Exposomics for childhood asthma

Geoffrey L. Winsor, Justin Cook, Karlie Edwards*, Erin E. Gill*, Charisse Petersen*, Karar Al-Mamaar, Amirthagowri Ambalavanan, Spencer Ames, Sonia S. Anand, Marie Arietta, Rilwan Azeez, Allan B. Becker, Lars Bode, Elissa Brookes, Darlene L.Y. Dai, Ruixue Dai, Bassel Dawod, Judah Denburg, Russel DeSouza, Dani Doiron, Qingling Duan, Thomas Eiwegger, Lauren Erdman, Kelsey Fehr, Catherine J. Field, Justin Foong, Emma Garlock, Anna Goldenberg, Emma Griffiths, Stan He, Michelle Helm, Perry Hystad, Meaghan J. Jones, Adrienne Kwan, Luis Ledesma Vega, Brenna Lee, Bryan Lee, Diana L. Lefebvre, Larisa Lotoski, Kelly McNagny, Luisa Mercada Mendoza, Garthika Navaranjan, Krista Pace, Jaclyn Parks, David M. Patrick, Martin Pham, Yu Chen Qian, Myrtha E. Reyna, Hind Sbihi, James A. Scott, Michael Surette, Kristina Szabo, Tim K. Takaro, Piush J. Mandhane, Kozeta Miliku, Elinor Simons, Jeffrey R. Brook, Michael Brudno, Michael S. Kobor, Anita L. Kozyrskyj, Wendy Y. W. Lou, Theo J. Moraes, Malcolm R. Sears, Meghan B. Azad, Stuart E. Turvey†, Padmaja Subbarao†, Fiona S.L. Brinkman†

NOTES:

* Co-third authors

† Co-senior authors

Abstract

Identification of early interventions to reduce/eliminate asthma - the most common chronic disease among children - could significantly reduce burden on the healthcare system. Large-scale asthma Exposome-Wide Association Studies (ExWAS) could identify potential interventions, however integration of diverse data is required to address association confounders. The CHILD Cohort Study has followed 3,454 healthy Canadian children and their families from early pregnancy, collecting exceptionally diverse data including 24,852 variables from participant questionnaires, clinical data, household and neighbourhood-level exposures, and sample-derived chemical analytic/omic datasets. Here, we report integration of these datasets into the CHILDdb database platform, and use these data to perform ExWAS and machine learning analyses, identifying and further characterizing associations between childhood asthma and 2,954 diverse early exposures (pregnancy to age 5). Significant asthma associations include antibiotic use, human milk components, DEHP phthalate, and mother's prenatal cleaning product/disinfectant exposure. Subsequent analysis revealed epigenetic changes in the cord blood at birth, after prenatal cleaner exposure, and different microbiome and/or inflammatory cytokine changes associated with different asthma-associated exposures in the child. Collective results support asthma as a heterogeneous condition involving multiple etiologies, with associated endotypes, including prenatal exposures with potential transgenerational effects, and suggest targets for early interventions.

Introduction

Non-communicable diseases rooted in early childhood often persist into later life[1–4], emphasizing the need to identify early preventative interventions[5,6]. Asthma, the most common chronic disease in children, is a significant burden on the healthcare system and can persist into adulthood, with associated declining pulmonary function, and heightened risk for comorbidities [7–12]. Asthma exhibits heterogeneity in phenotypes[10,11], with diet, antibiotic use, air quality including smoking, and other environmental and lifestyle factors encountered during critical developmental windows appearing to significantly influence asthma risk[13–20]. Emphasis has been placed on the study of early postnatal exposures, however some exposures during pregnancy, including maternal stress, medications, diet, and tobacco smoke, have also been associated with increased asthma risk in a child[21–23]. Additionally, family asthma history and genetic variants influence susceptibility[24].

Large-scale exposome studies, or exposomics (collective study of non-genetic factors impacting health or disease), have identified important asthma associations, including a French NutriNet-Santé cohort study of adults, an INMA cohort study focused only on phthalates, and a New Bedford study focused only on chemicals. However, these studies have been limited in terms of the variety of exposures examined, the diversity of the participants, or the size of the cohort[25–27]. Exposomics must also move toward identification of more causal factors of disease – more adequately factoring in the interrelationships between different disease-associated exposures with each other. This requires harmonized integration of very diverse exposure measurements for a suitably sized, relatively genetically diverse, population-based cohort[23,27–31].

The CHILD Cohort Study (CHILD) represents a unique opportunity to more integratively characterize a wider range of exposures and support more integrative exposomics[30]. CHILD is a large and long running Canadian prospective longitudinal birth cohort, tracking exceptionally diverse data types per participant, with 93.3 million data points to date - over 27,000 data points on average per participant for 3,454 children and their families. The families generally reflect the demographics of Canadians and represent

more ethnic diversity, or larger cohort size, than cohorts used in asthma exposome studies to date[32]. CHILD's recruitment of healthy participants at birth, and collection of data at very early critical development milestones is providing insights into the roles that environmental exposures, microbiomes, genetics, and epigenetics play in the development of asthma, allergy and other childhood diseases[33–36] [31,37].

We now report development of an integrative data platform called CHILDdb that combines harmonized data from the four CHILD sites, including participant questionnaire data, clinical data, household and neighbourhood-level exposures, and sample-derived biological data (chemical analytic and omic datasets). Spanning the 18th week of pregnancy to 8 years of age to date (with ongoing longitudinal data integration), CHILDdb captures rich measures of the exposome at critical windows of child development. Here, we use CHILDdb to perform an asthma Exposome-Wide Association Study (ExWAS), and machine learning analyses, using 2,954 environmental and lifestyle variables, including products used in the home, human milk analytes, nutritional information, chemical analytes, and neighbourhood-level environmental exposures. Asthma-associated environmental factors identified by the ExWAS, after correction for multiple testing, were integratively studied with gut microbiome, epigenetic, and inflammatory marker datasets, to further explore putative disease mechanisms and support hypothesis generation for causation testing.

The collective results highlight the importance of employing a more unbiased approach that integrates diverse prenatal and postnatal environmental factors, identifying early exposures and hypotheses for mechanisms for childhood asthma development - paving the way for timely early, appropriate interventions that can reduce healthcare burden due to asthma.

Results and Discussion

CHILDdb integrates diverse environmental, health, omics, and phenotypic datasets

To our knowledge, the CHILDdb.ca database, plus its web platform for viewing CHILD variables and making data requests, is one of the most diverse data platforms of its kind to date. CHILDdb currently integrates 27,006 variables per participant, from 242 diverse datasets, harmonized across four study centres across Canada (Figure 1; childdb.ca). Data has been collected primarily at 12 time points between the second trimester of pregnancy and age 8 of the child, plus a COVID-19-add-on study[38], and ongoing collection/integration of data at ages 13 and 16as the study continues. Data includes validated questionnaire responses, technician-based home assessments, clinical assessments by asthma/allergy specialists, geospatial data from CANUE: the Canadian Urban Environmental Health Research Consortium[39] linked by postal codes and then anonymized (including socioeconomic and neighbourhood environmental data), omics data (Supplementary figure 1; including microbiome, epigenome and genetic data), and sample-derived chemical analytic data. As a leading prospective, longitudinal birth cohort, CHILD has increased understanding of the early-life exposures impacting asthma and other common allergic diseases[17,40–42] (https://childcohort.ca/publications/). Compared to other well-established child cohorts worldwide[43–49], CHILD captures greater racial and ethnic diversity (approximately 28% of participants were identified as belonging to racialized groups), has a larger cohort size, or more early-life measurements, while reflecting Canadian demographics (Supplementary table 1). The CHILDdb.ca portal facilitates requests for its complex individual-level data, including elasticsearch (https://www.elastic.co/elasticsearch) to support advanced queries (with criteria to prevent participant identification), data visualizations which enable equitable studies with statistical power, review of data requests from researchers, and support for collaboration between

data requesters/analyzers and data generators. Source code is freely available under an open source licence.

An Exposome-Wide Association Study of prenatal and postnatal exposures supports a multifactorial etiology for childhood asthma

Capitalizing on CHILDdb, an Exposome-Wide Association Study (ExWAS) using logistic regression assayed 2,954 exposure variables, from 25 different categories of exposures in CHILDdb, for associations with specialist clinician-diagnosed asthma and recurrent wheeze endpoints in children aged 1 to 8 years (Figure 1; See supplementary table 2 for outcomes examined). After controlling for biological sex[50], study site and season of birth[51], 257 significant associations with 126 distinct exposures were identified as significant after correcting for multiple testing (Supplementary file S1). A second, more stringent ExWAS analysis was then performed on these significant exposures, controlling for previously-established covariates comprising family asthma history[52], antibiotic use[17], presence of older siblings[16], household income[53], breastfeeding[54], birth mode[55], any prenatal smoking in the household[56], study site, season of birth[51] and biological sex[50], to identify any novel associations. After correction for multiple testing, 112 exposures remained significantly associated with asthma and/or recurrent wheeze (Supplementary file S2).

As expected, more established risk factors for asthma, including antibiotic use, biological sex being male, and a family history of asthma were positively associated with specialist clinician-diagnosed “definite” asthma cases at 5 years of age (Figures 2a and 2b). Also as expected, the use of albuterol (salbutamol) or anti-allergenic drugs (medications given for treatment of recurrent wheeze/asthma[57] or allergies, respectively), were positively associated with an asthma diagnosis. The latter is likely indication bias - a form of “reverse causation” - where, for example, albuterol is prescribed because of asthma symptoms rather than causing asthma. Similarly, solid foods (“yes/no” questions) including eggs, peanuts, nuts and dairy products between 18 and 30 months of age was negatively associated with asthma, potentially

reflecting exposure to common food allergens, since food allergy is associated with asthma development[58]. However, though reverse causation may also be a factor, further study of the role of such food allergens is needed.

The ExWAS identified other exposures that were positively associated with 5-year asthma, but have not been as well characterized in the literature. These included the proportion of three omega-6 fatty acids in breast milk, DEHP (bis(2-ethylhexyl)) phthalate levels in house dust (found in certain plastics), at least one vitamin or supplement being given to the infant between birth and 3 months of age, and frequent maternal cleaning product use during the prenatal period (Figures 2a and 2b).

