Comparative proteomics analyses of whey proteins from breastmilk collected from two ethnic groups in northeast China

Cuina Wang [a], Yingcong Lu [a], Keyi He [a], Ru Zhao [a], Jianjun Cheng [b], Shilong Jiang [c], Mingruo Guo [d,*]

[a] *Department of Food Science, Jilin University, Changchun, China*
[b] *Department of Food Science, Northeast Agriculture University, Harbin, China*
[c] *R&D Center, Heilongjiang Feihe Dairy Co., Ltd, Beijing, China*
[d] *Department of Nutrition and Food Sciences, College of Agriculture and Life Sciences, University of Vermont, Burlington, VT 05405, USA*

ARTICLE INFO

ABSTRACT

The current study aims to investigate differences in whey protein of breastmilk of volunteered mother collected from two ethnic groups (Korean and Han) in China using data-independent acquisition (DIA) based proteomics technique. The total detected 624 proteins were principally allocated to cellular process of biological process (BP), cell and cell part of cell component (CC) and binding of molecular function (MF) according to Gene Ontology (GO) annotation; and carbohydrate metabolism of Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis. Among the 54 differently expressed proteins, 8 were related with immunity. Enrichment data showed that intracellular of GO functions and viral myocarditis KEGG pathways were most significantly enriched ($p < 0.05$). Protein-protein interaction (PPI) network suggested that 40S ribosomal protein S27a and 60S ribosomal protein L10a which interacted most with other proteins ranked the top two hub proteins by MCC (Maximal Clique Centrality) method. This study may have guiding role for development of infant formula powder for specific infants of Han or Korean groups according to responding breastmilk composition.

1. Introduction

Due to its role in providing immune protection and diverse biological functions for infants, human milk has attracted extensive attention. It is a biological colloidal fluid, which consists of all essential macro- and micro-nutrients, such as proteins, lipids, carbohydrates, minerals, vitamins, etc (Zhang et al., 2019) and some unique nutrients like immune components, anti-infective factors, and metabolic enzymes (Jin et al., 2021). Generally, human milk proteins are grouped into three major categories: casein, whey protein and milk fat globular membrane protein (Chen et al., 2022). Whey protein is one of human milk main components and accounts for the vast majority (60–80 %). It mainly includes α-lactalbumin (α-LA), lactoferrin (LF) and immunoglobulin (Ig) (Chen et al., 2022), and specific composition or abundance of proteins change dynamically with a variety of factors, such as lactation stage, delivery mode, gender, maternal environmental factors in geography, diet, ethnicity (Zhang et al., 2019).

Nowadays, the rapid development of various omics techniques has brought analysis of human milk into a new era. Omics approaches could detect protein profile efficiently and comprehensively, which could be helpful to understand impacts of influencing factors on human milk. Effects of region and nationality on human milk serum proteomes were studied using tandem mass tag (TMT) labeling combined with Nano-LC Q Exactive HF tandem mass spectrometry (MS/MS) proteomics in previous study (Zhang et al., 2019). Label-free quantification proteomics were applied to identify and quantify serum proteins in human, cow, goat, and yak milk (Lu et al., 2018). Shot-gun proteomics approach was employed to investigate variation in terms of host defense proteome between human and bovine milk (Hettinga et al., 2011).

Recently, a popular and promising quantitative proteomics technology, namely, data-independent acquisition (DIA) proteomics method appeared (Zhao et al., 2021). In one data acquisition cycle, DIA technique first acquires a full MS1 scan, followed by one or multiple MS2 acquisition scans. Different from data-dependent acquisition (DDA), DIA technique can acquire MS2 spectra for all precursor ions theoretically and select either MS1 or MS2 ions for metabolite quantification (Wang et al., 2019), which could effectively avoid data loss, resulting in more reliable quantitative results of low abundance signals (Basak et al.,

2015). Thus, DIA proteomic approach has been used in research with advantages of high accuracy, dynamic range, and reproducibility of selected reaction monitoring. Whey protein profiles of Saanen goat milk collected from different regions in China were characterized and compared using DIA quantitative proteomics technology (Zhao et al., 2021). DIA proteomic procedure was utilized to disclose hypotensive and immune-enhancing components in whey fraction collected from buffalo milk which was raised in locations at diverse altitudes (Zhang et al., 2020).

There are some differences in the abundance and composition of human milk whey protein from different ethnicities. Therefore, this study adopts DIA as the research technology to explore the dissimilarity between milk samples collected from ethnic groups of Korean and Han in China. Han is the largest group in China with a population of about 1.2 billion while Korean has a population of about 2 million and the Korean people mainly distribute in northeast China. Relevant results may contribute to improvement of human milk whey protein knowledge. Additionally, conclusions obtained may offer a theoretical underpinning for development of more accurate and refined infant formula which would be directed for infants of Han or Korean groups.

2. Materials and methods

2.1. Human milk sample collection

A total of 18 mother in each ethnicity of Korean and Han (Yanbian Autonomous Prefecture, Jilin Province) were selected based on preliminary questionnaire (including the diets) and written informed consent was signed. The breastmilk samples were collected from mothers with full-term babies in the lactation stage from 14 to 28 days postpartum. The age of mothers ranged from 20 to 35 and feed babies with breastmilk. The parity of most mothers is 1 or 2 and the milk samples were collected at 9:00 am to 11:00 am during the same time of day in May to July 2021.The selected volunteer mothers carefully cleansed their areola and nipples several times before taking samples, and then rinsed relevant parts with sterile water. Human milk was gently squeezed from the breast and first drops were discarded, and then about 30 mL was contained in 50-mL sterile centrifuge tube and labeled. The obtained samples were immediately placed in the refrigerator at −80 ℃ and then shipped to lab by cold chain transportation. Every-six samples were randomly mixed into one, and three biological replicates for milk from each nationality were determined. The human milk experiments were confirmed by the Ethics Committee of Northeast Agriculture University and experiments were conducted in accordance with Chinese laws and institutional guidelines.

2.2. Whey protein extraction and separation

Protein was extracted from human milk according to our previous study (Zhao et al., 2021). First, human milk sample was divided into two layers by centrifugation (4 °C, 3000 × rpm for 60 min) and the skimmed layer was further separated by adjusting the pH value to 4.6, where casein fraction can be sedimented. And then the serum layer was collected and added with precooled acetone at 4-fold volume and precipitated at −20℃ overnight. The sediments were obtained and washed using precooled acetone, and then dried in fume hood. BCA (bicinchoninic acid) was used to quantitatively analyze whey protein extracted from samples by establishing a standard curve ($R^2 > 0.99$) by determination of absorbance at 562 nm.

2.3. Filter-aided sample preparation (FASP) enzymatic hydrolysis

Protein sample (100 μg) was conducted for protein enzymatic hydrolysis based on FASP approach (Wisniewski et al., 2009). Briefly, the sample was added with 8 M urea to achieve a volume of 200 μL, dithiothreitol was introduced at a final concentration of 10 mM and the mixture was allowed to stand for 30 min at 56 ℃. Iodoacetamide was added to an ultimate level of 50 mM, and reacted at room temperature without light for 40 min. Samples were transferred to 10 K ultrafiltration tubes and then centrifuged at room temperature at 12,000*g*. The filtrate was dumped and this step was repeated for three times. Ammonium bicarbonate solution (200 μL, 50 mM) was added and then centrifuged at room temperature at 12,000*g*, and then the filtrate was discarded and this step was repeated for 3 times. Trypsin was added into the tube with the mass ratio of sample to enzyme at 50:1, and enzymolysis was performed at 37 ℃ for 16 h. Digested peptides were obtained by centrifugation at 12,000*g*, 4℃ and then lyophilized.

