compare with the other cell lines. Unique expressions in 231BR were highlighted as potential biomarkers in breast cancer brain metastasis. IPA of transcriptomic data suggested the highest levels of cell movement, invasion, and metastatic ability in 231BR. With agreement between transcriptomics and proteomics, it appears that the integrin signaling pathway with up-regulation of ITGA may play an important role in breast cancer brain metastasis. Alpha-V integrin (ITGAV) has been associated with brain metastasis, but the relationship between ITGA2/3 and breast cancer brain metastasis has not well known yet. The up-regulation of ITGA2 and ITGA3 observed in this study in brain seeking breast cancer cell line 231BR may provide a potential indicator for breast cancer metastasis, and better understand the role integrin plays in this critical bioprocess. 12 glycosylation genes that could affect glycan patterns were observed to have unique expressions in 231BR, and the protein-protein interactions of these genes showed their possible correlation to the integrin pathway. Based on glycomics results, the most glycans were identified in 231BR, denoting the significance of glycosylation in breast cancer brain metastasis. A higher sialylation level may be essential in breast cancer metastasis. Several highly sialylated glycans including HexNAc₃Hex₆DeoxyHex₁NeuAc₁, HexNAc₃Hex₆DeoxyHex₁NeuAc₂, HexNAc₃Hex₆DeoxyHex₁NeuAc₃, HexNAc₃Hex₄DeoxyHex₁NeuAc₁, HexNAc₄Hex₄DeoxyHex₁NeuAc₁, HexNAc₄Hex₅DeoxyHex₁NeuAc₁, and HexNAc₄Hex₅DeoxyHex₁NeuAc₂ were observed to be overexpressed in 231BR, suggesting an association of these glycans with breast cancer brain metastasis. The increase of sialylation level and some of the sialylated glycans such as HexNAc₃Hex₆DeoxyHex₁NeuAc₂ have been observed in breast cancer progression⁴⁷. In addition to breast cancer progression, in this study, the increase of sialylation and aforementioned seven glycans could be considered as potential indicators for breast cancer brain metastasis. These results provide important clues for a better understanding of breast cancer brain metastasis. Although these unique expressions of genes/proteins/glycans were observed in breast cancer brain seeking cell line, we cannot conclude that they are the biomarkers of breast cancer brain metastasis due to the limited sample size. They need to be further investigated and verified using clinical samples.

Methods

Materials and reagents. Breast cancer cell line MDA-MB-231, MDA-MB-361, HTB131, HTB22 and CRL-1620 were purchased from American Type Culture Collection (ATCC). Breast cancer cell line MDA-MB-231BR was a generous gift from Dr. Paul Lockman (Texas Tech University Health Sciences Center, School of Pharmacy, Amarillo, TX). RPMI-1640 Medium, Eagle's Minimum Essential Medium (EMEM), Dulbecco's Modified Eagle's Medium (DMEM), Fetal Bovine Serum (FBS) and Penicillin-Streptomycin Solution (100X) were purchased from ATCC. Corning Trypsin EDTA 1 × (0.25% Trypsin/2.21 mM EDTA) was purchased from Fisher. Cell culturing plates and flasks were purchased from Corning. Mammalian total RNA Extraction Kit was purchased from Sigma-Aldrich. Borane-ammonia complex, dimethyl sulfoxide (DMSO), sodium hydroxide beads, iodomethane, and formic acid (FA) were obtained from Sigma-Aldrich (St. Louis, MO). PNGase F and digestion buffers, including 10 × G7 reaction buffer (0.5 M sodium phosphate), RXN buffer, 10 × glycoprotein denaturing buffer (5% SDS, 0.4 M DTT), and NP-40 were acquired from New England Biolabs (Ipswich, MA). HPLC grade methanol, isopropanol, and acetic acid were procured from Fisher Scientific (Pittsburgh, PA). Acetonitrile was obtained from JT Baker (Phillipsburg, NJ). HPLC grade water and sodium hydroxide were purchased from Mallinckrodt Chemicals (Phillipsburg, NJ).

