

Human milk oligosaccharides are associated with maternal genetics and respiratory health of human milk-fed children

Received: 27 May 2022

Accepted: 14 August 2024

Published online: 04 September 2024

 Check for updates

Amirthagowri Ambalavanan¹, Le Chang ^{1,2}, Jihoon Choi¹, Yang Zhang², Sara A. Stickley ¹, Zhi Y. Fang¹, Kozeta Miliku^{3,4}, Bianca Robertson⁵, Chloe Yonemitsu⁵, Stuart E. Turvey⁶, Piushkumar J. Mandhane^{7,8}, Elinor Simons⁹, Theo J. Moraes¹⁰, Sonia S. Anand ¹¹, Guillaume Paré ¹², Janet E. Williams ¹³, Brenda M. Murdoch ¹³, Gloria E. Otoo¹⁴, Samwel Mbugua¹⁵, Elizabeth W. Kamau-Mbuthia¹⁵, Egidioh W. Kamundia¹⁵, Debela K. Gindola¹⁶, Juan M. Rodriguez¹⁷, Rossina G. Pareja¹⁸, Daniel W. Sellen¹⁹, Sophie E. Moore^{20,21}, Andrew M. Prentice²¹, James A. Foster²², Linda J. Kvist²³, Holly L. Neibergs²⁴, Mark A. McGuire¹³, Michelle K. McGuire ²⁵, Courtney L. Meehan²⁶, Malcolm R. Sears³, Padmaja Subbarao ^{3,10}, Meghan B. Azad ^{9,27,28} , Lars Bode ^{5,28}  & Qingling Duan ^{1,2,28} 

Breastfeeding provides many health benefits, but its impact on respiratory health remains unclear. This study addresses the complex and dynamic nature of the mother-milk-infant triad by investigating maternal genomic factors regulating human milk oligosaccharides (HMOs), and their associations with respiratory health among human milk-fed infants. Nineteen HMOs are quantified from 980 mothers of the CHILD Cohort Study. Genome-wide association studies identify HMO-associated loci on chromosome 19p13.3 and 19q13.33 (lowest $P = 2.4e-118$), spanning several fucosyltransferase (*FUT*) genes. We identify novel associations on chromosome 3q27.3 for 6'-sialyllactose ($P = 2.2e-9$) in the sialyltransferase (*ST6GAL1*) gene. These, plus additional associations on chromosomes 7q21.32, 7q31.32 and 13q33.3, are replicated in the independent INSPIRE Cohort. Moreover, gene-environment interaction analyses suggest that fucosylated HMOs may modulate overall risk of recurrent wheeze among preschoolers with variable genetic risk scores ($P < 0.01$). Thus, we report novel genetic factors associated with HMOs, some of which may protect the respiratory health of children.

Breastfeeding and consumption of human milk have many health benefits for infants such as a lower prevalence of infections and childhood obesity¹. It is unclear, however, whether breastfeeding protects against respiratory health outcomes such as asthma, which affects over 339 million individuals worldwide and is the most common chronic disease among children². A meta-analysis reported

protective effects of breastfeeding on asthma and wheeze³, but other studies reported conflicting results or no associations².

A major limitation of earlier research is that breastfeeding and human milk were typically considered as a single homogeneous exposure. This fails to acknowledge the complex and dynamic interactions of the mother-milk-infant “triad” as a coadapted system where

variations in each component could influence the others⁴. For example, human milk contains a myriad of bioactive components, which vary among women and change over the course of lactation in response to environmental stimuli and the changing requirements of the infant^{5,6}.

Human milk oligosaccharides (HMOs) constitute the third largest solid component of human milk (approximately 15 g/L), after lactose and lipids⁷, but most are absent from cow milk and infant formulas⁸. It has been suggested that HMOs modulate epithelial and immune cell responses of infants by influencing the microbiota and through direct interactions with pathogens and host immune cells³. Furthermore, HMOs were previously associated with early life infections in offspring⁹. More than 90% of total HMOs are made up of about 20 oligosaccharides, which vary considerably in concentration and composition among individuals⁸. This variability is known to be partially determined by maternal polymorphisms in the fucosyltransferase 2 gene (*FUT2*)¹⁰, which affects fucosylation of HMOs and other glycans¹¹. Little is known about other maternal genes that regulate the biosynthesis of HMOs among lactating mothers. Thus, the overall genomic regulation of HMOs and its potential influence on the health of human milk-fed infants remain poorly understood.

In this study, we investigated potential mechanisms underlying the complex and dynamic mother-milk-infant triad and explored how variations in this coadapted system may affect the respiratory health of human milk-fed children. Specifically, we conducted genome-wide association studies (GWASs) of HMO concentrations to evaluate maternal genomic regulation of HMO composition in human milk. Furthermore, we determined if exposure to variable HMO levels is associated with recurrent wheeze during early childhood among milk-fed children and dependent on their individual genetic risk. This is the first nuanced study of HMOs to examine their genetic regulation and potential impact on respiratory outcomes in children.

Results

Study design and participants

An overview of our experimental design is depicted in Supplementary Fig. S1. Of the 1206 mothers with HMO profiles in the CHILD Cohort Study, 980 had been genotyped and included in our GWASs of HMOs (names and abbreviations in Supplementary Table 1; distributions in Supplementary Figs. S2, S3; clustering in Supplementary Fig. S4). Characteristics of these 980 mothers are summarized and compared to 2370 mothers with genetics data only in Supplementary Table 2. The characteristics of their children are shown in Supplementary Table 3. Of the 980 mother-infant dyads, only 880 children were fed with human milk from their mothers until 6 months or longer. Therefore, these 880 kids were included in the gene-environment interaction (GxE) analyses that integrated their individual genetic risk and their mothers' HMO measures as the exposure variables. GWAS analyses including over 5.4 M SNPs identified both known and novel maternal genomic factors associated HMOs. Subsequent GxE analyses determined that exposure to specific HMOs (e.g., 2'FL, DFLNH) were associated with lower prevalence of recurrent wheeze among milk-fed infants, particularly among children with high genetic risk (calculated based on SNPs previously associated with asthma¹²).

GWASs of maternal HMOs identified known and novel genetic associations

In our GWAS of secretor status, 727 of the 980 mothers (73.7%) with both genomics and HMO profiles were classified as secretors, based on high abundance of 2'-fucosylactose (2'FL) in their milk (>200 nmol/mL). GWAS of this binary trait identified 265 SNP associations on chromosome 19q13.33, which harbors the *FUT2* gene (significant SNPs listed in Supplementary Data 1). These represent both novel and known associations, including two stop-gain variants (rs601338, rs681343 $P = 9.14e-51$, $\beta = 0.004$) previously associated with secretor

status in Caucasians, and a third missense variant (rs1047781), previously reported in East Asians ($P = 1.13e-10$, $\beta = 0.06$)¹³.

In our GWASs of individual and total HMO abundances, we identified 411 SNPs associated with continuous concentrations of the 19 individual HMOs, total HMOs, HMO-bound fucose and HMO-bound sialic acid. GWAS results were overlaid in Fig. 1a and shown individually for each HMO in Supplementary Fig. S5. SNP associations are listed in Supplementary Data 1, independent SNPs ($r^2 \leq 0.6$) included in Supplementary Data 2, and 23 lead SNPs ($r^2 \leq 0.1$) shown in Table 1, Supplementary Fig. S6 and annotated in Supplementary Data 3. These associated SNPs span 4 genomic regions on chromosomes 3, 10, and 19p and 19q, as described below. GWASs limited to central European participants showed similar association signals on chromosomes 3 and 19 (see Supplementary Fig. S7 and Supplementary Data 4). GWAS of HMOs stratified by secretor status identified similar associations on chromosome 19 (Supplementary Fig. S8).

We identified 66 independent ($r^2 < 0.6$) SNP associations at the chromosome 19q13.33 genomic region, which span multiple genes including *FUT2*, *FUT1*, *NTNS*, *CA11*, *MAMSTR*, *RASIP1*, *IZUMO1* (Fig. 1c). This region was associated with 18 of the 19 HMOs (all except 6'SL). These SNPs were generally associated with decreased concentrations of total HMOs and HMO-bound fucose; and increased concentrations of HMO-bound sialic acid (Table 1, Supplementary Data 1 and 2). The most significant association was a synonymous SNP (rs492602, $P = 2.38e-118$, $\beta = -0.96$) in the *FUT2* gene, correlated with decreased concentrations of LNFP1. Although *FUT2* encodes for an alpha1-2-fucosyltransferase, SNPs in this gene were not only associated with decreased concentrations of alpha1-2-fucosylated HMOs (e.g., 2'FL and LNFP1), but also with LSTc, 3'SL (Fig. 1d, Supplementary Data 1, 2). Notably, non-fucosylated precursors such as LNT as well as alternative products (LNFP2, LNFP3, LSTb and LSTc) were also associated with *FUT2* SNPs in both negative and positive directions; these observations align with the known biosynthetic interconnectivity of individual HMOs in vivo (Fig. 1d). This interconnectivity is further illustrated by visualizing the impact of *FUT2* secretor status on overall HMO composition (reflecting all 19 HMOs) using a PCoA analysis, where each milk sample was plotted in a three-dimensional space. For example, samples from women with the *FUT2* SNP rs601338 clustered in a significantly different part of the PCoA space with separation along the prime axis, which accounted for over 56% of the variation in overall HMO profiles (Fig. 1e).