2.4. High performance liquid chromatography (HPLC) classification

For the polypeptide samples after enzymatic hydrolysis, the same amount of each biological sample was mixed and performed for reverse phase chromatographic classification. Experimental conditions were as follows: Agilent 1100 (including Chemstation, 214 nm DAD detector and vacuum degasser); column of waters XBridge C18 (5 μm, 4.6 mm × 250 mm, 120 Å). Mobile phase was composed of phase A (98 % H_2O, 2 % acetonitrile, pH 10) and phase B (98 % acetonitrile, 2 % H_2O, pH 10). The preparative HPLC was run for a total of 60 min and the chromatographic gradient for classification was: 0–5 min: 3 % B at flow rate of 0.40 mL/min. For the rest time, the rate was elevated to 0.70 mL/min and the mobile phase composition was as follow: 5–5.10 min 3 % B; 5.10–10 min 5 % B; 10–35 min 18 % B; 35–45 min 34 % B; 45–58 min 95 % A; 58–60 min 3 % B. Samples were collected every 1 min for a total of 10 fractions based on polarity.

2.5. Construction of DDA reference spectrum library

The 10 fractions of peptides after classification were mixed into 3 mixtures and then analyzed by HPLC-MS. A total of 18 μL of the sample was loaded and gradient separation was performed using Chromatograph of UltiMate™ 3000 RSLCnano with column of Waters XBridge C18 (5 μm, 4.6 mm × 250 mm, 120 Å) within 60 min: 3 % to 100 % B; column flow was set at 600 nL/min. Orbitrap Fusion Lumos was used for MS and the relevant parameters were as follows: (1) MSn1: scan range 375–1500 *m/z*; maximum injection time (MIT) 50 ms; AGC target 400,000; orbitrap resolution 120,000. (2) MSn2: MIT 22 ms; AGC target 50,000; orbitrap resolution 15,000; collision energy 30 %.

DDA data collected were searched and analyzed with Proteome Discoverer 2.1.0182 (Thermo Fisher Scientific, Rockford, IL, USA). The detailed parameters were as follows: the polypeptide was digested by trypsin and the maximum allowable missing cleavages was 2. The peptide and fragment mass tolerance were required to ±10 ppm and ±0.02 Da. The results were compared with the human milk protein database (Uniprot TaxId: 9606) by Sequence HT.

2.6. DIA protein quantitative analysis

For DIA analysis, HPLC parameters were set as the same used in DDA procedure. The DIA MS parameters were as follows: (1) MSn1: scan range 350–1500 *m/z*; MIT 50 ms; AGC target 400,000; orbitrap resolution 60,000. (2) MSn2: scan range 200–2000 *m/z*; MIT 54 ms; AGC target 300,000; orbitrap resolution 30,000; collision energy 33 %.

DIA analysis was performed with skyline software, and peptide signal extraction and quantitative analysis were performed on the original DIA data by importing the above DDA spectrum library. The experiment adopted strict screening conditions to ensure the accuracy of quantitative results and the detailed parameters were as follows: the polypeptide length of 6–25 was reasonable; the selected sub ion M/Z was determined to be greater than parent ion and last ion-3; maximum number of sub ions was 5, while minimum was 2; 5 min was set as the condition of sub ion extraction window; the value of Dotp (dot product) was required to be larger than or equal to 0.6.
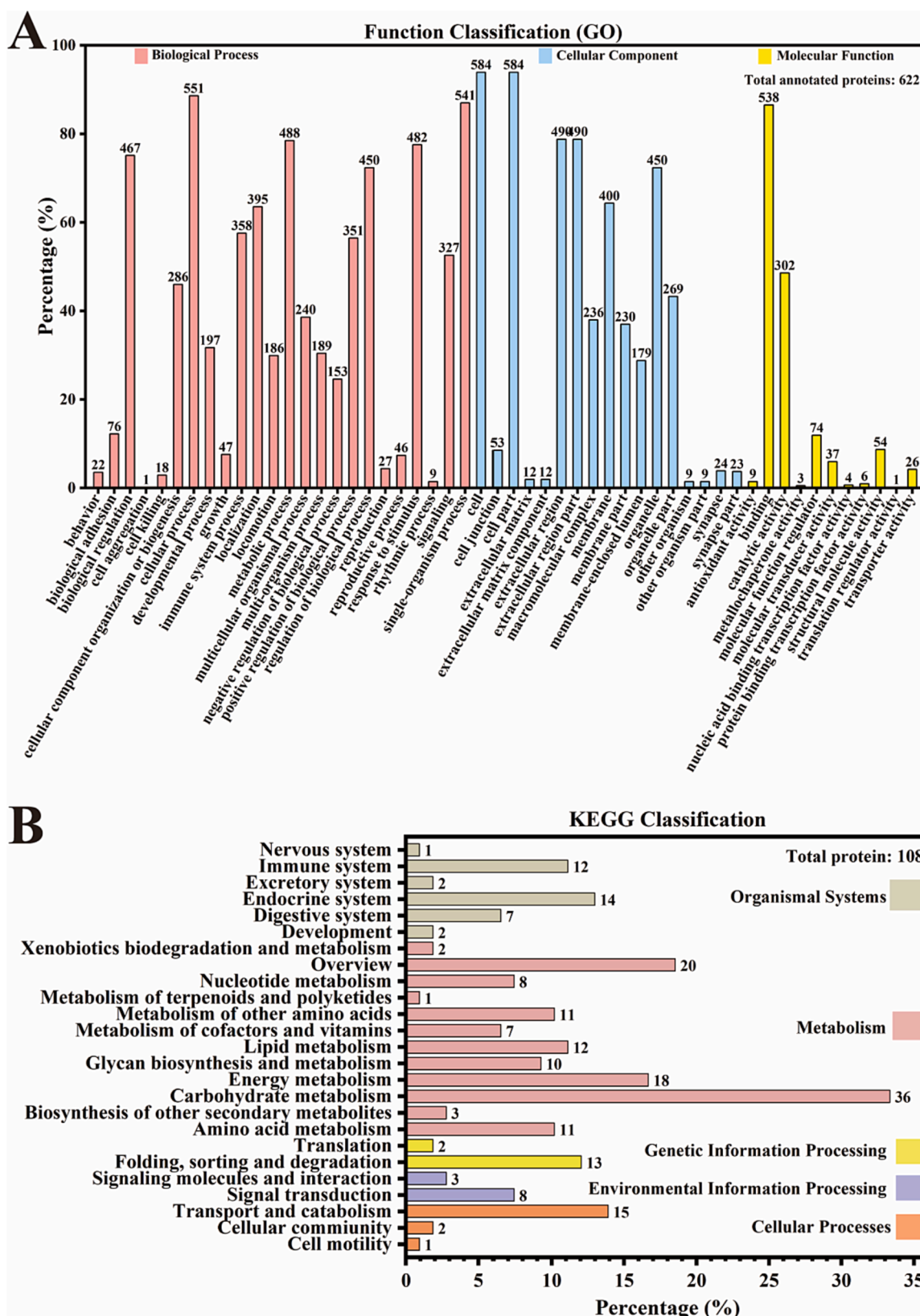
Fig. 1. Gene Ontology function annotation (A) and Kyoto Encyclopedia of Genes and Genomes metabolic pathway (B) of whey proteins from breastmilk collected from two ethnic groups in northeast China. Note: the values on and beside the columns was the protein numbers.

2.7. Bioinformatics analysis

ObaDIA was an analysis process integrating several mainstream bioinformatics tools and software, which can conduct comprehensive and systematic quantitative proteomic data analysis (Yan et al., 2021). The project mainly used obaDIA software for bioinformatics analysis, including quality control, abundance statistics, difference analysis, functional annotation, and enrichment analysis. UniProt database was applied to gain detailed information on all proteins identified. All identified proteins were annotated with Gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) databases respectively. KOBAS (2.0) was used for pathway enrichment analysis in obaDIA. The PPI network was analyzed using STRING (https://string-db.org/). The analysis results were imported into Cytoscape v3.7.2 (Shannon et al., 2003) and the top 10 hub proteins were selected using cytohubba.