Cell cultivation. MDA-MB-231 and HTB131 were grown in RPMI-1640 Medium (with 10% FBS and 2% Penicillin-Streptomycin); MDA-MB-231BR and CRL-1620 were grown in DMEM (with 10% FBS and 2% Penicillin-Streptomycin); MDA-MB-361 was grown in RPMI-1640 Medium (with 20% FBS and 2% Penicillin-Streptomycin); HTB22 was grown in EMEM (with 10% FBS and 2% Penicillin-streptomycin). Cell lines were taken out from liquid nitrogen and incubated at 37 °C with 5% CO₂ for 7 days. When cell confluence reached 80%, cells were subcultured into 175 cm² flask and incubated for another 3–7 days. After reaching 80% cell confluence, cells were washed by PBS and then detached by 5 mL trypsin-EDTA solution at 37 °C for 5 min. Then 10 mL of fresh medium was added to neutralize trypsin and cells were collected by centrifugation. These collected cells were used for transcriptomic, proteomic, and glycomic analysis.

Transcriptomics analysis. Fresh cells were washed by PBS and then total RNAs were extracted using the Mammalian total RNA Extraction Kit, following vendor's instruction. 5 μL of total RNA solution was taken out to do a concentration and quality measurement. RNA concentration and quality were measured by a Nanodrop 1000 instrument. Total RNAs which had the OD₂₆₀/OD₂₈₀ within 1.9–2.2 were considered to be good quality and were utilized for the further transcriptomic analysis. Each cell line had three biological replicates.

After quantification of total RNA, a multiplexing method was employed for the high throughput RNA sequencing. All 18 different samples from different cell lines (6 cell lines with 3 biological triplicates of each cell line) were ligated with different combinations of indexes. Adaptor ligated cDNA libraries were prepared using the NeoPrep automated microfluidic instrument with TruSeq Stranded RNA Library Prep Kit (Illumina). PCR enriched libraries were quantified by Qubit and equimolar indexed libraries (different samples had different indexes for multiplexing) were pooled. Pooled libraries were qualitatively checked using the Agilent TapeStation 2200 and quantified using Qubit. The libraries were then diluted to a final concentration of 9 pM and spiked with 2% phiX libraries (Illumina control). Illumina HiSeq. 2500 was used for next generation transcriptomic sequencing (NGS). The reads generated from different samples could be distinguished based on their different indexes. The quality of the reads were determined by FastQC software. The reads were quality filtered and mapped on the human genome using NGen software (DNA STAR, Madison, WI). RPKM normalization was then

performed prior to the quantification of gene expression by QSeq (DNA STAR). Genes showing >2.0 fold change and $p < 0.01$ were selected as differentially expressed genes and mapped into biological pathways using Ingenuity Pathway Analysis (IPA) software. Differentially expressed gene cluster annotation and function enrichment were achieved by DAVID. Glycosylation pathways were analyzed by KEGG.

Proteomic analysis. Three biological replicates of cells from different cell lines were suspended in 50 mM ABC buffer (containing 5% SDC) and then homogenized by a beads beater homogenizer. After lysing, 5 μ L of cell lysate was taken out for protein assay using the micro BCA method (Thermo Scientific/Pierce, Rockford, IL) according to the manufacturer's instruction. The remaining cell lysate was denatured in an 80 °C water bath for 30 min. Then the denatured proteins were reduced by 5 mM DTT at 60 °C for 45 min, alkylated by 20 mM IAA at 60 °C in darkness for 45 min, followed by quenching the reaction by adding another 5 mM DTT and incubating at 60 °C for 30 min. Next, protease Trypsin-Lys C (Promega, MS gold grade) was added to the sample with a 1:25 enzyme to protein ratio and incubated at 37.5 °C for 18 h. After tryptic digestion, formic acid was added (final concentration is 0.5%) to remove the SDC⁴⁸. After centrifugation at max speed for 10 min, supernatant which contained peptides was collected and dried. Samples were then resuspended with 5% ACN water and ready for proteomic analysis.