Strikingly, frequent prenatal use of hand sanitizer (OR: 1.71, 95% CI:1.14-2.56) and multiple different cleaning products were among the highest odds ratios. Cleaning products included prenatal weekly oven cleaner and prenatal weekly chemical hand cleaner (e.g. hand degreaser - chemicals used frequently in house cleaning and related occupations - see occupational analysis below) - the latter having a significantly higher odds ratio than antibiotic use, birth to 12 months (or other measures of antibiotic use; OR: 3.12, 95% CI: 1.29-7.56 for the cleaner, versus OR: 1.24, 95% CI: 1.05-1.46 for antibiotic use; Wald test for difference in coefficients $p < 0.05$). This frequent use, specifically prenatally, is correlated with the development of asthma, while cleaning product use at multiple other postnatal timepoints between birth and 5 years of age is not. Similar associations were observed if the mother used weekly chemical hand cleaner prenatally, but stopped the exposure after birth (as might occur if the exposure was occupational before maternity leave).

For an earlier endpoint of asthma at age 3, a greater variety of significant variables were found to be associated with asthma or recurrent wheeze, including frequent use of baby lotion, hours spent in a vehicle, and postnatal exposure to cleaning products/cleaning of homes 4 or more times per month (Supplementary figures 2 and 3). These observations are consistent with the association of vehicle exhaust and other factors with earlier wheeze and lower lung function,[59] and how it may resolve as the child gets older, likely depending on the level and duration of the exposure[60].

At age 8 years, only twelve exposures were positively associated and two exposures negatively associated with definite cases of asthma (Figures 2c and 2d), however, they all overlapped with exposures characterized for the 5-year analysis. Again, weekly prenatal chemical hand cleaner use was associated with higher odds of 8-year asthma than infant antibiotic use from birth to 12 months (OR = 4.73, 95% CI 2.11–10.63 vs. OR = 1.32, 95% CI 1.16–1.51; Wald $p<0.05$) - including mother's use prenatally but not after birth.

Conversely, later asthma diagnoses after age 5 are generally considered more "definitive"[61] and it's enticing to speculate that, in addition to family asthma history (potentially in part reflecting a genetic/epigenetic component of asthma), the reduced number of exposure associations for age 8 asthma may reflect key exposures that are potentially more causal, or associated with those that are casual, for more persistent age 8 asthma.

Frequent use of cleaning products and disinfectants during pregnancy is associated with downstream epigenetic changes in cord blood and enriched pathways characterized by neutrophilic asthma cytokine profiles later in the child

Despite assessment of cleaning product exposures from early pregnancy through 5 years of age, most significant associations with asthma at age 5 were observed during the prenatal period, particularly for maternal use of hand sanitizer, chemical hand cleaner, and household cleaning products including oven cleaner, tile cleaner, and polishes. These associations persisted after adjustment for multiple potential confounding factors, including other newly identified correlated asthma risk factors like DEHP phthalate and breast milk fatty acids, plus additionally when factoring in postnatal exposure to the same cleaning products. (Supplementary results, Supplementary Figures 4a, 4b, 5a, 5b).

Given the prenatal timing of these cleaning product exposures, DNA methylation profiles from cord blood were examined in an epigenome-wide association study where significant associations were observed

between prenatal use of hand sanitizer, chemical hand cleaner, oven cleaner, and other disinfectants (Figure 3, Supplementary file S4). The magnitude of these associations were similar to those previously described following prenatal smoke exposure[62,63].

Most notably, weekly or more chemical hand cleaner use by the mother during pregnancy, which was associated with increased odds of having asthma at 3, 5 and 8 years, was also significantly associated with reduced methylation at three CpG sites (Figure 3a), including cg09445240 in the CHIA gene, which encodes acidic mammalian chitinase, that co-localizes at sites of Th2-mediated inflammation and stimulates chemokine production by pulmonary epithelial cells[64]. Also significant was cg23403192 in the USP16 gene, encoding a ubiquitin specific peptidase involved in activation and proliferation of T cells.

A subcohort of mothers who used hand sanitizer daily during pregnancy, but not at all between birth and 3 months (suggesting occupational use), had multiple significant changes in DNA methylation (Figure 3d), including decreased methylation at a CpG site within CHSY1, a gene associated with increasing the levels of macrophages and neutrophils in mice[65]. This pattern of hand sanitizer use was associated with decreased methylation within DCST1, which plays a role in negative regulation of type 1 interferon signaling[66], and bicarbonate transporter SLC4A7, which regulates macrophage phagosome acidification[67]. To our knowledge, this is the first study reporting epigenetic changes in cord blood associated with prenatal hand sanitizer and other cleaning product use.

There was little evidence of gut microbiome dysbiosis in children at ages 3 months and 1 year (based on alpha diversity) after mother's prenatal cleaning product use. However, weekly or more prenatal use of oven cleaner (again, likely occupational) was associated with increased abundance of *Agathobaculum butyricproducens*, *Bifidobacterium animalis,* and *Sellimonas intestinalis* in the child's gut microbiome at 3 months (Supplementary figure 6a) and prenatal daily hand sanitizer use was associated with increased abundance in the child's gut microbiome of *Bifidobacterium dentium* at 3 months and decreased abundance of *Bifidobacterium pseudocatenulatum* at 1 year (Supplementary figure 6b).

For mothers who used hand sanitizer daily during pregnancy, their children had higher levels of serum adenosine deaminase at age 1 (Supplementary figures 7a-7b) and cytokine levels increased at 5 years

were enriched for pathways including maturation of monocyte-derived dendritic cells (a cell type that has been implicated in atopic asthma pathogenesis[68]), Gene Ontology (GO) biological processes including regulation of cell migration, chemotaxis, and response to interleukin-1, and the "IL-10-downregulated extracellular proteins" portion of the IL-10 signalling pathway – consistent with previous studies showing that reduced IL-10 signalling is associated with increased disease severity in asthmatics[69] (Supplementary figure 7c). Furthermore, cytokines making up the leading edge genes contributing most to these enriched pathways have been independently attributed to a Type 2 low (T2-low) profile that is characteristic of neutrophilic asthma, an asthma endotype that is associated with oxidizing chemical exposures and resistant to corticosteroid treatment [70,71]. While the above observations are associative, and do not demonstrate causation, they have been made possible due to the integration of CHILD Study data into CHILDdb, thus enabling hypotheses to test in future studies regarding the prenatal hand sanitizer-asthma association with asthma. As an example, when taken together, the data presented above allowed us to hypothesize a mechanism for the development of neutrophilic asthma in children whose mothers used daily hand sanitizer as presented in Figure 4.

Children of mothers who had used chemical hand cleaner weekly or more during pregnancy had a 1-year serum cytokine profile that was enriched for the toll-like receptor signalling pathway (KEGG:04620, WikiPathways:WP75), with a leading edge subset also predominantly represented by chemokines associated with a T2-low profile (Supplementary figure 8a). In contrast, weekly or more prenatal oven cleaner use by the mother was associated with higher child eosinophil-to-lymphocyte (β=0.264, 95% CI:0.013-0.514) and eosinophil-to-neutrophil ((β=0.344, 95% CI:0.0.063-0.625)) ratios (linked in other studies to eosinophilic asthma[72–74]) but not associated with enrichment of any inflammatory pathways (Supplementary figure 8b-8e).

The associations between child asthma and mother's use of prenatal cleaning products is supported by evidence from more limited studies in other cohorts: Using a Total Chemical Burden (TCB) approach, the ALSPAC cohort found a positive association between prenatal cleaner use and development of wheeze at 42 months, but did not rule out any contribution from postnatal exposure[75]. A later study by ALSPAC

used an index similar to TCB called Composite Household Chemical Exposure (CHCE) score to show that non-transient wheezing up to 7 years of age is associated with frequent use of chemicals in the home but did not determine if this was a proxy measure for cleaner use during the entire postnatal period, nor identify individual products[76]. Limited additional studies support associations between occupational use of more general 'medical disinfectants' during pregnancy and self-reported asthma or wheeze at early ages[77,78].

While most cleaning product associations with childhood asthma and wheeze in our study were prenatal in origin, CHILD researchers in another study used a Frequency of Use Score (FUS) to demonstrate use of multiple cleaning product during infancy (birth to 3 months) increases the odds of recurrent wheeze or asthma at 3 years, but did not examine prenatal exposure nor examine later time points[79]. Consistent with results from this study, we found that infants from homes that had 4 or more general cleanings per month (0-3m) did have increased odds of developing asthma and recurrent wheeze at 3 years of age (Supplementary figures 2, 3) and this was associated with increased serum IL-4, IL18R1, and Flt3L (Supplementary figures 9a, 9b) and enrichment of GO Biological Processes including "Response to Cytokines" and "Defense Response" at 5 years (Supplementary figure 9d).

Collectively, this work more comprehensively characterises associations between exposure to prenatal cleaners and subsequent childhood asthma or recurrent wheeze occurring between the ages of 1 and 8 years, factoring in also postnatal use, plus a diverse dataset of other exposures. This is also the first study to identify and describe epigenetic changes associated with this exposure that are highly significant and warrant further study. Notably, frequency of home cleaning was not correlated with hand sanitizer, oven cleaner or chemical hand cleaner use, consistent with these latter frequent cleaners' use in an occupational context (ie. house cleaners; example: Supplementary table 3). Similarly, frequent prenatal hand sanitizer was associated primarily with working in healthcare occupations (data not shown), a different group of mothers than those frequently using other cleaners. Venn diagrams of participants with asthma-associated frequent hand sanitizer use, chemical hand cleaner use, and antibiotic exposure do not commonly overlap (Supplementary figure 10).

It should be emphasized though that if these cleaning product exposures are actually playing a causative role for asthma, the specific components are unknown. However, bleach or related cleaners are known to cause inflammation in respiratory epithelium and direct DNA base chemical changes at CpG sites, along with other impacts[80]. Hand sanitizer fragrances can contain phthalates, and phthalates may leach from plastic containers[81], and prenatal exposure to hand sanitizer components such as triclosan have been linked to anthropometric differences in newborns[82].

