



OPEN

Multimodal machine learning for modeling infant head circumference, mothers' milk composition, and their shared environment

Martin Becker^{1,10}, Kelsey Fehr^{1,2,3,4}, Stephanie Goguen^{1,2,3,4}, Kozeta Miliku^{5,6}, Catherine Field⁷, Bianca Robertson¹¹, Chloe Yonemitsu¹¹, Lars Bode^{1,11}, Elinor Simons⁴, Jean Marshall⁸, Bassel Dawod⁸, Piushkumar Mandhane⁷, Stuart E. Turvey⁹, Theo J. Moraes¹², Padmaja Subbarao^{5,6,12}, Natalie Rodriguez^{1,2,3,4}, Nima Aghaeepour^{1,10} ✉ & Meghan B. Azad^{1,2,3,4} ✉

Links between human milk (HM) and infant development are poorly understood and often focus on individual HM components. Here we apply multi-modal predictive machine learning to study HM and head circumference (a proxy for brain development) among 1022 mother-infant dyads of the CHILD Cohort. We integrated HM data (19 oligosaccharides, 28 fatty acids, 3 hormones, 28 chemokines) with maternal and infant demographic, health, dietary and home environment data. Head circumference was significantly predictable at 3 and 12 months. Two of the most associated features were HM n3-polyunsaturated fatty acid C22:6n3 (docosahexaenoic acid, DHA; $p = 9.6e-05$) and maternal intake of fish ($p = 4.1e-03$), a key dietary source of DHA with established relationships to brain function. Thus, using a systems biology approach, we identified meaningful relationships between HM and brain development, which validates our statistical approach, gives credence to the novel associations we observed, and sets the foundation for further research with additional cohorts and HM analytes.

Physical growth and physiological development are complex processes influenced by a wide variety of factors across different domains (i.e., modalities), especially during the critical periods of gestation and infancy. These modalities include maternal dietary intake during pregnancy and/or lactation^{1–3}, sociodemographic characteristics (e.g., age, socioeconomic status)⁴, the home environment (e.g., smokers, pets, cleaning chemicals)⁵, infant morbidities (e.g., diarrhea)⁶, infant feeding patterns (e.g., breastfeeding consistency)⁷, as well as human milk composition⁸. Human milk (HM) is a particularly complex factor as it comprises thousands of nutritive and non-nutritive components that collectively support infant growth and development and are, in turn, influenced by many of the above-mentioned modalities. Analyzing and modeling human milk as a biological system that fundamentally connects mother-infant dyads is a challenging but essential task that will lead to a better understanding of healthy infant growth and help to prevent developmental disorders^{9–11}.

In this context, many studies have linked individual factors or modalities to infant growth and development—including head circumference (reflecting both physical growth and brain development^{12,13}) and cognitive or behavioral outcomes (such as the Bayley Scales of Infant and Toddler Development¹⁴ or the Ages and Stages Questionnaire¹⁵). For example, *maternal dietary intake* during lactation can directly influence the nutritional

¹International Milk Composition (IMiC) Consortium, Winnipeg, Canada. ²Manitoba Interdisciplinary Lactation Centre (MILC), Winnipeg, Canada. ³Children's Hospital Research Institute of Manitoba, Winnipeg, Canada. ⁴University of Manitoba, Winnipeg R3E3P4, Canada. ⁵University of Toronto, Toronto M5S 1A8, Canada. ⁶McMaster University, Hamilton M5S 1A8, Canada. ⁷University of Alberta, Edmonton T6G 2E1, Canada. ⁸Dalhousie University, Halifax B3H4R2, Canada. ⁹University of British Columbia and British Columbia Children's Hospital, Vancouver V5Z4H4, Canada. ¹⁰Stanford University, Stanford 94305, USA. ¹¹University of California, San Diego, La Jolla, CA 92093, USA. ¹²SickKids, Toronto M5G 0A4, Canada. ✉email: naghaeep@stanford.edu; meghan.azad@umanitoba.ca

composition of HM¹⁶ with long-term consequences for infants^{1–3}. Specific to cognition, maternal fish consumption leads to higher docosahexaenoic acid (DHA) concentrations in HM, which in turn has been linked to infant brain development¹⁷ even if it is unclear whether this relationship influences infant growth in general^{3,18}. *Maternal characteristics* such as age, body composition¹², and socioeconomic status (e.g., maternal education or income) have been associated with infant growth and head circumference⁴. Similarly, maternal smoking habits can negatively influence infant growth^{3,19}. *Home environment* factors like pets or cleaning chemicals can have an influence on infant development²⁰. Additionally, many studies find that *infant morbidities* like diarrhea or pneumonia can impair infant growth^{6,21}. Current evidence indicates that *infant feeding patterns* such as longer duration of exclusive or partial breastfeeding tended to be associated with healthier growth patterns during infancy (i.e., slower growth rate and earlier peak BMI in developed settings^{7,22}) and a reduced risk of overweight and obesity at ages 2 years and older²³. Similarly, some but not all observational studies of term infants fed HM have reported enhanced brain development (through imaging studies), higher intelligence quotient scores and increased cognitive and behavioral outcomes (using validated scales) compared to formula-fed infants^{8,24–28}, although this benefit has not been linked to any particular HM component.

Despite this large body of literature, the intricate relationships between the multitude of HM components and a plethora of maternal, infant, and environmental factors are poorly understood because research has typically focused on selected modalities and individual HM components or component types⁴². There is a growing recognition among researchers that multidisciplinary and systems biology approaches are required to decipher these complex relationships^{9–11}. The International Milk Composition (IMiC) Consortium (www.milcresearch.com/imic) was established to address these knowledge gaps by collecting a wide variety of data modalities, including the measurement of an extensive array of HM components. However, analyzing such large amounts of data is analytically challenging and requires advanced statistical tools.

Machine learning methods can simultaneously investigate a multitude of modalities and reveal clinically meaningful associations^{29,30}. Here, we used machine learning to assess whether multiple data modalities, including HM composition, can predict current and/or future head circumference (a common proxy measure for brain development^{31–34}) in breastfed infants. We integrated several existing datasets from the CHILD Cohort Study, a Canadian birth cohort with HM samples collected at 3 months. Specifically, HM data (oligosaccharides, fatty acids, cytokines, and hormones) were combined with maternal characteristics, diet, health and body composition, infant feeding and morbidities, as well as home environment information. This data was used to predict head circumference measured by study staff at 3 months and 1 year using machine learning (see Fig. 1 for an overview).

Results

Deep profiling of mother-infant dyads

The ongoing CHILD Cohort Study recruited 3624 pregnant Canadian women in 2009–2012 and has been tracking the growth and development of their children since birth. Additionally, a wide variety of data has been collected from periodic questionnaires, hospital records, as well as biological samples (see^{35,36} for more details). In the current study, we focus on a subcohort of 1022 CHILD mother-infant dyads with complete infant growth data through 1 year of age and available human milk data at 3–4 months (3 hormones measured by ELISA, 19 oligosaccharides measured by HPLC, 28 fatty acids measured by gas chromatography, and 28 immunomodulators including chemokines, cytokines, immunoglobulins and growth factors measured by Luminex assay or ELISA) (see “[Materials and methods](#)” for details). These data were combined with other modalities, including maternal characteristics (n = 36 features, e.g., age, and socioeconomic status), diet (n = 207, e.g., estimated nutrient and energy intakes, healthy eating index scores), body composition (n = 6, e.g., height, weight, BMI), and health (n = 84, e.g., depression, diabetes or asthma); infant morbidities (n = 87, e.g., diarrhea) and infant feeding (n = 27, e.g., breastfeeding duration and exclusivity); and home environment (n = 31, e.g., flooring and pets) (Supplementary Table S1).

As shown in Table 1, the mean age of mothers (n = 1022) was 32.99 (SD 4.23) years, and the mean infant age at milk collection was 3.73 (SD 1.07) months. Most mothers (91%) had a postsecondary degree, 27% were non-White, and 23% had asthma, while 34% were overweight (22%) or obese (12%). The median duration of exclusive breastfeeding in this subcohort was 4.5 months, the median duration of any breastfeeding was 1 year, and 77% of mothers reported feeding pumped milk to their infant before 3 months. Many houses had carpet installed (26%), and nearly half of all households had a pet (47%). Mean infant head circumference was 40.8 cm (SD 2.0) at 3 months (measured on the same date as HM sample collection) and 46.0 cm (SD 1.9) at 1 year. The Spearman rank correlation between the head circumference z-score at 3 months and 1 year was $r = 0.37$, $p = 9.57E-35$. Milk composition profiles have been described previously (separately) for human milk oligosaccharides (HMOs)³⁷, fatty acids³⁸, and hormones³⁹.

An integrated multi-omics model for head circumference

To analyze whether the collective information contained in the multi-modal data can jointly predict infant head circumference z-score by age (a key measure of growth and development^{12,13}), we generated machine learning models based on a fivefold cross-validation scheme. To account for a large number of involved variables while at the same time ensuring that more complex relationships between these variables are captured, we employed linear (ridge regression with nested parameter optimization) as well as non-linear models (support vector machines with radial basis function kernels), respectively. See “[Materials and methods](#)” for details.

Figure 1B shows a summary of model performances for predicting head circumference at 3 months (Panel B1) and 1 year (Panel B2) across the different data types and data modalities. Using all data modalities combined, infant head circumference was significantly predictable at 3 months (Panel B1.1, Spearman $r = 0.25$, $p = 3.5e-16$)