Proteomics analysis was performed by LC-MS/MS (Thermo Ultimate 3000 nano system coupled with LTQ Orbitrap Velos)⁴⁹. The online purification was performed using a C18 Acclaim PepMap 100 trapping column (75 μ m I.D. \times 2 cm, 3 μ m particle sizes, 100 Å pore sizes, Thermo Scientific, San Jose, CA). Peptide separation was achieved using a C18 Acclaim PepMap RSLC column (75 μ m I.D. \times 15 cm, 2 μ m particle sizes, 100 Å pore sizes, Thermo Scientific, San Jose, CA). 1 μ g of proteins was injected for the analysis under 60,000 resolution. A 120-minute LC method was used for the separation of the digested peptide sample. Briefly, the LC gradient of solvent B in LC-MS/MS analysis was: 5% over 10 min, 5–20% over 55 min, 20–30% over 25 min, 30–50% over 20 min, 50–80% over 1 min, 80% over 4 min, 80–5% over 1 min and 5% over 4 min. Two scan events were applied for the data-dependent acquisition. The first scan event was a full MS scan at a resolution of 60 K. The second scan event was CID MS/MS scan which fragmented the top 10 intense precursor ions that were selected from the first scan, using the normalized collision energy (CE) of 35%, Q value of 0.25, and 10 ms activation time.

Proteomic Data was then converted to a mascot file by Proteome Discover version 1.2 software (Thermo Scientific, San Jose, CA) and then searched against a UniProt database (2014_06, Homo sapiens, 20214 entries) in MASCOT version 2.4 (Matrix Science Inc., Boston, MA). The MASCOT result was verified using Scaffold (version Scaffold_3.6.3, Proteome Software Inc., Portland, OR). Peptide identification probability and protein identification probability were set to 95% and 99%, respectively, and the matched peptide number was 2. A normalized spectra count was applied for quantitation. Then an additional filter was utilized to remove the proteins which were detected only once in three replicates. Quantitative proteomic data sets from different cell lines were compared and submitted to IPA for pathway analysis and protein-protein interactions.

Glycomic analysis. Three tubes of independently harvested breast cancer cells from each cell line were applied in this research. Each tube contains approximately 10^6 cells. Then each tube is resuspended by a PBS buffer (with 5% SDC) to 200 μ L. Next, cell samples were sonicated in iced water for 60 minutes to open the cell and expose the cytosolic proteins. After sonication, 5 μ L solution was taken out for protein assay using a BSA protein assay kit (Thermo Scientific, Pittsburgh, PA), which is an essential step for normalization injection amount by protein amount. The remaining 95 μ L solution was subjected to the 80 °C water bath (Thermo Scientific, Pittsburgh, PA) for 30 minutes for denaturation. The solution was cooled down to ambient temperature and 5 μ L of ten-times diluted PNGase F enzyme was added to each vial. The samples were incubated in a 37 °C water bath for 18 hours. After PNGase F digestion, 2 μ L of formic acid was added to the sample solution to precipitate SDC. Then supernatants were collected after centrifuging at maximum speed for 10 min and dried. Dried samples were dissolved by 50 μ L of water and dialyzed overnight with the MWCO 500–1000 to remove salt. After dialysis, samples were dried and then resuspended in 90% ethanol to remove protein. After standing in –20 °C for 30 min, samples were centrifuged and supernatants (purified N-glycans) were collected and dried. The purified N-glycans were reduced as per the protocol previously reported⁵⁰, 20 μ L solution containing 10 μ g/ μ L ammonium-borane complex was added to each vial, the vials were incubated in a 60 °C water bath for 60 minutes, and then the mixture was washed by methanol to remove excess borane salts. After that, the reduced N-glycans were permethylated by utilizing solid-phase permethylation^{51–54}. Sodium Hydroxide beads were added to spin columns (Harvard Apparatus, Holliston, MA), dried samples were resuspended in 30 μ L Dimethyl sulfoxide (DMSO) and 1.2 μ L water mixed with 20 μ L Iodomethane, and the mixture was added to the spin column from the top. After 25 minutes' incubation, another 20 μ L Iodomethane was added to the column and incubated for another 15 minutes. The spin columns were centrifuged (Thermo Scientific, Pittsburgh, PA, USA) at 1.8k rpm for 2 minute, and then 30 μ L ACN was added to the columns for the final elution. These elutes were collected and dried under vacuum (Labconco, Kansas City, MO). Finally, the samples were resuspended by 20% ACN water solution (with 0.1% formic acid) to 24–48 μ L, corresponding to their protein amount as determined by BSA protein assay.