DEHP levels are positively associated with inflammation and childhood asthma

We observed a positive association between bis(2-ethylhexyl) phthalate (DEHP) measured in house dust collected during a 3-month home visit, and development of asthma at age 5 (OR: 1.86, 95% CI: 1.38-2.5; Figure 2b). Similar to the prenatal oven cleaner findings, dust DEHP concentrations above the 75th percentile were associated with significantly higher eosinophil-lymphocyte (β=0.267, 95% CI:0.076-0.458) and eosinophil-neutrophil (β=0.267, 95% CI:0.043-0.490) ratios in age 1 whole blood (Supplementary Figures 11a,b). GSEA analysis of upregulated cytokines in these samples showed enrichment for GO biological processes related to cation transport and metal ion transport (Supplementary figure 11c). In age 5 whole blood samples, enrichment analysis of downregulated cytokines identified pathways related to chemokine and chemokine receptor binding, genes upregulated by TNF, and granulocyte chemotaxis (Supplementary Fig. 12d) consistent with DEHP putatively acting through a different mechanism than the prenatal cleaners. More study is needed, however, DEHP and its metabolite MEHP have been previously implicated in the inhibition of chemotaxis,[83,84] and DEHP has been linked to eliciting a type 2 inflammatory response in human macrophages[85].

Unlike for cleaning product exposures described above, there were no associations between exposure to DEHP and any infant gut microbiome changes, nor significant epigenetic changes in cord blood DNA

(data not shown). However, given our findings that the primary DEHP exposure associated with asthma is after birth, cord blood DNA epigenetic changes at birth were not expected.

DEHP is positively correlated with monthly nitrogen dioxide (NO_2) levels mapped to the participant family postal code at 9 months prior to birth and 12 months post birth (Supplementary figure 4ab). Ambient NO_2 levels between the prenatal to first year postnatal period have been linked to childhood asthma[86] and increasing concentrations are correlated with higher DEHP levels in the home[87]. DEHP's association with asthma though is maintained when adjusting for a large set of asthma-associated variables (Supplementary figure 4b) including pollution/smoke exposure. Phthalates have been associated with increased lung inflammation and changes in inflammatory cell numbers[88]. Other cohort analyses of phthalates measured in urine, as well as mouse studies of phthalate exposure, support phthalate associations with asthma/impacts on asthma-associated genes, and even suggest a role for epigenetic changes with potential transgenerational effects[89–91], and this association is consistent with a more limited study with reduced confounders performed by the CHILD cohort[92].

Vitamin D supplementation is associated with decreased gut microbiome alpha diversity

Infants that were given at least one vitamin or supplement between birth and 3 months of age received vitamin D 99% of the time, a standard of care recommended by Health Canada for breastfed babies[93], and the initial ExWAS did show supplementation was positively associated with asthma (OR: 1.96, 95% CI: 1.19-3.22). However, a sensitivity analysis was conducted to evaluate the association between vitamin D supplementation (all sources) and asthma at age 5. After adjusting for an expanded set of covariates, including other asthma-associated exposures that correlated with vitamin D (including DEHP; Supplementary Figure 4a), the association was no longer observed (OR: 1.28, 95% CI: 0.87-1.87; Supplementary figure 4b, with similar analyses done for other asthma-associated exposures).

Further study though is required, since vitamin D use was associated with significant dysbiosis of the infant's gut microbiome at 3 months, including significantly reduced alpha diversity (Supplementary

figure 12a), particularly in exclusively breastfed infants (e.g. Chao1, β=-8.37, 95% CI: [-10.72,-6.02]; Shannon, OR:-0.26, 95% CI: [-0.38,-0.15], Inverse Simpson, OR: -1.49, 95% CI:[-2.04,-0.93]) after stratification by breastfeeding status (Supplementary figure 12b).

Sex-specific association between proportion of adrenic acid in human milk and later childhood asthma

Three human milk fatty acids were associated with asthma in our initial ExWAS analysis: adrenic acid (AdA; OR: 1.56, 95% CI: 1.25-1.94), gamma linolenic acid (GLA, OR: 2.12, 95% CI: 1.41-3.17) and dihomo-gamma linolenic acid (DGLA, OR: 2.9, 95% CI: 1.49-5.64). When an additional sensitivity analysis was conducted to address additional confounders not included in the original ExWAS (infant age at time of milk collection, and proportion of AdA, GLA, and DGLA components, combined), the analysis confirmed a positive association between AdA and development of asthma at 5 years (OR: 1.98, 95% CI: 1.12-3.49; Supplementary figures 13a, 13b).

Given that omega-6 fatty acid levels in human milk have been positively-associated with food allergy in a sex-dependent female manner[42], and AdA/DGLA levels are associated with maternal obesity and other overweight phenotypes[94], we stratified the cohort by sex and added an additional control for maternal BMI. Female infants exposed to breast milk AdA proportions in the highest quintile had increased odds of asthma at 5 years (OR:2.92, 95% CI:1.02-8.32) and 8 years (OR: 4.81, 95% CI: 1.48-15.66) , whereas no association was observed in males (Figures 5a, 5b). Because early-life wheeze is more common in boys (thought in part due to transient, airway size-related phenotypes that often resolve[95]) this sex-specific pattern may indicate that AdA is associated with more persistent asthma that continues throughout life, rather than recurrent early wheeze. An additional or complementary explanation could involve emerging sex-hormone influences in females by mid-childhood. While prior work has linked AdA to asthma-like symptoms[96], to our knowledge this is the first report of a sex-specific association between breast milk

AdA and asthma. Further GSEA analysis of upregulated cytokines is consistent with a T2-low pattern reported in neutrophilic asthma (Figure 5).

Gradient boosting–based machine learning identifies additional early-life asthma risk factors and interactions beyond ExWAS

Several established asthma severity factors, including tobacco smoke and traffic-related air pollution, were not statistically significant in our ExWAS.[60]. We hypothesized that they function in combination with other factors, and so we additionally performed machine learning analyses, which complement ExWAS by also identifying exposure combinations predictive of asthma..

Previously, models including wheezing status and atopy have been used to effectively predict asthma at 5 years of age for clinician use[41], but we further hypothesized that a model built using early-life environmental exposures and family health history could be used to also identify asthma-associated exposures. In this case we would not want to include intermediate measures of wheeze and atopy that are highly correlated with the asthma outcome and confound the identification of important early exposures. We constructed a 130-feature training dataset of early-life environmental exposures (those associated with asthma via ExWAS or other well known factors described in the literature), plus early child history including birth chart and weight for age between birth and 1 year, and parental health history (features listed in Supplementary file S6).

A holdout dataset used to test the best performing model had a recall of 0.818, Area Under the Receiver Operating Curve (AUROC) of 0.836 and an Area Under the Precision Recall Curve (AUPRC) of 0.161, consistent with multiple factors potentially causing asthma, but not with high precision due to a likely stochastic component to asthma development (for example, high cleaning product exposure would not definitively cause asthma, but increase the risk of epigenetic changes that impact asthma associated pathways). Determination of the top features contributing to the predictions was assessed by calculating SHAP (SHapley Additive exPlanations) values, which have the added benefit of providing insight into the

direction of effect (Figure 6a). Top features included DEHP levels in house dust, family asthma history, father history of atopic dermatitis, use of antibiotics, and use of chemical hand cleaner and hand sanitizer by the mother during pregnancy (all found in the ExWAS), plus additional features linked to primary home residence including pollution levels (several involving PM2.5 but also O3 and NO2), consistent with evidence that air pollution can exacerbate or otherwise play a role in asthma. SHAP interaction values indicating any combined effect of two features, after accounting for their individual effects, were calculated, and the highest ranking ones included interactions with DEHP or a family history of asthma (Figure 6b). Two-dimensional partial dependence plots used to visualize interactions between exposures suggest that exposure to higher levels of air pollution, particularly PM2.5, but also prenatal exposure to O3 and NO2, can enhance associations between a family history of asthma and odds of the child developing asthma (Figure 6c). Similar results were seen for the use of chemical hand cleaner and hand sanitizer during the first trimester of pregnancy. Results also suggested that exposure to higher levels of O3 during pregnancy can enhance the association between use of antibiotics and odds of developing asthma at age 5 (Supplementary figure 14). While this association was with prenatal O3, another study has shown ozone-induced changes in mouse serum metabolomes included metabolites capable of influencing pulmonary responsiveness, and that this profile was altered in mice that had microbiomes depleted using antibiotics[97].

There was a notable overlap in important features (e.g. DEHP, antibiotic use, prenatal cleaning product use, and family asthma history) identified by the machine learning and ExWAS approaches using different wheeze or asthma endpoints. Models like gradient boosting complement logistic regression/ExWAS by using a sequential, greedy approach to identifying multiple “weak learners” that can reduce the overall classification error[98] while the ExWAS allows you to test exposures independently. The overall performance of the different models is consistent with a stochastic process leading to asthma development. Some factors including family asthma or atopy history may represent genetically inherited factors, or transgenerational epigenetic effects, that greatly increase childhood asthma risk, with effects potentially enhanced through interactions with other factors including air quality. This evidence for

interactions may explain inconsistency of results between studies and the absence of exposures associated with air quality from our ExWAS analysis. In addition, use of prenatal cleaners may result in epigenetic changes that stochastically impact inflammatory or other genes that may not definitively result in progression to asthma.