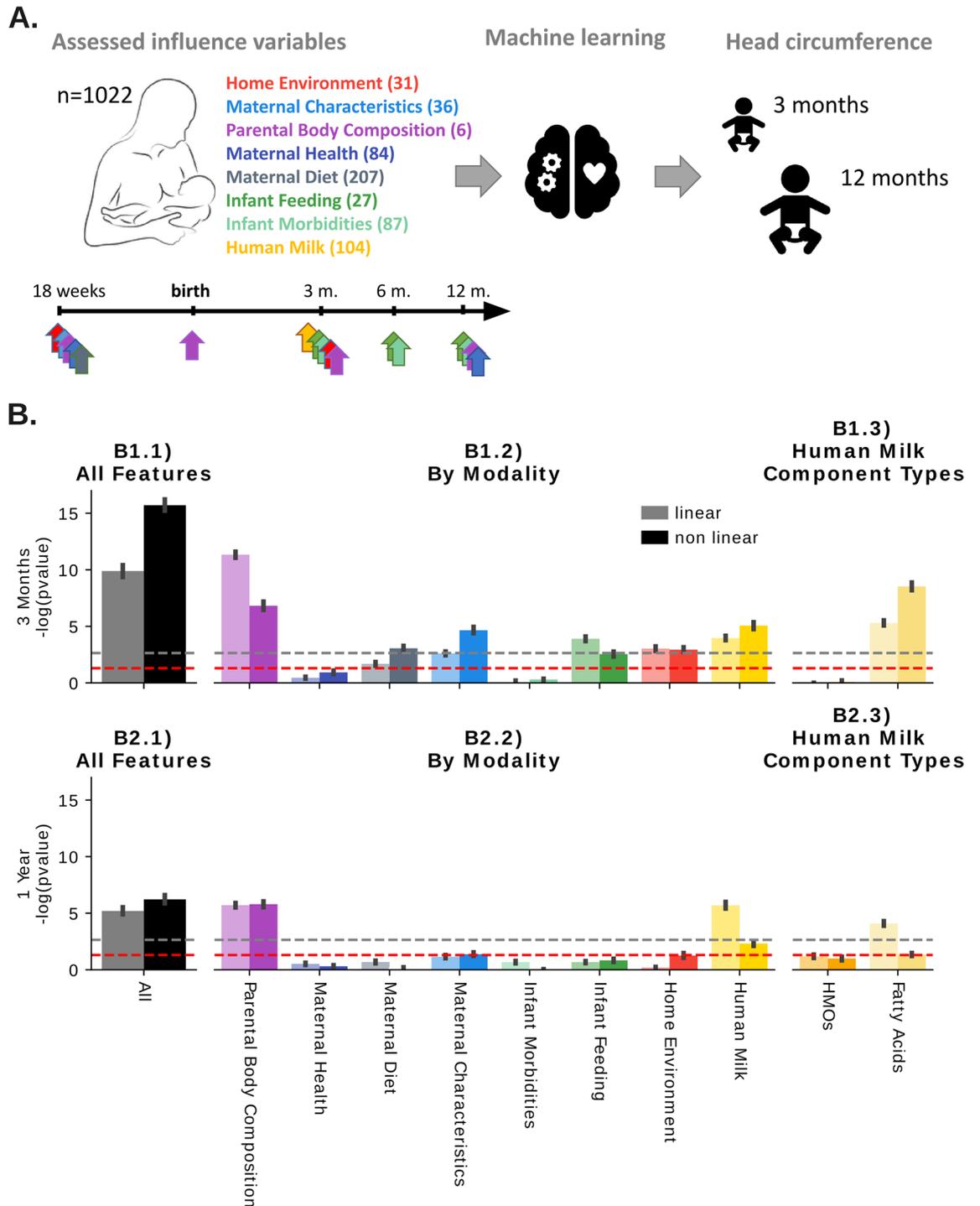


Figure 1. Study design and predictive modeling of head circumference at 3 months and 1 year using CHILd Cohort Study data (n = 1022 mother-infant dyads; n = 672 features). (A) A multitude of variables (features) across eight categories (modalities) were assessed at different time points before and after birth. This includes Home Environment (35 variables), Maternal Characteristics (42), Parental Body Composition (8), Maternal Health (149), Maternal Diet (207), Infant Feeding (45), Infant Morbidities (95), Human Milk (157). Using machine learning approaches, we then jointly integrate these data modalities to model infant head circumference (z-score for age) at 3 months and 1 year, respectively. For a full list of features, see Supplemental Table S1. Summarizes results from linear (ridge regression) and nonlinear (support vector machines) models using different combinations of features for prediction. Bars indicate the significance of the predictive power for each data subset and model measured by the negative log p-value of the association between the prediction and head circumference (Spearman), with 95% confidence intervals (black lines). For Spearman r values, see Supplemental Fig. S3. Dashed red lines denote p-value < 0.05 thresholds indicating the statistical significance of the predictions without multiple hypothesis comparison correction, the grey dashed lines denote a Bonferroni corrected statistical significance threshold assuming 22 experiments. Bar colors correspond to the modality color scheme in this figure. Combining all features into a multi-modal model (B1.1, B2.1) increases predictive power, particularly at 3 months. Predicting head circumference further into the future (1 year) is more challenging than short-term predictions (3 months). Human milk components are predictive (B1.2, B2.2), particularly fatty acids at 3 months (B1.3, B2.3). HMOs, human milk oligosaccharides. For a full list of features and modalities, see Supplemental Table S1.

	Features	Average	n (%)
Mother	Age (years)	33.0 ± 4.2	
	Height (cm)	165.1 ± 7.5	
	Weight (kg)	66.8 ± 14.8	
	BMI (kg/m ²)	24.5 ± 5.0	
	Overweight (%)		223 (22.5)
	Obese (%)		123 (12.4)
	Post-secondary education (%)		848 (90.6)
	Race (%)		
	Asian		210 (17.5)
	White		862 (71.9)
	First nations		46 (3.8)
	Other		79 (6.6)
	Asthma (%)		240 (23.5)
Infant	Gestational age at delivery (weeks)	39.7 ± 1.2	
	Age at milk collection (months)	3.8 ± 1.1	
	Sex (% female)		543 (53.1)
	Fever in first 3 months (%)		207 (20.3)
	Cold with diarrhea in first 3 months (%)		7 (0.7)
	Head circumference at 3 months (cm)	40.8 ± 2.0	
	Head circumference at 1 year (cm)	46.0 ± 1.9	
	Head circumference at 3 months (z-score)	0.0 ± 1.6	
	Head circumference at 1 year (z-score)	0.3 ± 1.4	
Feeding	Breastfeeding duration (max 12 months)	12.0 [8.0, 12.0]	
	Exclusive breastfeeding duration (months)	4.5 [1.0, 5.0]	
	Used pump at 3 months (%)		790 (77.3)
Home	Carpet (%)		262 (25.6)
	Pet (%)		479 (46.9)

Table 1. Key demographic characteristics of mother-infant dyads from the CHILd cohort study included in the current analysis (n = 1022). Values are n (%) or mean ± standard deviation or median [interquartile range]. By design, all dyads in this subset of the CHILd cohort breastfed for at least 3 months (the time of breast milk sample collection).

as well as 1 year (Panel B2.1, $r = 0.15$, $p = 1.3e-06$). While Fig. 1B shows a comparisons of p-values, Supplementary Fig. S3 visualizes the corresponding Spearman rank correlation coefficients.

At 3 months (Panel B1), maternal body composition ($r = 0.21$, $p = 5.1e-12$) followed by human milk composition ($r = 0.14$, $p = 1.0e-05$) and parental characteristics ($r = 0.13$, $p = 3.7e-05$) have the highest predictive association with head circumference across all modalities (Panel B1.2). Other predictive modalities included maternal diet ($r = 0.10$, $p = 9.3e-04$), infant feeding ($r = 0.11$, $p = 2.5e-04$), as well as home environment ($r = 0.11$, $p = 5.7e-04$). Only maternal health ($r = 0.05$, $p = 1.2e-01$) and infant morbidities ($r = 0.01$, $p = 5.6e-01$) did not show significant predictive power for head circumference at 3 months. Of the milk composition modalities included in model building (B1.3) fatty acids were the most associated with head circumference ($r = 0.18$, $p = 2.8e-09$) while HMOs did not show a collectively significant predictive signal.

Head circumference was less predictable at 1 year (Panel B2) compared to 3 months. This is to be expected as predicting further into the future adds uncertainty due to the increasing variety of factors that influence infant development which may not, or only incompletely, have been captured. Nevertheless, the combination of all modalities is still significantly predictive of head circumference at 1 year (Panel B2.1, $r = 0.15$, $p = 1.3e-06$). Furthermore, parental body composition (Panel B2.2, $r = 0.15$, $p = 1.4e-06$), as well as milk composition, remain significantly associated with head circumference (Panel B2.2, $r = 0.15$, $p = 2.6e-06$), and milk fatty acids are again the most predictive (Panel B2.3, $r = 0.12$, $p = 7.6e-05$). In contrast to the 3-month predictions, maternal characteristics, maternal diet, infant feeding, and home environment modalities are not significantly associated with head circumference at 1 year.

Overall, only parental body compositions and milk components are consistently predictive at 3 months as well as after 1 year (see Fig. 1B).

Individual associations of features with head circumference

We further investigated the relationship of individually measured variables and infant head circumference. To do this, we derived feature interdependency networks that visualize the correlation structure between variables while at the same time showing their association with head circumference at 3 months (Figs. 2, 4 and Tables 2, 4) and 1 year (Figs. 3, 5 and Tables 3, 5).

Modality	Category	Feature	na	unique	test	statistic	pvalue	pvalue (fdr_bh)
Parental Body Composition	general	mom_Height_cm	33	264	kendall	0.1234	0.0000	0.0000
Human Milk	Fatty Acids	C22:6n3 %	12	970	kendall	-0.1061	0.0000	0.0001
Human Milk	Fatty Acids	C20:0 %	12	991	kendall	0.0923	0.0000	0.0019
Maternal Diet	General dietary assessment	diet_pfa226	45	978	kendall	-0.0922	0.0000	0.0020
Parental Body Composition	general	mom_bmi_best	30	908	kendall	0.0871	0.0000	0.0033
Maternal Diet	General dietary assessment	diet_pfa205	45	978	kendall	-0.0877	0.0000	0.0033
Parental Body Composition	general	dad_Weight_kg	191	342	kendall	0.0947	0.0000	0.0033
Maternal Diet	General dietary assessment	diet_V_STARCH	45	969	kendall	0.0860	0.0001	0.0035
Maternal Diet	General dietary assessment	diet_M_FISH_HI	45	278	kendall	-0.0867	0.0001	0.0036
Parental Body Composition	general	dad_Height_cm	191	250	kendall	0.0926	0.0001	0.0036
Parental Body Composition	Postpartum BMI	mom_BMI_1y	32	927	kendall	0.0821	0.0001	0.0050
Maternal Diet	General dietary assessment	diet_pfa184	45	812	kendall	-0.0810	0.0002	0.0063
Maternal Diet	General dietary assessment	diet_pfa225	45	976	kendall	-0.0754	0.0004	0.0160
Maternal Diet	General dietary assessment	diet_mfa221	45	978	kendall	-0.0743	0.0005	0.0179
Maternal Diet	General dietary assessment	diet_tfa161t	45	978	kendall	0.0730	0.0006	0.0209
Maternal Diet	General dietary assessment	diet_daidzein	45	978	kendall	-0.0705	0.0010	0.0268
Maternal Diet	General dietary assessment	diet_A_BEV	45	42	kendall	-0.0830	0.0010	0.0268
Infant Morbidities	Diarrhea	*cold1_diarrheah_3m	914	3	ranksums	-3.2819	0.0010	0.0268
Human Milk	Fatty Acids	C20:3n6 %	12	1011	kendall	0.0690	0.0010	0.0268
Human Milk	Immunomodulators	milk_norm_TSLP	756	263	kendall	-0.1336	0.0012	0.0289
Maternal Characteristics	Age	mom_age	0	960	kendall	-0.0662	0.0015	0.0337
Maternal Diet	General dietary assessment	diet_calcium	45	978	kendall	0.0671	0.0017	0.0337
Maternal Diet	General dietary assessment	diet_D_TOT_SOYM	45	978	kendall	0.0670	0.0017	0.0337
Maternal Diet	General dietary assessment	diet_retinol	45	978	kendall	0.0668	0.0018	0.0337
Infant Feeding	Various	BF_9m	41	3	ranksums	3.1025	0.0019	0.0337
Maternal Diet	General dietary assessment	diet_D_CHEESE	45	977	kendall	0.0663	0.0019	0.0337
Maternal Diet	General dietary assessment	diet_D_TOTAL	45	978	kendall	0.0661	0.0020	0.0337
Maternal Diet	General dietary assessment	hei6_dairy	45	499	kendall	0.0712	0.0020	0.0337
Maternal Diet	General dietary assessment	diet_V_POTATO	45	975	kendall	0.0653	0.0023	0.0363
Maternal Diet	General dietary assessment	diet_zinc	45	978	kendall	0.0651	0.0024	0.0363
Maternal Diet	General dietary assessment	diet_mfa141	45	978	kendall	0.0650	0.0024	0.0363
Human Milk	Fatty Acids	C20:5n3 %	12	693	kendall	-0.0634	0.0026	0.0381
Home Environment	Flooring	floor_installed_carpet	4	3	ranksums	-2.9991	0.0027	0.0389
Maternal Diet	General dietary assessment	diet_gammtocho	45	978	kendall	0.0638	0.0028	0.0397
Maternal Characteristics	Socioeconomic status	data_edu	92	10	kendall	-0.0713	0.0029	0.0397
Human Milk	Fatty Acids	22:5n6 %	12	377	kendall	-0.0624	0.0031	0.0406
Maternal Health	Other health conditions	*mommed	886	4	kendall	-0.1965	0.0032	0.0413
Maternal Diet	General dietary assessment	diet_iron	45	978	kendall	0.0618	0.0039	0.0484
Maternal Diet	General dietary assessment	diet_vitb6	45	978	kendall	0.0612	0.0042	0.0513
Infant Feeding	Various	BF_12m	41	3	ranksums	2.8491	0.0044	0.0513

