

Comprehensive Human Milk Patterns Are Related to Infant Growth and Allergy in the CHMP Study

Shilong Jiang, Jiancun Pan, Yuanyuan Li, Mengnan Ju, Wei Zhang, Jing Lu,* Jiaping Lv,* and Kaifeng Li*

Scope: The aim of the present study is to identify human milk pattern using multi-omics datasets and to explore association between patterns, infant growth, and allergy using data from the Chinese Human Milk Project (CHMP) study.

Methods and Results: Three patterns are identified from integrative analysis of proteome, lipidome, and glycome profiles of 143 mature human milk samples. Factor 1 is positively associated with 128 proteins, phospholipids, and human milk oligosaccharides (HMOs) including lacto N-neohexaose (LNnH) and lacto-N-difucohexaose II (LNDFH II); factor 2 is negatively associated with α_1 -casein, phospholipids while positively associates with HMOs including LNnH, lactosialyl tetrasaccharide c (LSTc), and 2'-fucosyllactose (2'FL); factor 3 is positively associated with lysophospholipids while negatively associates with 27 proteins, triglycerides with two saturated fatty acids, 6'-sialyllactose (6'SL) and 2'FL. In general, factor 1 and factor 2 are associated with slower while factor 3 is associated with faster growth rate ($p < 0.044$). One unit higher in loadings of factor 2 is associated with 34% lower risk of allergies ($p \leq 0.017$). Associations are not significant after adjustment for city except for factor 1.

Conclusions: Three possible human milk patterns with varying degree of stability are identified. Future work is needed to understand these patterns in terms of generalization, biologic mechanisms, and genotype influences.

our understanding of the complex nature of HM. Importantly, recent studies have documented that differences in HM composition may further translate into heterogeneous trajectory of infant growth and altered risk of food allergy.^[6–9] However, endogenous metabolisms are complex and largely inter-connected.^[10] For example, studies have suggested that global lipid metabolism are related to protein signaling network, resulting in coregulation of both dynamic lipid and protein metabolism.^[11] On the other hand, evidence have shown that several proteins (such as β 1-4 galactosyltransferase and α -lactalbumin) are linked to the biosynthesis of (human milk oligosaccharide) HMOs despite the fact that the overall HMO biosynthesis remains largely unknown.^[12] Therefore, reliance on analyses of single omics data may not be sufficient to reveal the underlying complexity of molecular events. To help understand the impact of this complex interplay, it has been proposed that analysis of HM pattern may provide better opportunities to clarify the association between HM composition and health outcomes.^[13]

1. Introduction

Human milk (HM) is a complex and variable matrix of constituents each contributing to infant health outcomes,^[1] therefore is universally recommended as the optimal choice for infant feeding.^[2] Numerous studies have been carried out to characterize proteome,^[3] lipidome,^[4] and glycome^[5] profiles to deepen

Metabotype (metabolic phenotype) describes the categorization of individuals based on combination of their metabolic or phenotypic characteristics.^[14] Subjects within the same metabotype share similar metabolic profile therefore tend to present homogenous health outcome. This concept has been tested in some studies, whereby personalized nutrition interventions have been successfully delivered using the metabotype routine.^[15–17]

S. Jiang, J. Pan, Y. Li, M. Ju, W. Zhang, K. Li
Nutrition and Metabolism Research Division
Innovation Center
Heilongjiang Feihe Dairy Co., Ltd.
Beijing 100015, China
E-mail: likaifeng@feihe.com

S. Jiang, J. Pan, Y. Li, M. Ju, W. Zhang, K. Li
PKUHSC-China Feihe Joint Research Institute of Nutrition and Healthy
Lifespan Development
Xueyuan Road 38, Haidian, Beijing 100083, China
J. Lu, J. Lv
Key Laboratory of Agro-Food Processing and Quality Control
Institute of Food Science and Technology
Chinese Academy of Agricultural Sciences
Beijing, China
E-mail: lujing@btbu.edu.cn kjdairy@126.com

J. Lu
School of Food and Health
Beijing Technology and Business University
Fucheng Road 11, Haidian, Beijing 100048, China

The ORCID identification number(s) for the author(s) of this article can be found under <https://doi.org/10.1002/mnfr.202100011>

DOI: 10.1002/mnfr.202100011

Table 1. Maternal and infant characteristics (n = 143).

	Total (n = 143)	Beijing (n = 26)	Guangzhou (n = 21)	Lanzhou (n = 29)	Weihai (n = 14)	Jinhua (n = 30)	Chengdu (n = 23)	p
Maternal characteristics								
Age [years]	30.5 ± 5.0	33.1 ± 6.1	30.0 ± 3.2	29.6 ± 3.1	26.5 ± 8.6	31.1 ± 3.3	30.5 ± 4.3	0.002
BMI [kg m ⁻²]	22.4 ± 3.0	22.6 ± 2.8	23.2 ± 3.1	20.9 ± 2.9	24.0 ± 3.6	22.7 ± 2.4	21.6 ± 2.3	0.007
Gestational weight gain [kg]	13.8 ± 4.9	14.1 ± 4.1	12.6 ± 5.7	13.7 ± 4.7	16.3 ± 5.0	11.4 ± 12.1	13.7 ± 4.8	0.390
Parity								0.142
1	93 (65.0%)	18 (69.0%)	9 (43.0%)	24 (83.0%)	7 (70.0%)	18 (62.0%)	17 (81.0%)	
2	42 (29.4%)	8 (31.0%)	12 (57.0%)	5 (17.0%)	3 (30.0%)	10 (34.0%)	4 (19.0%)	
3	1 (0.7%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (4.0%)	0 (0.0%)	
Infant characteristics								
Sex [n]								0.439
Male	74 (52.1%)	14 (53.8%)	14 (70.0%)	14 (48.3%)	5 (35.7%)	14 (46.7%)	13 (52.1%)	
Female	68 (47.9%)	12 (46.2%)	6 (30.0%)	15 (51.7%)	9 (64.3%)	16 (53.3%)	10 (47.9%)	
Birth weight [g]	3343.0 ± 537.0	3541.9 ± 616.8	3371.9 ± 796.5	3336.7 ± 423.1	3428.6 ± 330.3	3243.1 ± 452.9	3180.0 ± 439.8	0.209
Birth length [cm]	50.0 ± 1.9	50.7 ± 1.6	49.2 ± 1.9	50.7 ± 2.2	49.3 ± 1.3	50.1 ± 1.6	49.5 ± 2.2	0.014
Age								0.288
1 mo	4 (2.8%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	3 (10.0%)	1 (4.3%)	
2 mo	23 (16.1%)	4 (15.4%)	2 (9.5%)	3 (10.3%)	3 (21.4%)	5 (16.7%)	6 (26.1%)	
3 mo	27 (18.9%)	6 (23.1%)	4 (19.0%)	5 (17.2%)	2 (14.3%)	32 (18.9%)	5 (21.7%)	
4 mo	35 (24.5%)	5 (19.2%)	4 (19.0%)	9 (31.0%)	2 (14.3%)	40 (24.5%)	9 (39.1%)	
5 mo	43 (30.0%)	9 (34.6%)	8 (38.1%)	11 (37.9%)	4 (28.6%)	48 (30.0%)	1 (4.3%)	
6 mo	11 (7.7%)	2 (7.7%)	3 (14.3%)	1 (3.4%)	3 (21.4%)	16 (7.7%)	1 (4.3%)	
Self-reported allergy, Yes	56 (39.2%)	12 (46.2%)	6 (28.6%)	9 (31.0%)	8 (57.1%)	9 (30.0%)	12 (52.2%)	0.227
Z Score								
WAZ	0.58 ± 1.62	1.2 ± 1.4	0.5 ± 2.0	0.6 ± 1.0	0.5 ± 1.2	-0.2 ± 1.4	0.9 ± 2.2	0.046
LAZ	0.22 ± 1.36	1.0 ± 1.3	-0.03 ± 1.4	0.6 ± 1.3	-0.13 ± 1.3	-0.2 ± 1.3	-0.2 ± 1.1	0.002
BAZ	0.66 ± 2.12	0.9 ± 2.2	0.7 ± 2.9	0.4 ± 1.3	0.8 ± 1.0	-0.04 ± 1.7	1.5 ± 3.2	0.218
WFL	0.73 ± 2.21	0.8 ± 2.1	0.9 ± 3.0	0.4 ± 1.4	1.0 ± 0.9	0.01 ± 1.9	1.6 ± 3.0	0.188

BAZ, BMI-for-age z score; LAZ, length-for-age z score; WAZ, weight-for-age z score; WFL, weight-for-length z score.