We identified 26 independent association signals in chromosome 19p13.3, which consists of genes *FUT3*, *FUT6*, and *NRTN* (Fig. 2a). The most significant association was for SNP rs708686 ($P = 1.46e-58$, $\beta = -0.75$) in *FUT6*, associated with a decrease in DFLNT and LNFP2 (Table 1). In contrast to SNPs on chromosome 19q13.33 above, SNPs in *FUT3* (e.g., rs28362459) and neighboring genes were associated with higher concentrations of 2'FL and LNFP1 (Fig. 2d, e, Supplementary Data 1 and 2). Interestingly, while the concentration of HMO-bound fucose was not affected by this *FUT3* SNP, that of HMO-bound sialic acid was significantly lower among homozygotes ($P < 0.0001$) (Fig. 2b). Specifically, this SNP was associated with lower 3'SL ($P < 0.0001$) but higher 6'SL concentrations ($P < 0.05$) (Supplementary Data 1, Fig. 2c); however, none of the extended sialylated HMOs such as LSTb, LSTc, DSLNT or DSLNH were affected by this SNP (Supplementary Data 1). Moreover, the alpha1-3-fucosylated HMO 3FL was not significantly associated with this *FUT3* SNP, but DFLac, which is both alpha1-2- and alpha1-3-fucosylated, was significantly reduced among homozygotes ($P < 0.001$) (Fig. 2d). Extended fucosylated HMOs were also affected by this *FUT3* SNP: concentrations of the type-1 alpha1-4-fucosylated HMO LNFP2 were lower ($P < 0.0001$), whereas concentrations of the alpha1-2-fucosylated structural isomer LNFP1 were higher ($P < 0.0001$) and there was no association with the type 2 isomer, the alpha1-3-fucosylated HMO LNFP3 (Fig. 2e, Supplementary Data 1). These findings further emphasize the interconnectivity of HMO biosynthesis,

Table 1 | Lead SNPs associated with 23 human milk oligosaccharide (HMO) features: secretor status, total HMO concentration, HMO-bound fucose, HMO-bound sialic acid, and 19 individual HMOs in 980 lactating mothers

Locus	Lead SNP	P value of lead SNP	Effect allele frequency	Nearst gene	HMO feature(s) lower with effect allele	HMO feature(s) higher with effect allele
3q27.3	rs4686843	2.15E-09	0.19	ST6GAL1		6'SL
10q22.1	rs3740444	4.83E-08	0.23	PALD1	LSTb	
19p13.3	rs8102202	4.31E-12	0.13	NRTN	DFLNT, LNH	
19p13.3	rs112742683	5.25E-18	0.07	NRTN	Sialylated , FDSLNH, LNFP2	
19p13.3	rs3859563	7.52E-26	0.31	FUT6	DFLNT, LNFP2, LNH	DFLNH, LNFP1
19p13.3	rs708686	1.24E-58	0.32	FUT6	Sialylated , 3'SL, DFLNH, DFLNT, FDSLNH, LNFP1, LNFP2, LNH	2'FL, LNT
19p13.3	rs2608894	1.27E-48	0.25	FUT3	Sialylated , 3'SL, DFLac, DFLNT, FDSLNH, LNFP2, LNH	2'FL, DFLNH, LNFP1
19q13.33	rs62130305	3.49E-13	0.2	SULT2B1	FDSLNH, LNFP2	Non-Sec, Fucosylated , 2'FL, DFLac, LNFP1
19q13.33	rs2302949	1.74E-11	0.07	SULT2B1		Fucosylated , 2'FL, LNFP1
19q13.33	rs41949455	2.61E-17	0.38	SULT2B1	LNFP2	Non-Sec, Fucosylated , 2'FL, 3FL, DFLac, DFLNT, LNFP1
19q13.33	rs2353017	2.43E-33	0.46	SEC1P	Fucosylated , 2'FL, 3FL, DFLac, DFLNH, DFLNT, LNFP1	Non-Sec, FDSLNH, LNFP2
19q13.33	rs8103121	4.06E-20	0.31	SEC1P	LNFP2, FDSLNH	Non-Sec, Fucosylated , 2'FL, 3FL, DFLac, DFLNH, LNFP1
19q13.33	rs62130341	1.22E-09	0.11	NTN5		Non-Sec, Fucosylated , 2'FL
19q13.33	rs3826846	9.23E-40	0.38	NTN5	Sialylated , FDSLNH, LNFP2, LNFP3, LNT, LSTb	Non-Sec, Total, Fucosylated , 2'FL, 3FL, DFLac, DFLNH, DFLNT, LNFP1, LSTc
19q13.33	rs35228311	3.78E-12	0.1	NTN5	FDSLNH, LNFP2	Non-Sec, Total, Fucosylated , 2'FL, DFLac, LNFP1
19q13.33	rs4926202	2.38E-118	0.41	FUT2	Total, Fucosylated , 2'FL, 3FL, 3'SL, DFLac, DFLNH, DFLNT, FLNH, LNFP1, LSTc	Non-Sec, Sialylated , DSLNH, FDSLNH, LNT, LNFP2, LNFP3, LSTb
19q13.33	rs281385	2.66E-13	0.15	MAMSTR	FDSLNH, LNFP2, LSTb	Non-Sec, Total, Fucosylated , 2'FL, 3FL, DFLac, LNFP1
19q13.33	rs62132778	3.28E-08	0.11	RASIP1		2'FL
19q13.33	rs28400012	4.65E-15	0.22	FUT1	FDSLNH, LNFP2	Non-Sec, Total, Fucosylated , 2'FL, LNFP1
19q13.33	rs838131	1.49E-21	0.49	FGF21	Total, Fucosylated , 2'FL, LNFP1	Non-Sec, DFLac, FDSLNH, LNFP2
19q13.33	rs669789	1.56E-16	0.35	FGF21	LNFP2, LSTb	Non-Sec, Total, Fucosylated , 2'FL, DFLac, LNFP1
19q13.33	rs4801776	3.76E-12	0.22	BCAT2	Total, Fucosylated , 2'FL, LNFP1	Non-Sec, FDSLNH, LNFP2
19q13.33	rs55973465	4.17E-09	0.13	PLEKH4A	2'FL	

Lead SNPs are significant SNPs identified by LD clumping ($r^2 < 0.1$, 250Kb window). Chromosomal positions are mapped to the GRCh37 genome. Association of the effect allele with HMO summary features (**bold**) are listed first, followed by individual HMOs (in alphabetical order). Non-Sec indicates Non-Secretors. See Supplementary Data 4 for additional information about gene functions and associated phenotypes.

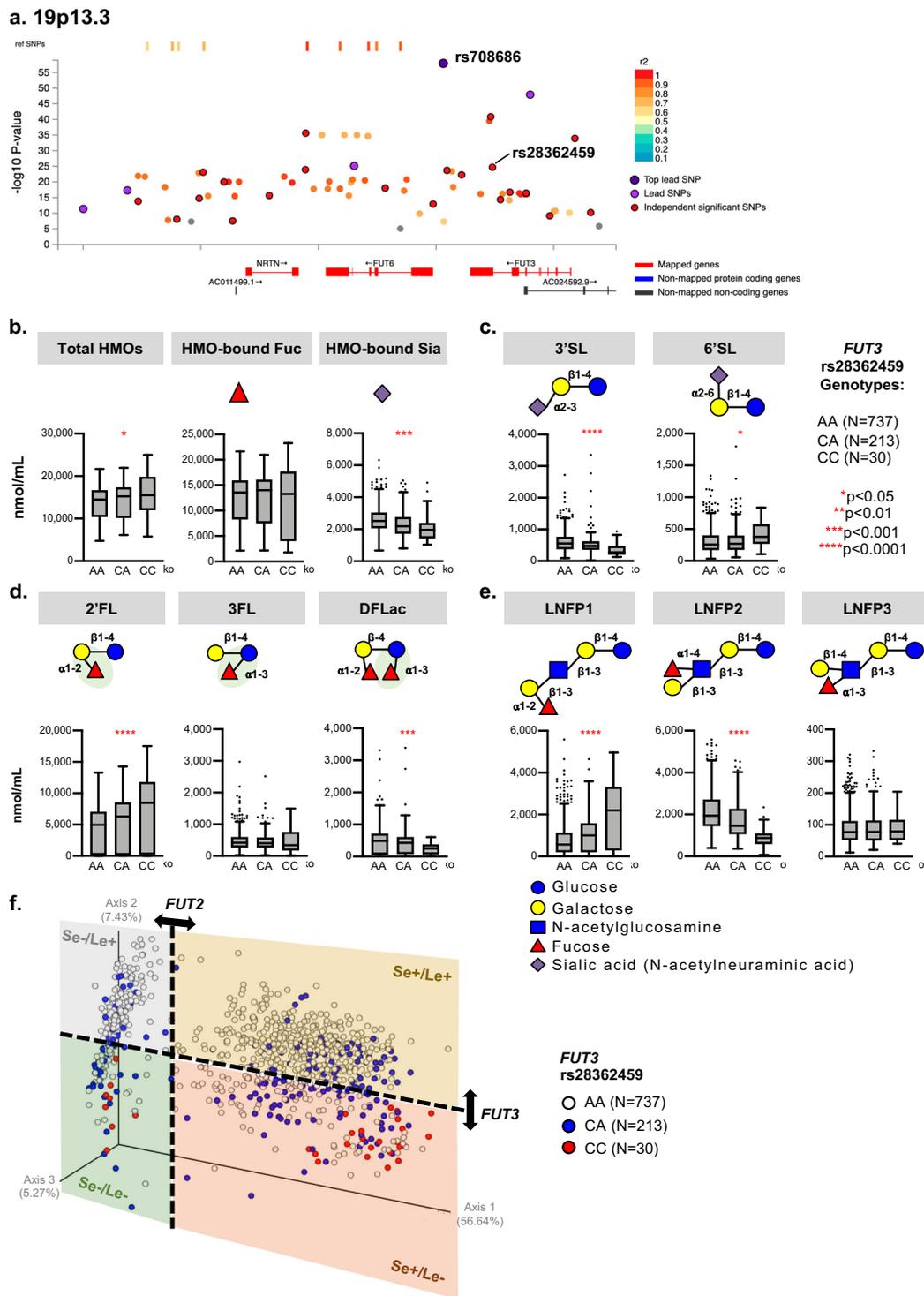


Fig. 2 | SNPs on chromosome 19p13.3 were associated with HMOs among 980 lactating mothers of the CHILD Cohort Study. **a** Overlaid Locus Zoom plots of chromosome 19p13.3: 70 SNPs were associated with individual HMOs (linear regression analysis in PLINK; 26 independent signals and 5 lead SNPs, defined by (LD) $r^2 < 0.6$ and $r^2 < 0.1$, respectively). The most significant SNP (rs708686, $P = 1.46e-58$), located in the *FUT6* gene, is represented by a dark purple dot. The x-axis identifies the genes mapped to this associated region and the y-axis indicates the significance of SNP associations ($P < 5e-8$). **b–e** HMO concentrations associated with rs28362459 in the *FUT3* gene using Kruskal-Wallis (non-parametric) tests. Box minimum: Q1, box maximum: Q3, box center: median, whiskers (farthest points that are not outliers (i.e., within 3/2 times of interquartile range)). **b** total HMO concentrations, $P = 0.006$; HMO-bound fucose, $P = 0.12$; and HMO-bound sialic acid, $P = 5.55e-6$; **c**) differently sialylated lactose HMOs: 3'SL, $P = 2.2e-11$; 6'SL,

$P = 0.68$; **d**) differently fucosylated lactose HMOs: 2'FL, $P = 9.83e-6$; 3FL, $P = 0.078$; DFLac, $P = 4.21e-4$; and **e**) different fucosylated pentasaccharides: LNFP1, $P = 1e-12$; LNFP2, $P = 1e-12$; LNFP3; $P = 0.6468$. Box minimum: Q1, box maximum: Q3, box center: median, whiskers (farthest points that are not outliers (i.e., within 3/2 times of interquartile range)). **f** Principal Co-ordinate Analysis (PCoA) plot of overall HMO profiles by the *FUT3* rs28362459 genotype using a Bray-Curtis distance matrix. Each dot represents the entire HMO profile of an individual mother. Variation along the secondary axis, which explains 7.43% of variation in overall HMO profiles, is partly driven by *FUT3* rs28362459 genotype. Variation along the primary axis is driven by *FUT2* secretor status (see Fig. 1e), dividing mothers into 4 groups based on *FUT2* Secretor (Se) and *FUT3* Lewis (Le) genotypes. Dashed lines are drawn arbitrarily to show the sectors with high or low numbers of Secretor (Se) and Lewis (Le) genotypes.

illustrating multiple alternate fucosylation or sialylation mechanisms, which contribute to the overall HMO balance (Fig. 2f).