Table 2. Univariate analysis of predictors of head circumference at 3 months in the CHILD cohort study: top 40 features sorted by p-value. Associations are measured by Kendall's Tau for numeric features or Wilcoxon rank sum for binary features. Feature colors correspond to modality types as in Fig. 1A. P-values in the last column are corrected using the Benjamini–Hochberg FDR procedure across all considered features. Significant features with and without correction are marked by bold font in the respective p-value column. A blue background in the p-value columns marks positive, and orange mark negative associations. A “*” indicate variables that depend on other variables (e.g., diarrhea is only assessed when the baby has a cold). The column “na” lists the number of missing values for the respective variable and “unique” denotes how many unique values the variable is represented by. For a detailed description of features, see Supplemental Table S1.

features are associated with infant head circumference at 3 months (Fig. 2; many large filled circles) than at 1 year (Fig. 3, few large filled circles). At the same time, they give an overview of the association strength between each feature and head circumference at 3 months and 1 year, respectively.

Table 2 lists the top variables associated with head circumference at 3 months in univariate analyses (filled circles in Fig. 2). Notably, besides parental body composition-related variables (height and BMI), the single most associated feature to head circumference after 3 months was the n3-polyunsaturated fatty acid C22:6n3 (docosahexaenoic acid, DHA; $p=9.6e-05$), which is widely known to influence infant brain development¹⁷. In addition,

univariate relation: 'zhc_1y'

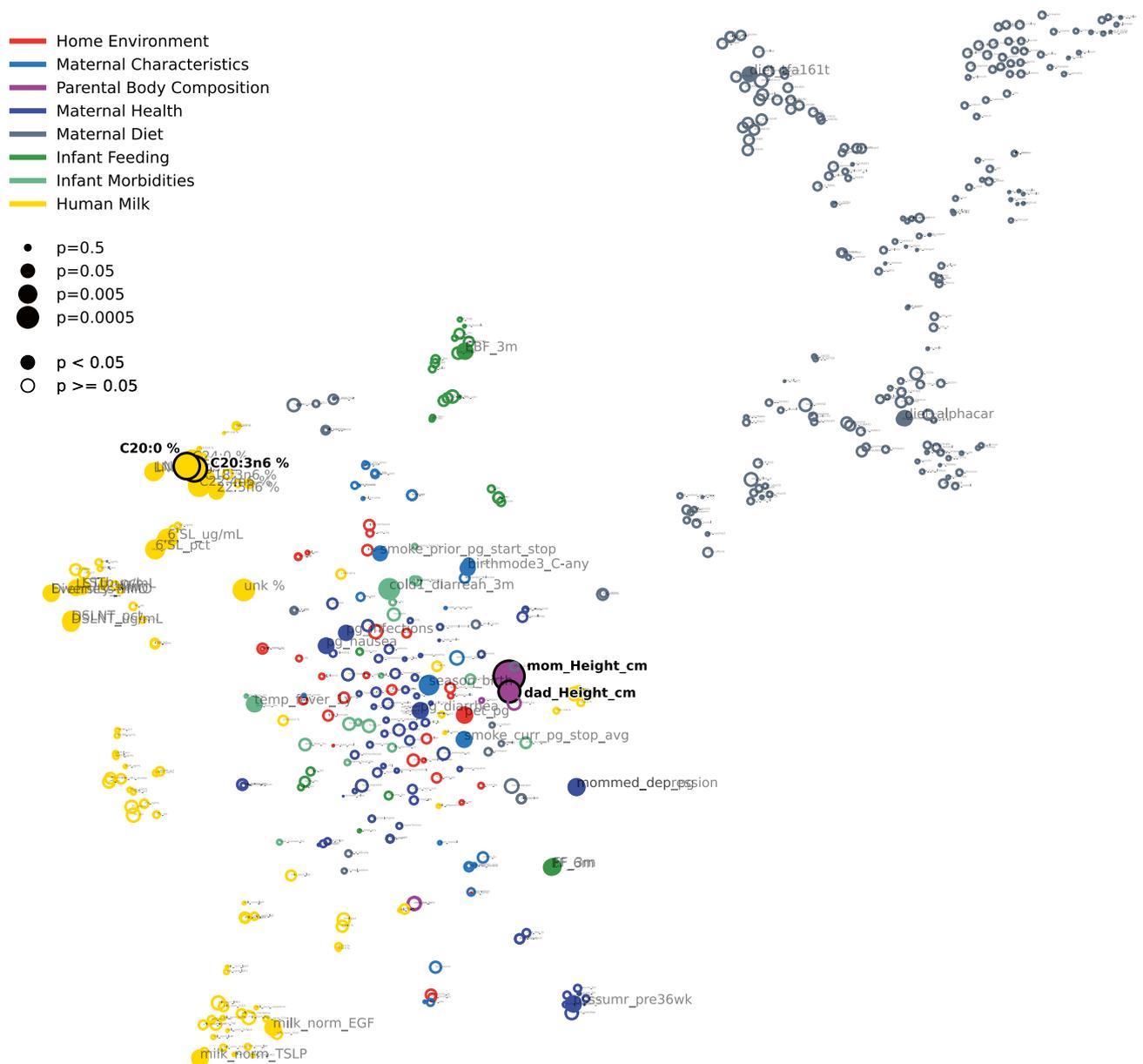


Figure 3. Feature interdependency network illustrating correlations among predictor variables and their associations with head circumference at 1 year. Analogously to Fig. 2, each node corresponds to a feature from various modalities (encoded by color). The closer the features the more similar they can be considered with regard to their correlation structure. Node sizes represent the strength of association between head circumference at twelve months and the corresponding feature based on the p-value of a respective significance test (Kendall's Tau for continuous variables, Wilcoxon rank sum test for binary variables). If this association passes a significance threshold of $p < 0.05$, the corresponding feature name is represented by a filled node (no multiple test correction for visualization purposes). Compared to head circumference at three months (Fig. 2), associations of features with head circumference at 1 year are much weaker (illustrated by fewer/smaller filled nodes). The remaining associations are concentrated in milk components. For a detailed description of features, see Supplemental Table S1.

the commonly investigated polyunsaturated fatty acid metrics DHA + EPA ($p = 2.7e-4$, where EPA represents eicosapentaenoic acid, i.e., C20:5n3) as well DHA/ARA ($p = 2.7e-4$, where ARA represents arachidonic acid, i.e., C20:4n6) were significantly associated with head circumference at 3 months (not shown in table). However, contrasting previously found connections between maternal fish oil or DHA supplementation and infant head circumference^{40,41}, we found increased DHA in HM to be associated with lower head circumference at 3 months ($\tau = -0.11$, $p = 9.6e-05$), pointing towards a negative relationship. The same is true for maternal fish intake

Modality	Category	Feature	na	unique	test	statistic	pvalue	pvalue (fdr_bh)
Parental Body Composition	general	mom_Height_cm	33	264	kendall	0.1127	0.0000	0.0001
Human Milk	Fatty Acids	C20:0 %	12	991	kendall	0.0921	0.0000	0.0030
Human Milk	Fatty Acids	C20:3n6 %	12	1011	kendall	0.0885	0.0000	0.0042
Parental Body Composition	general	dad_Height_cm	191	250	kendall	0.0826	0.0004	0.0489
Human Milk	Fatty Acids	C22:4n6 %	12	695	kendall	0.0685	0.0011	0.1067
Human Milk	Fatty Acids	unk %	12	704	kendall	-0.0676	0.0013	0.1067
Human Milk	Fatty Acids	C24:0 %	12	688	kendall	0.0660	0.0017	0.1128
Infant Morbidities	Diarrhea	*cold1_diarrhea_3m	914	3	ranksums	-3.1134	0.0018	0.1128
Human Milk	Fatty Acids	C18:3n6 %	12	843	kendall	0.0637	0.0025	0.1334
Maternal Characteristics	Season of birth	season_birth	0	4	kendall	-0.0667	0.0043	0.2104
Human Milk	HMOs	6'SL_ug/mL	4	1019	kendall	0.0591	0.0048	0.2121
Human Milk	HMOs	6'SL_pct	4	1019	kendall	0.0573	0.0062	0.2533
Human Milk	HMOs	DSLNT_pct	4	1019	kendall	0.0527	0.0118	0.4446
Human Milk	HMOs	LNnT_pct	4	1019	kendall	0.0517	0.0136	0.4726
Human Milk	Immunomodulators	milk_norm_TSLP	756	263	kendall	-0.0965	0.0192	0.5175
Maternal Health	Depression/Stress	*mommed_dep_pg	792	4	kendall	0.1210	0.0195	0.5175
Maternal Health	Other health conditions	*mommed_depression	792	4	kendall	0.1210	0.0195	0.5175
Human Milk	HMOs	DSLNT_ug/mL	4	1019	kendall	0.0473	0.0239	0.5175
Home Environment	Animals (pets)	pet_pg	29	3	ranksums	-2.2489	0.0245	0.5175
Infant Feeding	Various	EBF_3m	0	2	ranksums	2.2279	0.0259	0.5175
Infant Feeding	Various	FF_3m	0	2	ranksums	-2.2268	0.0260	0.5175
Maternal Health	Depression/Stress	psssumr_pre36wk	58	36	kendall	-0.0489	0.0264	0.5175
Maternal Health	Other pregnancy complications	pg_diarrhea	1	3	ranksums	2.1991	0.0279	0.5175
Human Milk	HMOs	LNnT_ug/mL	4	1019	kendall	0.0460	0.0282	0.5175
Human Milk	Immunomodulators	milk_norm_EGF	756	264	kendall	0.0899	0.0291	0.5175
Human Milk	HMOs	Evenness_HMO	4	1019	kendall	0.0457	0.0292	0.5175
Human Milk	HMOs	Diversity_HMO	4	1019	kendall	0.0457	0.0292	0.5175
Maternal Characteristics	Maternal smoking	*smoke_curr_pg_stop_a	822	20	kendall	0.1073	0.0310	0.5175
Infant Morbidities	Fever	*temp_fever_1y	422	3	ranksums	2.1536	0.0313	0.5175
Infant Feeding	Various	FF_6m	0	2	ranksums	-2.1468	0.0318	0.5175
Maternal Health	Other pregnancy complications	pg_nausea	24	3	ranksums	-2.1174	0.0342	0.5388
Maternal Diet	General dietary assessment	diet_alphacar	45	978	kendall	0.0444	0.0381	0.5605
Human Milk	HMOs	LSTb_ug/mL	4	1019	kendall	0.0434	0.0384	0.5605
Human Milk	Fatty Acids	22:5n6 %	12	377	kendall	-0.0435	0.0392	0.5605
Human Milk	HMOs	LSTb_pct	4	1019	kendall	0.0429	0.0405	0.5605
Maternal Health	Other pregnancy complications	pg_infections	24	3	ranksums	-2.0400	0.0414	0.5605
Maternal Characteristics	Delivery method	birthmode3_C-any	0	2	ranksums	-2.0123	0.0442	0.5829
Maternal Diet	General dietary assessment	diet_tfa161t	45	978	kendall	0.0423	0.0482	0.6186
Maternal Characteristics	Maternal smoking	*smoke_prior_pg_start_s	779	3	ranksums	1.9611	0.0499	0.6240
Maternal Diet	General dietary assessment	diet_aspartam	45	150	kendall	-0.0461	0.0515	0.6288