A Dionex Ultimate 3000 nano LC system (Thermo Scientific, Pittsburgh, PA, USA) was utilized in the LC-MS analysis. The separation was accomplished on an Acclaim C18 nano column (Thermo Scientific, Pittsburgh, PA, USA) under a flow rate of 350 μ L/min. An optimized gradient elution condition was applied. The LTQ Orbitrap Velos (Thermo Scientific, Pittsburgh, PA, USA) was utilized for the detection of N-glycans, and the mass spectrometry was operated at positive mode with m/z range of 700–2,000. The instrument was set at data dependent acquisition mode with 8 MS/MS scans.

The analyte peaks were extracted from the full MS scans by using Xcalibur Qual Browser (Thermo Fisher Scientific) with a 10 ppm mass tolerance. The identification and quantitation of N-glycans were performed by an in-house software, Multiglycan, with a 5 ppm mass tolerance and manually checked based on the full MS and MS2.

Integrated “Omics” analysis. Transcriptomic, proteomic, and glycomic data were correlated to each other to find the inner relationship among those different level changings and the possible biological mechanisms of breast cancer brain metastasis.

Data availability

Transcriptomics data were submitted to NCBI bioproject PRJNA383270; BioSample accessions: SAMN06806180, SAMN06806181, SAMN06806182, SAMN06806183, SAMN06806184, SAMN06806185. Proteomics data are available *via* ProteomeXchange with identifier PXD010506.

Received: 28 June 2019; Accepted: 7 November 2019;