Chromosome 3q27.3 consists of four independent SNPs associated with elevated concentrations of 6'SL ($3.92e-9 < P < 2.07e-8$), indicating gain-of-function mutations (Fig. 3a). The most significant was SNP rs4686843 (lead SNP), located within the intron of the β -galactoside alpha2-6-sialyltransferase 1 (*ST6GALI*) gene (Fig. 3b). This SNP was associated with higher concentrations of HMOs carrying an alpha2-6-linked sialic acid at the terminal galactose such as 6'SL and LSTc (Fig. 3b). This SNP was *not* associated with concentrations of 3'SL, indicating the enzyme only facilitates alpha2-6-, but not alpha2-3-sialylation. In addition, it was *not* associated with LSTb concentrations, indicating that the enzyme specifically catalyzes alpha2-6-sialylation to the terminal galactose, but not the internal N-acetyl-glucosamine (Fig. 3b). Figure 3c depicts associations with other sialylated HMOs such as DSLNH and LSTc ($P = 8.22e-8$, $\beta = 0.26$). Unlike the chromosome 19 SNPs, represented as clusters in the PCoA space, SNPs in *ST6GALI* were associated with specific sialylated HMOs (alpha2-6-linked sialic acid at the terminal galactose) but did not substantially drive overall HMO composition (Fig. 3d).

Chromosome 10q22.1 is a novel locus associated with lower LSTb concentrations ($P = 4e-8$, $\beta = -0.29$) and located within the phosphatase domain containing paladin 1 (*PALDI*) gene (Fig. 3e).

Conditional GWAS of HMOs identified novel associations

Conditional GWAS analyses using the 21 lead SNPs on chromosome 19 (Table 1) as covariates identified 28 additional novel SNPs on chromosomes 2q37.1, 7q31.32, 16p13.2 and 18q22.3 as well as independent associations in 3q27.3, 19p13.3 and 19q13.33 (Fig. 4a and Supplementary Data 5). For example, in the 2q37.1 region, SNP rs13025087 in gene *B3GNT7* was associated with increased concentration of 3'SL ($P = 4.64e-08$, $\beta = 0.26$). In region 7q31.32, 19 SNPs within the same LD block were associated with increased concentrations of LNT, including an intergenic SNP, rs1881374 ($P = 6.54e-10$, $\beta = 0.29$). In 16p13.2, 7 SNPs within the same LD block were associated with increased concentrations of fucose-bound HMOs (e.g., rs4578629 an intergenic variant, $P = 1.29e-08$, $\beta = 0.35$). Finally, we identified a SNP in 18q22.3, rs73472295, which was associated with decreased concentrations of DSLNT ($P = 4.06e-08$, $\beta = -0.42$).

Subsequent conditional GWAS analyses using the top lead SNP rs492602 on chromosome 19q13.3 identified 3 new significant loci on chromosomes 8q24.13, 17p13.1 and 22q12.3. These loci were not detected in our initial GWASs of HMOs, nor were they significant when we conditioned on all 21 lead SNPs (Fig. 4b). In the 8q24.13 region, rs2954165 was significantly associated with increased concentration of 3'FL ($P = 4.50e-08$, $\beta = 0.24$). In the 17p13.1 region, rs7209048 was significantly associated with decreased concentration of DSLNT ($P = 3.88e-08$, $\beta = -0.36$). In the 22q12.3 region, 5 SNPs within the same LD block were associated with increased concentrations of FLNH ($P = 5.66e-09$, $\beta = 0.28$). In addition, we detected significant signals on chromosomes 19p13.3, 19q13.33, 3q27.3 and 7q31.32, which had been identified in our previous conditional analysis using all lead SNPs.

In addition, conditioning on the lead SNP with the lowest P value (rs708686, $P = 1.24e-58$) on chromosome 19p13.3 resulted in one novel genome-wide significant signal at chromosome 3p24.3 as well as a signal at 18q22.3, which was identified in our earlier conditional analysis using all lead SNPs (Fig. 4c, Supplementary Data 5, Supplementary Fig. S9a). In the 3p24.3 region, rs4858536 was associated with increased concentration of DSLNH ($P = 3.27e-08$, $\beta = 0.24$). In the 18q22.3 region, rs73472295 was associated with decreased concentration of DSLNT ($P = 3.28e-08$, $\beta = -0.42$). Moreover, two associations at chromosomes 17p13.1 and 22q12.3 were identified, which were noted above in conditional analysis of the top SNP at chromosome 19q13.33.

Conditional GWAS analysis using each of the 21 lead SNPs from chromosome 19 (Table 1) yielded similar results as described above (Supplementary Data 5, Supplementary Fig. S9a). Similarly, step-wise conditional analysis using the top one, two, three and more SNPs in each locus did not detect any additional signals (Supplementary Fig. S9b).

HMO-associated loci replicated in the INSPIRE Study

To validate our results from the CHILD Cohort Study, we undertook targeted replication analyses in a sub-cohort of 395 mothers from the INSPIRE Study¹⁰. This replication sub-cohort used the same HMO analysis platform as CHILD, paired with a different genomics platform that was not imputed, resulting in relatively fewer SNP genotypes for analysis. Of the 2,669 SNPs associated with HMOs (suggestive $P < 1e-5$) in CHILD, 281 were genotyped in INSPIRE, of which 46 replicated including: one SNP in 3q27.3 (rs4234598 in the *ST6GALI* gene, $P < 0.05$) (Fig. 5a), four SNPs in the 7q31.32 region ($P < 0.0125$) (Fig. 5b), and 41 HMO-associated SNPs from the 19p13.3 and 19q13.33 regions, which span the *FUT2*, *FUT3*, and *FUT6* genes ($P < 0.0014$) (Fig. 5c).

Using a meta-analysis approach, we combined the association results from all 281 overlapping SNPs in the CHILD and INSPIRE sub-cohorts (total $N = 1375$). This identified 6 loci associated with HMOs ($P < 5e-08$) including chromosomes 3q27.3, 7q21.32, 7q31.32, 13q33.3, 19p13.3 and 19q13.33 (Fig. 5d and Supplementary Data 6). Two of these regions were not significantly associated with HMOs in either the CHILD or INSPIRE cohort studies alone, but only identified in the meta-analysis combining the two sub-cohorts. These include chromosome 7q21.32, which consists of an intronic SNP in *LMTK2* gene, rs16869462, associated with decreased concentrations of a neutral HMO LNH ($P = 9e-09$, $\beta = -5.75$). The second novel locus is on chromosome 13q33.3, which includes SNP rs79783730, significantly associated with increased concentrations of LNT ($2.21e-08$, $\beta = 5.59$).

Gene-by-environment interaction analyses identified specific HMOs that may protect respiratory health of human milk-fed children

Of the 980 mothers with both genetic and HMO profiles, 198 (20%) of their children had experienced recurrent wheeze between 2 and 5 years of age. Recurrent wheeze was strongly associated with two genetic risk scores (GRS): one generated based on 44 SNPs associated with asthma, regardless of age of onset¹² (denoted as *all-asthma* GRS, Supplementary Fig. S10), and the second based on 4 SNPs associated with childhood-onset asthma (denoted as *childhood-asthma* GRS). As shown in Fig. 6a, both GRSs were strongly associated with recurrent wheeze prevalence ($P = 2.48e-10$ and $P = 1e-7$).

In GxE analyses using the *all-asthma* GRS and restricting to children who were human milk-fed for at least 6 months ($N = 880$), we determined that exposure to specific HMOs was associated with prevalence of recurrent wheeze, particularly among those with high genetic risk (Fig. 6b, c, Supplementary Table 4). For example, among children with high GRS, exposure to high concentrations of certain HMOs (2'FL, DFLNH – which are higher in secretor milk) as well as total HMOs and total HMO-bound fucose was associated with reduced prevalence of recurrent wheeze (Fig. 6b) (interaction $P < 0.01$). Notably in this high-GRS group, exposure to high concentrations of certain other HMOs (LNFP2, LNFP3 – which are higher in non-secretor milk) were associated with higher prevalence of recurrent wheeze. Thus, these interactions suggest that among individuals with high genetic risk, exposure to specific HMOs modulate their overall risk of recurrent wheeze. We observed similar trends using the *childhood-asthma* GRS (Supplementary Fig. S11). These results contrast with GxE analyses using breastfeeding duration (3, 6, and 12 months) instead of HMO concentrations, which do not show significant modulating effects (Supplementary Fig. S12).

