Title: Early-life cytomegalovirus infection is associated with gut microbiota perturbations and increased risk of atopy

Conflict of interest statement

SG has received research funding and consulting fees from Moderna, Merck and VBI related to CMV vaccine development, and from Meridian Bioscience related to CMV diagnostics. All other authors have reported no conflicts of interest related to this manuscript.

Source of funding

Core funding for the CHILD Study was provided by the Canadian Institutes of Health Research (CIHR; AEC-85761, PJT-148484), the Allergy, Genes and Environment Network of Centres of Excellence (AllerGen NCE) (12CHILD) and Genome Canada/Genome BC (274CHI). HS was supported by Michael Smith Foundation for Health Research Postdoctoral Trainee Fellowship and CIHR postdoctoral Fellowship. SET holds the Aubrey J Tingle Professorship in Pediatric Immunology and Canada Research Chair in Pediatric Precision Health.

Abstract

Background: The 'old friends' hypothesis posits that reduced exposure to previously ubiquitous microorganisms is one factor involved in the increased rates of allergic diseases. Cytomegalovirus (CMV) may be one of the "old friends" hypothesized to help prevent allergic diseases. We sought to elucidate whether early-life CMV infection is associated with childhood atopy via perturbations of the gut microbiota.

Methods: Participants were recruited from a population-based birth cohort (CHILD study) and followed prospectively until age five years in four Canadian cities. A total of 928 participants provided stool microbiome data, urine for CMV testing, skin-prick tests, and questionnaires-based detailed environmental exposures.
CMV infection was assessed in the first year of life while the main outcome was defined by persistent sensitization to any allergen at ages 1, 3, and 5 years.

Results: Early CMV infection was associated with increased beta and decreased alpha diversity of the gut microbiota. Both changes in diversity measures and early CMV infection were associated with persistent allergic sensitization at age 5 years (aOR= 2.08; 95%CI: 1, 4.33). Mediation analysis demonstrated that perturbation of gut microbial composition explains 30% of the association.

Conclusions: Early-life CMV infection is associated with an alteration in the intestinal microbiota, which mediates the effect of the infection on childhood atopy. This work indicates that preventing CMV infection would not put children at increased risk of developing atopy. Rather, a CMV vaccine, in addition to preventing CMV-associated morbidity and mortality, might reduce the risk of childhood allergic diseases.

Key words

Cytomegalovirus, Immune development, Atopy, Allergy, Microbiome

INTRODUCTION

Cytomegalovirus (CMV) is one of the most prevalent human infections worldwide[1]. Although only approximately half of people are infected by adulthood in resource-rich settings, elsewhere most children are infected in the first few years of life[1,2]. Young children most commonly acquire CMV infection via breast milk or from contact with saliva or urine from another child[3,4]. Disease from CMV typically results only from congenital infection or in immunocompromised patients, such that the vast majority of CMV infections are asymptomatic[5]. Nevertheless, CMV is a powerful immune modulator[6,7], and it has been suggested that some of these effects may be beneficial[8,9]. Thus, CMV may be one of the "old friends" hypothesized to help prevent allergic diseases[10].

Allergy is defined as hypersensitivity to innocuous antigens, mediated by a dysregulated immune response to environmental stimuli. Atopy is a specific case of allergy where there is a tendency to generate dysregulated IgE-mediated responses, and is a precursor to the later development of diseases such as eczema and asthma[11–13]. The immune system and the intestinal microbiota develop concurrently in infancy, and increasing evidence indicates that early-life exposures guide the development of the microbiota, which is essential to educate the immune system and to prevent allergy and atopy[14–17]. In addition to its profound effects on the immune system, CMV can directly infect the intestinal tract of humans[18,19]. However, the potential impact of CMV infection on the gut microbiota of healthy individuals has not been described. We therefore sought to determine if CMV infection in early life has an impact on the composition of the intestinal microbiota, during the "critical window" of opportunity where the microbiota and immune system are developing contemporaneously.

In the present study, we hypothesized that CMV infection in infancy which has been a normal occurrence during human evolution, promotes healthy immunologic development that helps to prevent allergies, and that this effect may be mediated through an influence of CMV on the intestinal microbiota. To investigate the interactions between early life CMV infection, the microbiome, and subsequent development of atopy, we leveraged the CHILD cohort study, a population-based prospective Canadian birth cohort that involves collection of demographic and environmental information, stool microbiome analyses and rigorous determination of clinical outcomes, including the use of skin-prick allergy testing[20–23].

METHODS

Description of the Prospective CHILD Birth Cohort: The CHILD Study is a longitudinal birth cohort (n=3,264 eligible newborns in the General Cohort began the study – see Appendix Figure 1) after 3621 pregnant mothers were recruited from four Canadian cities, namely Vancouver, Edmonton, Winnipeg (includes Morden/Winkler, and thereafter referred to as Manitoba site), and Toronto. The current analyses included children with CMV infection status, microbiome data at 3 and 12 months of age, and who were assessed for sensitization at age 1, 3 and 5 years (n=928, Table 1). The primary outcome for this study was persistent sensitization defined as a positive skin test to any allergen at ages 1 and 3 and 5 years[24]. See Appendix for more details on skin prick testing.

Characterization of the infant biological samples: CMV infection status was defined by the detection of CMV DNA in urine. DNA extraction and qPCR amplification are described in detail in Appendix. Gut microbiota was defined by sequencing the V4 hypervariable region of the 16S

rRNA gene in DNA extracted from stool samples collected at age 3 months and 12 months. Methods of collection, DNA extraction and amplification have been previously described[23] and are detailed in Appendix.