Table 3. Univariate analysis of predictors of head circumference at 1 year in the CHILd cohort study: top 40 features sorted by p-value. Associations are measured by Kendall's Tau for numeric features or Wilcoxon rank sum for binary features. Feature colors correspond to modality types as in Fig. 1A. P-values in the last column are corrected using the Benjamini–Hochberg FDR procedure across all considered features. Significant features with and without correction are marked by bold font in the respective p-value column. A blue background in the p-value columns marks positive, and orange mark negative associations. A “*” indicate variables that depend on other variables (e.g., diarrhea is only assessed when the baby has a cold). The column “na” lists the number of missing values for the respective variable and “unique” denotes how many unique values the variable is represented by. For a detailed description of features, see Supplemental Table S1.

(tau = - 0.08, p = 4.1e-03), which is a key dietary source of DHA. Other significantly associated features (all p < 3.0e-02) observed in our analysis included the HM saturated fatty acid C20:0 (positive association), various

Modality	Category	Feature	na	unique	test	statistic	pvalue	pvalue (fdr_bh)
Human Milk	Fatty Acids	C22:6n3 %	12	970	kendall	-0.1061	0.0000	0.0000
Human Milk	Fatty Acids	C20:0 %	12	991	kendall	0.0923	0.0000	0.0006
Human Milk	Fatty Acids	C20:3n6 %	12	1011	kendall	0.0690	0.0010	0.0308
Human Milk	Immunomodulators	milk_norm_TSLP	756	263	kendall	-0.1336	0.0012	0.0308
Human Milk	Fatty Acids	C20:5n3 %	12	693	kendall	-0.0634	0.0026	0.0534
Human Milk	Fatty Acids	22:5n6 %	12	377	kendall	-0.0624	0.0031	0.0534
Human Milk	Fatty Acids	C24:0 %	12	688	kendall	0.0542	0.0100	0.1468
Human Milk	HMOs	LSTb_ug/mL	4	1019	kendall	0.0528	0.0118	0.1468
Human Milk	Fatty Acids	C22:5n3 %	12	922	kendall	-0.0524	0.0127	0.1468
Human Milk	Fatty Acids	C18:2n6 %	12	1011	kendall	-0.0510	0.0154	0.1604
Human Milk	Fatty Acids	C17:0 %	12	1011	kendall	0.0497	0.0182	0.1717
Human Milk	Immunomodulators	milk_norm_GRO	756	267	kendall	-0.0947	0.0215	0.1770
Human Milk	Fatty Acids	C18:3n6 %	12	843	kendall	0.0482	0.0221	0.1770
Human Milk	HMOs	LSTb_pct	4	1019	kendall	0.0463	0.0271	0.1816
Human Milk	Fatty Acids	C18:0 %	12	1011	kendall	0.0463	0.0276	0.1816
Human Milk	Fatty Acids	C24:1n9 %	12	543	kendall	0.0458	0.0297	0.1816
Human Milk	Hormones	milk_leptin	643	380	kendall	0.0742	0.0312	0.1816
Human Milk	Immunomodulators	milk_norm_CCL5	756	265	kendall	-0.0886	0.0314	0.1816
Human Milk	HMOs	LNH_ug/mL	4	1019	kendall	-0.0434	0.0382	0.2093
Human Milk	Immunomodulators	milk_norm_Lactoferrin	756	267	kendall	-0.0790	0.0552	0.2739
Human Milk	Fatty Acids	14:1n9 %	12	1007	kendall	0.0399	0.0579	0.2739
Human Milk	HMOs	LNH_pct	4	1019	kendall	-0.0397	0.0579	0.2739
Human Milk	HMOs	LNFP III_ug/mL	4	1019	kendall	-0.0388	0.0644	0.2913
Human Milk	Fatty Acids	C20:4n6 %	12	1011	kendall	-0.0382	0.0693	0.3002
Human Milk	Fatty Acids	C20:2n6 %	12	1011	kendall	-0.0363	0.0842	0.3502
Human Milk	Immunomodulators	milk_norm_IP-10	756	263	kendall	-0.0702	0.0883	0.3530
Human Milk	Immunomodulators	milk_norm_MCP-1	756	266	kendall	-0.0644	0.1180	0.4483
Human Milk	HMOs	3FL_ug/mL	4	1019	kendall	-0.0325	0.1207	0.4483
Human Milk	Fatty Acids	C16:1n9 %	12	1011	kendall	0.0301	0.1526	0.5473
Human Milk	Fatty Acids	C18:1C11 %	12	1011	kendall	-0.0286	0.1747	0.5911
Human Milk	Fatty Acids	TVA %	12	1011	kendall	0.0284	0.1768	0.5911
Human Milk	HMOs	3FL_pct	4	1019	kendall	-0.0274	0.1912	0.5911
Human Milk	HMOs	LNFP III_pct	4	1019	kendall	-0.0272	0.1941	0.5911
Human Milk	Immunomodulators	milk_norm_TLR2	756	263	kendall	-0.0532	0.1963	0.5911
Human Milk	Fatty Acids	unk %	12	704	kendall	0.0270	0.1989	0.5911
Human Milk	Fatty Acids	total lipid	149	151	kendall	0.0280	0.2186	0.6208
Human Milk	HMOs	DSLNH_ug/mL	4	1019	kendall	-0.0257	0.2209	0.6208
Human Milk	HMOs	Mom_HMO_secretor	4	3	ranksums	1.2069	0.2275	0.6226
Human Milk	HMOs	FDSLNH_ug/mL	4	1019	kendall	-0.0245	0.2422	0.6460
Human Milk	Fatty Acids	C10:0 %	12	1011	kendall	-0.0240	0.2535	0.6591

Table 4. Univariate analysis of human milk predictors of head circumference at 3 months in the CHILD cohort study: top 40 human milk components sorted by p-value. Associations are measured by Kendall's Tau for numeric features or Wilcoxon rank sum for binary features. Feature colors correspond to modality types as in Fig. 4. P-values in the last column are corrected using the Benjamini–Hochberg FDR procedure across all considered features. Significant features with and without correction are marked by bold font in the respective p-value column. A blue background in the p-value columns marks positive, and orange mark negative associations. The column “na” lists the number of missing values for the respective variable and “unique” denotes how many unique values the variable is represented by. For a detailed description of features, see Supplemental Table S1.

estimated PUFA intakes from FFQ data (negative), maternal height and body mass index (positive), infant diarrhea (negative), and maternal intake of alcohol (A_bev; negative) and starchy vegetables (M_STARCHY; positive). See Supplementary Fig. S1 for a visualization of relationships between selected variables and head circumference at 3 months.

By 1 year (Fig. 3, Table 3), many of the feature associations with head circumference observed at 3 months (represented by filled circles in Fig. 2) were no longer evident. The strongest associations to head circumference at 1 year were primarily among human milk components, although interestingly, these were mostly different from the components associated with head circumference at 3 months. Notably, only two fatty acids (C20:0 and C20:3n6) are significantly associated with head circumference at both 3 months and 1 year (see Supplementary