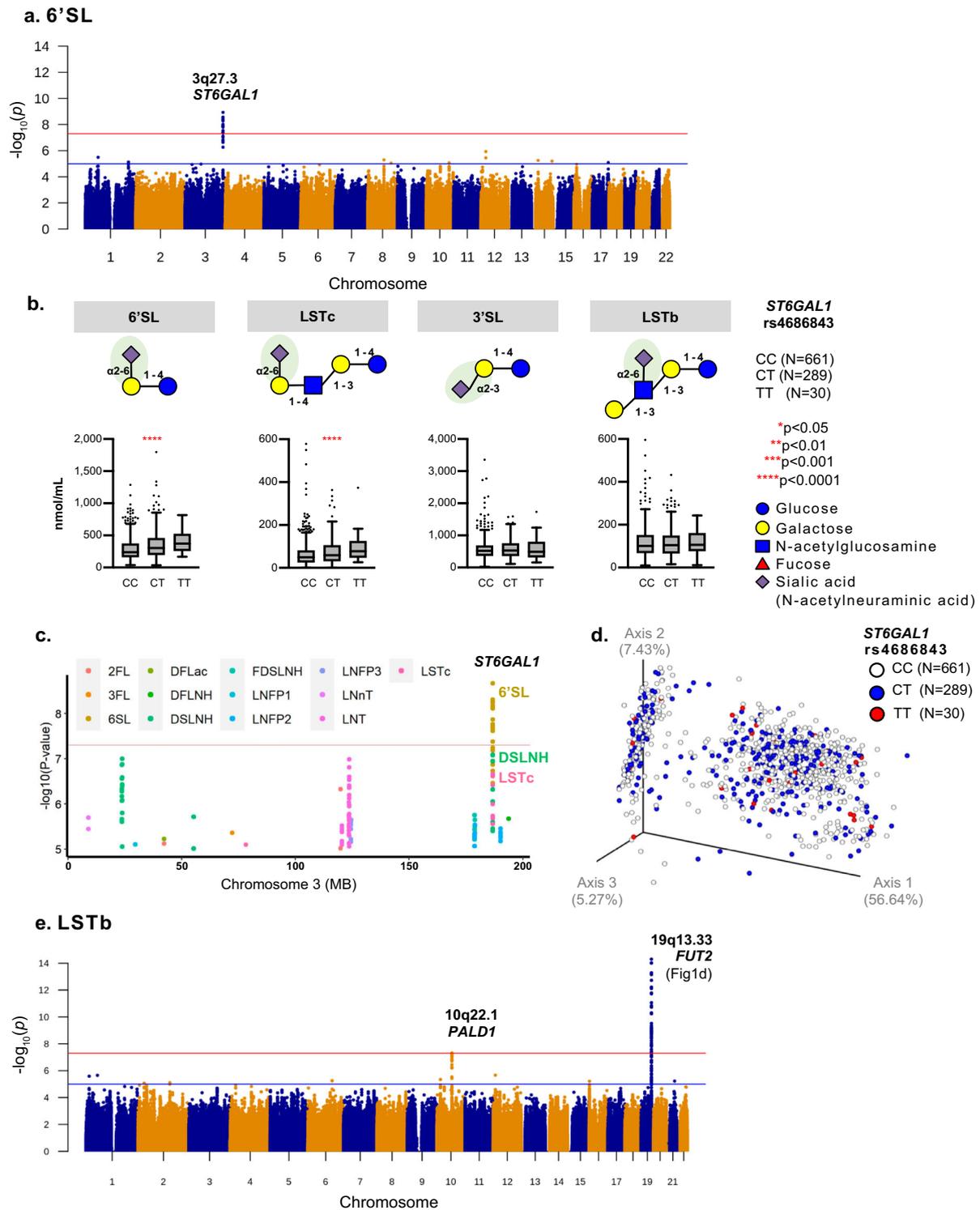


Fig. 3 | Novel HMO-associated loci in *ST6GAL1* and *PALD1* genes among lactating mothers of the CHILD Cohort Study. **a** Manhattan plot of GWAS for 6'SL: 19 SNPs on chromosome 3q27.3 in the *ST6GAL1* gene were significantly associated with 6'SL ($P < 5e-8$). The most significant lead SNP was rs4686843 located in the intron of this gene. The x-axis indicates the chromosomal position and the y-axis indicates the significance of SNP associations ($-\log_{10}(P)$). **b** HMO concentrations associated with rs4686843 using the Kruskal-Wallis test (non-parametric, not assuming normal distribution): SNP associated with higher concentrations of 6'SL ($P = 5.78e-9$) and LSTc ($P = 6.29e-6$), but not 3'SL ($P = 0.98$) or LSTb ($P = 0.98$), indicating role in alpha2-6-sialylation specific at the terminal galactose. Box minimum: Q1, box maximum: Q3, box center: median, whiskers (farthest points that are not outliers (i.e., within 3/2 times of interquartile range)). **c** Regional plot of genetic associations for 19 individual

HMOs within chromosome 3q27.3: Significant associations detected only for 6'SL. Trends detected for DSLNH and LSTc but these do not meet genome-wide significance as indicated by the red horizontal line ($P = 5e-8$). Blue line is indicated by the suggestive significance ($P = 1e-5$). **d** Principal Co-ordinate Analysis (PCoA) plot of overall HMO profiles by rs4686843 using a Bray-Curtis distance matrix: No evidence of substantial association between this SNP and overall HMO profile. Each dot represents the entire HMO profile of an individual mother. **e** Manhattan plot of LSTb: 135 SNPs on chromosome 19q13.33 (habouring genes such as *FUT2*, lowest $P = 2.4e-31$) and a single SNP on chromosome 10q22.1 in the *PALD1* locus ($P < 5e-8$) as indicated by the red line and were significantly associated with LSTb concentration. Blue line indicates the suggestive significance ($P = 1e-5$).

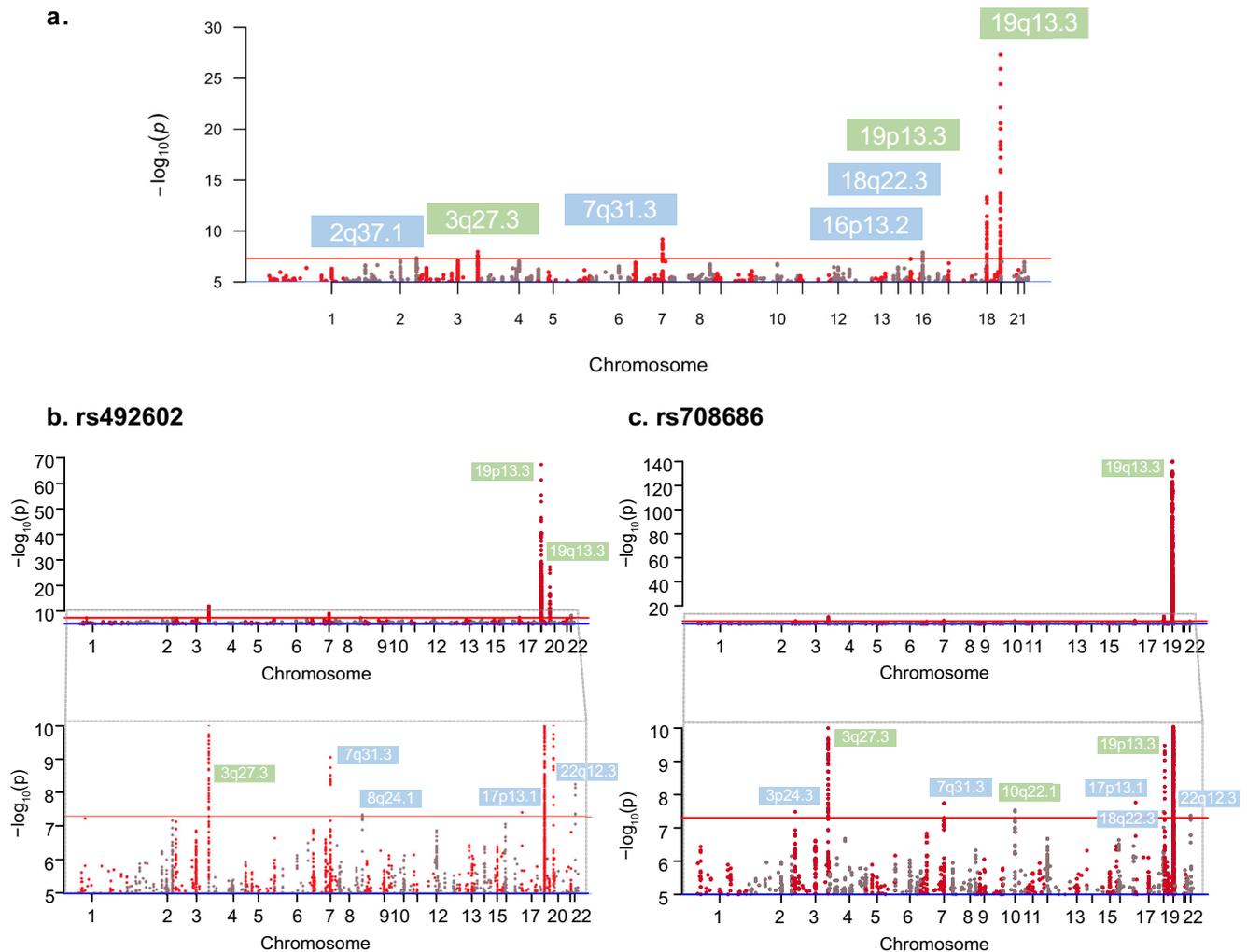


Fig. 4 | Genome-wide conditional analyses of HMOs. a Overlaid Manhattan plots from conditional GWAS analyses of 19 HMOs using linear regression to condition on 21 lead SNPs from Table 1 ($N=980$). Loci highlighted in blue were novel associated loci identified by conditioning on 21 lead SNPs from chromosome 19. Loci highlighted in green are primary loci previously identified from our GWAS. Red line represents the genome-wide significance ($P=5e-8$). **b** Overlaid Manhattan plots from conditional GWAS analyses on the lead SNP (rs492602, $P=2.38e-118$) with the lowest P value at the chromosome 19q13.33 locus, which identified 3 new significant loci on chromosomes 8q24.13, 17p13.1 and 22q12.3 (highlighted in blue). Loci highlighted in green are primary loci previously identified from our GWAS. Red

line represents the genome-wide significance threshold ($P=5e-8$). Zoomed panel shows the loci within significance between $P < 1e-05$ and $P < 5e-10$. **c** Overlaid Manhattan plots from conditional GWAS analyses on the lead SNP (rs708686, $P=1.24e-58$) with the lowest P value at the chromosome on 19p13.3 locus, which identified 4 new significant loci on chromosome 3p24.3, 17p13.1, 18q22.3 and 22q12.3 (highlighted in blue). Loci highlighted in green are primary loci previously identified from our GWAS. Red line represents the genome-wide significance ($P=5e-8$). Zoomed panel shows the loci within significance between $P < 1e-05$ and $P < 5e-10$.

Finally, we performed a sensitivity analysis to determine if these findings were driven by secretor status (which affects glycosylation throughout the entire body – not only in milk) rather than specific HMO exposures. When restricting our analyses to secretor mothers only ($N=640$ with children who were human milk-fed until 6 months or longer), we observed similar trends between the GRS and HMOs (Fig. 6b, Supplementary Table 5). For example, DFLNH interacts with the all-asthma GRS ($P < 0.05$) so that children with high genetic risk exposed to high concentrations of this HMO showed reduced prevalence of recurrent wheeze compared to children with similarly high GRS exposed to low DFLNH. Similar trends, albeit not significant, were also observed for 2'FL and LNFP1, but inverse interaction effects were observed for LNFP2 and LNFP3.

Discussion

We uniquely explored the mother-milk-infant triad, reporting how maternal genetics determined HMO composition, which in turn was

associated with respiratory health of human milk-fed infants in a manner dependent on their own genetic risk for recurrent wheeze. Our results implicate known and novel genes in the biosynthesis of HMOs, identifying associations to numerous loci across multiple chromosomal regions (e.g., 3q27.3, 7q21.32, 7q31.32, 19p13.3, and 19q13.33), many of which map to genes that encode enzymes involved in glycosylation (*TMTCL*, *GCNT3*, *ST8SIA2*, *FUT6*, *FUT3*, *FUT2*, *FUT1*, *B3GNT7*, *ST6GALI*, *B4GALTI*), and solute carrier proteins (*SLC39A8*, *SLCO3A1*, *SLC25A21*). Notably, these associations replicated in an independent cohort¹⁰. Furthermore, we found that among human milk-fed infants with high genetic risk, exposure to high concentrations of certain fucosylated HMOs was associated with lower prevalence of recurrent wheeze. Thus, our novel nuanced approach to studying the mother-milk-infant triad provides a deeper understanding of the protective effects of breastfeeding on the respiratory health of infants.