Defining the associations between CMV infection and persistent allergic sensitization: Conditional logistic regressions (stratified by study center) were performed to evaluate the association between the child's CMV infection status, which was categorized as Early CMV (infected by 3 months of age), Later CMV (infected between 3 and 12 months of age) or never infected (No CMV), and the diagnosis of persistent sensitization at age 5 years. Variable selection was performed by testing all available of covariates individually in bivariate analyses against our outcome of interest, persistent sensitization by age 5 years (Table 1). All models were adjusted for: parental atopy, child ethnicity and home occupancy. Adjusted odds ratios (aORs) and their 95% confidence intervals (CIs) were calculated.

Linking the infant gut microbiota with exposure and outcome: The link between exposure to CMV and gut microbiota structure, assessed via the within and between-subject variability (α- and β-diversity respectively) was estimated using a linear mixed effects models with study center as a random intercept for each sample collection time point (Table 2). The association between risk of being persistently sensitized and gut microbiota diversity was measured using conditional logistic regressions stratified by study center and controlling for the same set of covariates , namely, parental atopy, child ethnicity, home occupancy, in addition to exclusive breastfeeding duration (for variable selection see Table S1, in Appendix) .

Mediation analysis: Causal mediation analysis (CMA) was used to examine both direct and indirect causal relationships, testing the hypothesis that the effect of CMV infection on persistent sensitization would be through disruption of gut microbiota[25]. Gut microbiota dysbiosis, a latent variable in the mediation analysis was constructed based on confirmatory factor analysis of alpha and beta diversity indices at 3 months.

All statistical analyses were performed using R version 3.5.1 (Appendix for full detail) and this study was approved by the UBC Institutional Review Board (certificate H07-03120). Informed consent was obtained from all CHILD Cohort Study Participants directly or via a parent or legal guardian.

RESULTS

Description of study group

In total, 1151 children of the 3264 eligible children of the CHILD study (Figure S1, online) were tested for CMV infection by urine qPCR at 3 months and 1 year of age. These subjects were selected based on availability of both urine samples and stool microbiome data. After exclusion of samples with insufficient amplification of the internal control to ensure the absence of inhibition, CMV infection status was determined for 1100 children. Among the 1100 children with known CMV status, 172 did not have skin prick test data available at 1 and 3 and 5 years of age for determination of persistent atopic sensitization, therefore 928 children were included for subsequent analyses.

Prevalence of CMV infection among healthy Canadian infants. Among the 928 study participants, 13% (122) acquired CMV by 3 months of age (Early CMV), 7% (64) had acquisition between 3 months and 1 year of age (Later CMV) and 80% (742) were not infected in the first year of life (No CMV) (Table 1). Participants were from all four sites; 39%, 35%, 16% and 10% were from Manitoba, Vancouver, Edmonton and Toronto, respectively (Table S1, Appendix). Compared to the overall CHILD study population with skin prick testing data available (n=2350), all characteristics of the 928 children in this study were similar except maternal parity and study sites (Table S1 in Appendix).

Early CMV infection is associated with an increased risk of atopy. In this study, 6% (55) of children (n=928) were diagnosed with persistent atopic sensitization by age 5 years. From the demographic and behavioural data collected from the CHILD families, a list of possible risk factors for CMV infection and allergic diseases was created *a priori* based on literature. These covariates included child ethnicity (defined as both parents Caucasian white versus other parental ethnicities), antibiotic use in the first year, maternal parity, sex, age- and sex- standardized birth weight (z-score), number of occupants in the home, child activity patterns (time spent outside the home, a surrogate for childcare attendance), parental atopy, environmental tobacco smoke exposure in the first year of life, and exclusive breastfeeding duration. Parental atopic status, child ethnicity and home occupancy were significantly associated with atopy ($p<0.01$) and were used to adjust all subsequent analyses (Table 1).

In the multiple variable analysis, Early CMV infection was associated with a doubling of the risk of developing atopy by age 5 years compared to children without CMV infection during infancy

(aOR 2.08; 95%CI 1, 4.33; p=0.05; Figure 1), whereas children with Later CMV infection were not at a statistically significantly higher risk.

Gut microbiota diversity at 3 months of life is independently associated with CMV infection and with atopy. Early CMV infection was associated with changes in the structure of the gut microbiota at age 3 months. After adjustment for the same set of covariates (home occupancy, parental atopy, and child ethnicity) and controlling for breastfeeding status, an important determinant of early gut microbiota composition, we found in infant with early CMV infection a decreased diversity within samples as measured by α-diversity (estimated effect -0.3; 95%CI -0.55, -0.06; p=0.015) and increased variation between samples as quantified by β-diversity (estimated effect 0.4; 95%CI 0.17, 0.62; p<0.001) (Table 2). By 1 year of age, neither diversity metric was significantly different between CMV-infected and uninfected children (Table S2, Appendix).

An interquartile range (IQR) increase in α-diversity at 3 months of age (measured via Chao1 index which represents species richness) was associated with a 63% decreased risk of persistent sensitization (aOR 0.37; 95%CI 0.17, 0.8; p=0.011; Figure 2A). Conversely, an IQR increase in β-diversity was associated with a 3-fold increased risk of atopy (aOR=3.1; 95%CI 1.6, 6.03; p<0.001; Figure 2B). All analyses were stratified within study sites and adjusted for child ethnicity, parental atopy and home occupancy.

Gut microbiota mediates the association between early-life CMV infection and atopy. Using causal mediation analysis (CMA), we tested the hypothesis that CMV infection influences the

risk of persistent sensitization in part through its impact on the gut microbiota. By defining a latent variable that included both microbiota α-diversity and β-diversity measures, we found that the gut microbiota diversity at 3 months was a significant mediator between Early CMV infection and atopy at age 5 years (ACME β=0.042; $p<0.001$, after adjusting for confounders; Figure 3A). Thirty percent of the total effect of CMV infection on atopy was explained by the gut microbiota composition (Figure 3 and Table S3, online).