Toward identification of early interventions that reduce/eliminate asthma

Integration of CHILD's rich, diverse, and longitudinal datasets into CHILDdb enabled a more wholistic, unbiased approach that validated previously-described exposure associations with asthma, and identified other novel risk factors that we linked to epigenetic, inflammatory, and microbiome changes (Figure 7). Furthermore, GSEA identified exposure-specific pathway enrichment characterized by cytokine signatures linked to distinct asthma endotypes (Figure 7). Notably, prenatal cleaning product exposures and high levels of AdA in breast milk were associated with enriched pathways driven by a chemokine signature characteristic of a T2-low response. In contrast, infants given 3 or more systemic antibiotic courses between birth and 12 months had enriched pathways driven by chemokines that have been linked to T1-high responses in the literature (Supplementary figure 15), putatively representing an emerging T2-low subgroup versus what was observed for cleaning products and AdA, though the net effect can be similar (Figure 7). This work suggests targets for further study, such as the breast milk fatty acid AdA (potentially a marker of diet, genetic/epigenetic, or mother's inflammatory condition associated with her child's more persistent asthma). It further suggests early interventions to reduce the prevalence of asthma, such as avoiding certain cleaning product exposures during pregnancy (in particular reducing occupational use), in addition to actions already being taken to reduce child antibiotic use that we previously found had measurable benefits at the population level[17]. This work proposes potential mechanisms for future exploration, such as epigenetic changes in asthma-relevant genes due to prenatal cleaning product exposure. Such exposures can cause DNA methylation changes, then inflammation itself can impact the epigenome, with further potential transgenerational effects that may in part explain the family history association with asthma[80,99,100]). Notably, both cleaning product use and antibiotic use can

impact the microbiome, and changes in microbial diversity are a variable of importance in the machine learning analyses, supporting the importance of the microbiome in health, and asthmatic disease. In addition, this work further supports the role of air pollution in exacerbating asthma (Figure 6).

However, one must be careful about inferring causation, as for example hospital workers with frequent hand sanitizer use may be collectively exposed to other strong cleaners/disinfectants in such an environment that are the true causative agent. Later antibiotic use after 7 months, not birth to 6 months, is most associated with asthma, which could be a mix of causation and reverse causation (those with asthma may have more respiratory infections and require more antibiotics). The phthalate DEHP, which also correlates with other risk factors for asthma including NO2 and frequent oven cleaner use (Supplementary figure 16), may be both a causative factor, as supported by other mechanistic studies, as well as a marker for a lifestyle involving a collection of exposures that increase asthma risk. Regardless, DEHP exposure, antibiotic use, some cleaning product use, and breast milk components show up consistently in both ExWAS and machine learning-based analyses of factors of importance in asthma. While a limitation of our study is its focus primarily on one cohort, the ability of our study to address so many cofounders is a strength, as well as the cohort's representation of population demographic diversity over four harmonized sites.

This work validates other smaller studies and has important implications for occupational health policy regarding exposures of pregnant women, as well as child exposures. CHILDdb, through consistent, harmonized integration of such diverse data, is a resource that can be applied for other investigations of child health and disease, including other allergic diseases or diseases that have a combined environmental and genetic component.

Data and code availability

Data described in the manuscript are available by registration to the CHILD database https://childstudy.ca/childdb/ and the submission of a formal request. More information about data access for the CHILD Cohort Study can be found at https://childstudy.ca/for-researchers/dataaccess/. Researchers interested in accessing CHILD Cohort Study data for their research should contact child@mcmaster.ca.

Acknowledgements

We thank the CHILD Cohort Study (CHILD) participant families for their dedication and commitment to advancing health research. We thank the whole CHILD Cohort team, which includes interviewers, nurses, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers and receptionists. Key support for this work was provided by the Schroeder Allergy and Immunology Research Institute, Genome Canada, Genome BC, and CIHR, with additional support by Simon Fraser University and the Digital Research Alliance of Canada.

Methods

Cohort Description

The CHILD Cohort Study is a prospective longitudinal birth cohort study that recruited 3621 women with singleton pregnancies from Vancouver, Edmonton, Manitoba (Winnipeg and Morden/Winkler), and Toronto between 2008 and 2012, from which 3,454 delivered healthy, full-term infants, were eligible to commence the study (Supplementary table 1). Of the 3,454 infants eligible at birth, 3,133 remained enrolled in the study at the 8-year time point. Participants were phenotyped for conditions including specialist clinician-diagnosed asthma, multiple wheeze and body mass index (BMI) trajectories, food allergies, allergic rhinitis, atopic dermatitis, lung function, and measures of child and mother mental health and behaviour. Questionnaire and environmental datasets, clinical visits, genetic, microbiome and other omics data have been collected for multiple time points ranging from the 18th week of pregnancy to 13 years and beyond.

Ethical approval for the CHILD Cohort Study, including the oversight of the CHILD biological samples and the CHILD database (CHILDdb), was obtained from the local Research Ethics Board of each study site: the University of British Columbia, the University of Alberta, the University of Manitoba, the Hospital for Sick Children, and McMaster University.

The CHILD Cohort Study emphasizes participant and public involvement in all stages of its research. Research questions are informed and generated based on relevance to participants and their families to study those affected by childhood health conditions. Efforts are made to ensure participants' involvement is meaningful, comfortable, and aligned with their preferences. Study results will be shared with participants through various channels, including newsletters and online platforms. Academic and public health publications will also communicate findings. We acknowledge and thank all participant advisers and study participants for their valuable contributions.

Outcomes measured

A definite asthma diagnosis by a specialist physician was based on clinical assessments performed at 3, 5 and 8 years of age whereby the diagnosis considered a combination of the clinical history of symptoms and bronchodilator response[32,101]. Recurrent wheeze (assessed for ages 1, 3 and 5 years) was assigned to children having more than 2 episodes of wheeze in the past 12-month period based on parental responses to child health questionnaires that were administered at multiple time points between 3 months and 5 years of age.

Measurement of human milk components

Human milk fatty acids were measured by gas liquid chromatography as previously described[102], and human milk oligosaccharides (HMOs) were measured using high-performance liquid chromatography, also as previously described[103].

Exposome-Wide Association Study (ExWAS)

In order to identify exposures and other lifestyle factors associated with asthma and/or recurrent wheeze, a logistic regression-based exposome-wide association study (ExWAS) was performed. The ExWAS assessed the independent contribution of 2,954 variables representing common categories of exposures encountered between the 18th week of pregnancy and 5 years of age of the child or parental health history prior to birth, and measures of mothers' mental health between the 18th week of pregnancy up to 3 years post-birth (Supplementary methods table 1, Supplementary file S3), with the majority (2258/2954, 76%) representing exposures encountered between the prenatal period and age 1. Binary variables reflecting "daily", "weekly or more", or "monthly or more" use of cleaners, disinfectants, air fresheners, and solvents were derived specifically for this study from ordinal home environment questionnaire responses (more details in supplementary methods). The ExWAS analysis was performed in Python using[104]the StatsModels module[104] where two rounds of logistic regression iterated over the 2,954-variable exposure

dataset. The first ExWAS iteration (round 1) tested one independent exposure variable against each of the outcome phenotypes and controlled for study site, season of birth and biological sex of the child. Only participants with complete data for the independent exposure variable and the wheeze/asthma outcome were included. There were no missing values for the covariates used in the first round. The Benjamini-Hochberg method (based on implementation in StatsModels) controlled for false discovery rate (FDR) using a threshold of 0.05. Associations where the FDR < 0.05, the number of asthma or wheeze cases in the exposed group being >=10, and adjusted odds ratio values <0.9 or >1.1 were reported. Significant exposures from the first round of logistic regression were added to a list to be used for a second more stringent round of ExWAS controlling for family history of asthma, number of antibiotic courses between birth and 1 year of age, presence of an older sibling, household income, breastfeeding (exclusive breastfeeding at 6 months of age), birth mode, any exposure to cigarette smoke during pregnancy, study site, season of birth, biological sex and having two Caucasian parents. Only participants with complete data for the independent variable and wheeze/asthma outcome were included in round 2, however missing values for the other covariates were imputed (median for continuous variables and counts, mode for binary, ordinal variables). Significant associations involving exposures with adjusted odds ratios < 0.9 or > 1.1 and FDR < 0.05, where at least 10 asthma/wheeze cases were exposed, were reported.

Construction of an exposome correlation globe

Exposures that were significantly-associated with either 5- or 8-year asthma, and ExWAS covariates (family history of asthma, presence of an older sibling, household income, exclusive breastfeeding at 6 months of age, any breastfeeding at 3 months of age, birth mode, any exposure to cigarette smoke during pregnancy, biological sex and ethnicity) underwent an all-versus-all Spearman rank correlation analysis (scipy.stats, version 1.5.2). For each correlation, at least 100 participants were required after dropping participants with missing values. False discovery rate was controlled for using the Benjamini-Hochberg

method with FDR cutoff of $q<0.05$. The remaining pairs of correlations with an absolute value of $>= 0.1$ were visualized in an exposome correlation globe generated using Circos[105,106].

To provide more insight into other exposures that may be co-occurring with significant exposures identified by ExWAS, and justify the addition of additional covariates to more stringent downstream logistic regression analyses, we performed ad hoc correlation analyses between the following exposures found significant in the initial ExWAS: infant vitamin D supplementation (birth-3m), dust DEHP levels, breast milk fatty acids (GLA, DGLA, AdA), or prenatal cleaner use (hand sanitizer, oven cleaner, chemical hand cleaner), and the other 2983 variables in the starting ExWAS dataset. These correlation tests used the same cutoffs as the correlations performed for the correlation globe analysis.

Ad hoc multivariable logistic regression to validate associations with asthma

A logistic regression using an expanded list of 21 covariates was performed for further validation of significant 5-year asthma associations identified by the original ExWAS analysis. A more stringent model was adjusted for covariates included in the original ExWAS models: family history of asthma, presence of an older sibling, household income, breastfeeding status at 6 months, birth mode, cigarette smoke exposure, study site, season of birth, biological sex, and Caucasian ethnicity (both parents identified as Caucasian). This analysis also adjusted for DEHP (bis(2-ethylhexyl) phthalate), breast milk fatty acid proportions (AdA, GLA, and DGLA), number of antibiotic courses between birth and 1 year, prenatal exposure to daily hand sanitizer, weekly or more oven cleaner, and weekly or more chemical hand cleaner, vitamin D supplementation between birth and 3 months, and frequency of oven cleaner use between birth and 3 months. Odds ratios with 95% confidence intervals were estimated and exposures with a p-value < 0.05 were considered statistically significant. Missing values were imputed using the median value.