The most significant HMO-SNP association was located on chromosome 19q13.33, which harbors the *FUT2* gene, known to be involved

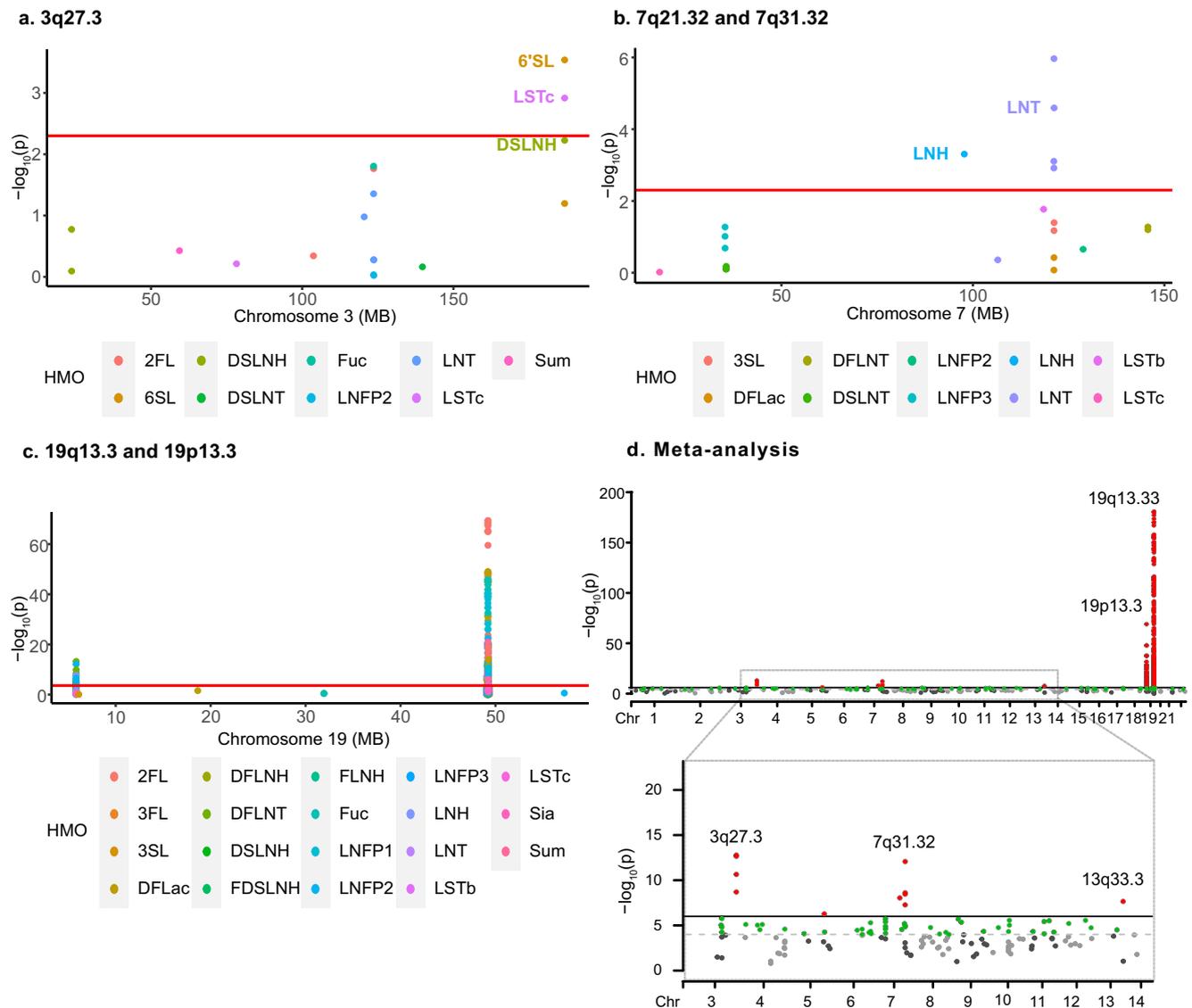


Fig. 5 | Replication analyses of HMO-associated loci in the INSPIRE Study.

a–c Overlaid regional plots show replication of HMO-associated loci in the INSPIRE sub-cohort ($N = 395$): significant replications after Bonferroni correction for multiple testing surpass the red horizontal line. **a** A SNP rs4234598 in locus 3q27.3 was associated with both 6'SL and LSTc. **b** Four SNPs in 7q31.32 were associated with LNT and one SNP (rs16869462) in 7q21.32 is associated with LNH. **c** A total of 41 SNPs in 19p13.2 and 19q13.33 loci were associated with all HMOs except 6'SL and

LNnT in the INSPIRE sub-cohort. **d** Overlaid Manhattan plot of the meta-analyses of 19 HMOs combining the CHILD and INSPIRE sub-cohorts ($N = 1375$). Loci in red exceeded the genome-wide significance threshold (meta- $P < 5e-08$) and loci in green were suggestively associated (meta- $P < 1e-05$). Out of the 6 associated genomic regions, two were novel (7q21.32 and 13q33.3), not previously identified in either the CHILD or INSPIRE sub-cohorts alone.

in glycosylation. Specifically, homozygosity for the loss of function *FUT2* rs601338 SNP was associated with near absence of alpha1-2-fucosylated HMOs: 2'FL, DFLac and LNFP1. Surprisingly, however, 3FL and 3'SL concentrations in the milk of these subjects were also lower. In contrast, structural isomers LNFP2 and LNFP3 were higher in concentration as well as the precursor LNT but not LNnT. Concentrations of the sialylated isoform LSTb was higher while those of LSTc was lower. These findings indicate that loss of *FUT2* not only blocks alpha1-2-fucosylation but shifts entire glycosylation pathways leading to an accumulation of non-fucosylated precursors such as LNT as well alternative products (e.g., LNFP2, LNFP3, and LSTb). The *FUT2* SNP also associated with lower concentrations of total HMO-bound fucose and higher total HMO-bound sialic acid. Overall, these results emphasize that a single SNP in *FUT2* may dramatically alter the HMO landscape, represented by distinct clustering when HMO composition profiles were plotted using PCoA plots.

The second most HMO-associated genomic region was located on chromosome 19p13.3, including candidate genes fucosyltransferases 3 and 6 (*FUT3/6*), and Neurturin (*NRTN*), which are known to regulate glycosylation of plasma proteins and IgG^{14,15}. SNPs in this region were associated with increased 2'FL and total HMOs, and lower concentrations of HMO-bound sialic acid. However, none of the extended sialylated HMOs (LSTb, LSTc, DSLNT, DSLNH), were associated with SNPs in this region. In contrast, we noted associations with extended fucosylated HMOs: lower LNFP2 (alpha1-4-fucosylated HMO) and higher LNFP1 (alpha1-2-fucosylated HMO) but no association with LNFP3 and 3FL (alpha1-3-fucosylated HMO). Interestingly, SNPs in *FUT3* were associated with decreased DFLac, which is both alpha1-2- and alpha1-3-fucosylated. Taken together, these findings improve our understanding of the complex biosynthesis of HMOs, identifying multiple mechanisms for fucosylation and sialylation.

Novel associations were reported for SNPs on chromosomes 3q27.3, 10q22.1 and 19p13.11, which map to the *ST6GALI*, *PALDI* and *DDAI* genes, respectively. SNPs in *ST6GALI* were correlated with higher concentrations of HMOs carrying an alpha2-6-linked sialic acid in the terminal galactose (6'SL and LSTc) but were not associated with 3'SL and LSTb, indicating that the ST6GALI enzyme does not play a role in alpha2-3-sialylation and specifically catalyzes the alpha2-6-sialylation to the terminal galactose (not internal N-acetyl-glucosamine). PALDI enables phosphatase tyrosine activity¹⁶ while DDAI is involved in protein ubiquitination¹⁷ - but their roles in HMO biosynthesis remain to be determined. Finally, our conditional GWAS analyses identified multiple novel associations on chromosomal regions: 2q37.1, 3p24.3, 3q31.3, 7q31.32, 8q24.1, 16p13.2, 17p13.1, 18q22.3 and 22q12.3. Future research is needed to further investigate genes within these genomic regions. Within 2q37.1 for example, the top SNP rs13025087 was associated with 3'SL and maps to *B3GNT7*, encoding beta-1.3-N-acetylglucosaminyltransferase 7 gene. While not much is known about *B3GNT7*, SNPs in the related gene *B3GNT5* were previously associated with bovine milk oligosaccharides such as LNT and LNH¹⁸.

Notably, many of the above associations from CHILD successfully replicated in the independent INSPIRE cohort. In addition, meta-analysis of these two cohorts identified SNPs on chromosome 7q21.32 associated with LNT concentrations. SNPs in this locus have been associated with expression of *FAM3C* in the esophagus mucosa and colon¹⁹, but have not been previously linked to milk synthesis or mammary tissue.

To our knowledge, this is the largest study to investigate individual HMOs and childhood respiratory health, involving nearly 1000 mother-child pairs. Sprenger *et al.* reported lower incidence of allergic reactions among breastfed infants ($N=266$) exposed to high concentrations of *FUT2*-dependent oligosaccharides, specifically among c-section born infants²⁰. Another study ($N=73$) found that early consumption of LNFP2 was associated with lower infant respiratory problems by 12 weeks²¹. In a smaller subset of the CHILD cohort ($N=421$), we previously reported that certain HMO profiles were associated with a reduced risk of allergic sensitization during infancy²², but maternal and infant genetics were not considered in these earlier studies. Our current study demonstrates that exposure to high concentrations of specific fucosylated HMOs (2'FL, LNFP1 and DFLNH) through breastfeeding was associated with reduced prevalence of recurrent wheeze during early childhood, specifically among infants with *high* genetic risk scores (GRS) for wheezing. Notably, our GxE analysis for "breastfeeding at 12 months" showed a protective association only among children with *low* GRS - indicating that specific milk components (such as HMOs) may have important health effects that are obscured when "breastfeeding" is considered as a single homogeneous exposure. Together, these results provide new evidence that a modifiable exposure (specific fucosylated HMOs) could mitigate genetic risk for poor respiratory health, and emphasize the importance of considering milk composition and infant genetic risk in the assessment of breastfeeding and infant health.