3. Results and discussion

3.1. Characterization of quantitatively identified proteins

Analysis on composition of human milk whey proteins using proteomics technique has been documented in previous studies. One hundred and fifteen low-abundance proteins in the whey fraction of human milk were identified by liquid chromatography tandem mass spectrometry (Liao et al., 2011). Abundance changes of whey protein in human milk produced from 1 week to 12 months were explored by utilizing a TMT approach in previous study. The authors reported that the number of proteins identified was 1333, among which 615 was quantified (Zhang et al., 2013). In another research, a total of 584 proteins in whey protein fraction of human milk was identified and quantified by isobaric tag for relative and absolute quantification (iTRAQ) proteomic method (Yang et al., 2017). In this study, 624 proteins were quantified in human milk samples collected from Korean and Han nationalities through DIA, which was higher than the quantified protein numbers in the studies aforementioned. However, in a study reported by Jin et al. (2021), a total of 1563 whey proteins were detected in human milk using the DIA method. The different results may be due to variation in FASP enzymatic hydrolysis, parameters set in DIA, and the softwires used which were different from this study. Elaborate information of identified proteins was exhibited in Table S1. Correlation matrix analysis was also performed for milk samples of Korean and Han groups. Fig. S1 showed that composition of whey proteins in human milk of Korean had a relative high similarity with those of Han at a correlation coefficient of 0.96.

LF and α-LA were the proteins with the largest peak area integral from the mass spectrum of the samples, indicating that the two had high abundance in the batch of samples. Previous study showed that LF is the second-most predominate protein and α-LA accounts for approximately 36 % of the whey protein in human milk (Auestad & Layman, 2021). LF is a cationic iron-binding glycoprotein and the concentration of free iron in biological fluid can be controlled and adjusted by the ability to dissolve or chelate iron ions (Fe^{3+}). LF possesses a wide range of biological activities and multiple functions. It exerts antibacterial activity by affecting the release of lipopolysaccharide (LPS) to unsteady cell wall through the binding of Ca^{2+} and binding with microorganisms leads to the instability of bacteria cell membrane (Ling & Schryvers, 2006). LF also plays the antiviral effect by degrading several viral factors and secretory proteins of some pathogenic bacteria through serine endopeptidase activity (Gomez et al., 2003), interfering with the attachment of virus and reduce the infection of host cells by combining to integrin and gag receptor on the surface of host cells (Hara et al., 2002). LF neutralize LPS and other microbial molecules by binding its lipid domain, preventing the further interaction between LPS binding protein (LPB) and endotoxin to block the binding between LPS and membrane CD14 (Britigan et al., 2001). α-LA is a small globular protein which can provide essential amino acids and bind metal cations, which is important to modulate interaction of α-LA with membranes, proteins, peptides

and low molecular weight substrates and products (Permyakov & Berliner, 2000). α-LA has important biological functions in mammary secretory cells and participates in lactose biosynthesis (Kamau et al., 2010).

3.2. GO analysis of identified whey proteins of Korean and Han human milk

GO was divided into three categories: biological process (BP), cell and cell part of cell component (CC) and binding of molecular function (MF). All proteins identified in human milk whey protein were annotated with GO terms in Fig. 1A. It was evident that 622 proteins identified were annotated into 24 BP, 17 CC and 11 MF under the GO database.

The term with the most protein participation in BP was cellular process (551, 88.59 %), with a total of 207 GO terms (Lev4). Additionally, term of metabolic process (488, 78.46 %) involving the third largest number of proteins was also worth analyzing. Metabolic process consisted of 142 GO terms (Lev4), among which macromolecular (414) and protein metabolisms (381) were the top two involved in the largest number of proteins. Analysis of CC can reveal the cellular environment of specific human milk whey proteins. It can be observed from Fig. 1A that cell (584, 93.9 %), cell part (584, 93.9 %), extracellular region (490, 78.78 %) and extracellular region part (490, 78.78 %) occupied the top four, with 26/88/12/15 GO terms (Lev4), respectively.

Speaking of MF, binding was the term involved with the most proteins (538, 86.5 %). There was no doubt that different proteins would bind to receptors/anions/cations, thus affecting life activities. Among all proteins identified, the number involved in receptor binding/cation binding/anion binding was 153/105/99, respectively. Besides, numerous proteins were also involved in the binding of nucleic acids/enzymes (72/59). Due to a relatively high percentage to the total whey proteins, catalytic activity should also be paid attention. It is known that infants cannot secrete enough enzymes in the gastrointestinal tract owning to their immature digestive system. However, infants can effectively digest and absorb milk proteins, which may be the consequence of endogenous enzymes in human milk, which can play the catalytic activity and degrade proteins into peptides (Zhang et al., 2019).

Under the overall GO annotation, the most functional protein which had the largest number of terms after analysis was amyloid-beta A4 protein (Aβ, A0A140VJC8). It was present in 645 (27.74 %) GO terms at Lev4 which had a total of 2325 specific terms. Existing in the form of amyloid-β 1–40 (Aβ40) and amyloid-β 1–42 peptides, Aβ is produced by the hydrolysis of amyloid precursor protein by β and γ secretases (Stamatelopoulos et al., 2015). Biliary atresia (BA) is a severe cholestatic liver disease in neonates. A previous study showed that Aβ expression increased in the plasma and livers of infants with BA. Liver organoids treated with Aβ had abnormal morphology and impaired growth (Tian et al., 2022).

3.4. KEGG pathway analysis of identified whey proteins from Korean and Han human milk

In general, KEGG pathway includes five parts: cellular processes, environmental information processing, genetic information processing, metabolism, and organismal systems. Proteins (108) that can be analyzed with KEGG annotation were summarized by function, and the results are shown in Fig. 1B. All the characterized proteins were mainly distributed in metabolism of KEGG analysis.