DISCUSSION

Given that lower incidence of CMV infection coincides with increased prevalence of allergic diseases in high-income countries[26,27], we hypothesized CMV may be an "old friend" that when acquired early in life helps to prevent development of allergic disease later in life. The rationale for this includes the 'anti-allergic' Th1-type responses induced by CMV infection, and reports suggesting the possibility that CMV might reduce the risk of hypersensitivity[18,29]. For example, CMV infection in the first 2 years of life was reported to decrease risk of IgE sensitization, albeit only in combination with Epstein-Barr virus (EBV) infection[29]. In addition, CMV-infected children have been found to have immune profiles consistent with protection against IgE sensitization (increased IFN-γ and decreased IL-4 producing cells)[30]. Furthermore, we expected that CMV might have a positive impact on the establishment of a healthy, diverse gut microbiota, by virtue of the fact that humans co-evolved with CMV and this virus is known to replicate in the intestine[18,19]. However, contrary to our *a priori* hypotheses, in a contemporary Canadian birth cohort we found that early CMV infection was associated with lower gut microbiome diversity and increased risk of atopy in this cohort.

206 CMV has a variety of immunomodulatory effects that may contribute to the risk of allergic
207 diseases, directly or indirectly through alteration of the gut microbiota[6,11,31]. The gut microbiota
208 and immune system develop contemporaneously in the first years of life, with the first three
209 months being postulated as critical time period, a "window of opportunity", for the establishment
210 of a healthy microbiota[14–17]. Lower overall microbial diversity and reduced abundance of
211 particular bacterial species early in life (1 – 3 months of age) have been associated with the later
212 development of atopy and asthma[23], an association that we replicated in the current study.
213 Further, the protective role of breastfeeding on allergies was also reproduced. Thus, the results of
214 our causal mediation analyses are plausible, that a component of the relationship between CMV
215 infection and allergy would be attributable to changes in the gut microbiota.
216
217 Strengths of this study include the relatively large number of children, the population-based
218 CHILD cohort, and use of allergy outcomes that were objectively assessed using validated and
219 standardized clinical skin prick testing. Limitations of this study include the fact that immune
220 responses to CMV could not be characterized or analyzed with respect to allergic responses. IgE
221 serology was not available, and numbers were too small to assess relationships with asthma,
222 eczema, or other diseases. However, the clinical relevance of our main outcome of persistent
223 allergic sensitization is emphasized by the facts that in the entire cohort, CHILD participants
224 with persistent atopy were 7 to 11-times more likely than healthy controls to be diagnosed with
225 one of asthma, allergic rhinitis, or atopic dermatitis (eczema) at age five years (Figure 2, online).
226 CMV infection was determined by urine PCR, which is extremely sensitive given the prolonged
227 high viral load shedding in urine that follows infection in young children and because maternal
228 antibody complicates serologic testing of infants[32]. Because a urine sample from birth was not

available, we estimate that 6 of the 122 infants with a positive CMV test at 3 months could have been infected congenitally, based on a prevalence of approximately 0.5% in North America. There was a lack of detailed information about other viral infections (e.g., rhinovirus, RSV, EBV) that might be correlated with risk of allergic diseases. Although the CHILD study is population-based, the samples specifically selected for microbiome sequencing possibly generated a skewed subset of children in whom atopy was more likely. Notwithstanding potential selection bias, the study subset had similar risk factors distribution as the rest of the general CHILD cohort.

In summary, early-life CMV infection was associated with evidence of gut microbial dysbiosis and with increased risk of subsequent childhood allergic disease. The relationship between CMV and atopy was estimated to be mediated in part through its effects on the gut microbiome. Additional research is necessary to understand the specific immune responses attributable to CMV infection on the microbiome and risk of atopy. Importantly, our findings indicate that, rather than promoting atopic diseases according to the “old friends” hypothesis, preventing CMV infection may actually reduce the risk of allergy and related problems. As such, the development of an effective CMV vaccine may have benefits over and above reducing the enormous burden of disease caused by CMV infection directly. At a minimum these new findings suggest that prevention of CMV infection does not appear to pose a risk to increasing the risk of allergic disease.

Acknowledgments

Funders played no role in design, analysis, or reporting of this work. We thank the CHILD study staff and the participating children and families. We would like to acknowledge the valuable contribution of Dr. Azad, CHILD study associated Investigator at the University of Manitoba, who provided detailed breastfeeding data information necessary to complete this work.

Impact statement

References

1. Cannon MJ, Schmid DS, Hyde TB. Review of cytomegalovirus seroprevalence and demographic characteristics associated with infection. *Rev Med Virol*. 2010;20(4):202-213. doi:10.1002/rmv.655

2. Gantt S, Orem J, Krantz EM, et al. Prospective Characterization of the Risk Factors for Transmission and Symptoms of Primary Human Herpesvirus Infections Among Ugandan Infants. *J Infect Dis*. 2016;214(1):36-44. doi:10.1093/infdis/jiw076

3. Cannon MJ, Stowell JD, Clark R, et al. Repeated measures study of weekly and daily cytomegalovirus shedding patterns in saliva and urine of healthy cytomegalovirus-seropositive children. *BMC Infect Dis*. 2014;14(1):1-10. doi:10.1186/s12879-014-0569-1

4. Schleiss MR. Role of breast milk in acquisition of cytomegalovirus infection: Recent advances. *Curr Opin Pediatr*. 2006;18(1):48-52. doi:10.1097/01.mop.0000192520.48411.fa

5. Manicklal S, Emery VC, Lazzarotto T, Boppana SB, Gupta RK. The "Silent" global burden of congenital cytomegalovirus. *Clin Microbiol Rev*. 2013;26(1):86-102. doi:10.1128/CMR.00062-12

6. Brodin P, Jojic V, Gao T, et al. Variation in the human immune system is largely driven by non-heritable influences. *Cell*. 2015;160(1-2):37-47. doi:10.1016/j.cell.2014.12.020

7. Moss P. 'From immunosenescence to immune modulation': a re-appraisal of the role of cytomegalovirus as major regulator of human immune function. *Med Microbiol Immunol*. 2019;208(3-4):271-280. doi:10.1007/s00430-019-00612-x

8. Barton ES, White DW, Cathelyn JS, et al. Herpesvirus latency confers symbiotic protection from bacterial infection. *Nature*. 2007;447(7142):326-329. doi:10.1038/nature05762

9. Furman D, Jojic V, Sharma S, et al. Cytomegalovirus infection enhances the immune response to influenza. *Sci Transl Med*. 2015;7(281).