A recent study determined urinary metabolome from over 1800 samples showed that urinary pattern may be used as objective measure to categorize subjects into either healthy versus unhealthy dietary pattern with a accuracy of 74–76%.^[18]

To our knowledge, only one pilot study explored preterm HM metabolome using integrated omics data, with distinct preterm HM patterns being derived which were associated with either “fast” or “slow” infant growth patterns.^[19] Beyond this, little is known in terms of what patterns/metabolome could be identified from multi-omics datasets of mature HM and how those patterns are associated with infant growth and risk of allergies. Against this backdrop, the objective of present study was to explore patterns using proteome, lipidome, and glycome data of mature HM and to investigate association between HM patterns, infant growth, and allergy based on the Chinese Human Milk Project (CHMP).

2. Results

In general, 143 participants (of which 93 were primiparous with mean maternal age of 30.5 ± 5.0 years and self-reported gesta-

tional weight gain of 13.8 ± 4.9 kg) from six cities of China were included in this analysis (Table 1). Over half of the infants (52.1%) were boys and mean (SD) birth weight and length were 3343 (537) g and 50.0 (1.9) cm, respectively.

2.1. Characteristics of Human Milk Patterns

Three HM patterns were identified using multiple factor analysis (MFA), which explained 39.6% of the total variance. Characteristics of three human milk patterns were summarized in Table 2. Factor 1 was positively associated with 128 proteins that are related to development of immune system while negatively associated with β -casein. In terms of lipidome profile, factor 1 was positively associated with phosphatidylethanolamines (PE), phosphatidylcholine (PC), ceramide, sphingomyelin (SM), and triglyceride (TG) with 2 or 3 unsaturated fatty acids (TG-SU2 and TG-UUU), while negatively associated with TG-SSS. Compared with factor 1, factor 2 was positively associated with TG-SSS while negatively associated with α ₁-casein, PE, PC, and SM. Factor 3 was negatively associated with 27 proteins that related to immunity while positively

Table 2. Summary of human milk metabolotypes obtained from multiple factor analysis ($n = 143$).

	Proteomics		Lipidomics		HMO		Variance explained
	Positive	Negative	Positive	Negative	Positive	Negative	
Factor 1	128 proteins related to platelet degranulation, endopeptidase activity, signal transduction, immune response, etc.	β -casein	Phospholipids including PE, PC, Ceramide, SM, TG-SU2, TG-UUU	TG-SSS	LNnH, LNDHF II	3'SL	16.30%
Factor 2	Vitronectin, CD81 molecule, complement C4A, fibroblast growth factor binding protein 1, milk fat globule-EGF Factor 8 protein, immunoglobulin heavy constant gamma 4	as ₁ -casein	TG-SSS	Phospholipids including PE, PC, SM	LNnH, LSTa, LSTb, LSTc, 3'FL, 2'FL	DSLNT	13.50%
Factor 3	None	27 proteins related to platelet degranulation, endopeptidase activity, innate immune response etc.	Phospholipids including PE, PC, PI, LPC, LPE, LPI, SM, Ceramide	TG-S2U	LSTC, LNFP III	6'SL, 2'FL	9.70%

Cb, Cerebroside; HMO, human milk oligosaccharides; LPC, lysophosphatidylcholine; LPE, lysophosphatidylethanolamine; LPI, lysophosphatidylinositol; PC, phosphatidylcholine; PE, phosphatidylethanolamine; PI, phosphatidylinositol; SM, sphingomyelin; TG, triglyceride; TG-S2U, triglyceride with two saturated fatty acids and one unsaturated fatty acid; TG-SSS, triglyceride with three saturated fatty acids; TG-SU2, triglyceride with one saturated fatty acid and two unsaturated fatty acids; TG-UUU, triglyceride with three unsaturated fatty acids; 2'FL, 2'-fucosyllactose; 3'FL, 3'-fucosyllactose; 3'SL, 3'-sialyllactose; 6'SL, 6'-sialyllactose; DSLNT, disialyllacto-N-tetraose; LNFP III, lacto N-fucopentaose III; LNDHF II, lacto-N-difuco-hexaose II; LNnH, lacto N-neohexaose; LSTa, lactosialyl tetrasaccharide a; LSTb, lactosialyl tetrasaccharide b; LSTc, lactosialyl tetrasaccharide c.

associated with lysophospholipids (LPL, including lysophosphatidylcholine [LPC], lysophosphatidylethanolamine [LPE], and lysophosphatidylinositols [LPI]). Regarding the HMO profile, factor 2 was positively associated with 2'-fucosyllactose (2'FL), lacto N-neohexaose (LNnH), lactosialyl tetrasaccharide a (LSTa), lactosialyl tetrasaccharide (LSTb), lactosialyl tetrasaccharide c (LSTc), and 3'-fucosyllactose (3'FL), while factor 3 was negatively associated with 2'FL and 6'-sialyllactose (6'SL).

2.2. Linear Regression Between Infant z Scores and Three Factor Loadings

Results of multiple linear regression between infant z scores and the factor loadings are presented in Table 3. Factor 1 was negatively (estimate = -0.28, $p = 0.002$) associated with infant length-for-age z scores (LAZ) under the crude model (no adjustment applied), and the associations remain significant (estimate = -0.20, $p = 0.020$) after further adjustment for infant age, infant sex, birth weight, birth length and maternal age (model 3). On the other hand, factor 2 was negatively associated with infant BMI-for-age z scores (BAZ, model 1) and weight-for-age z scores (WAZ, model 3). Importantly, breast milk with one unit higher of factor 2 loading was associated with 35% lower risk of infant allergy events (OR = 0.65, $p = 0.013$). In contrast, factor 3 was positively (estimate = 0.23, $p = 0.044$) associated with infant LAZ after adjustment for infant age and sex. However, none of these associations were significant after adjustment for city (model 2) except for factor 1.

2.3. Factor Loadings Across Cities and Infant Age

Three factor loadings across infant age are shown in Figure 1. Loadings of factor 1 was negatively associated with infant age

($p = 3.65 \times 10^{-5}$) after adjustment for cities, maternal BMI, parity, infant birth weight, infant birth length, and infant sex. No significant associations were observed for factor 2 or factor 3. Factor loadings of six cities are presented in Figure 2. In general, samples from Beijing showed the lowest loadings for factor 2 (-1.1 ± 0.6) while the highest loadings for factor 3 (1.2 ± 0.4). Samples from Weihai, on the other hand, showed the highest loadings for factor 1 (1.0 ± 0.4) while the lowest loadings (-1.0 ± 0.4) for factor 3. Samples of Lanzhou had the greatest loadings for factor 2 (1.2 ± 0.9). MFA score plot of multi-omics data and PCA score plot of individual dataset are presented in Figure S1, Supporting Information. City-specific effects were observed from MFA score plot, where samples from Beijing and Guangzhou are generally separated from others.

2.4. Robust Analysis

Results of robust analysis are shown in Figure S3 and Table S1, Supporting Information. In general, factor 1 was robust against bootstrap resampling, with mean bootstrap loadings (MBL) being overlapped with original loadings (Figure S3, Supporting Information). Factor 3 showed relatively fair robustness as the absolute MBL were generally lower than the original ones. In contrast, factor 2 tended to be less stable across bootstrap samples, with substantially lower absolute MBL and relatively high SD of bootstrap loadings being observed.

3. Discussion

In this study, three distinct human milk patterns were identified using multi-omics datasets from 143 Chinese mature human milk samples, of which factor 1 and factor 2 were associated with

Table 3. Association between human milk metabolotypes, infant growth, and allergy events (n = 143).