Three pathways were found under the term of cellular processes, and the hierarchy of transport and catabolism possessed the most proteins (15). For further subdivision, transport and catabolism was composed of lysosome (11), peroxisome (4) and phagosome (1). Compared with the other four classifications at the same level, terms (only two) fall into the category of environmental information processing and proteins (11)
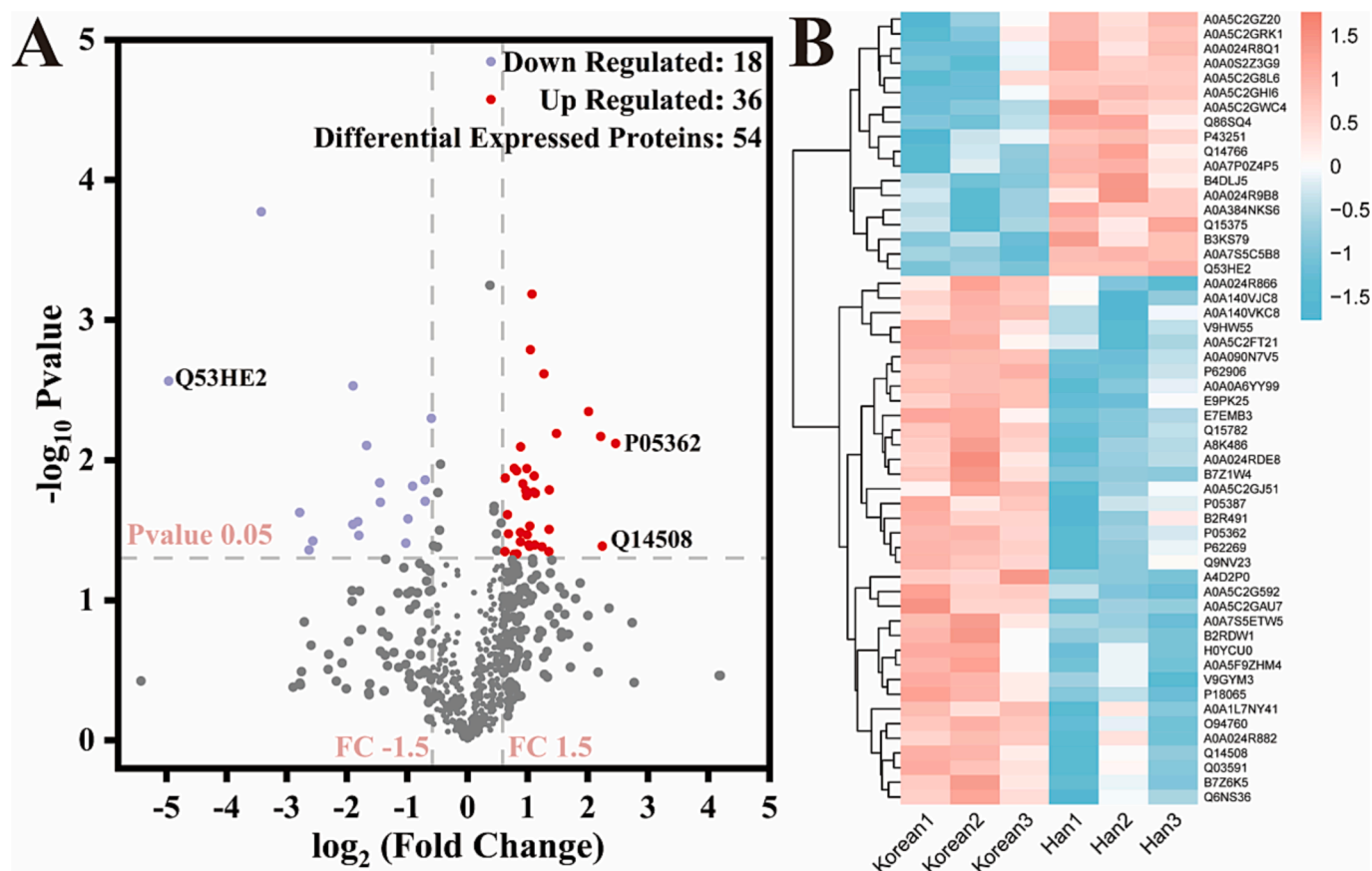
Fig. 2. Volcano diagram (A) and hierarchical clustering (B) of differentially expressed whey proteins from breastmilk collected from two ethnic groups in northeast China. Fig. 2B: Each column is a group. Each row is a protein. Red part represents upregulated protein, cyan part represents downregulated protein, and white part represents a protein that does not change significantly. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

involved were the least. Term of signal transduction had more proteins (8) than the other one, namely, signaling molecules and interaction. Signal transduction displayed five subordinate branches, which were signal pathways of AMPK (2), HIF-1 (2), MAPK (2), sphingolipid (2) and phospholipase D (1). In terms of genetic information processing, two pathway classifications with total of 15 proteins were observed and term of folding, sorting and degradation possesses 13 proteins. Folding, sorting and degradation included specific pathways of proteasome (6), endoplasmic reticulum protein processing (6) and RNA degradation (1). Metabolism undoubtedly deserved more attention as it involved more variety of branches with a larger quantity of proteins. Obviously, carbohydrate metabolism occupied the largest percentage in this column with 36 proteins, indicating the importance of this pathway hierarchy in human milk. Under carbohydrate metabolism, a total of 13 specific pathways were recorded, and glycolysis/gluconeogenesis (15), amino sugar and nucleotide sugar metabolism (9) and pentose phosphate pathway (8) ranked the top three. As for organismal systems, it was apparent to see from Fig. 1B that the term with maximum proteins was endocrine system (14 proteins). The specific downstream pathways of endocrine system were synthesis of various hormones and three signal pathways, namely glucagon, insulin and PPAR.

Under the KEGG annotation, the protein most involved in the specific pathways was aldehyde dehydrogenase (NAD(P) +) (A0A1B0GW77), involving 14 pathways. These 14 pathways belonged to amino acid metabolism, carbohydrate metabolism, lipid metabolism, metabolism of other amino acids and overview of metabolism, respectively. The aldehyde dehydrogenase superfamily of NAD(P)$^+$-dependent enzymes, in general, oxidize a wide range of endogenous and exogenous aliphatic and aromatic aldehydes to their corresponding carboxylic acids and play

an essential role in detoxification of reactive oxygen species accumulated under the stressed conditions (Yang et al., 2012).

3.5. Differentially expressed proteins (DEPs) in the whey fraction between Korean and Han

Identified proteins of Korean and Han were compared to explore the specific differences in whey fraction of human milk from different nationalities. DEPs were screened according to the conditions of fold change (FC) > 1.5 or < 0.65 and P value < 0.05. Results showed that 54 proteins out of the total 624 identified proteins were differently expressed in abundance between Korean and Han milk and the related information is displayed in Table S1. Volcanic map (Fig. 2A) revealed that 36 DEPs were upregulated (drawn in red color) and 18 were down regulated (in blue color).

Among the 36 increased DEPs, intercellular adhesion molecule-1 (ICAM-1, P05362) had the largest log2FC of 2.46, P = 0.0076, which was constitutively expressed on a variety of cells and was also one of the most important conserved human receptors required by numerous key physiological processes (Shukla et al., 2022). Had the second highest log2FC of 2.24, P = 0.04, WAP four-disulfide core domain protein 2 (WFDC2, Q14508) showed a significant difference between the two groups.

For the decreased DEPs, triosephosphate isomerase (TPI, Q53HE2) was the most differently protein with log2FC of −4.96 and P value of 0.0027. TPI is a key glycolytic enzyme to catalyze the interconversion of dihydroxyacetone phosphate and glyceraldehyde-3-phosphate during glycolysis and gluconeogenesis. In addition to catalysis, it also has an essential role in the carbohydrate metabolism and energy production in

Table 1
Immune-related differentially expressed whey proteins annotated by Gene Ontology (GO) terms in human milk whey protein.

Term (number of relevant proteins)	ID	Relevant proteins (Fold Change, P)
innate immune response activating cell surface receptor signaling pathway (1)	GO:0002220	V9HW55 (1.8916, 0.0147)
adaptive immune response (1)	GO:0002250	A0A384NKS6 (0.6607, 0.0050)
immune system process (4)	GO:0002376	Q03591 (2.5639, 0.0312)
		V9GYM3 (2.0487, 0.0295)
		A0A384NKS6
		A4D2P0 (1.9820, 0.0115)
immune response-activating cell surface receptor signaling pathway (1)	GO:0002429	A4D2P0
lymphocyte mediated immunity (1)	GO:0002449	A0A384NKS6
regulation of immune system process (2)	GO:0002682	A0A384NKS6 A4D2P0
positive regulation of immune system process (2)	GO:0002684	V9HW55 A0A384NKS6
regulation of cytokine production involved in immune response (1)	GO:0002718	V9GYM3
immune response-regulating cell surface receptor signaling pathway (1)	GO:0002768	B2RDW1 (2.1727, 0.0402)
immune response (1)	GO:0006955	B2RDW1
humoral immune response (2)	GO:0006959	A0A140VJC8 (2.1403, 0.0169)
		A0A384NKS6
innate immune response (1)	GO:0045087	P05362 (5.5050, 0.0076)
regulation of immune response (2)	GO:0050776	V9GYM3 P05362

all living cells (Pekel and Ari, 2020).