10. Daley D. The evolution of the hygiene hypothesis: The role of early-life exposures to viruses and microbes and their relationship to asthma and allergic diseases. *Curr Opin Allergy Clin Immunol*. 2014;14(5):390-396. doi:10.1097/ACI.0000000000000101

11. Thomsen SF. Epidemiology and natural history of atopic diseases. *Eur Clin Respir J*. 2015;2(1):24642. doi:10.3402/ecrj.v2.24642

12. Dharmage SC, Lowe AJ, Matheson MC, Burgess JA, Allen KJ, Abramson MJ. Atopic dermatitis and the atopic march revisited. *Allergy Eur J Allergy Clin Immunol*. 2014;69(1):17-27. doi:10.1111/all.12268

13. Bantz SK, Zhu Z, Zheng T. The Atopic March: Progression from Atopic Dermatitis to Allergic Rhinitis and Asthma. *J Clin Cell Immunol*. 2014;5(2). doi:10.4172/2155-9899.1000202.The

14. Martin R, Nauta A, Ben Amor K, Knippels L, Knol J, Garssen J. Early life: gut microbiota and immune development in infancy. *Benef Microbes*. 2010;1(4):367-382. doi:10.3920/BM2010.0027

15. Gensollen T, Iyer SS, Kasper DL, Blumberg RS, Medical H. How colonization by microbiota in early life shapes the immune system. *Science (80-)*. 2016;352(6285):539-

544. doi:10.1126/science.aad9378.How

16. Johnson CC, Ownby DR. The infant gut bacterial microbiota and risk of pediatric asthma and allergic diseases. *Transl Res*. 2017;179:60-70. doi:10.1016/j.trsl.2016.06.010

17. Stiemsma LT, Turvey SE. Asthma and the microbiome: Defining the critical window in early life. *Allergy, Asthma Clin Immunol*. 2017;13(1):1-9. doi:10.1186/s13223-016-0173-6

18. Nakase H, Onodera K. Targeting cytomegalovirus during ulcerative colitis flare-ups. *Expert Rev Gastroenterol Hepatol*. 2016;10(10):1119-1125. doi:10.1080/17474124.2016.1192461

19. You DM, Johnson MD. Cytomegalovirus Infection and the Gastrointestinal Tract. *Curr Gastroenterol Rep*. 2012;14(4):334-342. doi:10.1007/s11894-012-0266-4

20. Takaro TK, Scott JA, Allen RW, et al. The Canadian Healthy Infant Longitudinal Development (CHILD) birth cohort study: Assessment of environmental exposures. *J Expo Sci Environ Epidemiol*. 2015;25(6):580-592. doi:10.1038/jes.2015.7

21. Subbarao P, Anand SS, Becker AB, et al. The Canadian Healthy Infant Longitudinal Development (CHILD) Study: examining developmental origins of allergy and asthma. *Thorax*. 2015;70(10):998-1000. doi:10.1136/thoraxjnl-2015-207246

22. Moraes TJ, Lefebvre DL, Chooniedass R, et al. The canadian healthy infant longitudinal development birth cohort study: Biological samples and biobanking. *Paediatr Perinat Epidemiol*. 2015;29(1):84-92. doi:10.1111/ppe.12161

23. Arrieta M-CC, Stiemsma LT, Dimitriu PA, et al. Early infancy microbial and metabolic alterations affect risk of childhood asthma. *Sci Transl Med*. 2015;7(307):307ra152.

doi:10.1126/scitranslmed.aab2271

24. Tran MM, Lefebvre DL, Dharma C, et al. Predicting the atopic march: Results from the Canadian Healthy Infant Longitudinal Development Study. *J Allergy Clin Immunol*. 2018;141(2):601-607.e8. doi:10.1016/j.jaci.2017.08.024

25. Imai K, Keele L, Tingley D. A General Approach to Causal Mediation Analysis. *Psychol Methods*. 2010;15(4):309-334. doi:10.1037/a0020761

26. Embil JA, Haldane E V, MacKenzie RA, van Rooyen CE. Prevalence of cytomegalovirus infection in a normal urban population in Nova Scotia. *Can Med Assoc J*. 1969;101(12):78-81. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1946419&tool=pmcentrez&rendertype=abstract.

27. Pembrey L, Waiblinger D, Griffiths P, Patel M, Azad R, Wright J. Cytomegalovirus, Epstein-Barr virus and varicella zoster virus infection in the first two years of life: A cohort study in Bradford, UK. *BMC Infect Dis*. 2017;17(1):1-18. doi:10.1186/s12879-017-2319-7

28. Joseph SA, Beliveau C, Muecke CJ, et al. Risk factors for cytomegalovirus seropositivity in a population of day care educators in Montréal, Canada. *Occup Med (Chic Ill)*. 2005;55(7):564-567. doi:10.1093/occmed/kqi121

29. Nilsson C, Linde A, Montgomery SM, et al. Does early EBV infection protect against IgE sensitization? *J Allergy Clin Immunol*. 2005;116(2):438-444. doi:10.1016/j.jaci.2005.04.027

30. Nilsson C, Larsson Sigfrinius A-K, Montgomery SM, et al. Epstein-Barr virus and

cytomegalovirus are differentially associated with numbers of cytokine-producing cells and early atopy. *Clin Exp Allergy*. 2009;39(4):509-517. doi:10.1111/j.1365-2222.2008.03147.x

31. Mocarski E. Immunomodulation by cytomegaloviruses: manipulative strategies beyond evasion. *Trends Microbiol*. 2002. http://www.sciencedirect.com/science/article/pii/S0966842X02023934. Accessed May 18, 2016.