Measurement and analysis of cytokine levels

The Olink Target 96 Inflammation Panel (Thermo Fisher Scientific, Waltham, MA) was used to assay 92 protein biomarkers in serum collected from 1- and 5-year-old CHILD participants. Concentrations were analyzed in Olink's Normalized Protein eXpression (NPX) unit (Thermo Fisher Scientific, Waltham, MA). Differences in median between exposure groups were assessed using a non-parametric Mann-Whitney test and corrected for multiple testing using the Benjamini-Hochberg method with FDR=0.05. For select exposures, the OlinkAnalyze version 3.8.2) package[107] in R was used to perform a gene set enrichment analysis using the olink_pathway_enrichment function while specifying the "GSEA" method, a search against the "MSigDB" database and providing a dataframe of results from a two-sided t-test performed using the olink_ttest function. The top significant enriched pathways for each exposure, following correction for multiple testing using the Benjamini-Hochberg method with adjusted p-value (more conservative versus q value) less than 0.05 were returned.

Stool sample collection

Sample collection and sequencing were performed as previously described[108,109]. Specifically, stool samples from diapers were collected at a home visit at around 3 months [mean (SD), 3.8 (1.1) months] and a clinic visit at around 1 year [mean (SD), 12.5 (1.6) months]. Samples were aliquoted into four 2-mL cryovials using a stainless steel depyrogenated spatula and were frozen at −80°C.

Shotgun metagenomic sequencing

Shotgun metagenomic sequencing data were generated by Diversigen (Minneapolis, MN, USA) from fecal samples (average depth of 4.85 million reads per sample, SD[1.79 million]) [109] previously described[109]. DNA was extracted from samples using the MO BIO PowerSoil Pro with bead beating in 0.1mm glass bead plates, with high-quality input DNA verified using Quant-iT PicoGreen. Libraries were prepared and sequenced on an Illumina NextSeq using single-end 1 x 150 reads. Low-quality

(Q-Score<30) and length (<50) sequences were removed, and adapter sequences were trimmed. Host reads were removed, and 2,846 samples (1,425 from 3-month visit and 1,421 from 1-year visit) with at least 1 million remaining reads or more were retained for downstream analysis. The accession number for the shotgun metagenomic data reported in this paper is BioProject accession (NCBI): PRJNA838575. Quality filtered fastq files were processed via Humann3 (version 3.5) and Metaphlan3 (version 3.0.13) to produce functional profiles and taxonomic classifications, respectively, using the Chocophan database (v201901b)[110]. Species represented in fewer than 10% of total samples or less than 0.01% of total abundance were removed, resulting in the retention of 73 species.

Analysis of gut microbiota

For 3-month and 1-year stool samples, Shannon and Inverse Simpson diversity were calculated using the microbiome R package[111] while Chao1 diversity was calculated using Phyloseq[112]. Given the degree of variation in collection time for the 3-month and 1-year events, samples collected between 2.5 and 4 months of age were retained for the 3-month analysis while samples collected between 11 and 13 months of age were retained for the 1-year analysis. In order to test for associations between asthma- and wheeze-associated exposures and alpha diversity, mixed linear modelling was conducted using the statsmodels (version 0.12.1) mixed_linear_model module. For each regression performed, study site was added as a random effect, while breastfeeding status, delivery mode, presence of an older sibling, family asthma history, use of systemic antibiotics, sample processing time (time between collection and deep freezing the aliquots), and exact age of stool sample collection (weeks) were added as covariates. False discovery rate was controlled by adjusting P-values using the Benjamin-Hochberg method.
Relative species abundances were log-transformed, and the “Maaslin2” package was used to perform linear mixed-effects modelled (MaAsLin2 function)[113] to identify associations with prenatal cleaning product and hand sanitizer use while specifying study center location as a random effect and adjusting for stool sample of collection age, genetic ancestry of infants PC1-PC10[114], family income at prenatal enrollment, and biological sex. Centered log-ratio transformation (clr) was performed by first adding a

pseudocount of one half of the minimum relative abundance value following a clr transformation using the compositions package[115]. To identify associations between species and the "Healthy/Non-Allergic" phenotype, Maaslin2 was used to build linear mixed-effects models with study center location as random effect and adjusting for stool sample collection age (defined as time between collection and deep freezing the aliquots). For vitamin D supplementation, the Maaslin3 package[116] was used to identify associations with both abundance and prevalence of species with study center location as random effect while adjusting for birth mode, exclusive breastfeeding, older children in home, systemic antibiotic use, age of infant, and sample processing time. Only features that had an FDR <0.05 and an absolute coefficient value greater than 0.2 were retained for plotting in the heatmap. The heatmap was created using the complexHeatmap package[117].

Microarray genome-wide DNA methylation profiling

DNA was extracted from cord blood samples using the DNeasy Blood & Tissue Kit (Qiagen, Venlo, The Netherlands), samples were bisulfite converted using EZ-96 DNA Methylation kit (Zymo Research, Irvine, CA,) and DNA methylation profiles of the samples were measured with the Infinium MethylationEPC BeadChip array (EPIC) (Illumina, San Diego, CA). Raw intensity IDAT files for 866,836 data points encompassing 863,904 CpG sites were produced and subsequently preprocessed for downstream analysis.

DNA methylation preprocessing and data reduction

All available cord blood DNA methylation data (n = 838) were processed in RStudio (version 4.0.3) as previously described[118,119]. Briefly, the ewastools package[120] was employed for sample quality control to assess the technical parameters, the minfi package was used to assess intensities and check sex concordance and the lumi package was used to detect potential outliers[121–123]. Probe filtering was

performed based on the Pidsley annotation to remove cross hybridizing and polymorphic probes, and BMIQ normalization was done in conjunction with noob to account for probe type bias and background correction[124,125]. Fifteen of the 838 samples were removed due to failure of one or more quality metrics. Finally, to better account for known technical confounding factors, variation associated with batch (i.e., chip and row) was assessed and removed using the ComBat function from the sva package[126]. After all preprocessing steps, a total of 813 samples and 786,363 probes were included in the cleaned sample set. For further data reduction, DNA methylation probes were subjected to interquartile range filtering to subset only variable probes where the DNA methylation β value varied by at least 5% across samples in the 95th percentile[127,128]. A total of 356,037 variable probes were included in the Epigenome-wide association analysis. Of the 813 samples included in the DNA methylation data set, a further subset of 808 children had data regarding exposure to the cleaning variables of interest. A robust linear regression model using Huber M-estimation in the MASS package was used for site-by-site epigenome-wide association analysis[129]. The following formula was employed for each exposure variable:

$$\text{CpG} \sim [\text{Exposure Variable}] + \text{Biological sex} + \text{Gestational age} + \text{Birth mode} + \text{Genetic ancestry principal components} + \text{Cell type principal components} + \varepsilon$$

Biological characteristic covariates known to contribute significantly to DNA methylation differences, i.e., sex, gestational age, genetic ancestry, birth mode, and cell type proportions were included in the models[130]. CpG sites were determined to be significant if they met a Benjamini–Hochberg adjusted p-value threshold of < 0.05 and had a delta beta greater than 3%.

Prediction of childhood asthma and identification of most important features

A gradient boosting classifier was used to predict 5- year asthma cases using a training dataset of 130 including early environmental exposures, parental health history (including asthma, atopy and mental

health), child birth, chart and child anthropometric data (weight for age at birth, 3 months, and 1 year). In order to avoid using variables that reflect early manifestations of asthma that could potentially dominate the models, we purposely excluded variables describing children's wheeze, asthma and allergy.

We included a baseline variable dataset used previously to predict definite cases of 5-year asthma based on specialist clinician-diagnosis [41]. Baseline was supplemented with significant variables from the round 1 ExWAS analysis where the FDR cutoff was relaxed to 0.1, and a dataset of other variables previously linked to childhood asthma, including CANUE variables linked to the postal codes of primary residences (Supplementary file S6).

Prior to building the model, all features were confirmed to have Spearman rank correlations < 0.9 with each other. Participants containing contradictory parent asthma histories based on responses in 2 separate questionnaires were dropped. Rows where either the 3-year or 5-year asthma diagnosis were missing were also dropped, as well as any rows where 3-year or 5-year asthma was classified as 'probable' but not 'definite' based on criteria defined previously[32]. The number of pregnancies, duration of stay in hospital after the child was born, and implied duration of breastfeeding were transformed using the natural logarithm of one plus the input array in order to compress their distributions. Discrete variables were created from continuous data representing mother mental health Perceived Stress Scale[131] and Center for Epidemiologic Studies Depression Scale[132]). No further processing was performed on the other features prior to generating the training and validations sets.

The outcome variable was 5-year asthma based on diagnosis by a specialist clinician. Participants that had missing data in the baseline dataset were excluded from the analysis, leaving 1672 participants without asthma and 108 participants for the 5-year endpoint. Data was split into 70% training / 30% validation (holdout) datasets. The training set underwent stratified 3-fold cross-validation with missing values in each fold being imputed using mode for categorical variables and median for continuous variables. The gradient boosting classifier used log loss for internal optimization and hyperparameters were tuned on each fold. Cross validation of models was assessed using log loss and an optimal decision threshold based on a target recall of 0.8 applied in order to prioritize identification of true positive asthma predictions.

Recall, AUROC and AUPRC were calculated for the holdout dataset and SHAP values were calculated for the holdout dataset using the shap package (version 0.46.0) in Python. SHAP interaction values were calculated for the top 20 features using the shap package and the resulting mean absolute interaction for each pair was plotted in a heatmap.