		Factor 1		Factor 2		Factor 3	
		Estimates	p value	Estimates	p value	Estimates	p value
BAZ	Crude model	0.05	0.742	−0.35	0.031	−0.08	0.676
	Model 1	−0.01	0.938	−0.34	0.037	−0.09	0.621
	Model 2	0.25	0.433	−0.26	0.337	0.08	0.788
	Model 3	−0.03	0.838	−0.32	0.053	0.01	0.974
	Model 4	0.24	0.478	−0.17	0.561	0.14	0.649
WAZ	Crude model	−0.13	0.210	−0.22	0.057	0.07	0.620
	Model 1	−0.16	0.173	−0.25	0.039	0.05	0.698
	Model 2	−0.10	0.657	−0.12	0.557	0.003	0.989
	Model 3	−0.14	0.241	−0.26	0.033	0.01	0.928
	Model 4	−0.06	0.811	−0.09	0.679	−0.06	0.793
LAZ	Crude model	−0.28	0.002	0.01	0.896	0.24	0.040
	Model 1	−0.26	0.007	−0.02	0.814	0.23	0.044
	Model 2	−0.41	0.028	0.08	0.613	−0.06	0.737
	Model 3	−0.20	0.020	−0.08	0.391	0.04	0.693
	Model 4	−0.32	0.108	−0.05	0.783	−0.21	0.236
Allergies ¹	Crude model	1.01	0.936	0.67	0.014	0.86	0.404
	Model 1	0.93	0.617	0.67	0.017	0.84	0.318
	Model 2	0.53	0.077	0.74	0.281	0.68	0.271
	Model 3	0.91	0.527	0.65	0.013	0.84	0.368
	Model 4	0.53	0.091	0.97	0.993	0.67	0.218

¹) Estimates were presented as odds ratio. Model 1: adjusted for infant age and sex; Model 2 adjusted for infant age, sex, and city; Model 3 adjusted for infant age, infant sex, birth weight, birth length, and maternal age; Model 4 adjusted for infant age, infant sex, birth weight, birth length, maternal age, maternal BMI, city, and parity. BAZ, BMI-forSage z score; LAZ, length-for-age z score; WAZ, weight-for-age z score.

slower while factor 3 was associated with a greater infant growth during the first 6 months of life. Importantly, factor 2 was related to significantly lower risk of allergies of the infants.

Several recent analyses have been carried out to explore HM patterns/metabolotypes using metabolome,^[7] lipidome^[8] or multi-omics^[19] data. Those studies reported that HM patterns/metabolotypes are related to heterogenous infant growth trajectories and health outcome, which is in line with our study.

Interestingly, a human milk pattern (factor 3) being identified in the present study was associated with higher LAZ. No positive association was observed between factor 3 and proteome profile. In contrast, several proteins that were detected exclusively in human milk (including amyloid beta precursor protein, epidermal growth factor and immunoglobulin heavy constant alpha) were significantly lower in samples with high loadings of factor 3.^[20,21] Previous reports suggested that breastfed subjects had lower growth rates and lower risk of being overweight/obese later in life compared with their formula fed counterparts.^[2,22–24] It is possible that the altered growth pattern of infants may be partly due to the lacking of several essential bioactive proteins in factor 3.^[25] Besides, factor 3 also showed unique HMO profile compared with factor 1 and factor 2, with 6'SL and 2'FL being negatively associated with factor 3. In general, current evidence on HMO in relation to infant growth is inconclusive, with substantial variations being observed across geographic locations.^[26–29] Interestingly, however, two studies reported negative association

between 6'SL content and infant growth, which is in agreement with our results.^[27,28] It has been proposed that those variations may be partly due to the differing composition of gut microbiota as HMO might interact differently according to the intestine microbiome profile.^[28] Future studies are needed to take microbiota profile into account when exploring association between HMO content and infant growth.

It is worth noting that factor 3 was positively associated with LPC, LPE and LPI. Recently, positive associations were reported between umbilical cord blood LPC/LPE levels and infant birth weight.^[30] Animal studies also showed that diets rich in LPL promoted growth performance of sow and broilers by modulating intestine absorption of dietary fat.^[31,32] However, the associations between human milk LPL and infant growth remain unclear at the moment, and future work is needed to replicate our results and to explore potential health effects of human milk LPL.

The major characteristics that discriminate factor 1 and factor 2 were the opposite lipidome profile and the differing abundance in terms of HMO (Table 2). The associations between HMO and risk of allergies have been poorly documented, and current evidence tend to suggest that secretor status and Lewis blood type may play important role in the relationship between HMO and risk of allergies. For example, a recent longitudinal cohort study suggested that exposure to acidic Lewis HMO patterns (high 3'FL, 3'SL while low lacto N-fucopentaose I [LNFP I]) lead to increased risk of allergic events of the offspring.^[33] A study

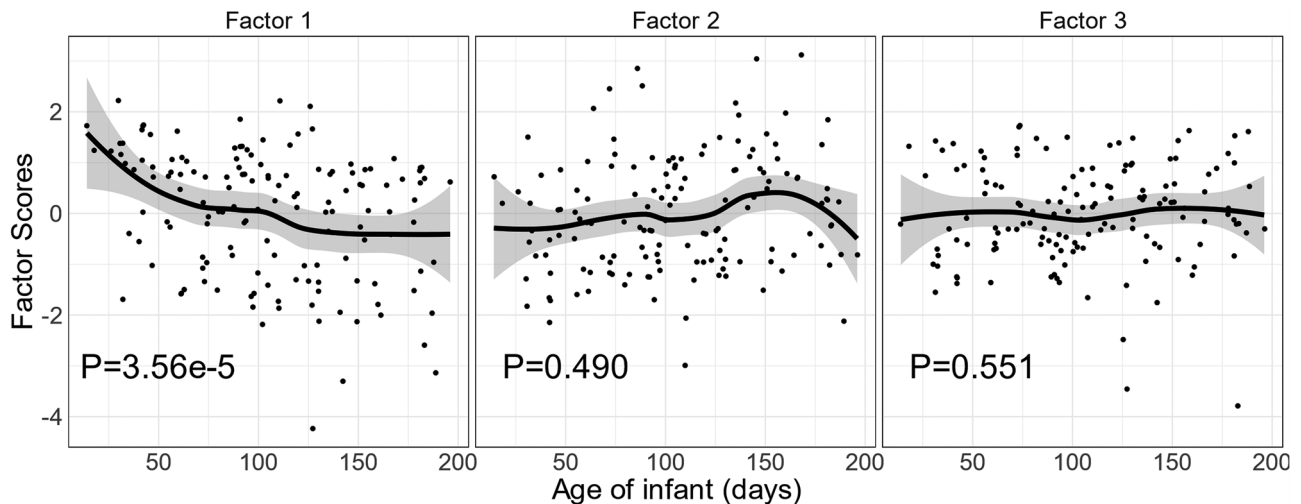


Figure 1. Three breast milk factor scores across age of infants ($n = 143$). Black dots in the figure represent factor score by infant age of each samples. Black lines and grey areas represent association and corresponding SE between factor scores and infant age, which were estimated and visualized using *geom_smooth* function of *ggplot2* R package. P values were based on multiple linear regression adjusted for cities, maternal BMI, parity, infant birth weight, infant birth length, and infant sex.