32. Gantt S, Bitnun A, Renaud C, Kakkar F, Vaudry W. Diagnosis and management of infants with congenital cytomegalovirus infection. *Paediatr Child Heal*. 2017;22(2):72-74. doi:10.1093/pch/pxx002

Table 1. Demographic and clinical characteristics among children with or without persistent sensitization in study sub-cohort (N= 928). To compare characteristics between groups, Wilcoxon rank sum test, and Fisher's exact test were used for continuous and categorical variables, respectively.

	Cohort	***No Sensitization***	***Persistent Sensitization***	***P-value***
No. participants	928	873	55	
Sites, n(%)				0.01
Edmonton	148 (15.9%)	138 (15.8%)	10 (18.2%)	
Toronto	88 (9.5%)	79 (9%)	9 (16.4%)	
Vancouver	329 (35.5%)	304 (34.8%)	25 (45.5%)	
Manitoba	363 (39.1%)	352 (40.3%)	11 (20%)	
CMV status, n(%)				0.0065
Early CMV	122 (13.1%)	109 (12.5%)	13 (23.6%)	
Later CMV	64 (6.9%)	57 (6.5%)	7 (12.7%)	
No CMV	742 (80%)	707 (81%)	35 (63.6%)	
Antibiotic use within 1 year, n(%)				0.35
	149 (16.1%)	143 (16.4%)	6 (10.9%)	
Child Ethnicity (ref: Caucasian white), n(%)				<0.001
	592 (64.4%)	573 (66.3%)	19 (34.5%)	
Mode of Delivery, n(%)				0.59
Vaginal	697 (76.3%)	659 (76.6%)	38 (71.7%)	
C-Section with labor	119 (13%)	110 (12.8%)	9 (17%)	
C-Section without labor	97 (10.6%)	91 (10.6%)	6 (11.3%)	
Having Older Sibling, n(%)				1
	396 (43.7%)	373 (43.7%)	23 (43.4%)	
Male, n(%)				0.32
	489 (54.8%)	456 (54.4%)	33 (62.3%)	
Birth Weight Z Score				0.030
Median (Range)	-0.1 (-3.1, 4.3)	-0.1 (-3.1, 4.3)	-0.3 (-2.1, 2.5)	
IQR (Q1,Q3)	-0.7, 0.6	-0.7, 0.6	-0.9, 0.2	
Parental Atopy, n(%)				0.021
	739 (81.2%)	688 (80.5%)	51 (92.7%)	
Exclusive Breastfeeding Duration (months)				0.52
Median (Range)	4 (0, 9)	4 (0, 9)	4.5 (0, 9)	
IQR (Q1,Q3)	0.2, 5	0.2, 5	0.2, 5	
Tobacco smoke exposure to age 1 year, n(%)				0.87
	212 (24.5%)	199 (24.4%)	13 (26%)	
Time activity pattern at 6 months (hours)				0.69
Median (Range)	8 (7, 56)	8 (7, 56)	8 (7, 25)	
IQR (Q1,Q3)	7, 16	7, 16	7, 12	
Home Occupancy (adults)				0.0016
Median (Range)	2 (1, 9)	2 (1, 9)	2 (2, 6)	
IQR (Q1,Q3)	2, 2	2, 2	2, 2	
Home Occupancy (children)				0.52
Median (Range)	1 (1, 6)	1 (1, 6)	1 (1, 4)	

IQR (Q1,Q3)	1, 2	1, 2	1, 2

Table 2. **Association between CMV infection and global gut microbiota diversity indices from 3-month stool samples**, estimated using a linear mixed effects model.

	Alpha-diversity			Beta-diversity		
Variables	**Estimated effect**	**p-value**	**95% CI**	**Estimated effect**	**p-value**	**95% CI**
CMV infection status						
Later CMV	0.1	0.53	(-0.21, 0.41)	-0.24	0.11	(-0.53, 0.05)
Early CMV	-0.3	0.015*	(-0.55, -0.06)	0.4	0.00089*	(0.17, 0.62)
Parental Atopy	-0.22	0.043*	(-0.44, -0.01)	0.25	0.018*	(0.04, 0.62)
Child Ethnicity (ref: Caucasian white)	-0.07	0.42	(-0.25, 0.1)	0.08	0.39	(-0.09, 0.24)
Home Occupancy (adults)	0.02	0.86	(-0.21, 0.26)	0.06	0.59	(-0.16, 0.28)
Exclusive Breastfeeding Duration (months)	-0.05	0.0034*	(-0.08, -0.02)	0.05	0.0012*	(0.02, 0.08)

Figure legends

Figure 1. Conditional logistic regression of risk of persistent sensitization by age 5 years in relation to CMV infection status, controlling for parental atopy, home occupancy and child ethnicity.

Figure 2. Conditional logistic regression of persistent sensitization in relation to (A) gut microbiota α-diversity assessed with Chao1 index and (B) β-diversity measured with NMDS ordination of Bray-Curtis distance at 3 months of age

Figure 3. Early-life gut microbiota alterations significantly mediate the relationship between X, CMV infection by 3 months of age, and Y, the outcome of persistent sensitization at age 5 years, as demonstrated (panel A) by a significant average causal mediation effect (ACME) through M, a latent variable constructed from diversity metrics, significantly associated with exposure to CMV and outcome (panel B).

Appendix: Early-life cytomegalovirus infection is associated with gut microbiota perturbations and increased risk of atopy

Description of the Prospective CHILD Birth Cohort:

The CHILD study recruited 3621 pregnant mothers generally in their second or third trimester, 216 in a Vanguard cohort and 3405 in the General cohort. Of 3405 recruited to the General cohort, 77 were ineligible at birth and 32 withdrew before any maternal or child data were collected, leaving 3296 eligible children with some maternal data. A further 32 withdrew before childbirth, leaving 3264 eligible children actually commencing the study.