Fig. 2B shows the heatmap of 54 significantly different proteins in two groups, revealing three biological replicates in each group. There were two main obviously different protein clusters according to hierarchical cluster based on protein abundance. It was obvious that the abundance of TPI and clusterin (A0A384NKS6) among the Han was higher than the Korean, while Korean contained a variety of higher abundance proteins related to ribosomes (40S ribosomal protein S27a (B2RDW1), 60S ribosomal protein L10a (P62906), 60S acidic ribosomal protein *P*2 (P05387), 40S ribosomal protein S4 (B2R491), 40S ribosomal protein S18 (P62269), 60S ribosomal protein L35 (A0A024R866)).

The hierarchical clustering results of DEPs can help us distinguish protein subsets with different expression patterns from differential protein sets. Proteins with similar expression patterns may have similar functions or participate in the same biological pathway, or be in adjacent regulatory positions in the pathway. Fig. 2B displays that WFCD2 and complement factor h-related protein 1(Q03591) were divided into a group, 60S ribosomal protein L35 and Aβ belonged to another group, while 60S acidic ribosomal protein *P*2, 40S ribosomal protein S4, ICAM-1, 40S ribosomal protein S18 were divided into a large group.

3.6. Immune related proteins in whey proteins between Korean and Han

Whey fraction of human milk contains various immune related proteins, and Ig was the most abundant and diverse. As a globulin with antibody activity, Ig is synthesized by the immune system after the antigen stimulates the human and will bind specifically to the antigen. An Ig monomer is a Y-type tetrapeptide chain structure composed of two identical light chains and two same heavy chains connected by several inter chain disulfide bonds. Both light and heavy chains have constant and variable regions. According to the different composition and sequence of amino acids in the constant region, the light chain is divided into two types: lambda and kappa, while heavy chain is five kinds, namely IgG (γ), IgM (μ), IgA (α), IgD (δ) And IgE (ε). As the receptor of

Ig, polymeric immunoglobulin receptor (PIGR) plays a role in transcytosis of IgA from the basolateral side to the top of epithelial cells (Kaetzel, 2005) and transporting IgA and IgM across epithelial cells to mucosal secretions (Johansen et al., 2000).

As shown in Table S1, Ig proteins detected in whey proteins of both Korean and Han in the top five abundance were PIGR (P01833), immunoglobulin heavy constant alpha 2 (IgA, P01877), immunoglobulin kappa light chain (P0DOX7), immunoglobulin J chain (P01591) and immunoglobulin mu heavy chain (IgM, P0DOX6), respectively. These proteins expressed similar in both samples. Previous study indicated that the distribution of Ig in human colostrum was the largest in IgA, followed by IgM, with low levels of IgG (2 %) (Hurley & Theil, 2011).

As exhibited in Table 1, 8 of the 54 DEPs were observed to be immune-related according to the subsequent GO enrichment analysis, sorted by FC values, namely ICAM-1(5.50), complement factor h-related protein 1(2.5639), 40S ribosomal protein S27a (2.1727), Aβ (2.1403), apolipoprotein A-II (2.0487, Apo A-II, V9GYM3), ras-related c3 botulinum toxin substrate 1 (1.9820, Rac1, A4D2P0), proteasome subunit alpha type-1(1.8916, V9HW55) and clusterin (0.6607). For the comparison of Korean vS Han, clusterin was down regulated, while the rest were upregulated, indicating that clusterin in whey protein of Korean human milk was less than that of Han, but the other proteins were more.

In addition to being the only down regulated protein among the above 8 proteins, clusterin was also the protein with the largest number of immune related GO terms. Clusterin is a highly conserved glycoprotein, which is involved in lipid transport, apoptosis, tissue remodeling, adhesion, and protection of nerve and cardiovascular cells (Peng et al., 2019). Clusterin exerts its extracellular chaperone function by acting as a scavenger or chelator. As a scavenger, it combines misfolded toxic proteins or plasmin-generated protein fragments (PGPFs) to form complexes, which are recognized by membrane receptors and transported in cells, where they are degraded by proteasomes or lysosomes. As for chelating ability, it acts as a chelating agent around misfolded toxic molecules in excess protein molecules to prevent their toxicity (Berdowska et al., 2022).

Rac1 showed the second largest number in immune related GO terms. As one of the most characteristic small GTPases, Rac1 participates in a wide spectrum of physiological processes, including gene expression, neurodevelopment, etc. Rac1 and related signaling pathways are prominently involved in the maintenance and regulation of basic nervous system functions. Aberrant Rac1 expression or activity regulation or even small alterations to its downstream signaling may lead to severe neurodevelopmental disorders such as schizophrenia, etc. (X. H. Wang et al., 2020). Rac1 also plays an important hub in the signal network to tumorigenesis and metastasis, because its overexpression or mutation can lead to invasive tumor phenotype and confer resistance to targeted therapy (De et al., 2020).

As the protein with the largest FC value, ICAM-1 showed that it was far more abundant in Korean human milk than Han. At the same time, it had the third largest number of immune related GO terms. ICAM-1 is a transmembrane glycoprotein, which has various immune functions, such as T cell activation, extravasation, inflammation and so on (Singh et al., 2021). As inducible cell adhesion molecules, they are expressed in various cells, including epithelial cells, mesothelial cells, endothelial cells, lymphocytes, monocytes, and fibroblasts (Igarashi et al., 2017). Moreover, ICAM-1 plays a pivotal role in maintaining intercellular interactions, and has been proved to be an indispensable part of leukocytes entering inflammatory sites. It promotes leukocyte adhesion and migration to inflammatory sites by interacting with leukocyte integrins (such as LFA and Mac 1) to activate lymphocytes (Singh et al., 2021).