Modality	Category	Feature	na	unique	test	statistic	pvalue	pvalue (fdr_bh)
Human Milk	Fatty Acids	C20:0 %	12	991	kendall	0.0921	0.0000	0.0013
Human Milk	Fatty Acids	C20:3n6 %	12	1011	kendall	0.0885	0.0000	0.0014
Human Milk	Fatty Acids	C22:4n6 %	12	695	kendall	0.0685	0.0011	0.0341
Human Milk	Fatty Acids	unk %	12	704	kendall	-0.0676	0.0013	0.0341
Human Milk	Fatty Acids	C24:0 %	12	688	kendall	0.0660	0.0017	0.0358
Human Milk	Fatty Acids	C18:3n6 %	12	843	kendall	0.0637	0.0025	0.0426
Human Milk	HMOs	6'SL_ug/mL	4	1019	kendall	0.0591	0.0048	0.0710
Human Milk	HMOs	6'SL_pct	4	1019	kendall	0.0573	0.0062	0.0810
Human Milk	HMOs	DSLNT_pct	4	1019	kendall	0.0527	0.0118	0.1369
Human Milk	HMOs	LNnT_pct	4	1019	kendall	0.0517	0.0136	0.1410
Human Milk	Immunomodulators	milk_norm_TSLP	756	263	kendall	-0.0965	0.0192	0.1814
Human Milk	HMOs	DSLNT_ug/mL	4	1019	kendall	0.0473	0.0239	0.1901
Human Milk	HMOs	LNnT_ug/mL	4	1019	kendall	0.0460	0.0282	0.1901
Human Milk	Immunomodulators	milk_norm_EGF	756	264	kendall	0.0899	0.0291	0.1901
Human Milk	HMOs	Evenness_HMO	4	1019	kendall	0.0457	0.0292	0.1901
Human Milk	HMOs	Diversity_HMO	4	1019	kendall	0.0457	0.0292	0.1901
Human Milk	HMOs	LSTb_ug/mL	4	1019	kendall	0.0434	0.0384	0.2218
Human Milk	Fatty Acids	22:5n6 %	12	377	kendall	-0.0435	0.0392	0.2218
Human Milk	HMOs	LSTb_pct	4	1019	kendall	0.0429	0.0405	0.2218
Human Milk	HMOs	3FL_pct	4	1019	kendall	-0.0394	0.0604	0.3062
Human Milk	Immunomodulators	milk_norm_IP-10	756	263	kendall	-0.0769	0.0618	0.3062
Human Milk	Immunomodulators	milk_norm_MDC	756	217	kendall	0.0762	0.0648	0.3062
Human Milk	Fatty Acids	14:1n9 %	12	1007	kendall	0.0374	0.0752	0.3398
Human Milk	HMOs	3FL_ug/mL	4	1019	kendall	-0.0356	0.0895	0.3817
Human Milk	Fatty Acids	total lipid	149	151	kendall	-0.0384	0.0918	0.3817
Human Milk	HMOs	2'FL_pct	4	1019	kendall	-0.0335	0.1102	0.4406
Human Milk	Fatty Acids	C10:0 %	12	1011	kendall	-0.0312	0.1383	0.5329
Human Milk	HMOs	LNT_ug/mL	4	1019	kendall	0.0305	0.1461	0.5428
Human Milk	Immunomodulators	milk_norm_TNF	756	163	kendall	-0.0591	0.1528	0.5480
Human Milk	HMOs	2'FL_ug/mL	4	1019	kendall	-0.0294	0.1613	0.5592
Human Milk	Immunomodulators	milk_norm_VEGF	756	264	kendall	0.0554	0.1789	0.5847
Human Milk	Immunomodulators	milk_norm_IgG2	756	249	kendall	0.0550	0.1816	0.5847
Human Milk	HMOs	LNT_pct	4	1019	kendall	0.0277	0.1855	0.5847
Human Milk	Immunomodulators	milk_norm_IgA	756	264	kendall	0.0521	0.2057	0.6293
Human Milk	HMOs	Mom_HMO_secretor	4	3	ranksums	1.1924	0.2331	0.6357
Human Milk	Immunomodulators	milk_norm_CCL5	756	265	kendall	-0.0485	0.2389	0.6357
Human Milk	Fatty Acids	C18:1C11 %	12	1011	kendall	-0.0247	0.2397	0.6357
Human Milk	Fatty Acids	20:4n3 %	12	713	kendall	0.0247	0.2397	0.6357
Human Milk	Hormones	milk_leptin	643	380	kendall	0.0402	0.2437	0.6357
Human Milk	Fatty Acids	TVA %	12	1011	kendall	0.0244	0.2464	0.6357

Table 5. Univariate analysis of human milk predictors of head circumference at 1 year in the CHILD cohort study: top 40 human milk components sorted by p-value. Associations are measured by Kendall's Tau for numeric features or Wilcoxon rank sum for binary features. Feature colors correspond to modality types as in Fig. 5. P-values in the last column are corrected using the Benjamini–Hochberg FDR procedure across all considered features. Significant features with and without correction are marked by bold font in the respective p-value column. A blue background in the p-value columns marks positive, and orange mark negative associations. The column “na” lists the number of missing values for the respective variable and “unique” denotes how many unique values the variable is represented by. For a detailed description of features, see Supplemental Table S1.

Fig. S2 for a visualization of C20:0). Additionally, HMOs are more prominent among the features significantly associated with head circumference at 1 year (without correction for multiple hypothesis comparison) compared at 3 month (cf. Tables 4 and 5).

Overall, the decreased association strengths of individual features with head circumference at 1 year could be explained by the more complex relationship of influence factors and developmental processes of infants as time progresses after pregnancy, which is also reflected by the multivariable modeling results discussed in the previous section and visualized in Fig. 1B where power decreases for the models at 1 year.

univariate relation: 'zhc_3m'

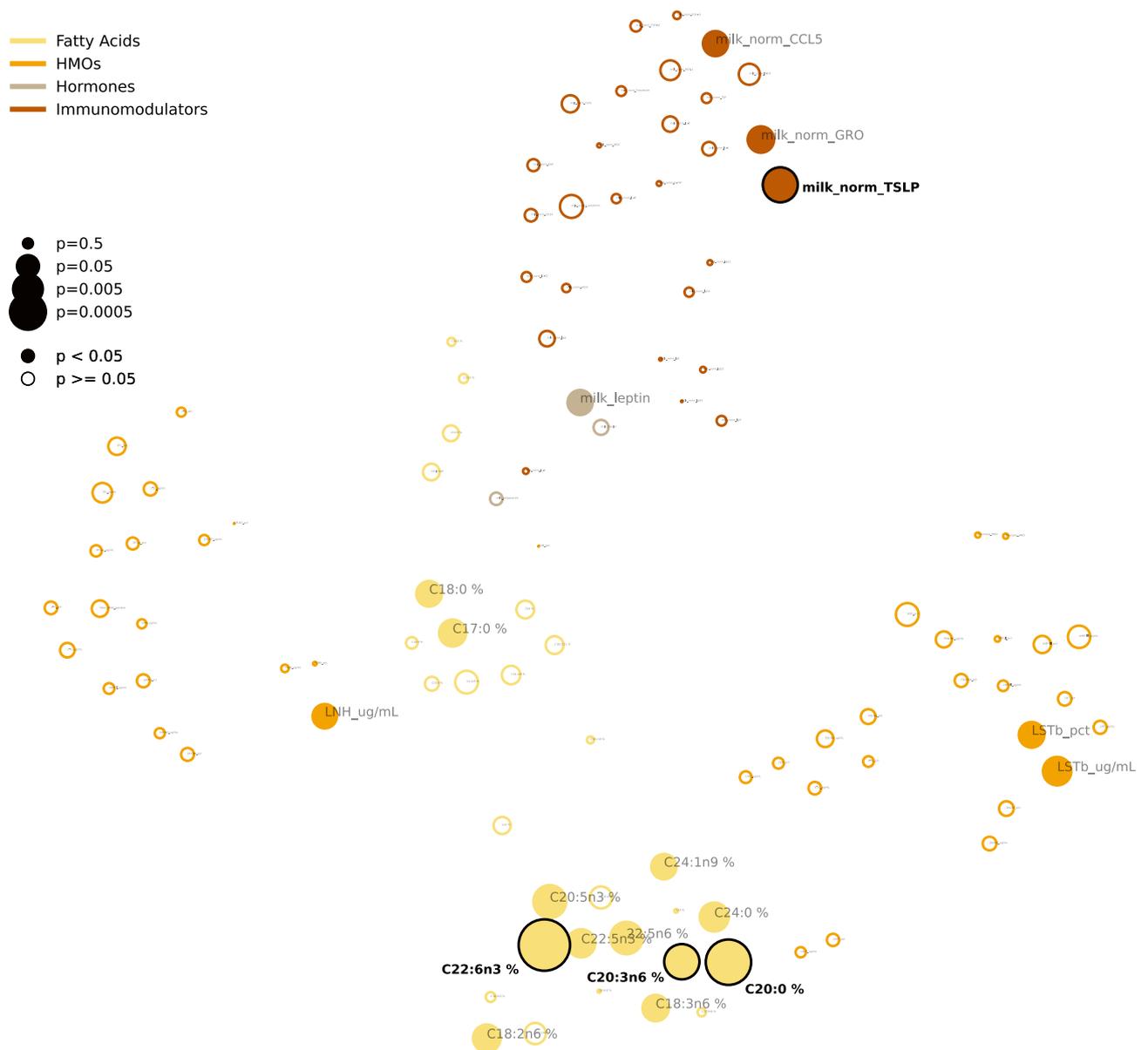


Figure 4. Feature interdependency network of human milk components illustrating correlations among predictor variables and their associations with head circumference at 3 months. Each node corresponds to a feature from various human milk modalities (encoded by color). The closer the features, the more similar they can be considered with regard to their correlation structure. Node sizes represent the strength of association between head circumference at 3 months and the corresponding feature based on the p-value of a respective significance test (Kendall's Tau for continuous variables, Wilcoxon rank sum test for binary variables). If this association passes a significance threshold of $p < 0.05$, the corresponding feature name is represented by a filled node (no multiple test correction for visualization purposes). Clusters emerge, grouping, for example, specific fatty acids and immunomodulators. For a detailed description of features, see Supplemental Table S1.

Discussion

In this study of over 1000 mother-infant dyads, we applied a unique systems biology approach to model the complex relationships between infant head circumference (a proxy for brain development) and hundreds of maternal, infant, environmental, and HM factors. Among others, we identified a well-established pathway with potential practical impact (i.e., maternal fish intake elevates HM DHA, which impacts infant brain development—albeit in an unexpected direction), which validates our statistical approach, gives credence to the additional novel associations we observed, and sets the foundation for further analyses with additional dyads, cohorts and HM analytes.