Of the 3,264 children eligible at birth, the current analysis included 928 children clinically assessed for atopy at age 1, 3, and 5 years of age, for whom CMV status was successfully ascertained and microbiome data was available. Pregnant women aged 18 years or older (19 or older in Vancouver) and infants born after 34 weeks of gestation with no congenital abnormalities were eligible. Parents/guardians completed child health questionnaires at 3, 6, 12, 18, 24, 30 months, 3, 4, and 5 years. Standardized skin-prick tests were completed at ages 1, 3, and 5 years for a panel of common inhalant and food allergens. Urine and Stool samples were collected at study visits scheduled for ages 3 months and 12 months. The primary outcome for this study was persistent sensitization at all three time points. The study sample did not differ significantly from overall CHILD population (Table S1) except for the distribution of children across study centers and maternal parity.

Characterization of child atopy: Children underwent three clinical assessments in the first 5 years of follow up: at ages 1, 3, and 5 years. Along with clinical assessments, children were administered epicutaneous skin-prick tests. At the first year visit, a battery of 6 inhalant (*Alternaria alternata*, cat hair, dog epithelium, house dust mites [*Dermatophagoides pteronyssinus* and *Dermatophagoides farinae*], and German cockroach) and 4 food (cow's milk, egg white, peanut, and soybean) allergens were tested. The

two subsequent visits, 6 additional inhalant allergens (*Cladosporium*, *Penicillum*, *Aspergillus*, Trees Midwest, Grass mix, weeds and ragweed mix) were included. Children were considered sensitized if they produced a wheal 2 mm or larger than that elicited by the negative control (glycerin) to at least 1 of the allergens. In cases in which skin tests were refused, some parents provided the results of external tests performed by other physicians, which were used to determine atopic status. Persistent sensitization, our main outcome of interest, was defined as any positive test at each age for all 3 time points.

Covariates considered for model building: Child sex, study centre, mode of delivery, age- and sex-standardized birth weight (z-score), time activity patterns, antibiotics use in the first year, having an older sibling, exclusive breastfeeding duration, home occupancy variables measured using the number of reported adults or children living in the home, exposure to tobacco smoke (maternal and second-hand), and ethnicity were derived from birth charts and questionnaires (Table S1). Child ethnicity was assigned from self-reported parental ethnicity. Parental atopy (either paternal or maternal) was assessed by standardized allergen skin-prick testing and defined by at least one positive skin-prick (wheal diameter≥2 mm greater than negative control). Parental atopy was selected as a covariate over each parent atopic status in this analysis in order to avoid multi-collinearity threat as these were highly correlated (Chi-squared p value $< 10^{-16}$). Similarly, home occupancy based on the number of adults and children were highly correlated, the former was chosen given a much stronger association with the persistent sensitization outcome.

Characterization of infant gut microbiota: The V4 hypervariable region of the 16S rRNA gene extracted from fecal DNA was amplified by PCR using universal bacterial primers (V4-515f: V4-806r) and sequenced on the Illumina MiSeq platform. Using VSEARCH and Deblur within the QIIME2 pipeline forward and reverse demultiplexed reads were assembled for a final length of 248 bp (unassembled sequences were discarded) and filtered against the GREENGENES reference database (v13.8). Taxonomic classification was achieved using a naïve Bayes classifier trained on reference reads extracted from the reference database at 97% sequence similarity. All subsequent microbiome analyses were

completed in R using the phyloseq package[14]. Quality control filtering resulted in 930 samples with an average reads of 10,430 (2 - 40,649). Samples were rarefied to 5,000 sequencing reads per sample prior to computing all diversity metrics. We used Chao1 index as a measure of the within-individual diversity (α-diversity) while the between-individual diversity metric (β-diversity) was computed as Bray-Curtis dissimilarity distances.

***Characterization of infant CMV infection*:** Using Qiagen's DNeasy® Blood and Tissue kits, DNA was extracted from urine samples. An unmodified pUC19 plasmid was used spiked into each sample as an internal control for quality of extraction and qPCR amplification. Primers and probes designed to amplify a unique section of CMV UL83 gene was used for quantification of viral genome copy number. A pUCIDT-AMP plasmid containing the same CMV-UL83 region was used to generate a quantitative standard curve from which the CMV copy numbers in CHILD study urine sample DNA were calculated. Individual qPCR reactions contained: 5 ul of IDT-PT 2X Master Mix, 1 ul of CMV-UL83 primer/probe mix, 1 ul pUC19 primer/probe mix and 4 ul of extracted DNA from the CHILD Study urine samples or extracted control DNA. The final concentration of primers per reaction was 900 nM, and 250 nM of the probes. The maximum volume of DNA possible (4 ul) was chosen to maximize detection of CMV and provide the greatest assay sensitivity. All qPCR reactions were run in a BIO-RAD C1000 Touch Thermal Cycler Chassis with a CFX96 Optical Reaction Module. Thermal cycling was programmed as recommended by the PrimeTime® Gene Expression Master Mix Protocol as follows: 1 cycle at 95°C for 3 minutes for polymerase activation, then 45 cycles of 95°C for 15 seconds followed by 60oC for 1 minute. Primers, probes, pUCIDT-AMP plasmid and PrimeTime® Gene Expression Master Mix 2X were ordered from Integrated DNA Technologies, Inc. (Coralville, Iowa, USA).