Published online: 22 November 2019

References

- Lin, N. U., Bellon, J. R. & Winer, E. P. CNS metastases in breast cancer. *J Clin Oncol* **22**, 3608–3617, <https://doi.org/10.1200/JCO.2004.01.175> (2004).
- Adamczyk, B., Tharmalingam, T. & Rudd, P. M. Glycans as cancer biomarkers. *Biochim Biophys Acta* **1820**, 1347–1353, <https://doi.org/10.1016/j.bbagen.2011.12.001> (2012).
- Rostami, R., Mittal, S., Rostami, P., Tavassoli, F. & Jabbari, B. Brain metastasis in breast cancer: a comprehensive literature review. *J Neurooncol* **127**, 407–414, <https://doi.org/10.1007/s11060-016-2075-3> (2016).
- Arshad, F., Wang, L., Sy, C., Avraham, S. & Avraham, H. K. Blood-brain barrier integrity and breast cancer metastasis to the brain. *Patholog Res Int* **2011**, 920509, <https://doi.org/10.4061/2011/920509> (2010).
- Bos, P. D. *et al.* Genes that mediate breast cancer metastasis to the brain. *Nature* **459**, 1005–1009, <https://doi.org/10.1038/nature08021> (2009).
- Gupta, P., Adkins, C., Lockman, P. & Srivastava, S. K. Metastasis of Breast Tumor Cells to Brain Is Suppressed by Phenethyl Isothiocyanate in a Novel *In Vivo* Metastasis Model. *PLoS One* **8**, e67278, <https://doi.org/10.1371/journal.pone.0067278> (2013).
- Kolbl, A. C., Andergassen, U. & Jeschke, U. The Role of Glycosylation in Breast Cancer Metastasis and Cancer Control. *Front Oncol* **5**, 219, <https://doi.org/10.3389/fonc.2015.00219> (2015).
- Leone, J. P. & Leone, B. A. Breast cancer brain metastases: the last frontier. *Exp Hematol Oncol* **4**, 33, <https://doi.org/10.1186/s40164-015-0028-8> (2015).
- Fidler, I. J., Balasubramanian, K., Lin, Q. T., Kim, S. W. & Kim, S. J. The brain microenvironment and cancer metastasis. *Mol Cells* **30**, 93–98 (2010).
- Abbott, N. J., Ronnback, L. & Hansson, E. Astrocyte-endothelial interactions at the blood-brain barrier. *Nat Rev Neurosci* **7**, 41–53, <https://doi.org/10.1038/nrn1824> (2006).
- Dejana, E. Endothelial cell-cell junctions: happy together. *Nat Rev Mol Cell Biol* **5**, 261–270, <https://doi.org/10.1038/nrm1357> (2004).
- Yonemori, K. *et al.* Disruption of the blood brain barrier by brain metastases of triple-negative and basal-type breast cancer but not HER2/neu-positive breast cancer. *Cancer* **116**, 302–308, <https://doi.org/10.1002/cncr.24735> (2010).
- Sirkisoon, S. R. *et al.* EGFR and HER2 signaling in breast cancer brain metastasis. *Front Biosci* **8**, 245–263 (2016).
- Lee, K. Y. *et al.* Human brain endothelial cell-derived COX-2 facilitates extravasation of breast cancer cells across the blood-brain barrier. *Anticancer Res* **31**, 4307–4313 (2011).
- Nam, B. H. *et al.* Breast cancer subtypes and survival in patients with brain metastases. *Breast Cancer Res* **10**, R20, <https://doi.org/10.1186/bcr1870> (2008).
- Dawood, S. *et al.* Defining prognosis for women with breast cancer and CNS metastases by HER2 status. *Ann Oncol* **19**, 1242–1248, <https://doi.org/10.1093/annonc/mdn036> (2008).
- Melisko, M. E., Moore, D. H., Sneed, P. K., De Franco, J. & Rugo, H. S. Brain metastases in breast cancer: clinical and pathologic characteristics associated with improvements in survival. *J Neurooncol* **88**, 359–365, <https://doi.org/10.1007/s11060-008-9578-5> (2008).