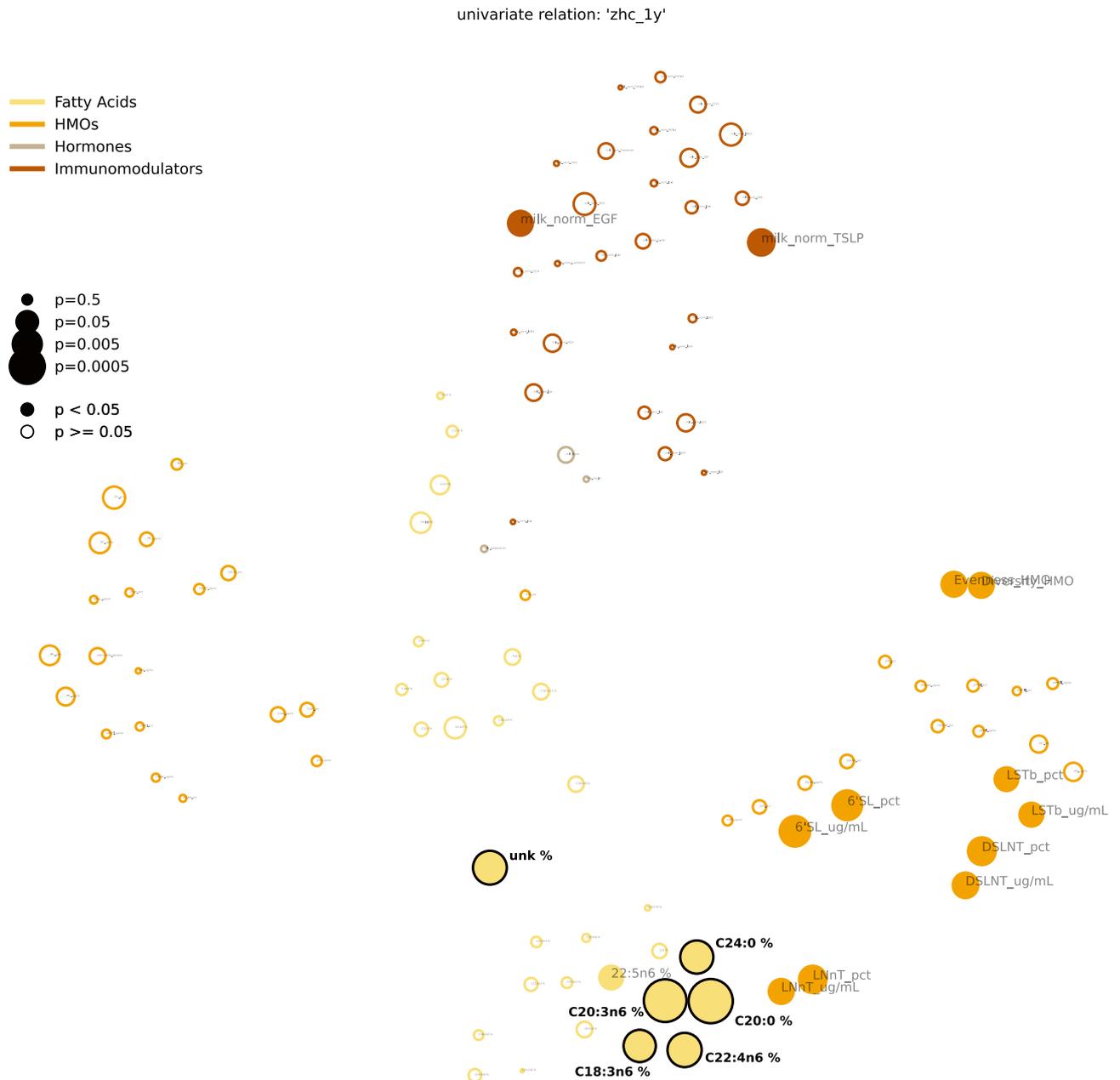


Figure 5. Feature interdependency network of human milk components illustrating correlations among predictor variables and their associations with head circumference at 1 year. Each node corresponds to a feature from various human milk modalities (encoded by color). The closer the features, the more similar they can be considered with regard to their correlation structure. Node sizes represent the strength of association between head circumference at 3 months and the corresponding feature based on the p-value of a respective significance test (Kendall's Tau for continuous variables, Wilcoxon rank sum test for binary variables). If this association passes a significance threshold of $p < 0.05$, the corresponding feature name is represented by a filled node (no multiple test correction for visualization purposes). Compared to associations at three months (Fig. 4), associations with head circumference appear increasingly in an HMO cluster. For a detailed description of features, see Supplemental Table S1.

Integrated models increase predictive power and highlight the complexity of infant growth

Infant growth has been studied intensively in relation to many different factors, including feeding practices and HM components⁴². However, few, if any, have taken a systems biology approach. In this study, we jointly analyzed many different HM components together with other modalities in an integrated multivariate predictive model. This allowed us to simultaneously consider all included factors; an approach that has shown to be effective with regard to predictive power in other research areas. For example, integrating multiple modalities and omics has been successfully applied to model pregnancy progression⁴³, predicting onset of labor⁴⁴ and preeclampsia^{45–47}. Our current findings highlight the complexity of infant growth and brain development and emphasize the need

for further studies in diverse populations accounting for additional HM components and more sociodemographic factors and their interplay.

Key predictors of infant head circumference

Consistent with previous research⁴⁸, we found that maternal height was significantly associated with infant head circumference. After maternal height, the most associated feature to head circumference at 3 months was the HM n3-polyunsaturated fatty acid C22:6n3 (docosahexaenoic acid, DHA; $p = 9.6e-05$), which has a well-established role in supporting brain development. DHA is the most abundant omega-3 fatty acid in the brain and is implicated in several neuronal functions, including neurogenesis and neurotransmission^{17,49,50}, although postnatal DHA supplementation trials in preterm⁵¹ and term⁵² infants have not shown consistent benefits for neurodevelopment, perhaps because DHA is mainly accumulated by the fetus during the last trimester of pregnancy¹⁷. Notably, in our study, increased DHA in HM seemed to be associated with *lower* head circumference at 3 months ($\text{tau} = -0.11$, $p = 9.6e-05$), suggesting a potentially negative relationship. However, it is important to note that this finding is based on relative DHA proportions (not absolute concentrations), since information on total milk fat, calories, and volume was not available and therefore we could not determine the total “dose” of milk fatty acids delivered to each infant³⁸. Further investigation is needed to confirm and understand the unexpected direction of this relationship.

Other features significantly associated with infant head circumference in our analysis included other HM fatty acids (e.g., C20:0, positive relationship), various estimated PUFA intakes from maternal FFQ data (negative), and maternal intake of alcohol (negative), fish (negative) and starchy vegetables (positive). This supports evidence that maternal diet during pregnancy and/or lactation can influence fetal/infant brain development^{53–55}. Infant colds with diarrhea were associated with smaller head circumference at 3 months, suggesting that gut health and intestinal infections in early life could influence brain development.

Complexity of modeling growth factors over time

The complexity of the different factors and data modalities and their relationship to infant growth is further highlighted by the reduced predictive power for head circumference after 1 year. Only parental body composition and milk components are consistently predictive at 3 months as well as after 1 year (see Fig. 1B), however, they still decrease in predictive power. This shows that predicting further into the future (1 year vs. 3 months) can be challenging and may require additional information or data to make accurate predictions or more powerful models. This also points towards external variables and influencing factors that continuously change and may require additional longitudinal monitoring of key variables during the growth period (for example, HM composition changes over time, but was only measured once in this study). In addition, for HM data in general and HM fatty acids in particular, we observed that non-linear models performed better at 3 months while linear models were superior at 1 year, hinting at more complex functional relationships that are picked up by these models at 3 months. As well, while HMOs did not predict head circumference at 3 months, they approached significance as predictors of head circumference at 1 year, suggesting that different HM components (in this case, fatty acids vs. HMOs) may contribute differentially to head growth or brain development at different stages of infancy, thus emphasizing the importance of a) broadly considering many HM components and b) examining child outcomes longitudinally.

Limitations and future work

This proof-of-concept analysis is a starting point to explore the complex relationships among different predictors of HM composition and infant growth. For example, the microbiome of HM⁵⁶ could actively modify (metabolize or synthesize) other HM components and directly or indirectly influence infant growth. Studying these intricate functional relationships may allow for more powerful predictive models and a deeper understanding of the underlying processes⁵⁷. A limitation of our study is that some data were available for only a subset of dyads (e.g., immunomodulator and hormone data were only available for roughly 25% of dyads), and some key HM components were not analyzed (e.g., growth factors, micronutrients, and macronutrients). This is because we used an existing dataset that was assembled primarily to study associations with immune development and allergic disease, where nutrients were not prioritized. These gaps will be addressed in future studies within the International Milk Composition (IMiC) Consortium (www.milcresearch.com/imic), established specifically to study the complex associations between HM and infant growth. It is also important to note that, while significant, the associations detected are mostly weak to moderate, suggesting that additional data, as well as more advanced models, will be required to model the increasingly complex relationships between external factors and infant growth in a more effective manner. In future work, it may be useful to apply more intricate multiomics modeling approaches²⁹ to account for the different information densities within the different data modalities (for example, the microbiome is more sparse than other modalities). Finally, in this work, we focused solely on infant head circumference; however, the approach could equally be applied to other anthropometric outcomes such as infant weight, height, or growth trajectories. As these outcomes are tightly related, state-of-the-art multitask models⁵⁸, as well as further concentrating on the interaction of mothers and their infants⁵⁹ may be of particular interest.

Conclusion

Using a systems biology approach to investigate multiple HM components simultaneously together with maternal, infant, and environmental data, we identified well-established pathways with potential practical implications (e.g., an association between maternal fish intake and HM DHA, which is connected to infant head circumference). These pathways, as well as our holistic machine learning based approach to understanding infant development in the context of head circumference, set the foundation for further analyses with additional dyads, HM analytes,

and clinical outcomes. Within the CHILD Cohort Study and IMiC Consortium, this will include additional features such as the HM proteome, metabolome, and microbiome, as well as the gut microbiome and additional health outcomes, including linear growth, weight gain, asthma, and obesity. These additional modalities, outcomes, and increased sample sizes will enable the application of state-of-the-art multi-modal multi-task machine learning⁵⁸ to jointly model, integrate, exploit, and understand relationships between the different HM components, other modalities and infant outcomes, paving the way for unprecedented insights into infant development.

Materials and methods

Study design and data

We used a systems biology approach to investigate multiple HM components together with maternal, infant, and environmental data from mother-infant dyads in the CHILD cohort. CHILD is an ongoing general population pregnancy cohort of 3624 families recruited in 2009–2012 across four Canadian centers (Vancouver, Edmonton, Manitoba, and Toronto)³⁵. The study was approved by the Human Research Ethics Boards at McMaster University, University of Manitoba, University of Alberta, University of British Columbia, and SickKids Hospital, and carried out in accordance with relevant guidelines and regulations. All participants provided written informed consent at enrollment. Raw data and processed data will be available with appropriate permissions from the CHILD Cohort Study: <https://childstudy.ca/for-researchers/data-access/>.

HM samples were collected at 3–4 months postpartum during a home visit³⁶. Briefly, mothers collected (hand expression preferred; pump expression accepted) and mixed foremilk and hindmilk from multiple feedings over a 24-h period and kept the sample refrigerated (for no more than 24 h) until it was collected by study staff and transported on ice to the laboratory for aliquoting and storage at $-80\text{ }^{\circ}\text{C}$ until analysis. While some degradation of some components is possible within this 24-h period, current literature indicates that, aside from nucleic acids (which were not analyzed in our study), “refrigeration for up to 72 h keeps most constituents intact and limits lipolysis and bacterial growth”⁶⁰. This sampling protocol could have increased the risk of (potentially selective) degradation of some HM components. However, the CHILD study opted to have mothers collect and refrigerate samples over a 24 h period in order to collect a “daily average” profile, which is important for HM components that are known to fluctuate diurnally. The HM subset ($n = 1200$) was originally selected to enrich for dyads with allergy and obesity phenotypes, plus healthy controls³⁸. A subset of 1022 dyads with available HM data and head circumference measurements at both 3 months and 1 year was included in this analysis (see Supplementary Fig. S4 for a corresponding flow chart).

Head circumference was measured by trained and certified study staff at a 3 month home visit and at a 1 year clinical assessment. Head circumference was measured by taking a maximum of three repetitions. We used the z-score for age to represent head circumference in our study according to WHO standards⁶¹. The CHILD study excluded premature infants and those born with congenital anomalies, including macrocephaly and microcephaly. Genetic conditions associated with HC were not considered for exclusion.