Collectively our data suggest that fucosylated HMOs may enhance immunity and protect respiratory health during early-life development, which could occur through several mechanisms. First, HMOs may act directly as immune modulators on the mucosa of the esophagus or colon, or systemically after absorption. For example, fucosylated HMOs are reported to have anti-inflammatory effects, reducing cytokine and CD14 expression in intestinal epithelial cells^{23,24}. Second, HMOs may affect immune development indirectly via the microbiome. For example, HMOs are known to promote beneficial bacteria, such as *Bifidobacterium*, while inhibiting harmful bacteria, such as *Clostridium difficile*²⁵, and we have previously shown in the CHILD cohort that gut microbiota dysbiosis is associated with asthma²⁵.

A limitation of our current study, which is common to many published GWASs, is the focus on subjects of central European

ancestry who make up approximately 78% of the CHILD cohort. This is important because the SNPs associated with *FUT2* secretor status (and potentially other HMO synthesis pathways) vary geographically. For example, SNP rs1047781 in *FUT2* is prevalent among East Asians but not found in Europeans and Africans, whereas the reverse is true for another SNP (rs601338) in the same gene. Future studies including more diverse populations may identify novel variants and reveal additional biosynthesis pathways of HMOs. Another limitation is our single timepoint analysis, as HMOs can vary longitudinally across lactation^{5,6}. In addition, although the 19 HMOs analyzed here are the most abundant, it is possible that other HMOs have important physiological functions that may warrant future GWAS analyses. Finally, further studies are necessary to identify causal SNP variants and determine how they contribute to HMO regulation, and to replicate our GxE findings and investigate molecular mechanisms by which HMOs could impact lung health.

Overall, this investigation significantly advances knowledge about the genetics of HMO biosynthesis in the human mammary gland. A deeper mechanistic understanding of this process will not only allow us to develop maternal interventions to improve HMO composition (once we know which HMO compositions are indeed beneficial) but will also help optimize HMO synthesis strategies in cell-free enzyme systems or bioengineered cells to make HMOs available for research and application. In addition, our study illustrates how a nuanced approach to studying the mother-milk-infant triad can provide a deeper understanding of the benefits of breastfeeding and help identify potential 'personalized' interventions such as supplemental HMOs to protect infant health among those at high risk. Collectively, our findings can help inform new mechanistic research into the development of HMO-based therapies to curb the lifelong risk of chronic respiratory diseases.

Methods

Study protocols were reviewed and approved by Human Research Ethics Boards at Queen's University, McMaster University, the Universities of Manitoba, Alberta and British Columbia, and the Hospital for Sick Children. All participants or their legal guardians provided informed consent. GWAS analyses of HMO profiles from breast milk samples were conducted in lactating mothers who reported as female and their biological sex was confirmed using genomics data from the X-chromosome. Genomics analysis in their children included reported sex, which was also confirmed for each participant using data from their X-chromosomes, as a co-variate. Gender was not available for the children given their young ages (five years and younger) and not considered in the mothers as it is not known to affect HMO profiles nor genomics. An overview of our experimental design is depicted in Supplementary Fig. S1.

CHILD study participants

The CHILD Cohort Study recruited 3455 pregnant women between 2009 and 2012 from predominantly urban areas in four Canadian provinces²⁶. Infants born ≥ 35 weeks of gestation and with birthweights ≥ 2500 g remained eligible ($N=3264$). These children were followed from birth onwards, with a home visit at 3-4 months postpartum, repeated questionnaires examining environmental exposures, and clinic visits at ages 1, 3, and 5 years for detailed assessments and obtaining biospecimens. Characteristics of the 980 mothers with HMO data are summarized and compared to 2,370 mothers with genetics data in Supplementary Table 2. The characteristics of their children are shown in Supplementary Table 3.

Genomics data

DNA was isolated from peripheral blood collected from 2552 mothers and cord blood collected from 2967 infants of the CHILD Study and used for genotyping 557,006 single nucleotide variants (SNVs) on the

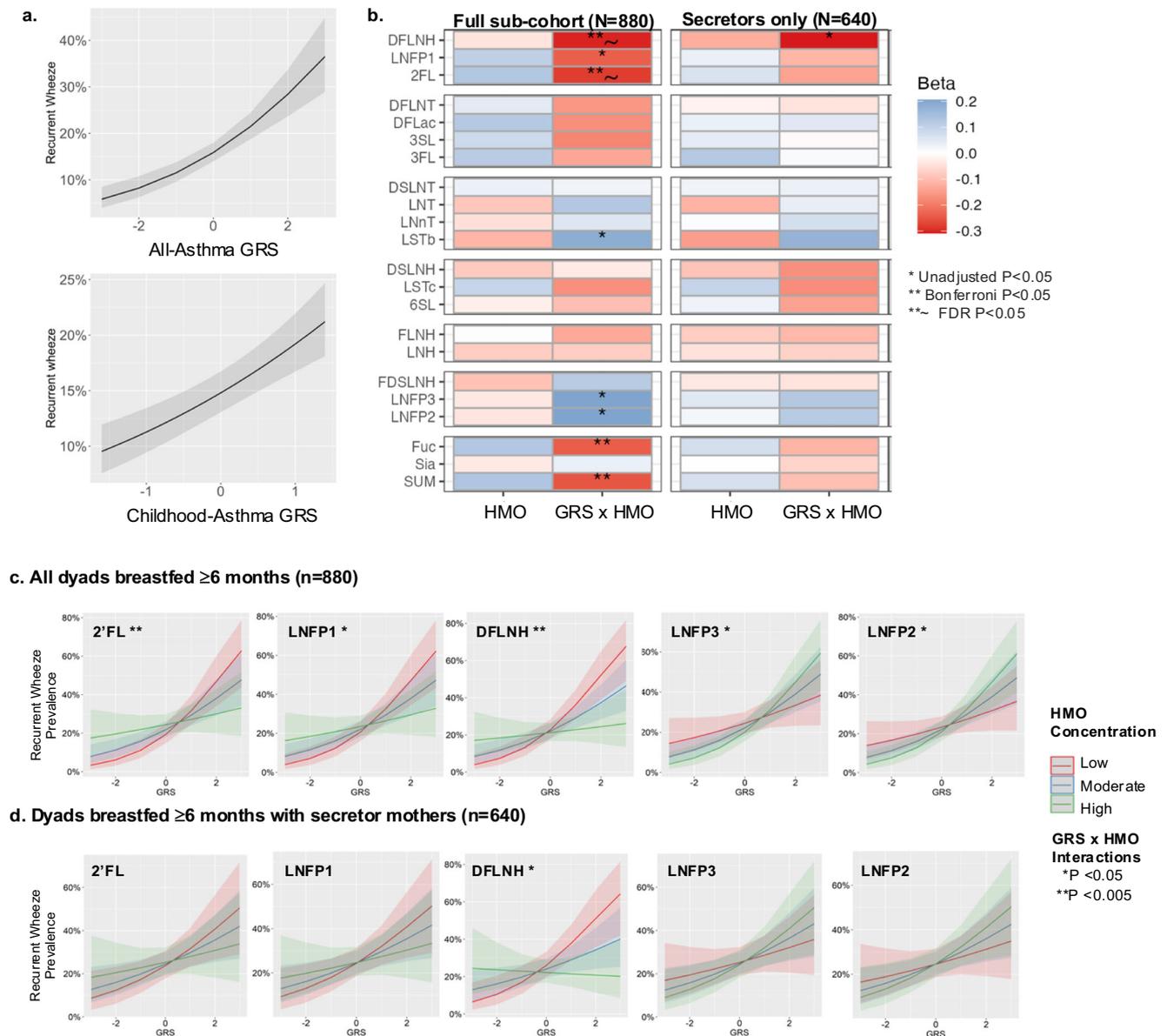


Fig. 6 | Gene-environment interaction analyses: Exposure to certain fucosylated HMOs interacts with individual genetic risk to modulate risk of recurrent wheeze during childhood in the CHILD cohort. **a** Association of all-asthma GRS and childhood-asthma GRS with recurrent wheeze between ages 2 and 5 years ($N = 2835$; $P = 2.48 \times 10^{-10}$ and $P = 1 \times 10^{-7}$, respectively). The x-axes indicate GRS Z-scores and the y-axes indicate the prevalence of recurrent wheeze. **b–d** Gene-environment interaction (GxE) analyses using logistic regression analyses indicate that exposure to specific HMOs modulate risk of recurrent wheeze among human milk-fed children in the sub-cohort ($N = 880$ children of all mothers with HMOs and genomics data) as well as in the sub-cohort of 640 secretor mothers. Unadjusted $*P < 0.05$;

Bonferroni adjusted $**P < 0.05$; FDR adjusted $**~P < 0.05$. Normalized HMO concentrations were categorized as Low (< -1 SD), Moderate (-1 to $+1$ SD) and High ($> +1$ SD) represented as bands and colors. Exposure to low levels of three HMOs (e.g., 2'FL, LNFP1, DFLNH) is associated with increased prevalence of recurrent wheeze among children with high GRS ($P < 0.05$). GxE unadjusted P values using generalized linear model: 2'FL; $P = 0.004$, LNFP1; $P = 0.01$, DFLNH; $P = 0.002$). In contrast, exposure to high levels of other HMOs (e.g., LNFP2, LNFP3) is associated with increased prevalence of recurrent wheeze among children with high GRS (GxE Unadjusted p values: LNFP2; $P = 0.03$, LNFP3; $P = 0.02$). Full summary statistics are provided in Supplementary Tables 4, 5.

Illumina HumanCoreExome BeadChip. Quality control (QC) measures were applied using PLINK (version 1.9)²⁷, which employed filters as previously described²⁸. Briefly, subjects were omitted using the following filters: X chromosome F-coefficient > 0.2 for females and < 0.8 for males (validated using Y chromosome counts); excess heterozygosity ± 2 standard deviations; genotype missingness $> 10\%$; relatedness using inbreeding coefficient > 0.185 . Next, SNVs were excluded with missingness > 0.05 and Hardy-Weinberg equilibrium $P < 10 \times 10^{-7}$ in a sub-cohort of Central European Ancestry. A total of 515,033 SNVs in 2,426 mothers and 2835 children were then carried forward for whole genome imputations using data from the Haplotype

Reference Consortium (HRC; r1.1 2016) and the Michigan server²⁹. Following imputations, variants with minor allele frequency > 0.01 and imputation quality score $R^2 > 0.3$ were retained for further analysis. QC and imputation steps are detailed in Supplementary Information.