3.7. GO enrichment analysis of the DEPs for Korean vS Han

GO enrichment analysis was an effective tool to observe the functions of related proteins, to explore their functions and biological significance. The three most important parameters in enrichment analysis were P
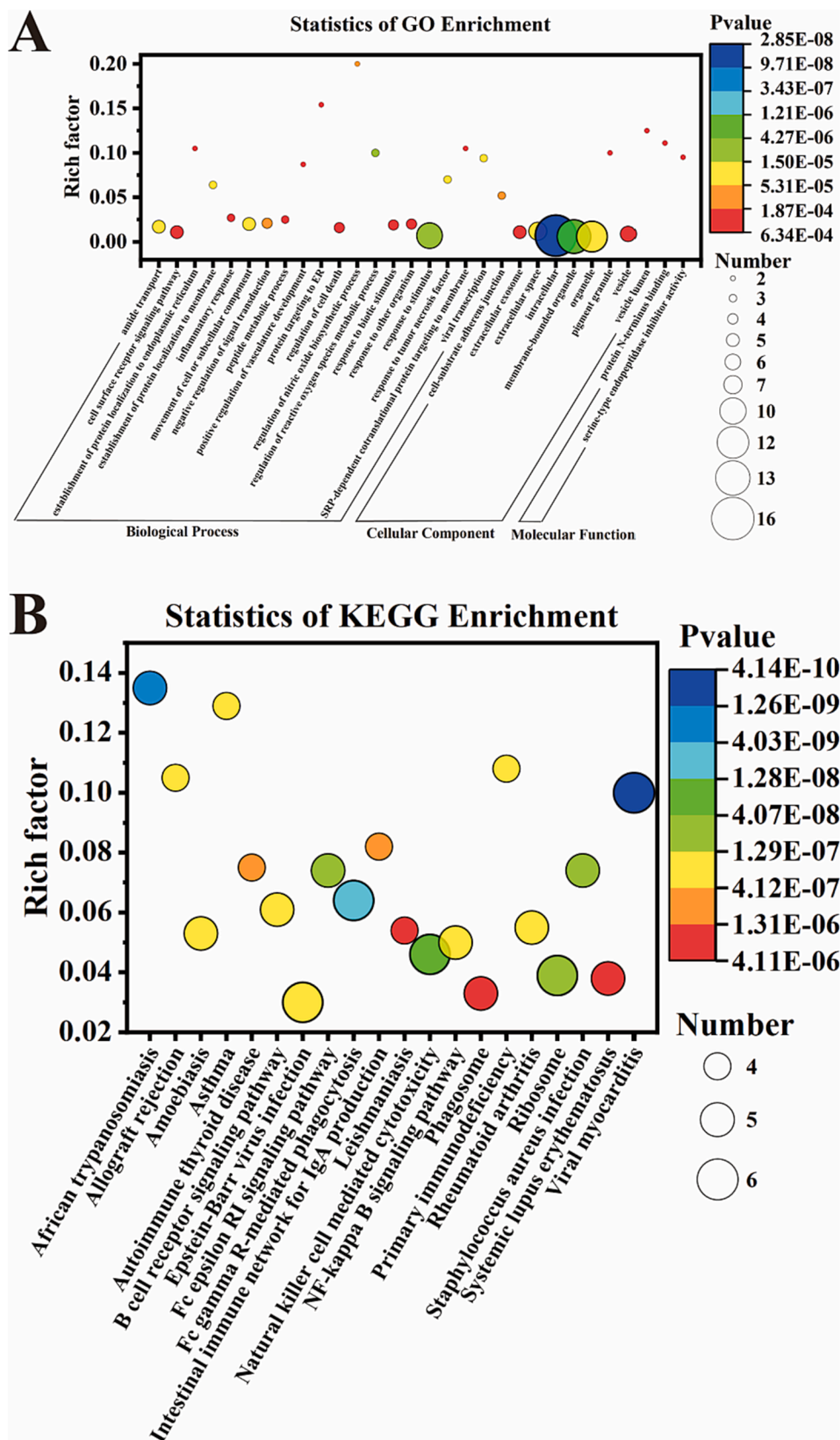
Fig. 3. Gene Ontology enrichment classification (A) and Kyoto Encyclopedia of Genes and Genomes pathway enrichment (B) of differentially expressed whey proteins from breastmilk collected from two ethnic groups in northeast China.
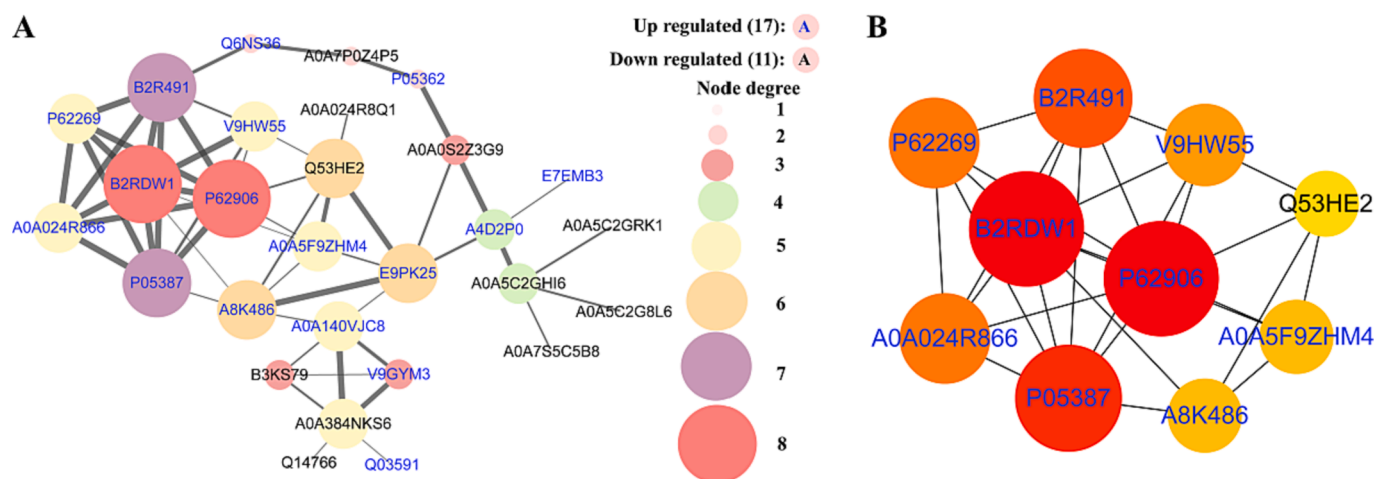
Fig. 4. Protein-protein interaction network map (A) of differentially expressed whey proteins from breastmilk collected from two ethnic groups in China. Each node represents a protein, and each edge represents the direct interaction between proteins. The larger the combine score value, the bolder the edge. The top 10 hub proteins network (B) ranked by Maximal Clique Centrality (MCC) method. The closer the color of protein to red, the higher its MCC score. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

value, rich factor, and the number of related proteins. P value represents the significance of enrichment, and the smaller it is, the more significant the enrichment of DEPs in this GO term. Rich factor is the ratio of the percentage of DEPs enriched in one pathway to the percentage of proteins enriched in the same pathway in the background. The higher the value represents higher enrichment degree.

DEPs between Korean and Han human milk were studied by GO enrichment analysis, and the result showed that a total of 357 of the 722 GO terms were significantly enriched (P < 0.05). The first 30 GO terms with significant enrichment are shown in Fig. 3A in the form of bubble chart, including 19 BP, 9 CC and 2 MF. As the term with the largest number of participating proteins (16), intracellular had a rich factor of 0.007 and a minimum P value of 2.85E-08. Among the 54 DEPs, Rac1 was involved in the largest number of terms in GO enrichment analysis, with a total of 101, followed by 40S ribosomal protein S27a (96) and Aβ (95).

It was obvious that DEPs showed a significant enrichment in only two functions under the classification of MF, which were protein *N*-terminal binding (2 proteins) and serine-type endopeptidase inhibitor activity (2 proteins). However, serine hydrolase activity cannot be ignored as it was major in the catalytic activity of all proteins identified. WFDC2 and cDNA flj35730 FIS (down regulated, b3ks79) had serine-type endopeptidase inhibitor activity among DEPs. Serine hydrolases has been identified for more than 200 numbers and involved in many physiological and pathological processes including inflammation, peptide/protein processing, and protein/lipid digestion (Shahiduzzaman & Coombs, 2012).

3.8. KEGG pathway enrichment analysis of DEPs from Korean and Han human milk

DEPs between whey fraction collected from Korean and Han human milk were investigated by KEGG enrichment analysis and the result showed that a total of 58 out of the 129 KEGG pathways were significantly enriched (P < 0.05). The 20 pathways with the most significant enrichment are displayed by bubble diagram (Fig. 3B). Viral myocarditis was the largest enrichment pathway, with a rich factor of 0.1, involving 6 proteins, and a P value of 4.14E-10.

When analyzing proteins involved in enrichment pathway, IGH + IGL c301_heavy_IGHV3-7_IGHD3-9_IGHJ2 (Fragment, upregulated, A0A5C2GAU7), IG c437_heavy_IGHV3-74_IGHD5-24_IGHJ6 (Fragment, down regulated, A0A5C2GHI6), IGH c3394_heavy_IGHV3-53_IGHD3-10_IGHJ3 (Fragment, down regulated, A0A7S5C5B8) and IGH

c1201_heavy_IGHV3-33_IGHD6-19_IGHJ4 (Fragment, upregulated, A0A7S5ETW5) should be paid more attention because they all took part in 19 pathways except ribosome among the 20 most significant enrichment pathways. The above proteins in the pathway corresponded to lgA, lgM, lgG, and BCR, respectively. The results indicated that the percentage of fragment in DEPs was much greater than the proportion of Ig in all identified proteins. Aside from these proteins, Rac1 and ICAM-1 were both involved in 7 metabolic pathways.