HM data included: 3 hormones (leptin, insulin, and adiponectin) measured using enzyme-linked immunosorbent assay³⁹, 19 HMOs measured using high-performance liquid chromatography^{37,62}, 28 fatty acids measured using gas chromatography³⁸, and 28 immunomodulators measured using immunoassays. Immunoglobulin levels and selected cytokines and chemokines were measured by Luminex multiplex assays: a panel of 24 analytes was analyzed using premixed multianalyte kits according to the manufacturers’ recommendations and acquired by Luminex 200 (Bio-Rad), with calibration and standard controls. The Luminex kits used were R&D LXSAHM and Thermofisher EPX070-10818-901, EPX010-12283-901, and PPX-09-MX2W79V. Sandwich ELISA assays were used to assess total IgA (e-bioscience 88-50600-88), TGF- β 1 (R&D Dy240), and TGF- β 2 (R&D Dy241). We integrated these HM datasets with data reflecting maternal sociodemographic characteristics (e.g., age, marital status, education; $n = 36$ features), health (e.g., past and present chronic conditions; $n = 84$), diet (e.g., food and nutrient intakes and dietary patterns; $n = 207$)^{63,64}, and body composition (e.g., height, weight, body mass index; $n = 6$); infant health (e.g., infections, colds, fevers, chronic conditions, medical visits; $n = 87$) and feeding practices (e.g., breastfeeding exclusivity and duration, introduction of formula milks and solid foods; $n = 27$); and the home environment (e.g., types of flooring, furniture and cleaning products; $n = 31$). For maternal BMI we refer to a best estimate of pre-pregnancy BMI. This estimate is based on height at 1 year postpartum (measured by study staff) and either self-reported pre-pregnancy weight or (if the mother could not recall) measured weight at 1 year postpartum. A complete list of features is provided in Supplemental Table S1. These features were used collectively to predict head circumference (z-score for age) measured by study staff at 3 months and 1 year using a flexible tape measure wrapped snugly around the widest possible circumference (average of three repeat measurements). The final number of features used to predict head circumference was 498 at 3 months and 582 at 1 year (some features like “the number of people in the house at 1 year” were not valid to use for predicting head circumference at 3 months).

Predictive multi-omics modeling

We aimed to combine the previously mentioned data sources (HM oligosaccharides, fatty acids, hormones and immunomodulators; maternal demographic, health and dietary information, infant morbidities, feeding data, and home environment information) to collectively predict infant head circumference at 3 months and 1 year. To achieve this, we first align the different modalities across all infants and concatenate them into an integrated feature matrix. We then built models and evaluated them for both timepoints separately. For each timepoint, we employ a fivefold cross-validation scheme and collect predictions across all five folds to then calculate the significance of the prediction measured by Spearman’s rank correlation. The cross-validation scheme ensures that, even if overfitting on the training set occurs, the scores are reported on out of sample instances from the test set. In each fold, individually, we independently impute missing values using the median, and standard scale

features. We kept all variables independent of their missing value count to preserve rare and seldomly recorded events like epilepsy or previous cancer therapy for the model to use. Imputation and scaling are derived from each fold's training data and applied to the respective test data. We repeated this procedure 50 times and visualized the mean and standard deviation of the corresponding negative \log_{10} p-values in Fig. 1B. We also use the same set up to predict head circumference from individual data modalities to understand their contribution and association with head circumference. Note that milk immunomodulator and hormone modalities were excluded from model building since they were only available for less than one-third of the data. Additionally, we only include variables that have been measured at 3 months or before. For the models, we used linear and non-linear machine learning algorithms (Ridge regression and Support Vector Machines). The former has the potential to cope better with large amounts of features, even if features are highly correlated. The latter allows us to model more complex relationships between features and outcomes. For both approaches, the data matrix X contains all features concatenated across all data modalities for all available mother-infant-dyads, while y represents the head circumference either after three months or 1 year.

For ridge regression⁶⁵, the goal is to derive coefficients β for each feature in X minimize the overall difference from y :

$$L(\beta) = \|y - X\beta\|_2^2.$$

However, this approach is not ideal for the analysis of the highly interrelated multi-modality data set, because it would select only representatives of communities of correlated features while disregarding highly correlated but potentially relevant features. To address this limitation, L_2 regularization is applied on β to allow the inclusion of highly correlated measurements:

$$\beta_{\text{ridge}} = \operatorname{argmin}_{\beta} \|y - X\beta\|_2^2 + \lambda \|\beta\|_2^2$$

Here λ specifies the regularization strength and is selected via nested cross-validation in each fold separately. That is, for each outer fold, the training set is split into inner folds via a leave-one-out procedure with a respective training and testing set each. For each inner fold we fit (inner training set) and test (inner test set/sample) ridge regression models with a set of parameters $\lambda \in \{0.1, 1.0, 10.0\}$. The best parameter λ according to mean performance (negative mean squared error) across the inner fold is selected to train the final model for the outer fold (the model is retrained on the complete outer training set).

In addition to ridge regression, we employ support vector regression⁶⁶ with a radial basis function (RBF) kernel K in order to capture more complex relationships between features and head circumference:

$$K(x_1, x_2) = \exp\left(-\frac{\|x_1 - x_2\|^2}{2\sigma^2}\right) = \exp(-\gamma\|x_1 - x_2\|^2)$$

Here, σ^2 represents the variance of the Gaussian distribution underlying the kernel, and $\gamma = \frac{1}{2\sigma^2}$. The use of this kernel allows the SVR model to project data points into an infinite dimensional space that enables learning nonlinear relationships between features in X and outcome, where γ specifies the radius of influence of each support vector/training sample on the final function. Here, γ was set to $\frac{1}{n_{\text{features}} \operatorname{VAR}(X)}$. In addition, SVMs allow to specify a parameter C that allows adjusting how strongly outliers are taken into account. In our experiments C is set to 1.

On a programmatic level, the Scikit-Learn (version 0.23.2) Python (version 3.7.8) package was used to train the models⁶⁷.

Univariate analysis and visualization

To understand univariate associations of individual features with head circumference, we calculated Kendall's Tau for numeric features and Wilcoxon rank sum test for binary variables. The top associated features at 3 months are shown in Table 2, and the top features at 1 year are shown in Table 3. The corresponding p-values were corrected for false discovery rates using the Benjamini–Hochberg procedure. We furthermore give an overview of the association of head circumference to all features as correlation networks (cf. Figs. 2 and 3). To calculate 2-d coordinates for each variable, we first imputed all missing values using the respective medians. Then we calculated the correlation matrix between all features using Spearman correlation. Based on the absolute values of this correlation matrix, features are then placed in a 2-d plane using t-distributed stochastic neighbor embeddings (t-SNE)⁶⁸. Thus, the closer to features, the more similar they are with regard to their correlation structure. We provide the same statistics for milk components, specifically in Tables 4 and 5, as well as Figs. 4 and 5.

Data availability

Raw data and processed data will be made available with appropriate permissions from the CHILd Cohort Study: <https://childstudy.ca/for-researchers/data-access/>. The source code is publicly available at <http://nalab.stanford.edu/infant-growth-multiomics>.