- Pangeni, R. P. *et al.* The GALNT9, BNC1 and CCDC8 genes are frequently epigenetically dysregulated in breast tumours that metastasise to the brain. *Clin Epigenetics* **7**, 57, <https://doi.org/10.1186/s13148-015-0089-x> (2015).
- Meany, D. L. & Chan, D. W. Aberrant glycosylation associated with enzymes as cancer biomarkers. *Clin Proteomics* **8**, 7, <https://doi.org/10.1186/1559-0275-8-7> (2011).
- Saldova, R., Fan, Y., Fitzpatrick, J. M., Watson, R. W. & Rudd, P. M. Core fucosylation and alpha2-3 sialylation in serum N-glycome is significantly increased in prostate cancer comparing to benign prostate hyperplasia. *Glycobiology* **21**, 195–205, <https://doi.org/10.1093/glycob/cwq147> (2011).
- Sarrats, A. *et al.* Glycan characterization of PSA 2-DE subforms from serum and seminal plasma. *OMICS* **14**, 465–474, <https://doi.org/10.1089/omi.2010.0050> (2010).
- Pierce, A. *et al.* Levels of specific glycans significantly distinguish lymph node-positive from lymph node-negative breast cancer patients. *Glycobiology* **20**, 1283–1288, <https://doi.org/10.1093/glycob/cwq090> (2010).
- Arnold, J. N. *et al.* Novel glycan biomarkers for the detection of lung cancer. *J Proteome Res* **10**, 1755–1764, <https://doi.org/10.1021/pr101034t> (2011).
- Zhu, R., Zacharias, L., Wooding, K. M., Peng, W. & Mechref, Y. Glycoprotein Enrichment Analytical Techniques: Advantages and Disadvantages. *Methods Enzymol* **585**, 397–429, <https://doi.org/10.1016/bs.mie.2016.11.009> (2017).
- Wooding, K. M., Peng, W. & Mechref, Y. Characterization of Pharmaceutical IgG and Biosimilars Using Miniaturized Platforms and LC-MS/MS. *Curr Pharm Biotechnol* **17**, 788–801 (2016).
- Veillon, L. *et al.* Characterization of isomeric glycan structures by LC-MS/MS. *Electrophoresis* **38**, 2100–2114, <https://doi.org/10.1002/elps.201700042> (2017).
- Saldova, R., Reuben, J. M., Abd Hamid, U. M., Rudd, P. M. & Cristofanilli, M. Levels of specific serum N-glycans identify breast cancer patients with higher circulating tumor cell counts. *Ann Oncol* **22**, 1113–1119, <https://doi.org/10.1093/annonc/mdq570> (2011).
- Shirure, V. S. *et al.* CD44 variant isoforms expressed by breast cancer cells are functional E-selectin ligands under flow conditions. *Am J Physiol Cell Physiol* **308**, C68–78, <https://doi.org/10.1152/ajpcell.00094.2014> (2015).
- Schnegelsberg, B., Schumacher, U. & Valentiner, U. Lectin histochemistry of metastasizing and non-metastasizing breast and colon cancer cells. *Anticancer Res* **31**, 1589–1597 (2011).
- Rambaruth, N. D., Greenwell, P. & Dwek, M. V. The lectin Helix pomatia agglutinin recognizes O-GlcNAc containing glycoproteins in human breast cancer. *Glycobiology* **22**, 839–848, <https://doi.org/10.1093/glycob/cws051> (2012).
- Mayoral, M. A. *et al.* Identification of galectin-3 and mucin-type O-glycans in breast cancer and its metastasis to brain. *Cancer Invest* **26**, 615–623, <https://doi.org/10.1080/07357900701837051> (2008).
- Yoneda, T., Williams, P. J., Hiraga, T., Niewolna, M. & Nishimura, R. A Bone-Seeking Clone Exhibits Different Biological Properties from the MDA-MB-231 Parental Human Breast Cancer Cells and a Brain-Seeking Clone *In Vivo* and *In Vitro*. *J Bone Miner Res* **16**(8), 1486–1495 (2001).