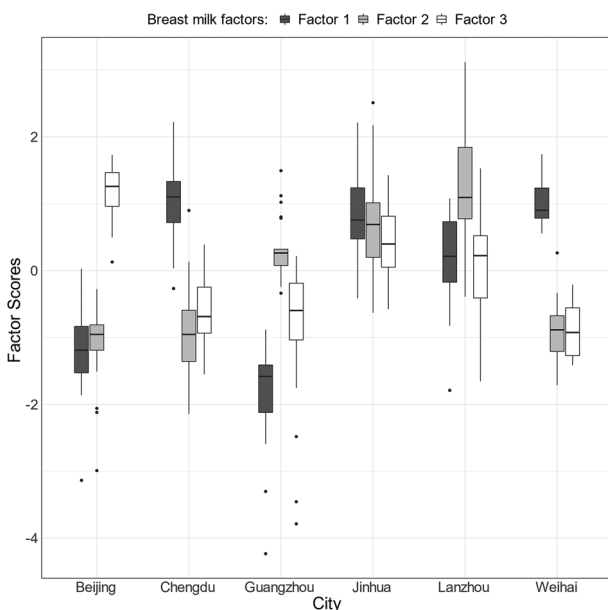


Figure 2. Three breast milk factor scores across six cities ($n = 143$). Individual box plots show each factor scores the median, 25th and 75th percentile across cities. Different factors were colored in discrete greyscales.

based on the CHILD cohort reported negative associations between 2'FL, LNFP I and risk of food sensitization.^[9] In addition, LNFP I and 2'FL have also been reported to be protective against cow's milk allergy or eczema.^[34,35] Similarly, factor 2 in our study with a clear secretor phenotype (positive loading for 2'FL) was associated with reduced risk of allergies.

It has been documented that gut microbiota may alter immune tolerance and therefore may be able to regulate immune response to dietary antigens.^[36] It is likely that the HMO profile of factor 2 modulated the gut bacterial of the infants through

their prebiotic effects which in turn lead to enhanced immune tolerance to food antigen. Notably, factor 2 was negatively associated with α_1 -casein, which is one of the most common allergens in cow's milk.^[37] It is possible that lower α_1 -casein of factor 2 also contribute to the reduced risk of allergies of infants. Generally, relationship between HM α_1 -casein and risk of allergies has been poorly studied, which calls for future investigation. In addition, factor 2 was negatively associated with phospholipid including PC and PE. It has been reported that lipids (especially phospholipids) facilitate intestinal uptake of food allergens by protecting them from gastrointestinal degradation, which may promote allergic process.^[38] This is supported by results showing that PC inhibits digestion of milk allergens β -lactoglobulin and α -lactalbumin, which in turn leads to increased allergenic activities of those allergens.^[39,40] Collectively, the profile of factor 2 strongly supports its protective effect against risk of allergies, despite that the negative association was not significant after adjusting for cities. Future studies are needed to replicate this result in different population.

Bootstrap resampling technic was used to elucidate stability of each factors. Factor 1 showed great stability under bootstrap resampling, this is supported by the result that association between factor 1 and LAZ is robust against confounder adjustment (Table 3). In contrast, factor 2 tended to be less stable under the bootstrap scheme, however, the features with absolute original loadings over 0.4 generally remain significant across bootstrap samples (819-998 out of 1000 bootstrap samples as shown in Table S1, Supporting Information). Therefore, it seems that the overall structure of factor 2 still exist beyond resampling even though loadings have been substantially weakened.

To our knowledge, only one study explored human milk pattern using integrated omics datasets.^[19] However, this study was focused on preterm infants with a relatively small sample size ($n = 26$). Our study included 143 mother-infant dyads which is the largest study to date to examine HM patterns using a comprehensive multi-omics approach. The limitation of

our study is the cross-sectional study design which only reveal correlation rather than causality. Another limitation is that the HM samples included in our study distributed evenly across one to 6 months postpartum which might affect the association between patterns and infant growth. However, only the loadings for factor 1 decreased across infant age, and we carefully adjusted for confounders including infant age for regression analyses which should minimize the impact of infant age on the final results. Of note, city-specific effects were observed in present study, with some of the associations being not significant after further adjustment for cities. Given that factor 2 was sensitive to resampling, indicating that the hypotheses generated here should be re-investigated in different population groups. Lastly, FUT2 and FUT3 genotype and Lewis blood group were not determined in our study which might be one of the underlying driving factors of each HM pattern. In addition, disialyllacto-N-tetraose (DSLNT) was the prevalent HMO in our study. Considering that neither of the two biggest studies investigating Chinese HMOs reported DSLNT concentrations so far^[41,42] and the fact that DSLNT concentrations vary substantially across geographic locations,^[43] future studies are required to investigate DSLNT concentrations in Chinese mature human milk.

To conclude, the present study identified three possible human milk patterns using multi-omics datasets which showed varying degree of stability. Further analyses showed heterogeneous effects of HM patterns on infant growth patterns and risk of allergies. Future work is needed to understand these HM patterns in terms of generalization, biologic mechanisms, and genotype influences.

4. Experimental Section

The Chinese Human Milk Project: The CHMP study was a cross-sectional study which recruited 1800 participants from eight cities of Mainland China (including Guangzhou, Weihai, Chengdu, Beijing, Harbin, Jinhua, Zhengzhou, and Lanzhou) to evaluate human milk composition in a Chinese population (NCT03675204). Healthy lactating mothers aged 25–35 years who were breast-feeding infants of 15–180 days and who were not smoking and not alcohol drinkers were recruited in the study. Exclusion criteria includes mothers who were being treated for gastrointestinal symptoms; who were suffering from mastitis, infectious diseases, cardiovascular disease, metabolic diseases, mental system diseases, cancer and other malignant diseases; history of antibiotics taking; who were not able to answer the study questions; who participated in any nutritional or pharmaceutical intervention study recently.

Participants were asked to collect a human milk sample at 9:00–11:00 in the morning using an electric pump by expressing one breast fully, weight and volume of milk were then record. The milk was gently mixed to ensure homogenization and then aliquoted to 6 × 5 mL tubes first and the remaining samples were aliquoted to 10 mL tubes as needed. Aliquoted samples were stored at -20 °C immediately and transferred to -80 °C refrigerator. Samples from different centers were delivered to the lab in Beijing using dried ice. Temperatures during delivery were carefully monitored and all the delivery were completed within 3 days. Final samples were stored under -80 °C until further analysis. A questionnaire of maternal and infant characteristics was administrated for each participant. The study was approved by the review board of Shanghai Nutrition Society Ethic Committee (Ethic Approval [2016] No. 006). Informed written consent was obtained from all participants.

Analysis Sample: The data of present study was based on the CHMP study, 150 mother-infant dyads from six cities (Guangzhou, Weihai,

Chengdu, Beijing, Jinhua and Lanzhou) were randomly selected. Subjects who gave birth to a preterm infant (<37-week gestation) or with a birth weight of <2500 g (n = 7) were further excluded, leaving a final sample size of 143.

Maternal and Infant Characteristics: Maternal education, occupation, parity, delivery mode, family annual income, pre-pregnancy weight, postpartum weight, sleeping duration, smoking and drinking habits, infants' weight, and length were assessed by questionnaire. Allergy was defined as having one of the following diseases since birth of the infant: eczema, rhinitis, asthma, and atopic dermatitis, which was estimated by a self-reported questionnaire. Maternal body weight and height were measured upon sampling by calibrated electronic scales while participants were only wearing indoor clothing and no shoes. Weight and length of infant were measured using a tared scale and a length board with a sliding foot piece, respectively. LAZ, WAZ, BAZ, and weight-for-length z scores were calculated based on the WHO Child Growth Standards.^[44]

Determination of Proteomics: The human milk samples were centrifuged at 1500 × g for 10 mins at 10 °C to remove fat. BCA assay was used to determine the protein concentrations in milk samples. Milk samples were further digested as described previously.^[45] In brief, 10 µL of 0.1 M dithiothreitol was added to 100 µL sample solution containing 10 µg of protein in 0.05 M NH₄HCO₃. The mixture was incubated for 30 min at 56 °C, which was then alkylated by adding 15 µL of 0.5 M iodoacetamide and incubated for 30 min in the dark at room temperature. The proteins were digested using trypsin (1:100 enzyme/protein) overnight at 37 °C and 1% formic acid was added to stop the reaction. Digested protein samples were injected to C18 analytical column (75 µm inner-diameter, 360 µm outer-diameter 15 cm, 3 µm 150 Å pore size, Durashell C18) using 0.2% formic acid in water and 0.2% formic acid in acetonitrile as mobile phase.^[46] The conditions of LTQ-Orbitrap Velos were set as the followings: the source was operated at 2.0 kV, with no sheath gas flow and with the ion transfer tube at 350 °C. The mass spectrometer was programmed to acquire in a data dependent acquisition mode. The survey scan was from m/z 375 to 1300 with resolution 60000 at m/z 400. The MS/MS spectra were acquired in the LTQ normal scan mode. All MS/MS spectra obtained in each run were analyzed by MaxQuant 1.5.0.12, with Andromeda as peptide search engine.^[47] The Homo sapiens proteome database was used for identification and was downloaded from Uniprot (<https://www.uniprot.org>), with 71 544 sequences being downloaded on November 8, 2017. The gene ontology enrichment of proteins was performed by using DAVID Bioinformatics Resources 6.8.^[48]