Human milk collection and quantification of human milk oligosaccharides (HMOs)

Human milk was collected from 1206 lactating CHILD mothers during their 3–4-month home visit as described previously³⁰. Nineteen of the most abundant HMOs were quantified in two batches ($N = 427$ and $N = 779$, HMO names and abbreviations are listed in Supplementary

Table 1)³⁰. In total, 980 mothers had both HMO and genomics data (summarized in Supplementary Table 2). HMOs were isolated by high-throughput solid-phase extraction, fluorescently labeled, and analyzed by high-performance liquid chromatography (HPLC) as previously described³¹. Total HMO concentrations, and HMO-bound fucose or sialic acid concentrations were calculated as the sum of all measured HMOs and of fucose or sialic acid residues, respectively. HMO concentrations were normalized using the inverse rank normalization score method³² (Supplementary Figs. S2 and S3 show HMO distributions pre- and post-normalization). Normalized HMO concentrations were categorized as Low (<-1 SD), Moderate (-1 to +1 SD) and High (> +1 SD). Phenotypic secretor status was determined by the presence or near absence (<200 nmol/mL) of 2'-fucosylactose (2'FL)³⁰.

Genome-wide association study (GWAS) analyses of HMOs

GWAS analyses of HMO profiles in the 980 CHILD mothers included 5.4 M common single nucleotide polymorphisms (SNPs), which are SNVs with minor allele frequencies > 0.05, adjusted for population substructure (principal components (PCs) 1–3) and two HMO batches. Secondary GWAS analyses included 764 mothers of central European (CEU) ancestry only (see details in Supplementary Information). Using PLINK 1.9²⁷, we performed GWAS analysis using logistic regression for binary phenotypic secretor status (based on presence of 2'FL), or linear regression for continuous concentrations of individual HMOs, total HMOs, fucosylated HMOs and sialylated HMOs. Manhattan plots employed the “qqman” R-package³³ and EASY STRATA³⁴. Lead and independent signals were identified from significant SNPs ($P < 5e-8$) by linkage disequilibrium (LD) clumping using $r^2 \leq 0.1$ and $r^2 \leq 0.6$, respectively, within 250 kb windows using the SNP2GENE function in the FUMA GWAS web interface³⁵. FUMA GWAS was also employed to generate locus zoom plots. Boxplots and the principal coordinates analysis (PCoA) plots, using QIIME³⁶, depict the allelic effects of select SNVs among the secretors and non-secretors. PCoA analysis for the genotypes were estimated using Simpson dissimilarity³⁷.

Following our primary GWAS analyses of HMOs above, we performed a joint conditional GWAS analyses³⁸ using linear regression models in PLINK that included 21 lead SNPs from our primary GWAS results (all on chromosome 19) as covariates. In addition, we conducted conditional GWAS analyses using each of the 21 lead SNPs from our primary GWAS as covariates. Finally, we applied a stepwise approach using one, two and three of the top lead SNPs at each the chromosome 19 loci from Table 1.

Replication of GWAS findings in the INSPIRE Cohort Study

SNPs identified from the primary and conditional GWAS analyses of HMOs in the CHILD Study ($P < 1e-05$) were carried forward for replication analyses in the independent INSPIRE cohort¹⁰. Genome-wide SNV genotyping data from the INSPIRE cohort were obtained using Illumina Multi-Ethnic Global-8 v1.0 arrays as previously described¹⁰. Briefly, the INSPIRE Study was an international, intercultural investigation of factors associated with global variation in human milk composition. Details regarding experimental design, milk collection, HMO analyses, and maternal genotyping have been previously published^{10,39}. SNP QC exclusion criteria included: variant call rates <89.5%, minor allele frequencies <0.0095, and Hardy-Weinberg equilibrium (HWE) of $P < 10e-7$. SNPs were tested for association with HMO phenotypes using an additive model. A total of 281 SNPs identified in CHILD overlapped with those genotyped in INSPIRE. First, we performed a locus specific replication, which focused on SNPs associated with HMOs in CHILD ($P < 1e-05$). Multiple testing corrections were applied using the Bonferroni method by adjusting for the number of independent SNPs in each genomic region (LD, $r^2 \leq 0.6$). Second, we performed a meta-analysis using METAL⁴⁰ to combine association results for all SNPs from the CHILD ($N = 980$) and INSPIRE cohorts ($N = 395$). Significant SNPs for meta-

analysis were identified using genome-wide significance threshold ($P < 5e-08$).

Recurrent wheeze and genetic risk score (GRS) analysis

Recurrent wheeze was selected as the primary respiratory outcome during early childhood, which was defined as two or more episodes of wheeze in one year between ages 2 and 5 years. We opted to omit early wheezing episodes prior to age 2 years as these are more likely to be transient and result of an infection⁴¹. Moreover, we selected recurrent wheeze instead of asthma because the latter is a highly heterogeneous disease^{42,43} that is difficult to diagnose in young children^{44,45}. This heterogeneity is known to delay diagnosis and is a major challenge for studying asthma in early childhood⁴⁶. In addition, wheeze is known to be a key defining trait of asthma, and persistent symptoms during early life, even if associated with later remission, has been linked to lower lung function and chronic lung disease later in life⁴⁷.

Using recurrent wheeze as our primary outcome, we conducted two GRS analyses: one utilizing summary statistics from one of the largest published GWAS of asthma, regardless of age of onset¹²; and another from the largest GWAS childhood-onset asthma⁴⁸. GRS analyses were conducted in four steps: (1) Identify SNPs associated with asthma from the published GWAS ($P < 5e-8$) in the CHILD cohort; (2) Eliminate redundant signals by pruning SNPs in linkage disequilibrium (LD; $r^2 > 0.8$) using a window size of 50 kb and shift distance of 2; (3) Beginning with the most significant SNV, apply a stepwise forward regression model by summing the risk alleles weighted by the effect sizes obtained from the published study; (4) Determine the association of the resulting GRS with recurrent wheeze using logistic regression analysis, adjusted for sex and population substructure (e.g. PCs 1–10). We further converted this GRS into Z-scores ranging from -3 to +3 for each of the CHILD infants. Of the 980 mothers with both HMOs and genomics profiles, 880 of their children had genomics data for GRS analysis (see Supplementary Table 3 for characteristics of the infants including prevalence of asthma and recurrent wheeze).

Gene-environment interaction analyses

We determined gene-environment interactions (GxE) between each child's GRS for recurrent wheeze (G; Z-score) and their exposure to maternal HMOs (E; nmol/mL) using a generalized linear model in R (logistic regression test for each of the 19 HMOs individually), which adjusted for sex, two HMO batches, and 10 PCs (see equation in Supplementary Information). This GxE analysis was undertaken among the 880 children with available GRS data, maternal HMO data, maternal genetic data, who been human milk-fed for 6 months or longer. Multiple testing corrections were applied to adjust for 19 HMOs using the false discovery rate (FDR) as well as the Bonferroni method, which adjusted for six clusters of correlated HMOs (see Supplementary Fig. S4). The interaction plots were generated using the R package sjPlot (<https://CRAN.R-project.org/package=sjPlot>). GRS were normalized (Z-scores ranging from -3 to +3) and used as continuous variables without any grouping. Normalized HMO concentrations were categorized as Low (<-1 SD), Moderate (-1 to +1 SD) and High (> +1 SD). The lines were fitted by estimating the effects using generalized linear model (glm) with interaction term in R and plotted using the plot_model function in sjPlot.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The GWAS summary statistics generated in this study have been deposited in the GWAS Catalog database under accession number GCP ID GCP000844, and are also provided in the Supplementary Table 6.

Data from the CHILD Cohort Study are available from the CHILD database: <https://childstudy.ca/childdb/>.

Code availability

No custom code was generated for this manuscript, however, all code has been shared: https://github.com/ComputationalGenomicsLaboratory/HMO_GWAS.

References

- Victora, C. G. et al. Breastfeeding in the 21st century: epidemiology, mechanisms, and lifelong effect. *Lancet* **387**, 475–490 (2016).
- Miliku, K. & Azad, M. B. Breastfeeding and the developmental origins of asthma: Current evidence, possible mechanisms, and future research priorities. *Nutrients* **10**, 995 (2018).
- Lodge, C. J. et al. Breastfeeding and asthma and allergies: a systematic review and meta-analysis. *Acta Paediatr.* **104**, 38–53 (2015).
- 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).
- Witkowska-Zimny, M. & Kaminska-El-Hassan, E. Cells of human breast milk. *Cell Mol. Biol. Lett.* **22**, 11 (2017).
- Boix-Amoros, A., Collado, M. C. & Mira, A. Relationship between milk microbiota, bacterial load, macronutrients, and human cells during lactation. *Front Microbiol* **7**, 492 (2016).
- Mychaleckyj, J. C. et al. Multiplex genomewide association analysis of breast milk fatty acid composition extends the phenotypic association and potential selection of FADS1 variants to arachidonic acid, a critical infant micronutrient. *J. Med. Genet* **55**, 459–468 (2018).
- Bode, L. Human milk oligosaccharides: every baby needs a sugar mama. *Glycobiology* **22**, 1147–1162 (2012).
- Doherty, A. M. et al. Human milk oligosaccharides and associations with immune-mediated disease and infection in childhood: A systematic review. *Front Pediatr.* **6**, 91 (2018).
- Williams, J. E. et al. Key genetic variants associated with variation of milk oligosaccharides from diverse human populations. *Genomics* **113**, 1867–1875 (2021).
- Bode, L. The functional biology of human milk oligosaccharides. *Early Hum. Dev.* **91**, 619–622 (2015).
- Demenais, F. et al. Multiancestry association study identifies new asthma risk loci that colocalize with immune-cell enhancer marks. *Nat. Genet* **50**, 42–53 (2018).
- Ferrer-Admetlla, A. et al. A natural history of FUT2 polymorphism in humans. *Mol. Biol. Evol.* **26**, 1993–2003 (2009).
- Shen, X. et al. Multivariate discovery and replication of five novel loci associated with Immunoglobulin G N-glycosylation. *Nat. Commun.* **8**, 447 (2017).
- Lauc, G. et al. Genomics meets glycomics—the first GWAS study of human N-Glycome identifies HNF1alpha as a master regulator of plasma protein fucosylation. *PLoS Genet* **6**, e1001256 (2010).
- Dominguez-Cruz, M. G. et al. Pilot genome-wide association study identifying novel risk loci for type 2 diabetes in a Maya population. *Gene* **677**, 324–331 (2018).
- Sinnott-Armstrong, N. et al. Genetics of 35 blood and urine biomarkers in the UK Biobank. *Nat. Genet* **53**, 185–194 (2021).
- Poulsen, N. A., Robinson, R. C., Barile, D., Larsen, L. B. & Buitenhuis, B. A genome-wide association study reveals specific transferases as candidate loci for bovine milk oligosaccharides synthesis. *BMC Genomics* **20**, 404 (2019).
- Consortium, G. T. The genotype-tissue expression (GTEx) project. *Nat. Genet* **45**, 580–585 (2013).
- Sprenger, N. et al. FUT2-dependent breast milk oligosaccharides and allergy at 2 and 5 years of age in infants with high hereditary allergy risk. *Eur. J. Nutr.* **56**, 1293–1301 (2017).
- Stepans, M. B. et al. Early consumption of human milk oligosaccharides is inversely related to subsequent risk of respiratory and enteric disease in infants. *Breastfeed. Med.* **1**, 207–215 (2006).
- Miliku, K. et al. Human milk oligosaccharide profiles and food sensitization among infants in the CHILD Study. *Allergy* **73**, 2070–2073 (2018).
- Walsh, C., Lane, J. A., van Sinderen, D. & Hickey, R. M. Human milk oligosaccharides: Shaping the infant gut microbiota and supporting health. *J. Funct. Foods* **72**, 104074 (2020).
- He, Y. et al. The human milk oligosaccharide 2'-fucosyllactose modulates CD14 expression in human enterocytes, thereby attenuating LPS-induced inflammation. *Gut* **65**, 33–46 (2016).
- Moossavi, S., Miliku, K., Sepehri, S., Khafipour, E. & Azad, M. B. The prebiotic and probiotic properties of human milk: implications for infant immune development and pediatric asthma. *Front Pediatr.* **6**, 197 (2018).
- Subbarao, P. et al. The Canadian Healthy Infant Longitudinal Development (CHILD) Study: examining developmental origins of allergy and asthma. *Thorax* **70**, 998–1000 (2015).
- Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet* **81**, 559–575 (2007).
- Marees, A. T. et al. A tutorial on conducting genome-wide association studies: Quality control and statistical analysis. *Int. J. Methods Psychiatr. Res.* **27**, e1608 (2018).
- Das, S. et al. Next-generation genotype imputation service and methods. *Nat. Genet* **48**, 1284–1287 (2016).
- 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).
- 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).
- McCaw, Z. Inverse rank normalization, rnmomi: rank normal transformation omnibus, <https://CRAN.R-project.org/package=RNOmni>. R package (2020).
- SD, T. qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. *Journal of Open Source Software*, 3(25) (2018).
- Winkler, T. W. et al. EasyStrata: evaluation and visualization of stratified genome-wide association meta-analysis data. *Bioinformatics* **31**, 259–261 (2015).
- Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* **8**, 1826 (2017).
- Caporaso, J. G. et al. QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* **7**, 335–336 (2010).
- Masi, A. C. et al. Human milk oligosaccharide DSLNT and gut microbiome in preterm infants predicts necrotising enterocolitis. *Gut* **70**, 2273–2282 (2021).
- Yang, J. et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat. Genet* **44**, 369–375 (2012).
- McGuire, M. K. et al. What's normal? Oligosaccharide concentrations and profiles in milk produced by healthy women vary geographically. *Am. J. Clin. Nutr.* **105**, 1086–1100 (2017).
- Willer, C. J., Li, Y. & Abecasis, G. R. METAL: Fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
- Taussig, L. M. et al. Tucson children's respiratory study: 1980 to present. *J. Allergy Clin. Immunol.* **111**, 661–675 (2003). quiz 676.
- Comberiati, P., Di Cicco, M. E., D'Elios, S. & Peroni, D. G. How Much Asthma Is Atopic in Children? *Front Pediatr.* **5**, 122 (2017).