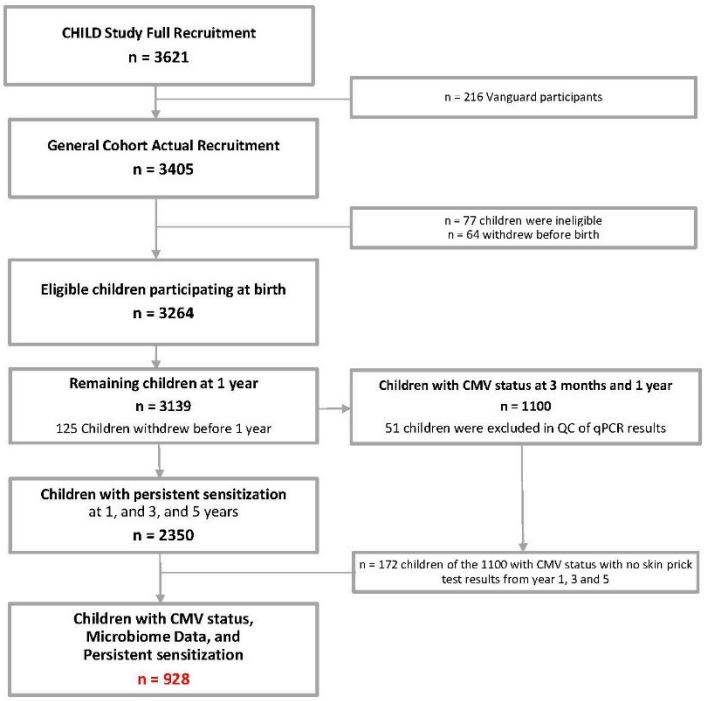

Figure S1: Consort diagram of the CHILD study cohort

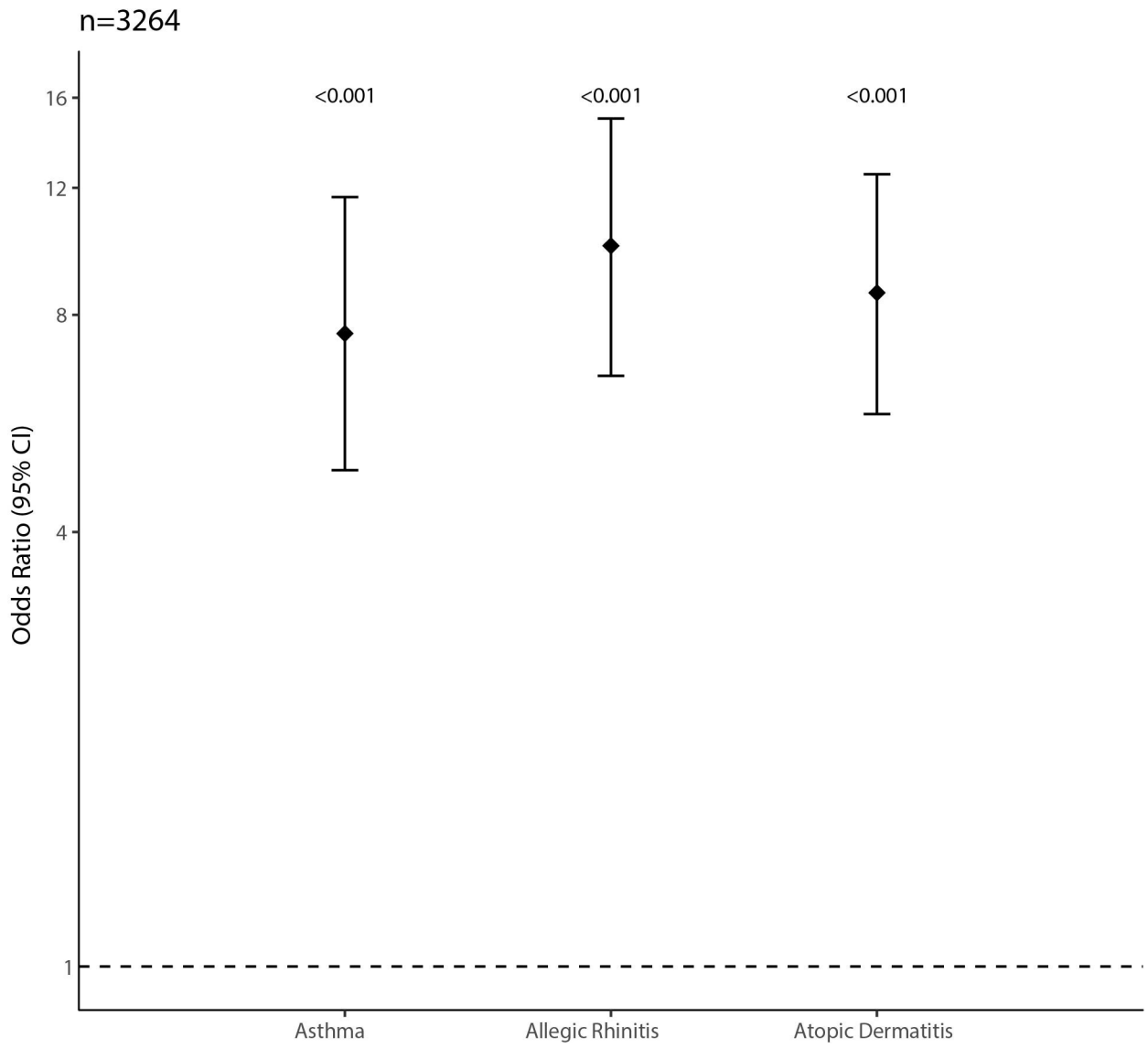

Figure S2: Odds Ratio in the CHILD study of developing physician-diagnosed asthma, atopic dermatitis and allergic rhinitis at five years of age in children with persistent sensitization.

Table S1: Demographic characteristics in the overall CHILD cohort, the sample of children with available sensitization skin prick tests at age 1, and 3, and 5 years and the subset with CMV infection status ascertained by age 1 year and with gut microbiota data.