Tables and figures

Figure 1

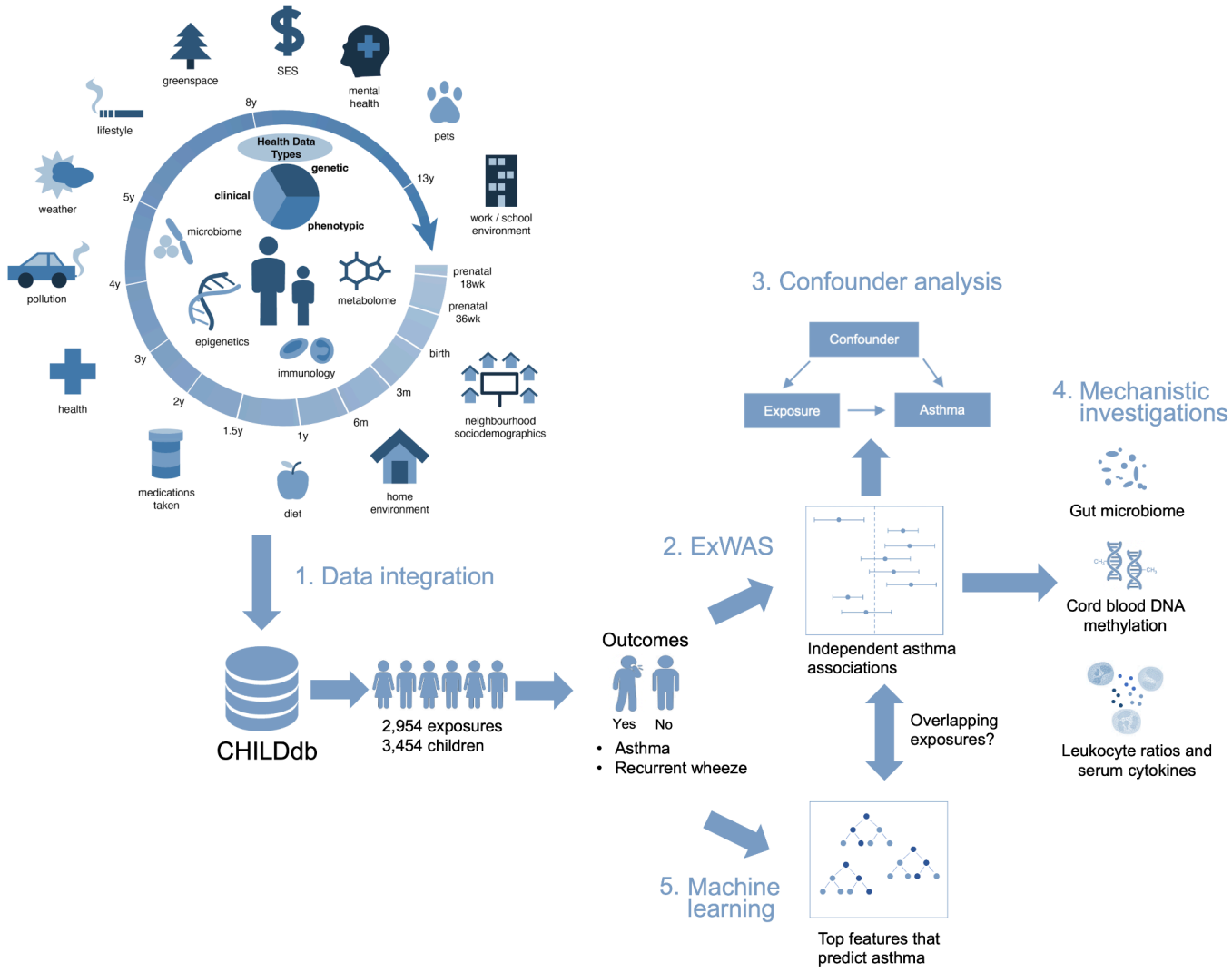
greenspace
SES
mental health
lifestyle
pets
weather
work / school environment
pollution
health
medications taken
diet
home environment
neighbourhood sociodemographics
Health Data Types
genetic
clinical
phenotypic
microbiome
metabolome
epigenetics
immunology
8y
13y
5y
4y
3y
2y
1.5y
1y
6m
3m
birth
prenatal 36wk
prenatal 18wk
1. Data integration
CHILDdb
2,954 exposures
3,454 children
Outcomes
Yes
No
• Asthma
• Recurrent wheeze
2. ExWAS
Independent asthma associations
3. Confounder analysis
Confounder
Exposure
Asthma
Overlapping exposures?
4. Mechanistic investigations
Gut microbiome
Cord blood DNA methylation
Leukocyte ratios and serum cytokines
5. Machine learning
Top features that predict asthma

Fig.1 | CHILD Cohort Study data integration enables characterization of associations with the early-life exposome. Data collected at time points ranging from enrollment during pregnancy, to the 8th year visit were integrated into CHILDdb (13-year data collected and currently undergoing cleaning). Sample-derived data (inside the ring) includes gut, nasal, and breast milk microbiome, epigenetic, metabolomic, and immunological data, along with clinical test results, participant phenotypes and genetic variant data. In addition it contains diverse questionnaire-derived datasets and external-derived socioeconomic and environmental datasets linked to securely stored, anonymized participant postal codes. After integration in CHILDdb, a dataset consisting of 2954 exposures (most exposures encountered between pregnancy and 3 months of age) for 3454 study participants underwent an exposome-wide association study (ExWAS) analysis to identify independent associations with asthma and recurrent wheeze. Significant exposures underwent more stringent analysis with additional covariates added in order to further confirm the associations. These exposures were then subjected to additional mechanistic investigation to provide evidence for putative mechanisms of action. In order to provide more support for exposures identified by ExWAS, additional machine learning approaches were used to predict asthma from a dataset of primary exposures (hypothetical initial exposures) and features that were most important for the model were identified.

Figure 2

a
5-year asthma
Chemical hand cleaner use by mother (prenatal)
DEHP
Family asthma history
Given eggs and nuts (or peanuts) between 18m-2y
Adrenic acid (breast milk)
Oven cleaner use by mother (prenatal)
Dairy products before 2y
Poly-vi-sol multivitamin use (24m-30m)
Number of antibiotic courses (birth-1y)
← Decreased odds of asthma
Increased odds of asthma →
Cleaners
Nutrition
Breast milk
Medication / Antibiotic
Other
Prenatal
Not prenatal
-log10(P-value)
log2(OR)
b
5-year asthma
← Decreased odds of asthma
Increased odds of asthma →
Breast milk
Adrenic acid (birth-3m) 856/110
Dihomo-gamma-linolenic acid (birth-3m) 856/110
Gamma-linolenic acid (birth-3m) 856/110
Cleaners and disinfectants
Chemical hand cleaner - frequency of use by mother (prenatal) 246/34
Chemical hand cleaner - monthly or more use by mother (prenatal) 98/21
Chemical hand cleaner - weekly use by anyone in home (prenatal) 127/23
Chemical hand cleaner - weekly use by mother (prenatal) 47/17
Chemical hand cleaner - weekly use by mother (prenatal, not 6m-12m) 37/10
Chemical hand cleaner - weekly use by mother (prenatal, not birth-3m) 34/13
Floor polish - frequency used by mother (prenatal) 214/23
Furniture, floor or dusting polish - monthly or more use (prenatal) 574/63
Hand sanitizer or disinfectant cleaner - daily use by mother (prenatal) 361/43
Hand sanitizer and disinfectant used monthly (prenatal-3m) 64/12
Hand sanitizer used daily by mother (prenatal) 323/37
Oven cleaner - frequency of use in home (birth-3m) 475/52
Oven cleaner - monthly or more use (prenatal) 140/21
Oven cleaner - monthly or more use by anyone (prenatal) 211/28
Oven cleaner - weekly or more use by anyone (prenatal) 62/14
Oven cleaner - weekly or more use by mother (prenatal) 39/12
Biological sex is male 1167/107
Ethnicity is caucasian 1444/86
Has family asthma history 674/91
Phthalates
Bis(2-ethylhexyl) phthalate (DEHP) in house dust (birth-3m) 453/126
Solid food
Frequency cookies/brownies consumed (4y-5y) 1769/113
Given eggs (24m-30m) 1570/91
Given nuts (not peanuts) (24m-30m) 1329/74
Dairy products (e.g., yogurt, cheese, ice cream) before 2 years 1748/114
Eggs before 2 years 1603/95
Frequency peanut butter consumed (30-36m) 1531/85
Frequency peanut butter consumed (4y-5y) 1517/83
Given eggs and nuts or peanuts (18m-2y) 1449/78
Given eggs and nuts and peanuts (18m-2y) 1026/48
Given peanuts or peanut butter (24m-30m) 1454/81
Nuts (other than peanuts) before 2 years 1213/65
Peanuts/peanut butter before 2 years 1438/81
Medications
Number of antibiotic courses (7m-1y) 322/45
Number of antibiotic courses (birth-12m) 584/65
Given antibiotic (7m-1y) 322/45
Used albuterol (7m-1y) 37/13
Used anti-allergic drug (7m-1y) 211/29
Vitamins and supplements
Poly-vi-sol multivitamin use (24m-30m) 83/14
Child given any vitamins or supplements (birth-3m) 1292/111
99% vitamin D
1.0e-01
1.0
10.0
Adjusted odds ratio and 95% confidence intervals
c
8-year asthma
Family asthma history
Weekly chemical hand cleaner use (prenatal)
Number of antibiotic courses (birth-1y)
Peanuts/peanut butter before 2y
← Decreased odds of asthma
Increased odds of asthma →
Cleaners
Nutrition
Breast milk
Medication / Antibiotic
Other
Prenatal
Not prenatal
-log10(P-value)
log2(OR)
d
8-year asthma
← Decreased odds of asthma
Increased odds of asthma →
Breast milk
Gamma-linolenic acid (birth-3m) 856/71
Cleaners and disinfectants
Chemical hand cleaner - frequency of use by mother (prenatal) 241/21
Chemical hand cleaner - monthly or more use by mother (prenatal) 95/14
Chemical hand cleaner - weekly use by mother (prenatal) 47/11
Chemical hand cleaner - weekly use by mother (prenatal, not birth-3m) 30/11
Oven cleaner - weekly or more use by anyone (prenatal) 62/12
Oven cleaner - weekly or more use by mother (prenatal) 40/10
Has family asthma history 677/70
Solid food
Given eggs and nuts and peanuts (18m-2y) 991/41
Peanuts/peanut butter before 2 years 1402/65
Medications
Number of antibiotic courses (birth-12m) 554/52
Used albuterol (7m-1y) 45/10
Used anti-allergic drug (4m-6m) 123/16
Used anti-allergic drug (7m-1y) 205/25
1.0
10.0
Adjusted odds ratio and 95% confidence intervals
Prenatal exposure
Postnatal, family history, socioeconomic
n/n = Non-asthmatics with exposure / Asthmatics with exposure

Fig. 2 | Exposures identified by ExWAS as associated with specialist clinician-diagnosed asthma at 5 and 8 years of age.
a. and **c.** Volcano plots showing exposures that are significantly associated with 5- and 8-year year asthma, respectively, after a second more stringent ExWAS. Logistic regressions controlling for additional covariates were conducted on the 112 variables significantly associated with asthma or wheeze in the first round. Significant prenatal exposures are coloured orange while other significant exposures are blue. Shapes reflect the main categories of exposure. **b.** and **d.** Forest plots of the same exposures that were significantly-associated with 5- and 8-year year asthma, respectively, after the second round of ExWAS. Orange and blue horizontal plots are adjusted odds ratios and 95% confidence intervals for prenatal and postnatal exposures, respectively. Positive associations are over 1 and negative associations are less than 1. Numbers to the right of each plot represent the total number non-asthmatic cases / asthmatic cases subjected to each exposure, respectively. All were significant after correcting for multiple testing (see methods). Major groups of exposures are labelled and coloured for additional clarity (Grey).