43. Papi, A., Brightling, C., Pedersen, S. E. & Reddel, H. K. Asthma. *Lancet* **391**, 783–800 (2018).
44. Carr, T. F. & Bleecker, E. Asthma heterogeneity and severity. *World Allergy Organ J.* **9**, 41 (2016).
45. Reyna, M. E. et al. Development of a symptom-based tool for screening of children at high risk of preschool asthma. *JAMA Netw. Open* **5**, e2234714 (2022).
46. Ducharme, F. M. et al. Diagnosis and management of asthma in preschoolers: A Canadian Thoracic Society and Canadian Paediatric Society position paper. *Paediatr. Child Health* **20**, 353–371 (2015).
47. Dai, R. et al. Wheeze trajectories: Determinants and outcomes in the CHILd cohort study. *J. Allergy Clin. Immunol.* **149**, 2153–2165 (2022).
48. Pividori, M., Schoettler, N., Nicolae, D. L., Ober, C. & Im, H. K. Shared and distinct genetic risk factors for childhood-onset and adult-onset asthma: genome-wide and transcriptome-wide studies. *Lancet Respir. Med.* **7**, 509–522 (2019).

Acknowledgements

We are grateful to all the CHILd families who took part in this study, and the whole CHILd team, which includes interviewers, nurses, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, and receptionists. For a list of investigators and enrolling centers, visit www.childcohort.ca. The Canadian Institutes of Health Research (CIHR) and the Allergy, Genes and Environment Network of Centers of Excellence (AllerGen) provided core support for CHILd. This study was funded by operating grants from CIHR: PJT-178390, PI-Q.D.; MRT-160844, Pls-P.S. and M.B.A.; and the Team Grant RFA#201301FH6, PI-S.S.A. HMO analysis was funded by a New Investigator Grant to M.B.A. from Research Manitoba. M.B.A. holds a Tier 2 Canada Research Chair in Early Nutrition and the Developmental Origins of Health and Disease and is a Fellow of the CIFAR Humans and the Microbiome Program. L.B. is Chair of Collaborative Human Milk Research, endowed by the Family Larsson-Rosenquist Foundation (FLRF) in Switzerland. The INSPIRE study was supported by the National Science Foundation (award 1344288 to M.K.M. and C.M.), the Ministry of Economy and Competitiveness, Spain (project AGL2013-4190-P; to L.R.), and the European Commission [grant 624773 (FP-7-PEOPLE-2013-IEF); to L.R.]. Sterile, single-use milk-collection kits used in the INSPIRE study were provided by Medela Inc. P.J.M. receives funding from the Women's and Children's Health Research Institute (WCHRI). Computational analyses were performed on resources and with support provided by the Center for Advanced Computing (CAC) at Queen's University in Kingston, Ontario. The CAC is funded by the Canada Foundation for Innovation, the Government of Ontario, and Queen's University.

Author contributions

Q.L.D., M.B.A. and L.B. conceived and directed the study; A.A. led the cleaning of the genomics data and statistical analyses (GWASs, GRS, GxE) with help from L.C., J.C., Y.Z., S.A.S. and Z.Y.F.; L.B. led the quantification of the HMOs with help from B.R. and C.Y.; M.B.A. and K.M. led processing of the human milk samples and analysis of the HMOs; S.E.T., P.J.M., E.S., T.J.M., M.R.S. and P.S. led the recruitment of CHILd Study participants and collection of biospecimens, clinical data and assessments; M.R.S., S.S.A. and G.P. led collection of

genomics data; J.E.W., B.M.M., G.E.O., S.M., E.W.K.M., E.W.K., D.K.G., J.M.R., R.G.P., D.W.S., S.E.M., A.M.P., J.A.F., L.J.K., H.L.N., M.A.M., M.K.M., C.L.M. led data collection and analysis of the INSPIRE Cohort; Q.D., M.B.A. and L.B. and A.A. wrote the manuscript that was reviewed and revised by all coauthors. All coauthors contributed to the interpretation of the data or results and approved the submitted version of the manuscript.

Competing interests

J.C. is currently an employee of F. Hoffman-La Roche Ltd., however, the published work was done prior to this employment and does not involve/promote any of Roche's materials or point of view. M.B.A. has consulted for DSM Nutritional Products (a food ingredient company) and serves on the Scientific Advisory Board for TinyHealth (a microbiome testing company). She has received research funding (unrelated to this project) and speaking honoraria from Prolacta Biosciences (a human milk fortifier company). L.B. is a co-inventor on patent applications related to the use of HMOs in preventing NEC and other inflammatory diseases. The remaining authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-024-51743-6>.

Correspondence and requests for materials should be addressed to Meghan B. Azad, Lars Bode or Qingling Duan.

Peer review information *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://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-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, 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 you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. 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-nc-nd/4.0/>.

© The Author(s) 2024

¹Department of Biomedical and Molecular Sciences, Queen's University, Kingston, ON, Canada. ²School of Computing, Queen's University, Kingston, ON, Canada. ³Department of Medicine, McMaster University, Hamilton, ON, Canada. ⁴Department of Nutritional Sciences, University of Toronto, Toronto, ON, Canada. ⁵Department of Pediatrics and Larsson-Rosenquist Foundation Mother-Milk-Infant Center of Research Excellence (MOMI CORE), Human Milk Institute (HMI), University of California San Diego, La Jolla, CA, USA. ⁶Department of Pediatrics, Division of Allergy and Immunology, University of British Columbia, Vancouver, BC, Canada. ⁷Department of Pediatrics, University of Alberta, Edmonton, AB, Canada. ⁸Faculty of Medicine, USCI University, Kuala

Lumpur, Malaysia. ⁹Department of Pediatrics and Child Health, University of Manitoba, Winnipeg, MB, Canada. ¹⁰Department of Pediatrics, Hospital for Sick Children and University of Toronto, Toronto, ON, Canada. ¹¹Chanchlani Research Centre, Dept. of Medicine, McMaster University, Hamilton, ON, Canada. ¹²Pathology and Molecular Medicine, McMaster University, Hamilton, ON, Canada. ¹³Department of Animal, Veterinary and Food Sciences, University of Idaho, Moscow, Idaho, USA. ¹⁴Department of Nutrition & Food Science, University of Ghana, Accra, Ghana. ¹⁵Department of Human Nutrition, Egerton University, Nakuru, Kenya. ¹⁶Department of Anthropology, Hawassa University, Hawassa, Ethiopia. ¹⁷Department of Nutrition and Food Science, Complutense University of Madrid, Madrid, Spain. ¹⁸Nutrition Research Institute, Lima, Peru. ¹⁹Department of Anthropology, University of Toronto, Toronto, ON, Canada. ²⁰Department of Women and Children's Health, King's College London, London, UK. ²¹The Medical Research Council Unit, The Gambia at the London School of Hygiene & Tropical Medicine, Fajara, Gambia. ²²Department of Biological Sciences, University of Idaho, Moscow, ID, USA. ²³Faculty of Medicine, Lund University, Lund, Sweden. ²⁴Department of Animal Sciences, Washington State University, Pullman, WA, USA. ²⁵Margaret Ritchie School of Family and Consumer Sciences, University of Idaho, Moscow, ID, USA. ²⁶Department of Anthropology, Washington State University, Pullman, WA, USA. ²⁷Manitoba Interdisciplinary Lactation Centre (MILC), Children's Hospital Research Institute of Manitoba, Winnipeg, MB, Canada. ²⁸These authors jointly supervised this work: Meghan B. Azad, Lars Bode, Qingling Duan ✉ e-mail: meghan.azad@umanitoba.ca; lbode@health.ucsd.edu; qingling.duan@queensu.ca