3.9. PPI network analysis of DEPs from Korean and Han human milk

PPI network between DEPs was analyzed using STRING, an online search database with numerous species and abundant interactive information and the results are exhibited in Fig. 4A. The combined score represented the degree of credibility for interactions between proteins, ranging from 0.402 to 0.999. Degree represents the number of protein interactions. The network contained 28 nodes and 54 edges (combined score >0.4), indicating that 28 DEPs (11 down regulated and 7 upregulated) were involved in 54 interactions. The most interacting proteins were 40S ribosomal protein S27a and 60S ribosomal protein L10a, both of which interacted with 8 other proteins.

Displayed in Fig. 4B, the top 10 hub proteins ranked by MCC (Maximal Clique Centrality) method were 40S ribosomal protein S27a (150), 60S ribosomal protein L10a (150), 60S acidic ribosomal protein P2 (146), 40S ribosomal protein S4 (1 4 5), 40S ribosomal protein S18 (120), 60S ribosomal protein L35 (1 2 0), proteasome subunit alpha type-1 (26), peptidyl-prolyl cis–trans isomerase (PPIases, A8K486, 12), L-lactate dehydrogenase (A0A5F9ZHM4, 12) and TPI (11). For the comparison of Korean vS Han groups, TPI was down regulated while the rest were upregulated.

PPIases are a highly conserved family of immunophilins (Li et al., 2021) and play an important function in biological processes like protein folding, trafficking, assembly (Perrucci et al., 2015). PPIases can catalyze the isomerization of peptide bonds from trans to cis conformation to accelerate protein folding (Perrucci et al., 2015). These subfamilies are also reported to be related with cardiac and vascular diseases such as heart failure, arrhythmias, vascular stenosis (Perrucci et al., 2015). Lactate dehydrogenases (LDH) is an enzyme in the cytosol of various eukaryotic and prokaryotic organisms. LDH can assist the interconversion of pyruvate and lactate and oxidize nicotinamide adenine dinucleotide (NADH) to NAD$^+$, which is critical for cell survival and development (Kayamba et al., 2021). As mentioned in Section 3.5, TPI plays its catalytic role in glycolysis and gluconeogenesis, which may be a

potential target for anticancer studies by inhibition of this pathway since glycolysis plays a central role in cancer energy metabolism (Pekel & Ari, 2020).

4. Conclusions

In conclusion, DIA proteomics method can be effectively used to quantitatively study whey protein profile of human milk collected from Korean and Han nationality in China. A total of 624 proteins were detected in the investigated samples, of which 54 were differently expressed and deserve special attention. These different proteins included some immune related ones. Differences in the functions and protein interactions were observed due to variation in abundance of whey proteins. Results of this proteomic study may further deepen understanding of composition of mature human milk collected from different ethnic groups. Additionally, it is well known that the purpose of developing infant formula powder is that its composition is infinitely close to breast milk, and most of the breast milk is based on Han mothers. The results in this study showed that there were some differences in the breastmilk composition in the mothers of two different ethnic groups, which may help develop targeted infant formula for infants of Han or Korean groups according to responding breastmilk composition.

CRediT authorship contribution statement

Cuina Wang: Methodology, Writing – original draft, Writing – review & editing. **Yingcong Lu:** Methodology, Validation, Data curation, Writing – original draft. **Keyi He:** Software, Data curation. **Ru Zhao:** Software, Data curation. **Jianjun Cheng:** Validation, Data curation. **Shilong Jiang:** Validation. **Mingruo Guo:** Funding acquisition, Project administration, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.fochx.2023.100568.

References

Auestad, N., & Layman, D. K. (2021). Dairy bioactive proteins and peptides: A narrative review. *Nutrition Reviews, 79*(Suppl 2), 36–47. https://doi.org/10.1093/nutrit/nuab097

Basak, T., Bhat, A., Malakar, D., Pillai, M., & Sengupta, S. (2015). In-depth comparative proteomic analysis of yeast proteome using iTRAQ and SWATH based MS. *Molecular Biosystems, 11*(8), 2135–2143. https://doi.org/10.1039/c5mb00234f

Berdowska, I., Matusiewicz, M., & Krzystek-Korpacka, M. (2022). HDL accessory proteins in Parkinson's disease-focusing on clusterin (Apolipoprotein J) in regard to its involvement in pathology and diagnostics-A review. *Antioxidants, 11*(3). https://doi.org/10.3390/antiox11030524

Britigan, B. E., Lewis, T. S., Waldschmidt, M., McCormick, M. L., & Krieg, A. M. (2001). Lactoferrin binds CpG-containing oligonucleotides and inhibits their immunostimulatory effects on human B cells. *Journal of Immunology, 167*(5), 2921–2928. https://doi.org/10.4049/jimmunol.167.5.2921

Chen, J. Y., Liu, Y., Zhao, J. Y., Jiang, T. M., Xu, T., Liu, B., … Chen, L. J. (2022). Three urban areas of China: A cohort study of maternal-infant factors and HM protein components. *Food & Function, 13*(9), 5202–5214. https://doi.org/10.1039/d1fo04123a

De, P., Rozeboom, B. J., Aske, J. C., & Dey, N. (2020). Active RAC1 promotes tumorigenic phenotypes and therapy resistance in solid tumors. *Cancers, 12*(6). https://doi.org/10.3390/cancers12061541

Gomez, H. F., Ochoa, T. J., Carlin, L. G., & Cleary, T. G. (2003). Human lactoferrin impairs virulence of Shigella flexneri. *Journal of Infectious Diseases, 187*(1), 87–95. https://doi.org/10.1086/345875

Hara, K., Ikeda, M., Saito, S., Matsumoto, S., Numata, K., Kato, N., … Sekihara, H. (2002). Lactoferrin inhibits hepatitis B virus infection in cultured human hepatocytes. *Hepatology Research, 24*(3), 228–235. https://doi.org/10.1016/S1386-6346(02)00088-8

Hettinga, K., van Valenberg, H., de Vries, S., Boeren, S., van Hooijdonk, T., van Arendonk, J., & Vervoort, J. (2011). The host defense proteome of human and bovine milk. *PLoS One1, 6*(4). https://doi.org/10.1371/journal.pone.0019433

Hurley, W. L., & Theil, P. K. (2011). Perspectives on immunoglobulins in colostrum and milk. *Nutrients, 3*(4), 442–474. https://doi.org/10.3390/nu3040442

Igarashi, Y., Morishita, Y., Yoshizawa, H., Imai, R., Imai, T., Hirahara, I., … Nagata, D. (2017). The association between soluble intercellular adhesion molecule-1 levels in drained dialysate and peritoneal injury in peritoneal dialysis. *Renal Failure, 39*(1), 392–399. https://doi.org/10.1080/0886022x.2017.1287735

Jin, D. P., Liu, H., Bu, L. L., Ke, Q. H., Li, Z. Y., Han, W. N., … Liu, C. H. (2021). Comparative analysis of whey proteins in human milk using a data-independent acquisition proteomics approach during the lactation period. *Journal of Agricultural and Food Chemistry, 69*(14), 4319–4330. https://doi.org/10.1021/acs.jafc.1c00186

Johansen, F. E., Braathen, R., & Brandtzaeg, P. (2000). Role of J chain in secretory immunoglobulin formation. *Scandinavian Journal of Immunology, 52*(3), 240–248. <Go to ISI>://WOS:000088805900004.