Figure 3

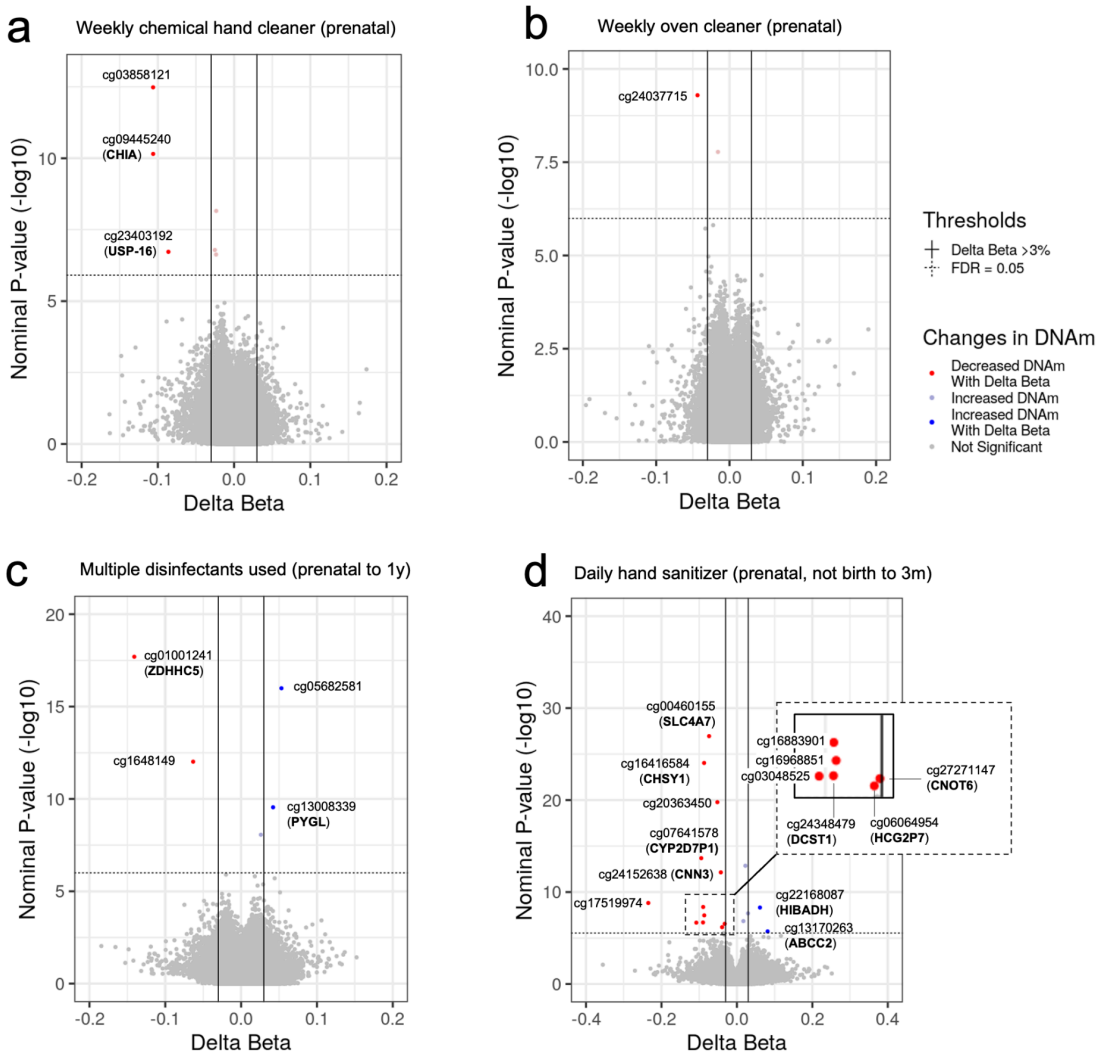
a
Weekly chemical hand cleaner (prenatal)
cg03858121
cg09445240
(CHIA)
cg23403192
(USP-16)
Nominal P-value (-log10)
Delta Beta
b
Weekly oven cleaner (prenatal)
cg24037715
Nominal P-value (-log10)
Delta Beta
Thresholds
Delta Beta >3%
FDR = 0.05
Changes in DNAm
Decreased DNAm With Delta Beta
Increased DNAm
Increased DNAm With Delta Beta
Not Significant
c
Multiple disinfectants used (prenatal to 1y)
cg01001241
(ZDHHC5)
cg05682581
cg1648149
cg13008339
(PYGL)
Nominal P-value (-log10)
Delta Beta
d
Daily hand sanitizer (prenatal, not birth to 3m)
cg00460155
(SLC4A7)
cg16883901
cg16968851
cg03048525
cg27271147
(CNOT6)
cg16416584
(CHSY1)
cg20363450
cg24348479
(DCST1)
cg06064954
(HCG2P7)
cg07641578
(CYP2D7P1)
cg24152638 (CNN3)
cg17519974
cg22168087
(HIBADH)
cg13170263
(ABCC2)
Nominal P-value (-log10)
Delta Beta

Fig. 3 | Use of cleaning/disinfecting products corresponds with differential methylation at CpG sites within the genome. a. Weekly chemical hand cleaner used by mother (prenatal). **b.** Weekly oven cleaner used during the prenatal period. **c.** Use of multiple disinfectants on a monthly or more basis during pregnancy and up to 1 year of age (defined by affirmative responses to having used hand sanitizer, disinfectant cleaner for general house cleaning, and disinfectant cleaner for the bedroom). **d**. Daily hand sanitizer used prenatally but monthly or less from birth to 3 months. Red dots are CpG sites with significantly decreased methylation while blue dots are CpG sites with significantly increased methylation. CpG sites are labelled with the gene they are located in, if applicable.

Figure 4

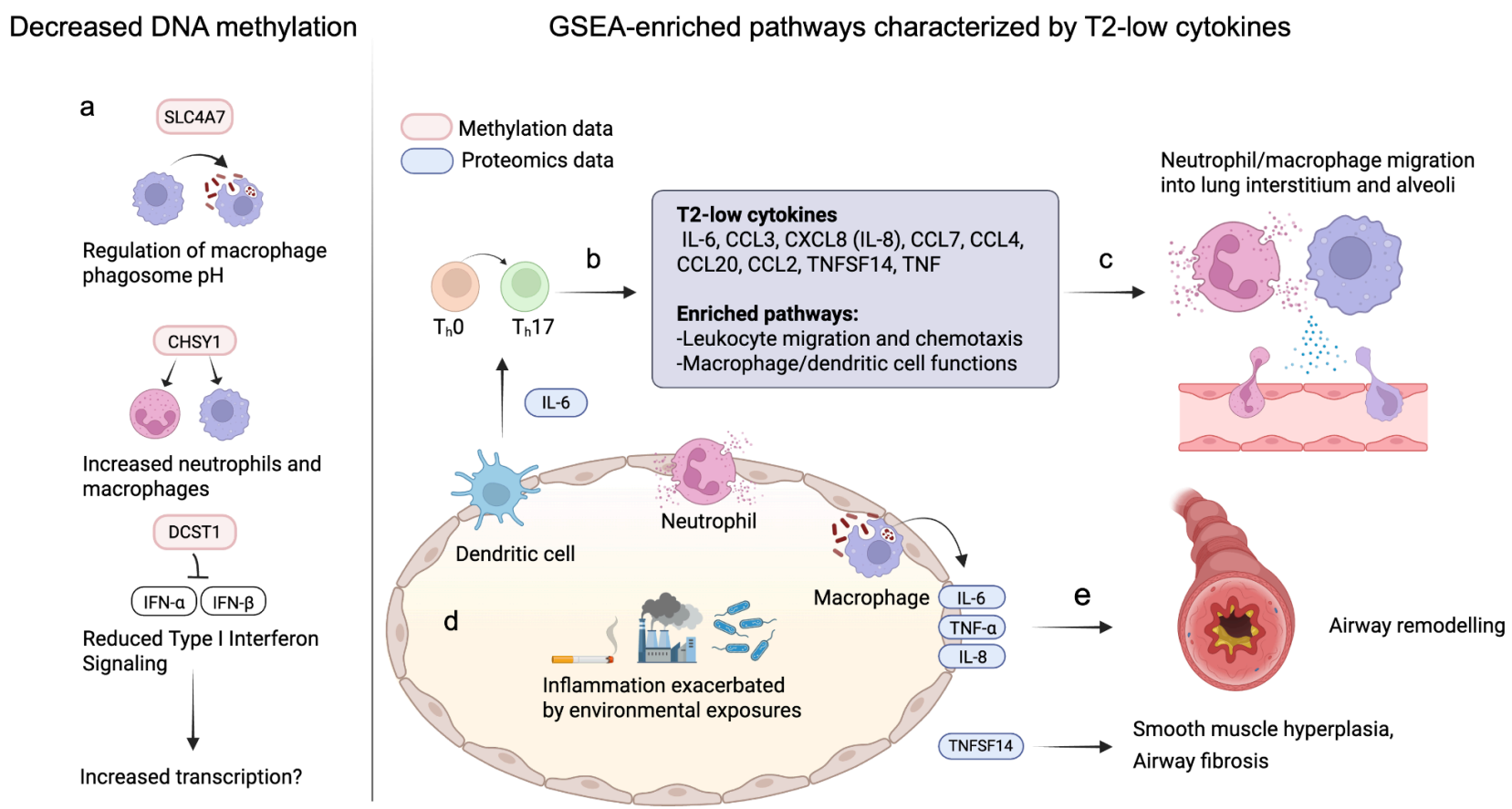
Decreased DNA methylation
GSEA-enriched pathways characterized by T2-low cytokines
a
SLC4A7
Regulation of macrophage phagosome pH
CHSY1
Increased neutrophils and macrophages
DCST1
IFN-α
IFN-β
Reduced Type I Interferon Signaling
Increased transcription?
Methylation data
Proteomics data
T2-low cytokines
IL-6, CCL3, CXCL8 (IL-8), CCL7, CCL4, CCL20, CCL2, TNFSF14, TNF
Enriched pathways:
-Leukocyte migration and chemotaxis
-Macrophage/dendritic cell functions
b
Th0
Th17
IL-6
c
Neutrophil/macrophage migration into lung interstitium and alveoli
Dendritic cell
Neutrophil
Macrophage
d
Inflammation exacerbated by environmental exposures
IL-6
TNF-α
IL-8
TNFSF14
e
Airway remodelling
Smooth muscle hyperplasia,
Airway fibrosis