Variable	Overall CHILD Population	Outcome: sensitized cohort	p value (Outcome vs. Overall)	Study: Microbiota+CMV Subsample	p value (Study vs. Outcome)
No. patients	3264	2350		928	
Sensitization at age 5 years, n(%)			1		0.39
No sensitization	2228 (68.3%)	2228 (94.8%)		873 (94.1%)	
Persistent sensitization	122 (3.7%)	122 (5.2%)		55 (5.9%)	
No phenotype available	914	0		0	
Institution, n(%)			<0.001		**<0.001**
Edmonton	769 (23.6%)	498 (21.2%)		148 (15.9%)	
Toronto	770 (23.6%)	451 (19.2%)		88 (9.5%)	
Vancouver	730 (22.4%)	574 (24.4%)		329 (35.5%)	
Manitoba	995 (30.5%)	827 (35.2%)		363 (39.1%)	
CMV status, n(%)			0.96		1
Early CMV	140 (12.7%)	122 (13%)		122 (13.1%)	
Later CMV	76 (6.9%)	64 (6.9%)		64 (6.9%)	
No CMV	884 (80.4%)	742 (80%)		742 (80%)	
No samples available	2164	1422		0	
Antibiotic use within 1 year (Systemic), n(%)			0.057		0.52
	435 (13.3%)	356 (15.1%)		149 (16.1%)	
Mode of Delivery, n(%)			0.74		0.39
Vaginal	2408 (75%)	1745 (75%)		697 (75%)	
C-Section with labor	425 (13%)	290 (13%)		119 (14%)	
C-Section without labor	390 (12%)	286 (12.2%)		97 (11%)	
No Record	41	29		15	
Having Older Sibling, n(%)			0.91		**0.025**
Yes	1548 (47.4%)	1110 (47.2%)		489(52.7%)	
No Record	59 (1.8%)				
Male, n(%)			0.66		0.43
Yes	1720 (52.6%)	1252 (53.3%)		509 (54.7)	
No record	2				
Birth Weight Z Score			0.72		0.59
Median (Range)	-0.1 (-3.1, 4.3)	-0.1(-3.1, 4.3)		-0.1(-3.1, 3.1)	
IQR (Q1,Q3)	-0.7, 0.5	-0.7, 0.6		-0.7, 0.6	
No Record	182	142		60	
Parental Atopy, n(%)			1		0.96
Yes	2447 (81%)	1863 (81.1%)		739 (81.2%)	
No Record	246	53		18	
Duration of exclusive breastfeeding (Months)			0.16		0.4
Median (Range)	4 (0, 9)	4.2 (0, 9)		4 (0, 9)	
IQR (Q1,Q3)	0.5, 5	0.5, 5		0.2, 5	
No Record	209	41		12	
Tobacco smoke exposure to age 1 year, n(%)			0.006		0.41
Yes	808 (29.6%)	553 (26%)		212 (24.5%)	
No Record	530	223		63	
Home Occupancy (adults)			0.67		0.95
Median (Range)	2 (1,9)	2 (1,7)		2 (1,7)	
IQR (Q1,Q3)	(2,2)	(2,2)		(2,2)	
No Record	687	326		89	
Child Ethnicity, n(%)			0.35		0.78

Caucasian White	2046 (64%)	1525 (65%)	592 (65%)
East Asian	102 (3·2%)	76 (3·3%)	33 (3·6%)
Multiracial	745 (23%)	541 (23%)	230 (25%)
South Asian	78 (2·4%)	47 (2%)	14 (1·5%)
South East Asian	82	63 (2.7%)	21 (2.3%)
Other	140 (4·4%)	77 (3%)	29 (3.2%)
No record	71	21	9

Table S2: Associations between CMV infection and gut microbiota diversity from first year stool samples (N=323), estimated using a linear mixed effects model.

	Alpha-diversity			*Beta-diversity*		
Variables	**Estimated effect**	**p-val**	**95% CI**	**Estimated effect**	**p-val**	**95% CI**
Infection (ref: no CMV)						
Later CMV	0.14	0.55	-0.27, 0.61	-0.02	0.9	-0.39, 0.32
Early CMV	0.13	0.32	-0.11, 0.40	-0.06	0.57	-0.27, 0.15
Parental atopy	-0.06	0.58	-0.44, -0.007	0.11	0.21	-0.081, 0.28
Child Ethnicity (ref: Caucasian white)	-0.11	0.26	-0.25, 0.10	0.08	0.29	-0.079, 0.23
Home occupancy	0.06	0.64	-0.22, 0.26	0.12	0.32	-0.08, 0.32

Table S3: Output of the Causal Mediation Analysis Results. M: mediating variable (gut microbial diversity at 3 months), X: Exposure (CMV infection), Y: Outcome (Persistent Sensitization)

	Estimate	***95% CI***	***p-value***
Mediation results			
Total Effect	0.14	0.012 , 0.27	0.032
ACME	0.042	0.00058 , 0.05	<0.001
ADE	0.097	-0.021 , 0.22	0.12
Prop. Mediated	0.30	0.046 , 1.38	0.032
Regressions	***Estimate***	***Std. Error***	***p-value***
Mediation (M~X)			
Early CMV	-0.47	0.15	0.0018
Parental Atopy	0.31	0.13	0.021
Child Ethnicity (ref: Caucasian white)	-0.082	0.11	0.46
Home Occupancy (adults)	-0.013	0.14	0.93
Exclusive Breastfeeding Duration (months)	-0.067	0.021	0.0013
Edmonton	0.059	0.16	0.70
Toronto	0.0055	0.14	0.97
Manitoba	0.044	0.12	0.71
Full(Y~X+M)			
Early CMV	0.52	0.28	0.066
Later CMV	0.53	0.37	0.15
Latent variable for gut microbiome dysbiosis	-0.48	0.17	0.0061
Parental Atopy	0.91	0.51	0.07
Child Ethnicity (ref: Caucasian white)	0.72	0.24	0.0027
Home Occupancy (adults)	0.45	0.27	0.095
Edmonton	0.20	0.34	0.56
Toronto	0.0099	0.30	0.97
Manitoba	-0.43	0.29	0.14