Kaetzel, C. S. (2005). The polymeric immunoglobulin receptor: Bridging innate and adaptive immune responses at mucosal surfaces. *Immunological Reviews, 206*, 83–99. https://doi.org/10.1111/j.0105-2896.2005.00278.x

Kamau, S. M., Cheison, S. C., Chen, W., Liu, X. M., & Lu, R. R. (2010). Alpha-lactalbumin: Its production technologies and bioactive peptides. *Comprehensive Reviews in Food Science and Food Safety, 9*(2), 197–212. https://doi.org/10.1111/j.1541-4337.2009.00100.x

Kayamba, F., Faya, M., Pooe, O. J., Kushwaha, B., Kushwaha, N. D., Obakachi, V. A., … Karpoormath, R. (2021). Lactate dehydrogenase and malate dehydrogenase: Potential antiparasitic targets for drug development studies. *Bioorganic & Medicinal Chemistry, 50*. https://doi.org/10.1016/j.bmc.2021.116458

Li, J. Y., Mo, C. F., Guo, Y. F., Zhang, B. W., Feng, X., Si, Q. Y., … Shao, J. C. (2021). Roles of peptidyl-prolyl isomerase Pin1 in disease pathogenesis. *Theranostics, 11*(7), 3348–3358. https://doi.org/10.7150/thno.45889

Liao, Y., Alvarado, R., Phinney, B., & Lonnerdal, B. (2011). Proteomic characterization of human milk whey proteins during a twelve-month lactation period. *Journal of Proteome Research, 10*(4), 1746–1754. https://doi.org/10.1021/pr101028k

Ling, J. M. L., & Schryvers, A. B. (2006). Perspectives on interactions between lactoferrin and bacteria. *Biochemistry and Cell Biology, 84*(3), 275–281. https://doi.org/10.1139/O06-044

Lu, J., Zhang, S. W., Liu, L., Pang, X. Y., Ma, C. L., Jiang, S. L., & Lv, J. P. (2018). Comparative proteomics analysis of human and ruminant milk serum reveals variation in protection and nutrition. *Food Chemistry, 261*, 274–282. https://doi.org/10.1016/j.foodchem.2018.04.065

Pekel, G., & Ari, F. (2020). Therapeutic targeting of cancer metabolism with triosephosphate isomerase. *Chemistry & Biodiversity, 17*(5). https://doi.org/10.1002/cbdv.202000012

Peng, M., Deng, J., Zhou, S. C., Tao, T., Su, Q. L., Yang, X., & Yang, X. P. (2019). The role of Clusterin in cancer metastasis. *Cancer Management and Research, 11*, 2405–2414. https://doi.org/10.2147/Cmar.S196273

Permyakov, E. A., & Berliner, L. J. (2000). alpha-Lactalbumin: Structure and function. *FEBS Letters, 473*(3), 269–274. https://doi.org/10.1016/s0014-5793(00)01546-5

Perrucci, G. L., Gowran, A., Zanobini, M., Capogrossi, M. C., Pompilio, G., & Nigro, P. (2015). Peptidyl-prolyl isomerases: A full cast of critical actors in cardiovascular diseases. *Cardiovascular Research, 106*(3), 353–364. https://doi.org/10.1093/cvr/cvv096

Shahiduzzaman, M., & Coombs, K. M. (2012). Activity based protein profiling to detect serine hydrolase alterations in virus infected cells. *Frontiers in Microbiology, 3*. https://doi.org/10.3389/fmicb.2012.00308

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., … Ideker, T. (2003). Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Research, 13*(11), 2498–2504. https://doi.org/10.1101/gr.1239303

Shukla, S. D., Shastri, M. D., Vanka, S. K., Jha, N. K., Dureja, H., Gupta, G., … Walters, E. H. (2022). Targeting intercellular adhesion molecule-1 (ICAM-1) to reduce rhinovirus-induced acute exacerbations in chronic respiratory diseases. *Inflammopharmacology, 30*(3), 725–735. https://doi.org/10.1007/s10787-022-00968-2

Singh, R., Thakur, M., Mishra, M., Yadav, M., Vibhuti, R., Menon, A. M., … Yadav, V. (2021). Gene regulation of intracellular adhesion molecule-1 (ICAM-1): A molecule with multiple functions. *Immunology Letters, 240*, 123–136. https://doi.org/10.1016/j.imlet.2021.10.007

Stamatelopoulos, K., Sibbing, D., Rallidis, L. S., Georgiopoulos, G., Stakos, D., Braun, S., … Stellos, K. (2015). Amyloid-beta (1–40) and the risk of death from cardiovascular

causes in patients with coronary heart disease. *Journal of the American College of Cardiology, 65*(9), 904–916. https://doi.org/10.1016/j.jacc.2014.12.035

Tian, X., Wang, Y., Zhou, Y., Wu, B., Lu, Y., Du, J., … Xiao, Y. (2022). Beta-amyloid deposition in biliary atresia reduces liver regeneration by inhibiting energy metabolism and mammalian target of rapamycin signaling. *Clinical and Translational Gastroenterology, 13*(11), e00536.

Wang, R. H., Yin, Y. D., & Zhu, Z. J. (2019). Advancing untargeted metabolomics using data-independent acquisition mass spectrometry technology. *Analytical and Bioanalytical Chemistry, 411*(19), 4349–4357. https://doi.org/10.1007/s00216-019-01709-1

Wang, X. H., Liu, D. B., Wei, F. Z., Li, Y., Wang, X. F., Li, L. J., … Zhang, L. (2020). Stress-sensitive protein Rac1 and its involvement in neurodevelopmental disorders. *Neural Plasticity, 2020*. https://doi.org/10.1155/2020/8894372

Wisniewski, J. R., Zougman, A., Nagaraj, N., & Mann, M. (2009). Universal sample preparation method for proteome analysis. *Nature Methods, 6*(5), 359–U360. https://doi.org/10.1038/Nmeth.1322

Yan, J., Zhai, H. N., Zhu, L., Sa, S. S., & Ding, X. J. (2021). obaDIA: One-step biological analysis pipeline for data-independent acquisition and other quantitative proteomics data. *Bioinformatics, 37*(14), 2066–2067. https://doi.org/10.1093/bioinformatics/btaa893

Yang, M., Cao, X. Y., Wu, R. N., Liu, B. A., Ye, W. H., Yue, X. Q., & Wu, J. R. (2017). Comparative proteomic exploration of whey proteins in human and bovine colostrum and mature milk using iTRAQ-coupled LC-MS/MS. *International Journal of Food Sciences and Nutrition, 68*(6), 671–681. https://doi.org/10.1080/09637486.2017.1279129

Yang, S. H., Niu, X. L., Luo, D., Chen, C. D., Yu, X., Tang, W., … Liu, Y. S. (2012). Functional characterization of an aldehyde dehydrogenase homologue in rice. *Journal of Integrative Agriculture, 11*(9), 1434–1444. <Go to ISI>://WOS:000308856400008.

Zhang, F. L., Pu, J. H., Gu, Z. B., & Mao, H. M. (2020). DIA proteomics reveals hypotensive and immune-enhancing constituents in buffalo whey from different altitudes. *International Journal of Biological Macromolecules, 164*, 4146–4154. https://doi.org/10.1016/j.ijbiomac.2020.08.213

Zhang, L. N., Ma, Y., Yang, Z. Y., Jiang, S., Liu, J., Hettinga, K. A., … Zhou, P. (2019). Geography and ethnicity related variation in the Chinese human milk serum proteome. *Food & Function, 10*(12), 7818–7827. https://doi.org/10.1039/c9fo01591d

Zhang, Q., Cundiff, J. K., Maria, S. D., McMahon, R. J., Woo, J. G., Davidson, B. S., & Morrow, A. L. (2013). Quantitative analysis of the human milk whey proteome reveals developing milk and mammary-gland functions across the first year of lactation. *Proteomes, 1*(2), 128–158. https://doi.org/10.3390/proteomes1020128

Zhao, Z. X., Liu, N., Wang, C. N., Cheng, J. J., & Guo, M. R. (2021). Proteomic analysis of differentially expressed whey proteins in Saanen goat milk from different provinces in China using a data-independent acquisition technique. *Journal of Dairy Science, 104*(10), 10513–10527. https://doi.org/10.3168/jds.2020-19877