Received: 28 September 2023; Accepted: 17 January 2024

Published online: 05 February 2024

References

1. Rush, D. Maternal nutrition and perinatal survival. *J. Health Popul. Nutr.* **19**, S217–S264 (2001).

2. Hermoso, M., Vollhardt, C., Bergmann, K. & Koletzko, B. Critical micronutrients in pregnancy, lactation, and infancy: Considerations on vitamin D, folic acid, and iron, and priorities for future research. *Ann. Nutr. Metab.* **59**, 5–9 (2011).
3. de Waard, M. *et al.* Optimal nutrition in lactating women and its effect on later health of offspring: A systematic review of current evidence and recommendations (EarlyNutrition project). *Crit. Rev. Food Sci. Nutr.* **57**, 4003–4016 (2017).
4. Phuphaibul, R. *et al.* Socioeconomic determinants of infant growth: The perspective cohort study of Thai children. *Jpn J. Nurs. Sci.* **11**, 16–22 (2014).
5. Vielwerth, S. E., Jensen, R. B., Larsen, T. & Greisen, G. The impact of maternal smoking on fetal and infant growth. *Early Hum. Dev.* **83**, 491–495 (2007).
6. Schlaudecker, E. P., Steinhoff, M. C. & Moore, S. R. Interactions of diarrhea, pneumonia, and malnutrition in childhood: Recent evidence from developing countries. *Curr. Opin. Infect. Dis.* **24**, 496–502 (2011).
7. Patro-Goląb, B., Zalewski, B. M., Polaczek, A. & Szajewska, H. Duration of breastfeeding and early growth: A systematic review of current evidence. *Breastfeed Med.* **14**, 218–229 (2019).
8. Perrella, S. *et al.* Human milk composition promotes optimal infant growth, development and health. *Semin. Perinatol.* **45**, 151380 (2021).
9. Shenhav, L. & Azad, M. B. Using community ecology theory and computational microbiome methods to study human milk as a biological system. *mSystems.* **7**, 3221 (2022).
10. Christian, P. *et al.* The need to study human milk as a biological system. *Am. J. Clin. Nutr.* **113**, 1063–1072 (2021).
11. Bode, L., Raman, A. S., Murch, S. H., Rollins, N. C. & Gordon, J. I. Understanding the mother–breastmilk–infant “triad”. *Science.* **367**, 1070–1072 (2020).
12. Abrego Del Castillo, K. Y. *et al.* Maternal BMI, breastfeeding and perinatal factors that influence early childhood growth trajectories: A scoping review. *J. Dev. Org. Health Dis.* **13**, 541–549 (2022).
13. Turck, D. *et al.* World Health Organization 2006 child growth standards and 2007 growth reference charts: A discussion paper by the committee on Nutrition of the European Society for Pediatric Gastroenterology, Hepatology, and Nutrition. *J. Pediatr. Gastroenterol. Nutr.* **57**, 258–264 (2013).
14. Bayley, N. Bayley scales of infant and toddler development, third edition [Internet]. PsycTESTS Dataset. American Psychological Association (APA). <https://doi.org/10.1037/t14978-000> (2012).
15. Squires, J., & Bricker, D. Ages and stages questionnaires”, third edition [Internet]. PsycTESTS Dataset. American Psychological Association (APA). <https://doi.org/10.1037/t11523-000> (2012).
16. Ministry of Health. Food and nutrition guidelines for healthy pregnant and breastfeeding women: a background paper. Ministry of Health, Wellington (2006).
17. Innis, S. M. Dietary (n-3) Fatty Acids and Brain Development. *J Nutr.* **137**, 855–859 (2007).
18. Berti, C. *et al.* Micronutrients in pregnancy: current knowledge and unresolved questions. *Clin Nutr.* **30**, 689–701 (2011).
19. Källén, K. Maternal smoking during pregnancy and infant head circumference at birth. *Early Hum. Dev.* **58**, 197–204 (2000).
20. Goldman, L. R. & Koduru, S. Chemicals in the environment and developmental toxicity to children: A public health and policy perspective. *Environ. Health Perspect.* **108**(Suppl 3), 443–448 (2000).
21. George, C. M. *et al.* Diarrhea prevalence and child growth faltering are associated with subsequent adverse child developmental outcomes in Bangladesh (CHoBI7 Program). *Am. J. Trop. Med. Hyg.* **106**, 233–238 (2021).
22. Jiang, S. *et al.* The determinants of growth failure in children under five in 25 low- and middle-income countries. *J. Glob. Health.* **13**, 04077 (2023).
23. Snetselaar, L. G., de Jesus, J. M., DeSilva, D. M. & Stoozy, E. E. Dietary Guidelines for Americans, 2020–2025: Understanding the scientific process, guidelines, and key recommendations. *Nutr. Today.* **56**, 287–295 (2021).
24. Anderson, J. W., Johnstone, B. M. & Remley, D. T. Breast-feeding and cognitive development: A meta-analysis. *Am. J. Clin. Nutr.* **70**, 525–535 (1999).
25. Andres, A. *et al.* Developmental status of 1-year-old infants fed breast milk, cow’s milk formula, or soy formula. *Pediatrics.* **129**, 1134–1140 (2012).
26. Ou, X. *et al.* Voxel-based morphometry and fMRI revealed differences in brain gray matter in breastfed and milk formula-fed children. *AJNR Am. J. Neuroradiol.* **37**, 713–719 (2016).
27. Deoni, S., Dean, D. 3rd., Joelson, S., O’Regan, J. & Schneider, N. Early nutrition influences developmental myelination and cognition in infants and young children. *Neuroimage.* **178**, 649–659 (2018).
28. de Weerth, C. *et al.* Human milk: From complex tailored nutrition to bioactive impact on child cognition and behavior. *Crit. Rev. Food Sci. Nutr.* **1**, 1–38 (2022).
29. Ghaemi, M. S. *et al.* Multiomics modeling of the immunome, transcriptome, microbiome, proteome and metabolome adaptations during human pregnancy. *Bioinformatics.* **35**, 95–103 (2019).
30. Kline, A. *et al.* Multimodal machine learning in precision health: A scoping review. *NPJ Digit. Med.* **5**, 1–14 (2022).
31. Lindley, A. A., Benson, J. E., Grimes, C., Cole, T. M. 3rd. & Herman, A. A. The relationship in neonates between clinically measured head circumference and brain volume estimated from head CT-scans. *Early Hum. Dev.* **56**, 17–29 (1999).
32. Qian, L. *et al.* Mendelian randomization suggests that head circumference, but not birth weight and length, associates with intelligence. *Brain Behav.* **11**, e02183 (2021).
33. Gale, C. R., O’Callaghan, F. J., Bredow, M. & Martyn, C. N. Avon longitudinal study of parents and children study team: The influence of head growth in fetal life, infancy, and childhood on intelligence at the ages of 4 and 8 years. *Pediatrics.* **118**, 1486–1492 (2006).
34. Bartholomeusz, H. H., Courchesne, E. & Karns, C. M. Relationship between head circumference and brain volume in healthy normal toddlers, children, and adults. *Neuropediatrics.* **33**, 239–241 (2002).
35. Subbarao, P. *et al.* The Canadian Healthy Infant Longitudinal Development (CHILd) Study: examining developmental origins of allergy and asthma. *Thorax.* **70**, 998–1000 (2015).
36. Moraes, T. J. *et al.* The Canadian healthy infant longitudinal development birth cohort study: biological samples and biobanking. *Paediatr. Perinat. Epidemiol.* **29**, 84–92 (2015).
37. Azad, M. B. *et al.* Human milk oligosaccharide concentrations are associated with multiple fixed and modifiable maternal characteristics, environmental factors, and feeding practices. *J. Nutr.* **148**, 1733–1742 (2018).
38. Miliku, K. *et al.* Human milk fatty acid composition is associated with dietary, genetic, sociodemographic, and environmental factors in the CHILd Cohort Study. *Am. J. Clin. Nutr.* **110**, 1370–1383 (2019).
39. Chan, D. *et al.* Adiponectin, leptin and insulin in breast milk: Associations with maternal characteristics and infant body composition in the first year of life. *Int. J. Obes.* **42**, 36–43 (2018).
40. Lauritzen, L., Hoppe, C., Straarup, E. M. & Michaelsen, K. F. Maternal fish oil supplementation in lactation and growth during the first 2.5 years of life. *Pediatr. Res.* **58**, 235–242 (2005).
41. Carlson, S. E. *et al.* DHA supplementation and pregnancy outcomes. *Am. J. Clin. Nutr.* **97**, 808–815 (2013).
42. Reyes, S. M. *et al.* Human milk micronutrients and child growth and body composition in the first 2 years: A systematic review. *Adv. Nutr.* <https://doi.org/10.1016/j.advnut.2023.06.005> (2023).
43. Aghaeepour, N. *et al.* An immune clock of human pregnancy. *Sci. Immunol. [Internet]*. **2**, 1. <https://doi.org/10.1126/sciimmunol.aan2946> (2017).

44. Stelzer, I. A. *et al.* Integrated trajectories of the maternal metabolome, proteome, and immunome predict labor onset. *Sci. Transl. Med. [Internet]*. **13**, 1. <https://doi.org/10.1126/scitranslmed.abd9898> (2021).
45. Marić, I. *et al.* Early prediction and longitudinal modeling of preeclampsia from multiomics. *Patterns (N Y)*. **3**, 100655 (2022).
46. Caly, H. *et al.* Machine learning analysis of pregnancy data enables early identification of a subpopulation of newborns with ASD. *Sci. Rep.* **11**, 6877 (2021).
47. Pammi, M., Aghaeepour, N. & Neu, J. Multiomics, artificial intelligence, and precision medicine in perinatology. *Pediatr. Res.* **93**, 308–315 (2023).
48. Nicolaou, L. *et al.* Factors associated with head circumference and indices of cognitive development in early childhood. *BMJ Glob. Health [Internet]*. **5**, 1. <https://doi.org/10.1136/bmjgh-2020-003427> (2020).
49. Sidhu, V. K., Huang, B. X. & Kim, H.-Y. Effects of docosahexaenoic acid on mouse brain synaptic plasma membrane proteome analyzed by mass spectrometry and (16)O/(18)O labeling. *J. Proteome Res.* **10**, 5472–5480 (2011).
50. Layé, S., Nadjar, A., Joffre, C. & Bazinet, R. P. Anti-inflammatory effects of omega-3 fatty acids in the brain: physiological mechanisms and relevance to pharmacology. *Pharmacol. Rev.* **70**, 12–38 (2018).
51. Simmer, K. Long-chain polyunsaturated fatty acid supplementation in preterm infants. *Cochrane Database Syst. Rev.* **1**, 375 (2000).
52. Jasani, B., Simmer, K., Patole, S. K. & Rao, S. C. Long chain polyunsaturated fatty acid supplementation in infants born at term. *Cochrane Database Syst. Rev.* **3**, CD000376 (2017).
53. Cheatham, C. L. Nutritional factors in fetal and infant brain development. *ANM.* **75**, 20–32 (2019).
54. Arijia, V. & Canals, J. Effect of maternal nutrition on cognitive function of children. *Nutrients [Internet]*. **13**, 1. <https://doi.org/10.3390/nu13051644> (2021).
55. Cortés-Albornoz, M. C., García-Guáqueta, D. P., Velez-van-Meerbeke, A. & Talero-Gutiérrez, C. Maternal nutrition and neurodevelopment: A scoping review. *Nutrients [Internet]*. **13**, 1. <https://doi.org/10.3390/nu13103530> (2021).
56. Selma-Royo, M., Calvo Lerma, J., Cortés-Macías, E. & Collado, M. C. Human milk microbiome: From actual knowledge to future perspective. *Semin. Perinatol.* **45**, 151450 (2021).
57. Becker, M. *et al.* Large-scale correlation network construction for unraveling the coordination of complex biological systems. *Nat. Comput. Sci.* **3**, 346–359 (2023).
58. Baltrusaitis, T., Ahuja, C. & Morency, L.-P. Multimodal machine learning: A survey and taxonomy. *IEEE Trans. Pattern Anal. Mach. Intell.* **41**, 423–443 (2019).
59. De Francesco, D. *et al.* Data-driven longitudinal characterization of neonatal health and morbidity. *Sci. Transl. Med.* **15**, 683 (2023). Available from: <https://www.science.org/doi/10.1126/scitranslmed.adc9854>
60. Azad, M. B., Atkinson, S. & Geddes, D. Collection and storage of human milk for macronutrient and macromolecule analysis—an overview. *Human Milk* **1**, 3–33 (2021).
61. Rodd, C. *et al.* World Health Organization growth standards: How do Canadian children measure up?. *Paediatr* **9854**. *Child Health.* **26**, e208–e214 (2021).
62. Berger, P. K. *et al.* Stability of human-milk oligosaccharide concentrations over 1 week of lactation and over 6 hours following a standard meal. *J. Nutr.* **152**, 2727–2733 (2023).
63. Patterson, R. E. *et al.* Measurement characteristics of the Women's Health Initiative food frequency questionnaire. *Ann. Epidemiol.* **9**, 178–187 (1999).
64. Guenther, P. M. *et al.* Update of the healthy eating index: HEI-2010. *J. Acad. Nutr. Diet.* **113**, 569–580 (2013).
65. Hoerl, A. E. & Kennard, R. W. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics.* **12**, 55–67 (1970).
66. Smola, A. J. & Schölkopf, B. A tutorial on support vector regression. *Stat. Comput.* **14**, 199–222 (2004).
67. Garreta, R., & Moncecchi, G. Learning scikit-learn: Machine Learning in Python (Packt Publishing Ltd, 2013).
68. Van der Maaten, L., & Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res. [Internet]*. **9**, 2579–2605 (2008). Available from: <https://www.jmlr.org/papers/volume9/vandermaaten08a/vandermaaten08a.pdf?fbclid>

Author contributions

Based on the CRediT model, M.B., M.A., and N.A. were responsible for conceptualization; M.B., K.F., and S.G. for data curation; M.B. for formal analysis; M.A. and N.A. for funding acquisition; M.B. for investigation; M.B. and N.A. for methodology development and design; M.B., and M.A., N.A. for project administration; N.A. and M.A. for providing resources; M.B. for software; N.A. and M.A. for supervision; M.B. for validation; M.B. for visualization; M.B. and M.A. for writing the original draft; and all authors for reviewing and editing. P.S., T.M., S.T., P.M., and E.S. oversaw participant recruitment and data collection for the CHILD Cohort Study. NR managed operations for the IMiC Consortium. L.B., B.R., and C.Y. generated HMO data. J.M. and B.D. generated milk immunomodulator data. C.F. generated, and K.M. curated milk fatty acid data.

Funding

This secondary analysis of CHILD Cohort Study data, through the International Milk Composition (IMiC) Consortium, was funded by the Bill and Melinda Gates Foundation (INV-001734). Funding for the CHILD Cohort Study and analysis of human milk samples was provided by: National Institutes of Health (R35GM138353), Bundesministerium für Bildung und Forschung (01IS22077), Canadian Institutes of Health Research, AllerGen Network of Centers of Excellence, Research Manitoba, Children's Hospital Research Institute of Manitoba, Canadian Respiratory Research Network, Manitoba Medical Services Foundation, Canada Research Chairs Program, Alfred E. Mann Foundation, Don and Debbie Morrison, SickKids Foundation.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-52323-w>.

Correspondence and requests for materials should be addressed to N.A. or M.B.A.

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

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



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

© The Author(s) 2024