Determination of Human Milk Lipidome: The extraction of human milk lipids was carried out using the Bligh and Dyer method,^[33] and TG compositions were determined using method as reported previously.^[49] Briefly, 5 mL of human milk sample was mixed with CHCl₃, MeOH, and Milli-Q water with a ratio of 1:1:2:0.8 (v/v/v/v). The mixture was oscillated for 30 min, CHCl₃ and Na₂SO₄ (2%, w/v) were then added with a final volume ratio of 2:2:1.8 (MeOH/CHCl₃/water, v/v/v). The mixture was oscillated again for 15 min and centrifuged at 2250 × g for 30 min. The organic fraction was filtered through filter paper containing 2 g of Na₂SO₄ which was then dried under nitrogen. The extracted lipids were dissolved in CHCl₃/MeOH (1:2, v/v) at 5 mg mL⁻¹ and was diluted with a ratio of 1:50 using MeOH with an internal standard (d5-(17:0/17:1/17:0) TAG). The MeOH method was used to extract human milk PLs.^[50] In general, the lipids were separated by a UPLC system (I-class Acquity UPLC, Waters) equipped with a BEH C18 reversed-phase column (1.7 µm, 2.1 mm i.d. × 100 mm, Waters Corporation). Samples were then detected by MS (API 4500 Q-Trap, AB SCIEX) with an ESI source in multi-reaction monitoring mode. The typical operating parameters were set as follows: curtain gas (CUR) 25; collision gas (CAD) medium; ion source gas 1 (GS1) 45; ion source gas 2 (GS2) 50; electrospray voltage 5500 V; and a temperature of 550 °C. The peak areas of TG isomers were compared with the that of internal standard (d5-17:0/17:1/17:0 TG) to calculate the concentrations of TGs. All the TGs were detected in positive ionization mode. Internal standards 12:0 LPC, and 14:0 LPA and 17:0 Ceramide were used as for choline-containing PL and negatively charged PL, respectively.

Determination of HMOs: Absolute quantification of 12 HMOs was carried out as reported previously.^[51] In brief, 1 mL of human milk was mixed

with 1 mL ultra-pure water (Milli-QPlus System, >18.3 M Ω). The samples were vortexed and centrifuged at 4600 $\times$ g, 4 $^{\circ}$ C for 10 min. The supernatant consisting predominantly of fat was removed. Then, 3 mL acetonitrile was added, vortexed, and then treated with ultrasonic for 10 min and centrifuged at 11000 $\times$ g, 4 $^{\circ}$ C for 10 min. This operation was repeated twice. The supernatant containing oligosaccharides and residual lactose were filtered using a 0.22 μ m filter and then analyzed using ultra-performance liquid chromatography (UPLC) tandem mass spectrometry. The samples were separated using UPLC firstly, and the column (ACQUITY UPLC BEH Amide Column, 130 $\text{\AA}$, 1.7 μ m, 2.1 mm $\times$ 100 mm (Waters, USA)) was equilibrated with solution A (10 mmol L $^{-1}$ ammonium formate solution) until remaining balance. The separation of 12 oligosaccharides was performed on amino UPLC column (2.1 $\times$ 100 mm), with a flow rate of 0.3 mL min $^{-1}$ and column temperature of 50 $^{\circ}$ C. The HMOs were analyzed by Agilent 1290 series LC system coupled with Agilent 6495A triple quadrupole mass spectrometer (Agilent Technologies, Santa Clara, CA) under multi-reaction monitoring mode. HMOs were introduced to ion traps via ESI chip interface in positive-ion mode. The air-dried temperature was 150 $^{\circ}$ C with a flow rate of 16 L min $^{-1}$, while the insulation air temperature was 300 $^{\circ}$ C at 12 L min $^{-1}$. Standard curves of 12 HMOs (2'FL, 3'FL, 3'SL, 6'SL, LNFP I, LNFP III, LNDFH II, DSLNT, LNH, LSTa, LSTb and LSTc) were determined as described elsewhere.^[51] The absolute concentration of each HMO in human milk samples was then calculated according to the standard curves.

Performance of Multiple Factor Analysis and Principal Component Analysis: In general, 473 proteins, 158 lipids, and 12 HMOs were detected. Proteins and lipids not being detected in 20% of the samples were excluded, leaving final protein number and lipid number of 193 and 138 (64 TG and 74 PL), respectively. For lipidomics and proteomics dataset, missing values were estimated using k-nearest neighbor imputation. Quantile normalization was then applied to reduce the batch effect using *MetaboAnalystR* R package, followed by log-transformation and auto-scaling.^[52] After data pre-processing, three omics datasets were merged as per each sample and were jointly analyzed using MFA^[53] by *FactoMineR* R package. Three to six factor solutions were considered, and 3-factor solution was chosen based on scree plot and the interpretability of the factors. Factor score of each sample, factor loadings, and contributions of each feature were calculated. *P* values of correlation between features and factor loadings were adjusted using the Benjamini & Hochberg method.^[54] PCA was also carried out for each individual omics data using the same protocol as MFA.

Robust Analysis: To evaluate stability of each factor, 1000 bootstrap samples were drawn and MFA was performed for each bootstrap sample. Mean and SD of loadings of all bootstrap samples were calculated and compared with original factor loadings. Correlations between features and factor loadings were also assessed for each bootstrap sample, frequencies of significant correlation (false discovery rate controlled with *q* of 0.05) were recorded for each feature across 1000 bootstrap samples.

Statistical Analysis: Multiple linear regression adjusted for infant age, infant sex, birth weight, birth length, maternal age, maternal BMI, city, and parity were carried out to explore association between factor loadings and infant growth *z* scores. Odds ratios of allergies against three factor loadings were calculated using generalized linear model. All statistical analyses were carried out using R software (<https://cran.r-project.org/>, version 4.0.2).

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements

This study was funded by the Government of Heilongjiang Province of the People's Republic of China under the "Bai-Qian-Wan Engineering and Technology Master Project" (Grant # 2019ZX07B01). This study also re-

ceived funding from the National Nature Science Foundation of China (Grant # 82003459).

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

S.J.: Conceptualization, Methodology, Validation, Investigation, Visualization, Writing - review & editing. J.P.: Methodology, Validation, Formal analysis. Y.L.: Investigation, Validation. M.J.: Validation, Formal analysis. W.Z.: Resources, Writing - review & editing. J.L.: Methodology, Formal analysis. J.L.: Methodology, Formal analysis. K.L.: Conceptualization, Methodology, Formal analysis, Supervision, Project administration, Funding acquisition, Writing - review & editing.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Keywords

glycome, human milk patterns, lipidome, proteome, the Chinese Human Milk Project

Received: January 7, 2021

Revised: June 10, 2021

Published online:

- [1] D. Munblit, V. Verhasselt, J. O. Warner, *Front. Pediatr.* **2019**, *7*, 319.
- [2] C. G. Victora, R. Bahl, A. J. D. Barros, G. V. A. Frana, S. Horton, J. Krasevec, S. Murch, M. J. Sankar, N. Walker, N. C. Rollins, *Lancet* **2016**, *387*, 475.
- [3] J. Zhu, K. A. Dingess, *Nutrients* **2019**, *11*, 1834.
- [4] A. D. George, M. C. L. Gay, R. D. Trengove, D. T. Geddes, *Nutrients* **2018**, *10*, 1169.
- [5] S. Thurl, M. Munzert, G. Boehm, C. Matthews, B. Stahl, *Nutr. Rev.* **2017**, *75*, 920.
- [6] J. L. Saben, C. R. Sims, B. D. Piccolo, A. Andres, *Am. J. Clin. Nutr.* **2020**, *112*, 1228.
- [7] C. R. Sims, M. E. Lipsmeyer, D. E. Turner, A. Andres, *Am. J. Clin. Nutr.* **2020**, *112*, 548.
- [8] M.-C. Alexandre-Gouabau, T. Moyon, V. Cariou, J.-P. Antignac, E. M. Qannari, M. Croyal, M. Soumah, Y. Guitton, A. David-Sochard, H. Billard, A. Legrand, C. Boscher, D. Darmaun, J.-C. Roz , C.-Y. Boquien, *Nutrients* **2018**, *10*, 164.
- [9] K. Miliku, B. Robertson, A. K. Sharma, P. Subbarao, A. B. Becker, P. J. Mandhane, S. E. Turvey, D. L. Lefebvre, M. R. Sears, L. Bode, M. B. Azad, *Allergy* **2018**, *73*, 2070.
- [10] K. J. Karczewski, M. P. Snyder, *Nat. Rev. Genet.* **2018**, *19*, 299.
- [11] A. Casanovas, R. R. Sprenger, K. Tarasov, D. E. Ruckerbauer, H. K. Hannibal-Bach, J. Zanghellini, O. N. Jensen, C. S. Ejlsing, *Chem. Biol.* **2015**, *22*, 412.
- [12] L. Bode, *Glycobiology* **2012**, *22*, 1147.
- [13] D. Munblit, D. G. Peroni, A. Boix-Amor s, P. S. Hsu, B. V. t Land, M. C. L. Gay, A. Kolotilina, C. Skevaki, R. J. Boyle, M. C. Collado, J. Garssen, D. T. Geddes, R. Nanan, C. Slupsky, G. Wegienka, A. L. Kozyskyj, J. O. Warner, *Nutrients* **2017**, *9*, 894.

- [14] A. Riedl, C. Gieger, H. Hauner, H. Daniel, J. Linseisen, *Br. J. Nutr.* **2017**, *117*, 1631.
- [15] E. Hillesheim, L. Brennan, *Nutr. Res. Rev.* **2020**, *33*, 33.
- [16] I. Garcia-Perez, J. M. Posma, E. S. Chambers, J. C. Mathers, J. Draper, M. Beckmann, J. K. Nicholson, E. Holmes, G. Frost, *Nat. Food* **2020**, *1*, 355.
- [17] C. B. O'Donovan, M. C. Walsh, A. P. Nugent, B. McNulty, J. Walton, A. Flynn, M. J. Gibney, E. R. Gibney, L. Brennan, *Mol. Nutr. Food Res.* **2015**, *59*, 377.
- [18] J. M. Posma, I. Garcia-Perez, G. Frost, G. S. Aljuraiban, Q. Chan, L. Van Horn, M. Daviglus, J. Stamler, E. Holmes, P. Elliott, J. K. Nicholson, *Nat. Food* **2020**, *1*, 426.
- [19] M.-C. Alexandre-Gouabau, T. Moyon, A. David-Sochard, F. Fenaille, S. Cholet, A.-L. Royer, Y. Guitton, H. Billard, D. Darmaun, J.-C. Rozé, C.-Y. Boquien, *Nutrients* **2019**, *11*, 528.
- [20] K. Hettinga, H. v. Valenberg, S. d. Vries, S. Boeren, T. v. Hooijdonk, J. v. Arendonk, J. Vervoort, *PLoS One* **2011**, *6*, e19433.
- [21] J. Lu, S. Zhang, L. Liu, X. Pang, C. Ma, S. Jiang, J. Lv, *Food Chem.* **2018**, *261*, 274.
- [22] B. L. Horta, C. L. d. Mola, C. G. Victora, *Acta Paediatr.* **2015**, *104*, 30.
- [23] M. B. Azad, L. Vehling, D. Chan, A. Klopp, N. C. Nickel, J. M. McGavock, A. B. Becker, P. J. Mandhane, S. E. Turvey, T. J. Moraes, M. S. Taylor, D. L. Lefebvre, M. R. Sears, P. Subbarao, o. b. o. t. C. S. Investigators, *Pediatrics* **2018**, *142*, e20181092.
- [24] K. S. d. Fluiter, G. F. Kerkhof, I. A. L. P. v. Beijsterveldt, L. M. Breij, B. J. M. v. d. Heijning, M. Abrahamse-Berkeveld, A. C. S. Hokken-Koelega, *Clin. Nutr.* **2020**.
- [25] O. Ballard, A. L. Morrow, *Pediatr. Clin.* **2013**, *60*, 49.
- [26] H. Lagström, S. Rautava, H. Ollila, A. Kaljonen, O. Turta, J. Mäkelä, C. Yonemitsu, J. Gupta, L. Bode, *Am. J. Clin. Nutr.* **2020**, *111*, 769.
- [27] K. M. Tonon, M. B. de Moraes, A. C. F. V. Abrão, A. Miranda, T. B. Moraes, *Nutrients* **2019**, *11*, 1358.
- [28] J. M. Jorgensen, R. Young, P. Ashorn, U. Ashorn, D. Chaima, J. C. C. Davis, E. Goonatilleke, C. Kumwenda, C. B. Lebrilla, K. Maleta, E. L. Prado, J. Sadalaki, S. M. Totten, L. D. Wu, A. M. Zivkovic, *Am. J. Clin. Nutr.* **2021**, *113*, 209.
- [29] N. Sprenger, L. Y. Lee, C. A. D. Castro, P. Steenhout, S. K. Thakkar, *PLoS One* **2017**, *12*, e0171814.
- [30] J. L. LaBarre, M. Puttabyatappa, P. X. K. Song, J. M. Goodrich, L. Zhou, T. M. Rajendiran, T. Soni, S. E. Domino, M. C. Treadwell, D. C. Dolinoy, V. Padmanabhan, C. F. Burant, *Sci. Rep.* **2020**, *10*, 14209.
- [31] C. Chen, B. Jung, W. K. Kim, *Poult. Sci.* **2019**, *98*, 3902.
- [32] K. B. Jang, J. M. Purvis, S. W. Kim, *J. Anim. Sci.* **2020**, *98*, 1.
- [33] E. G. Bligh, W. J. Dyer, *Can. J. Biochem. Physiol.* **1959**, *37*, 911.
- [34] A. E. Seppo, C. A. Autran, L. Bode, K. M. Järvinen, *J. Allergy Clin. Immunol.* **2017**, *139*, 708.
- [35] N. Sprenger, H. Odenwald, A. K. Kukkonen, M. Kuitunen, E. Savilahti, C. Kunz, *Eur. J. Nutr.* **2017**, *56*, 1293.
- [36] N. Lunjani, P. Satitsuksanoa, Z. Lukasik, M. Sokolowska, T. Eiwegger, L. O'Mahony, *Allergy* **2018**, *73*, 2314.
- [37] M. Natale, C. Bisson, G. Monti, A. Peltran, L. Perono Garoffo, S. Valentini, C. Fabris, E. Bertino, A. Coscia, A. Conti, *Mol. Nutr. Food Res.* **2004**, *48*, 363.
- [38] M. Bublin, T. Eiwegger, H. Breiteneder, *J. Allergy Clin. Immunol.* **2014**, *134*, 521.
- [39] A. Bossios, M. Theodoropoulou, L. Mondoulet, N. M. Rigby, N. G. Papadopoulos, H. Bernard, K. Adel-Patient, J.-M. Wal, C. E. Mills, P. Papageorgiou, *Clin. Transl. Allergy* **2011**, *1*, 1.
- [40] F. J. Moreno, A. R. Mackie, E. C. Mills, *J. Agric. Food Chem.* **2005**, *53*, 9810.
- [41] M. Wang, Z. Zhao, A. Zhao, J. Zhang, W. Wu, Z. Ren, P. Wang, Y. Zhang, *Nutrients* **2020**, *12*, 826.
- [42] S. Liu, Y. Liu, Y.-y. Mao, X.-k. Cai, F. Tian, Y.-r. Zhao, S. Wang, *Acta Nutrimenta Sinica* **2020**, *42*, 219.
- [43] M. K. McGuire, C. L. Meehan, M. A. McGuire, J. E. Williams, J. Foster, D. W. Sellen, E. W. Kamau-Mbuthia, E. W. Kamundia, S. Mbugua, S. E. Moore, *Am. J. Clin. Nutr.* **2017**, *105*, 1086.
- [44] W. H. Organization, *WHO child growth standards : length/height-for-age, weight-for-age, weight-for-length, weight-for-height and body mass index-for-age : methods and development*, World Health Organization, **2006**.
- [45] J. Lu, X. Wang, W. Zhang, L. Liu, X. Pang, S. Zhang, J. Lv, *Food Chem.* **2016**, *196*, 665.
- [46] C. Ding, J. Jiang, J. Wei, W. Liu, W. Zhang, M. Liu, T. Fu, T. Lu, L. Song, W. Ying, C. Chang, Y. Zhang, J. Ma, L. Wei, A. Malovannaya, L. Jia, B. Zhen, Y. Wang, F. He, X. Qian, J. Qin, *Mol. Cell. Proteomics* **2013**, *12*, 2370.
- [47] J. Cox, N. Neuhauser, A. Michalski, R. A. Scheltema, J. V. Olsen, M. Mann, *J. Proteome Res.* **2011**, *10*, 1794.
- [48] D. W. Huang, B. T. Sherman, R. A. Lempicki, *Nat. Protoc.* **2009**, *4*, 44.
- [49] P. Zhao, S. Zhang, L. Liu, X. Pang, Y. Yang, J. Lu, J. Lv, *J. Agric. Food Chem.* **2018**, *66*, 4571.
- [50] Z. Zhao, Y. Xu, *J. Lipid Res.* **2010**, *51*, 652.
- [51] W. Zhang, T. Wang, X. Chen, X. Pang, S. Zhang, J. U. Obaroakpo, J. Shilong, J. Lu, J. Lv, *Carbohydr. Polym.* **2019**, *219*, 328.
- [52] J. Chong, D. S. Wishart, J. Xia, *Curr. Protoc. Bioinform.* **2019**, *68*, e86.
- [53] M. de Tayrac, S. Lê, M. Aubry, J. Mosser, F. Husson, *BMC Genomics* **2009**, *10*, 32.
- [54] Y. Benjamini, Y. Hochberg, *J. R. Stat. Soc. Ser. B (Methodological)* **1995**, *57*, 289.