33. Giancotti, F. G. & Ruoslahti, E. Integrin signaling. *Science* **285**, 1028–1032 (1999).
34. Desgrosellier, J. S. & Cheresch, D. A. Integrins in cancer: biological implications and therapeutic opportunities. *Nat Rev Cancer* **10**, 9–22, <https://doi.org/10.1038/nrc2748> (2010).
35. Goel, H. L. & Mercurio, A. M. VEGF targets the tumour cell. *Nat Rev Cancer* **13**, 871–882, <https://doi.org/10.1038/nrc3627> (2013).
36. Shinomiya, H. Platin family of actin-bundling proteins: its functions in leukocytes, neurons, intestines, and cancer. *Int J Cell Biol* **2012**, 213492, <https://doi.org/10.1155/2012/213492> (2012).
37. Riplinger, S. M. *et al.* Metastasis of prostate cancer and melanoma cells in a preclinical *in vivo* mouse model is enhanced by L-plastin expression and phosphorylation. *Mol Cancer* **13**, 10, <https://doi.org/10.1186/1476-4598-13-10> (2014).
38. Royle, S. J. The cellular functions of clathrin. *Cell Mol Life Sci* **63**, 1823–1832, <https://doi.org/10.1007/s00018-005-5587-0> (2006).
39. Tung, K. H. *et al.* CHC promotes tumor growth and angiogenesis through regulation of HIF-1 α and VEGF signaling. *Cancer Lett* **331**, 58–67, <https://doi.org/10.1016/j.canlet.2012.12.001> (2013).
40. Abdi, H. & Williams, L. J. Principal component analysis. *Wiley Interdiscip Rev Comput Stat* **2**, 433–459, <https://doi.org/10.1002/wics.101> (2010).
41. Fagerberg, L. *et al.* Analysis of the Human Tissue-specific Expression by Genome-wide Integration of Transcriptomics and Antibody-based Proteomics. *Mol Cell Proteomics* **13**, 397–406, [10.1074/](https://doi.org/10.1074/) (2014).
42. Zorniak, M. *et al.* Differential expression of 2',3'-cyclic-nucleotide 3'-phosphodiesterase and neural lineage markers correlate with glioblastoma xenograft infiltration and patient survival. *Clin Cancer Res* **18**, 3628–3636, <https://doi.org/10.1158/1078-0432.CCR-12-0339> (2012).
43. Kyselova, Z. *et al.* Breast cancer diagnosis and prognosis through quantitative measurements of serum glycan profiles. *Clin Chem* **54**, 1166–1175, <https://doi.org/10.1373/clinchem.2007.087148> (2008).
44. Chen, W., Smeeckens, J. M. & Wu, R. Systematic and site-specific analysis of N-sialoglycosylated proteins on the cell surface by integrating click chemistry and MS-based proteomics. *Chem. Sci.* **6**, 4681–4689, <https://doi.org/10.1039/c5sc01124h> (2015).
45. Murugaesu, N. *et al.* An *in vivo* functional screen identifies ST6GalNAc2 sialyltransferase as a breast cancer metastasis suppressor. *Cancer discovery* **4**, 304–317, <https://doi.org/10.1158/2159-8290.cd-13-0287> (2014).
46. Dall'Olio, F. & Chiricolo, M. Sialyltransferases in cancer. *Glycoconj J* **18**, 841–850 (2001).
47. Peng, W. *et al.* Clinical application of quantitative glycomics. *Expert review of proteomics* **15**, 1007–1031, <https://doi.org/10.1080/14789450.2018.1543594> (2018).
48. Peng, W., Zhang, Y., Zhu, R. & Mechref, Y. Comparative membrane proteomics analyses of breast cancer cell lines to understand the molecular mechanism of breast cancer brain metastasis. *Electrophoresis* **38**, 2124–2134, <https://doi.org/10.1002/elps.201700027> (2017).
49. Elmasri, W. A. *et al.* Multitargeted Flavonoid Inhibition of the Pathogenic Bacterium *Staphylococcus aureus*: A Proteomic Characterization. *J Proteome Res* **16**, 2579–2586, <https://doi.org/10.1021/acs.jproteome.7b00137> (2017).
50. Dong, X., Zhou, S. & Mechref, Y. LC-MS/MS analysis of permethylated free oligosaccharides and N-glycans derived from human, bovine, and goat milk samples. *Electrophoresis* **37**, 1532–1548 (2016).
51. Mechref, Y., Kang, P. & Novotny, M. V. Solid-phase permethylation for glycomic analysis. *Methods Mol Biol* **534**, 53–64, https://doi.org/10.1007/978-1-59745-022-5_4 (2009).
52. Kang, P., Mechref, Y., Klouckova, I. & Novotny, M. V. Solid-phase permethylation of glycans for mass spectrometric analysis. *Rapid Commun Mass Spectrom* **19**, 3421–3428, <https://doi.org/10.1002/rcm.2210> (2005).
53. Kang, P., Mechref, Y. & Novotny, M. V. High-throughput solid-phase permethylation of glycans prior to mass spectrometry. *Rapid Commun Mass Spectrom* **22**, 721–734, <https://doi.org/10.1002/rcm.3395> (2008).
54. Zhou, S. *et al.* Isomeric Separation of Permethylated Glycans by Porous Graphitic Carbon (PGC)-LC-MS/MS at High Temperatures. *Anal Chem* **89**, 6590–6597, <https://doi.org/10.1021/acs.analchem.7b00747> (2017).

Acknowledgements

This work was supported by grants from Cancer Prevention & Research Institute of Texas, CPRIT (RP130624) and The National Institutes of Health, NIH (1R01GM112490-04 and 1R01GM130091-01).

Author contributions

W.P. performed transcriptomic/proteomic/glycomic analysis and drafted the manuscript. R.Z. and P.M. performed proteomic analysis. S.Z. performed glycomic analysis. Y.M. as the P.I., conceived, designed, supervised the study, and finalized the writing of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41598-019-53984-8>.

Correspondence and requests for materials should be addressed to Y.M.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019