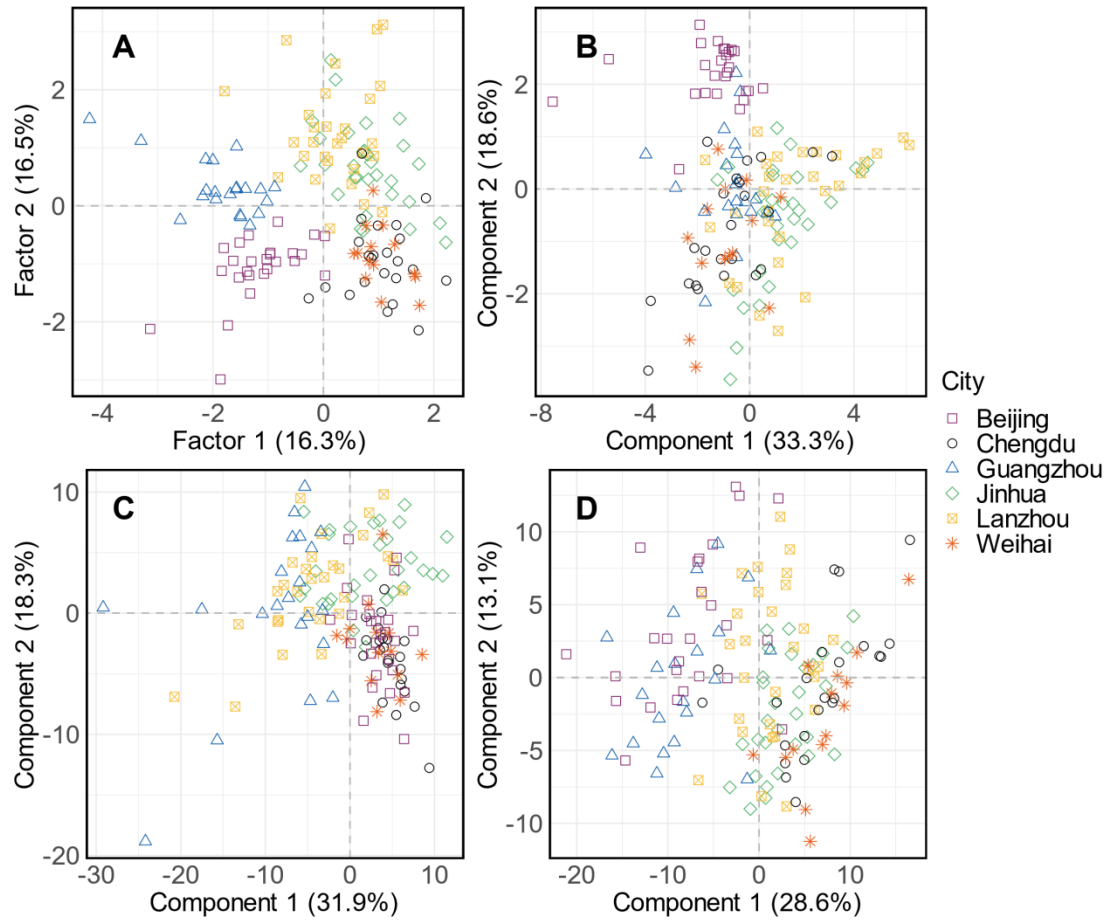


Figure S1 MFA and PCA score plot across six cities (n=143).

A: MFA score plot; B: PCA score plot of HMO profile; C: PCA score plot for lipidomic profile; D: PCA score plot of proteomic profile.

Samples collected from different cities are presented in different color and shape as indicated by the legend.

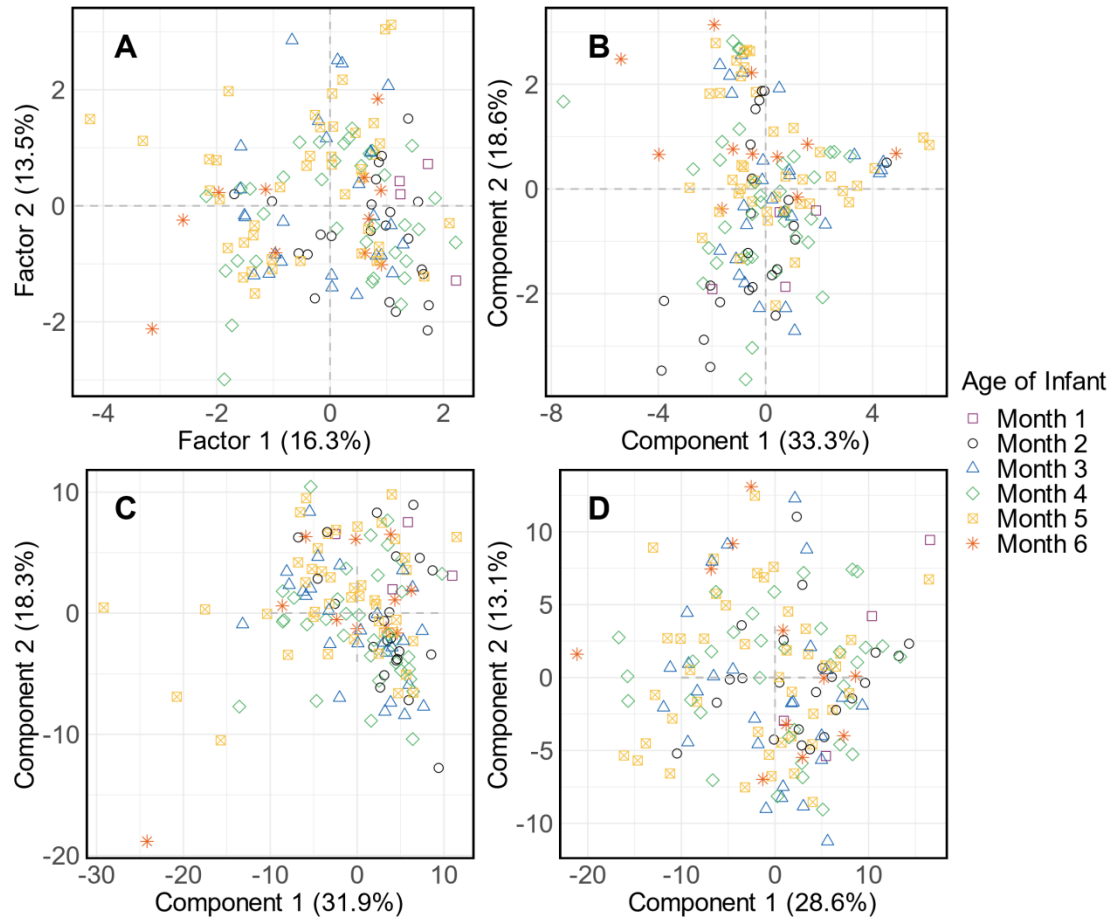


Figure S2 MFA and PCA score plot across lactation (n=143).

A: MFA score plot; B: PCA score plot of HMO profile; C: PCA score plot for lipidomic profile; D: PCA score plot of proteomic profile.

Samples related to different infant age are presented in different color and shape as indicated by the legend.

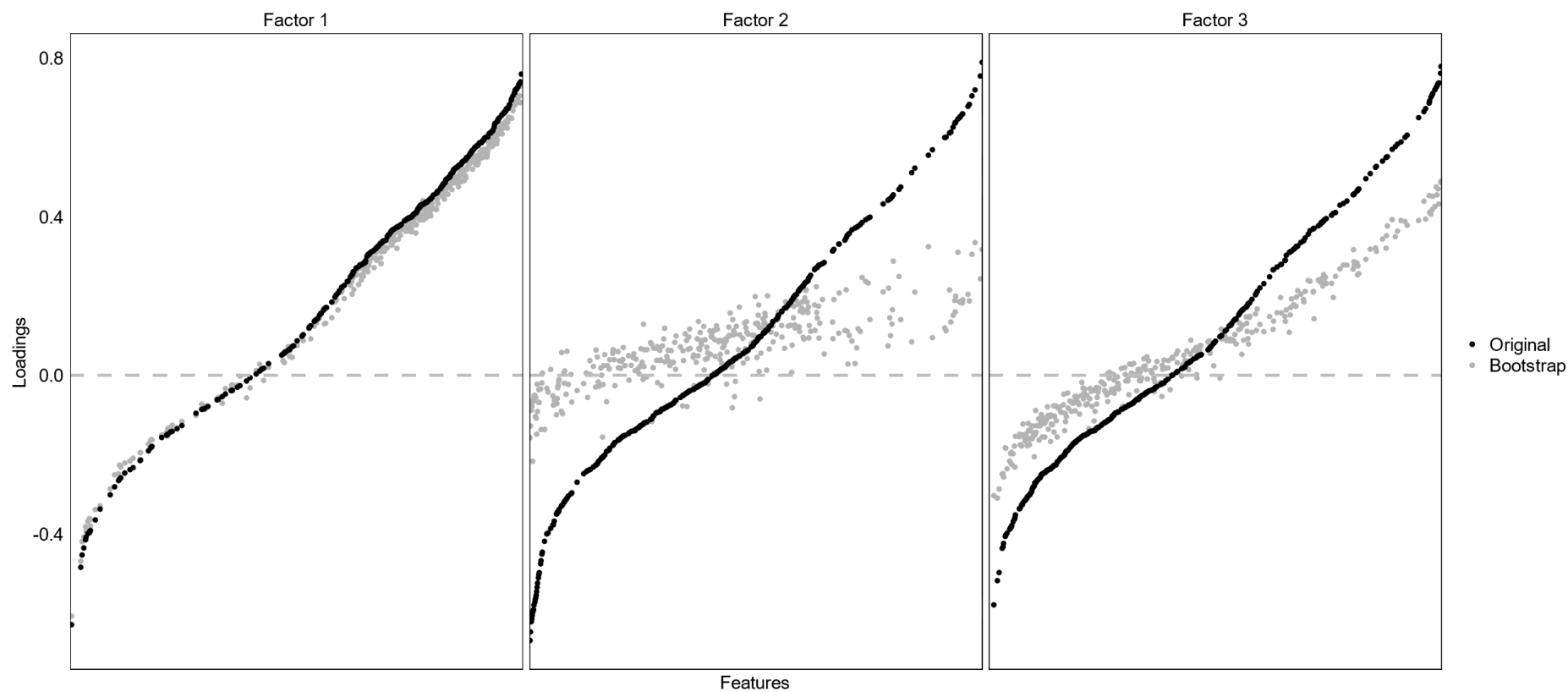


Figure S3 Robust analysis of MFA.

To evaluate the robustness of each factor obtained using MFA, 1000 bootstrap samples were drawn and MFA was performed over each bootstrap sample. Bootstrap loadings were then averaged and plotted as shown above. The original MFA loadings are colored in black, while mean bootstrap loadings are colored in grey.

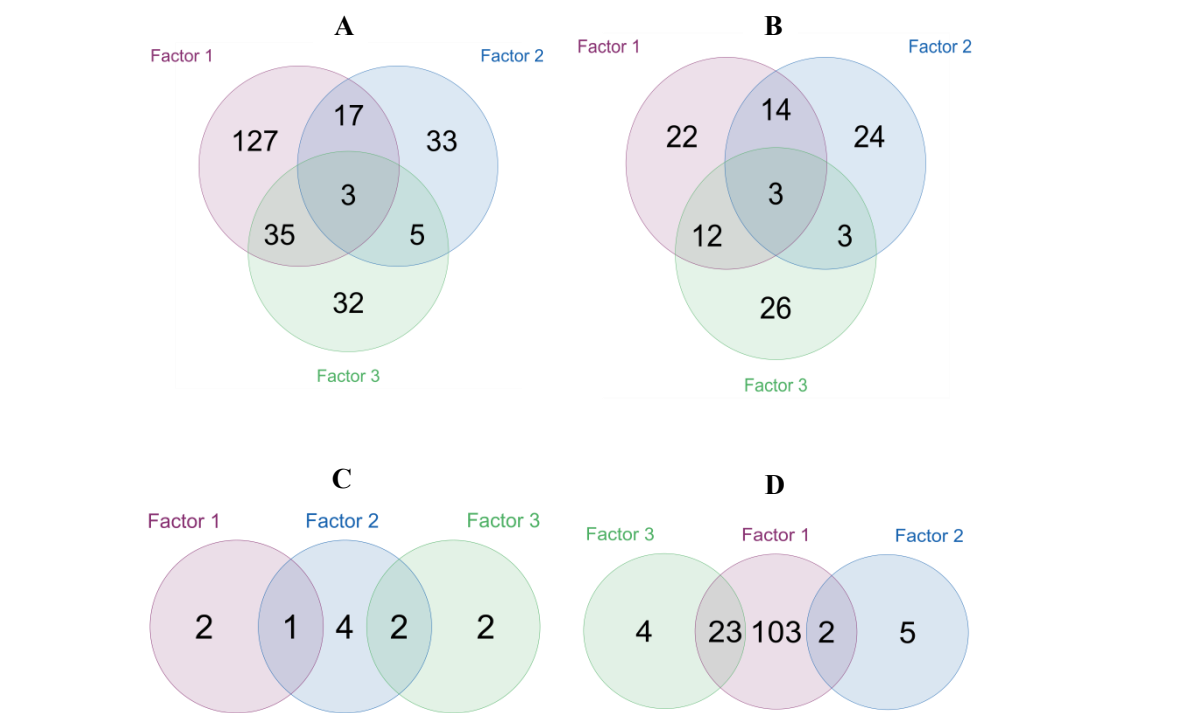


Figure S4 Venn diagram of MFA

A: Venn diagram based on three omics data combined; B: Venn diagram of lipidomic profile; C: Venn diagram of HMOs; D: Venn diagram of proteomic profile.