Fig. 4 | Combining ExWAS, DNA methylation, and Gene Set Enrichment Analysis (GSEA) facilitates generation of example hypotheses explaining the association between prenatal hand sanitizer and later development of asthma in a child

a. Daily prenatal use of hand sanitizer is associated with decreased methylation in birth cord blood at sites within SLC47 CHSY1, AND DCST1 and these genes have been linked to regulation of macrophage phagosome function, increased numbers of neutrophils and macrophages, and associations with negative regulation of Type I Interferon Signaling, respectively. Decreased methylation could be putatively accompanied by increased transcription of these genes leading to inflammation and a neutrophilic asthma phenotype in more severe cases. **b.** The leading edge cytokines that contribute most to the enrichment of pathways related to leukocyte migration and chemotaxis, macrophage, and dendritic cell functions (blue box) have been linked to a T2-low asthma endotype response that is often associated with dendritic cell induction of naive CD4+ T cells into Th17 cells. **c.** Elevated T2-low cytokines could lead to increased migration of neutrophils and macrophage migration into the lung interstitium and alveoli. **d.** Migration of neutrophils and macrophages into the alveoli could be followed by inflammation which is further exacerbated by common environmental exposures associated with T2-low cytokines. **e.** Uncontrolled inflammation could continue to the point that smooth muscle hyperplasia and airway fibrosis (further regulated by cytokines like TNFSF14) lead to airway remodelling in those that have progressed to the point of being diagnosed with asthma. Figure generated using Biorender.

Figure 5

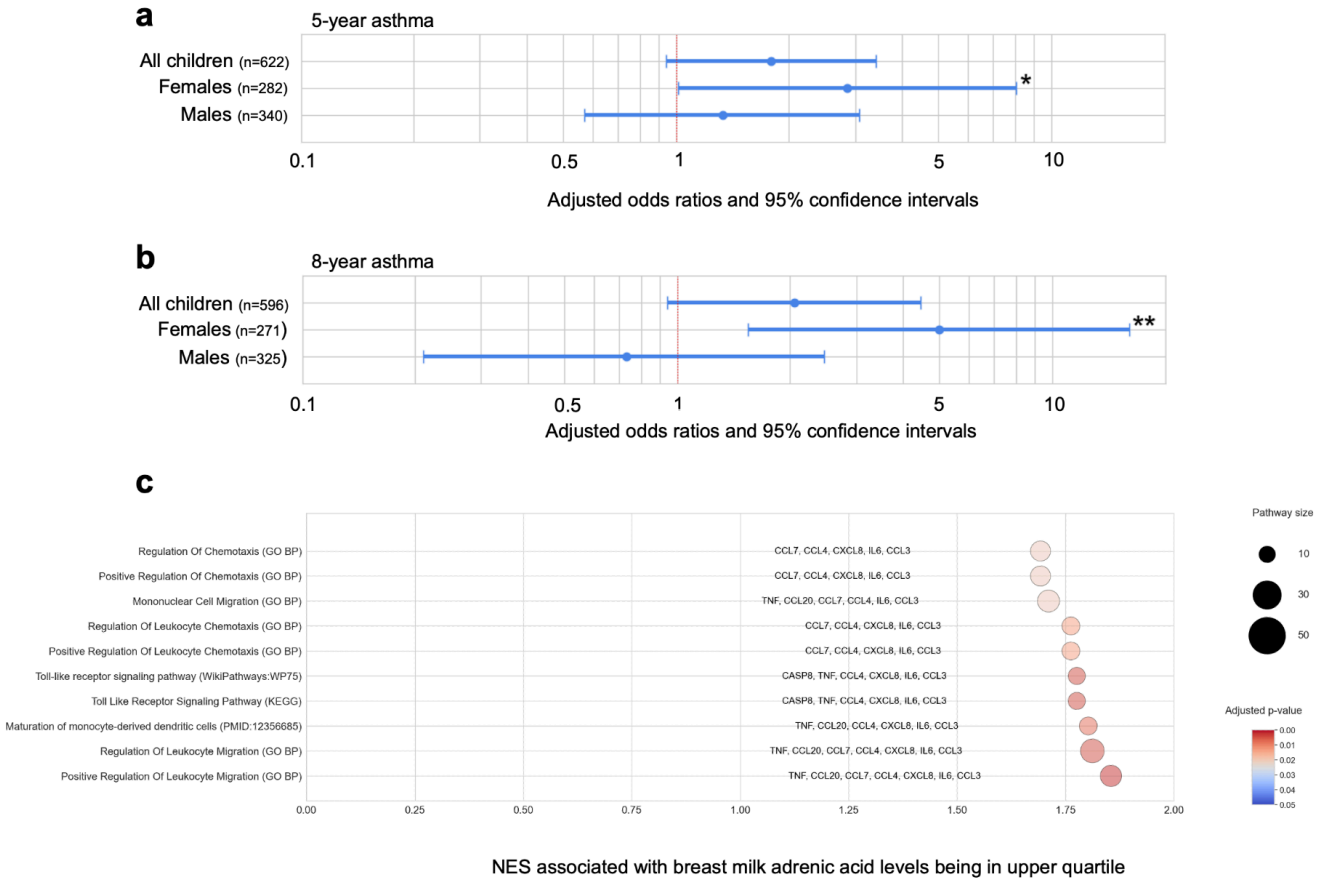
a
5-year asthma
All children (n=622)
Females (n=282)
Males (n=340)
*
0.1
0.5
1
5
10
Adjusted odds ratios and 95% confidence intervals
b
8-year asthma
All children (n=596)
Females (n=271)
Males (n=325)
**
0.1
0.5
1
5
10
Adjusted odds ratios and 95% confidence intervals
c
Regulation Of Chemotaxis (GO BP)
Positive Regulation Of Chemotaxis (GO BP)
Mononuclear Cell Migration (GO BP)
Regulation Of Leukocyte Chemotaxis (GO BP)
Positive Regulation Of Leukocyte Chemotaxis (GO BP)
Toll-like receptor signaling pathway (WikiPathways:WP75)
Toll Like Receptor Signaling Pathway (KEGG)
Maturation of monocyte-derived dendritic cells (PMID:12356685)
Regulation Of Leukocyte Migration (GO BP)
Positive Regulation Of Leukocyte Migration (GO BP)
CCL7, CCL4, CXCL8, IL6, CCL3
CCL7, CCL4, CXCL8, IL6, CCL3
TNF, CCL20, CCL7, CCL4, IL6, CCL3
CCL7, CCL4, CXCL8, IL6, CCL3
CCL7, CCL4, CXCL8, IL6, CCL3
CASP8, TNF, CCL4, CXCL8, IL6, CCL3
CASP8, TNF, CCL4, CXCL8, IL6, CCL3
TNF, CCL20, CCL4, CXCL8, IL6, CCL3
TNF, CCL20, CCL7, CCL4, CXCL8, IL6, CCL3
TNF, CCL20, CCL7, CCL4, CXCL8, IL6, CCL3
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
Pathway size
10
30
50
Adjusted p-value
0.00
0.01
0.02
0.03
0.04
0.05
NES associated with breast milk adrenic acid levels being in upper quartile

Fig. 5 | Breast milk AdA levels in the upper quintile are associated with increased odds of girls being diagnosed with asthma, enrichment of serum biomarkers of leukocyte migration and the Toll-like receptor signaling pathway. a. Forest plot of sex-stratified association between adrenic acid (AdA) levels being in the upper quintile and odds of asthma at age 5. **b.** Sex-stratified association between AdA levels being in the upper quintile and odds of asthma at age 8. Both logistic regression models were adjusted for study site, season of birth, biological sex, birth mode, exclusive breastfeeding status, prenatal smoking, family asthma history, household income, Caucasian ethnicity, number of antibiotic courses (birth - 1year), breast milk GLA and DGLA levels, infant age at time of milk sample collection, and maternal BMI. P-values: * $p < 0.05$, ** $p < 0.01$. **c.** Gene set enrichment analysis (GSEA) of biological pathways associated with the children's 5-year serum cytokine dataset in response to AdA levels being in the upper quartile. Figure shows the top significantly-enriched pathways which are primarily related to regulation of leukocyte migration, Toll-like signaling pathway, and chemotaxis. The positive normalized enrichment scores (NES) indicate upregulation of each pathway in children of mothers who had high AdA proportions in breastmilk. The diameter of each circle indicates set size that reflects the number of proteins in the specific molecular signatures database gene set after filtering out genes not included in the serum cytokine dataset. Core enrichment genes that contribute most to the enrichment signal are listed to the side of each bubble.

Figure 6

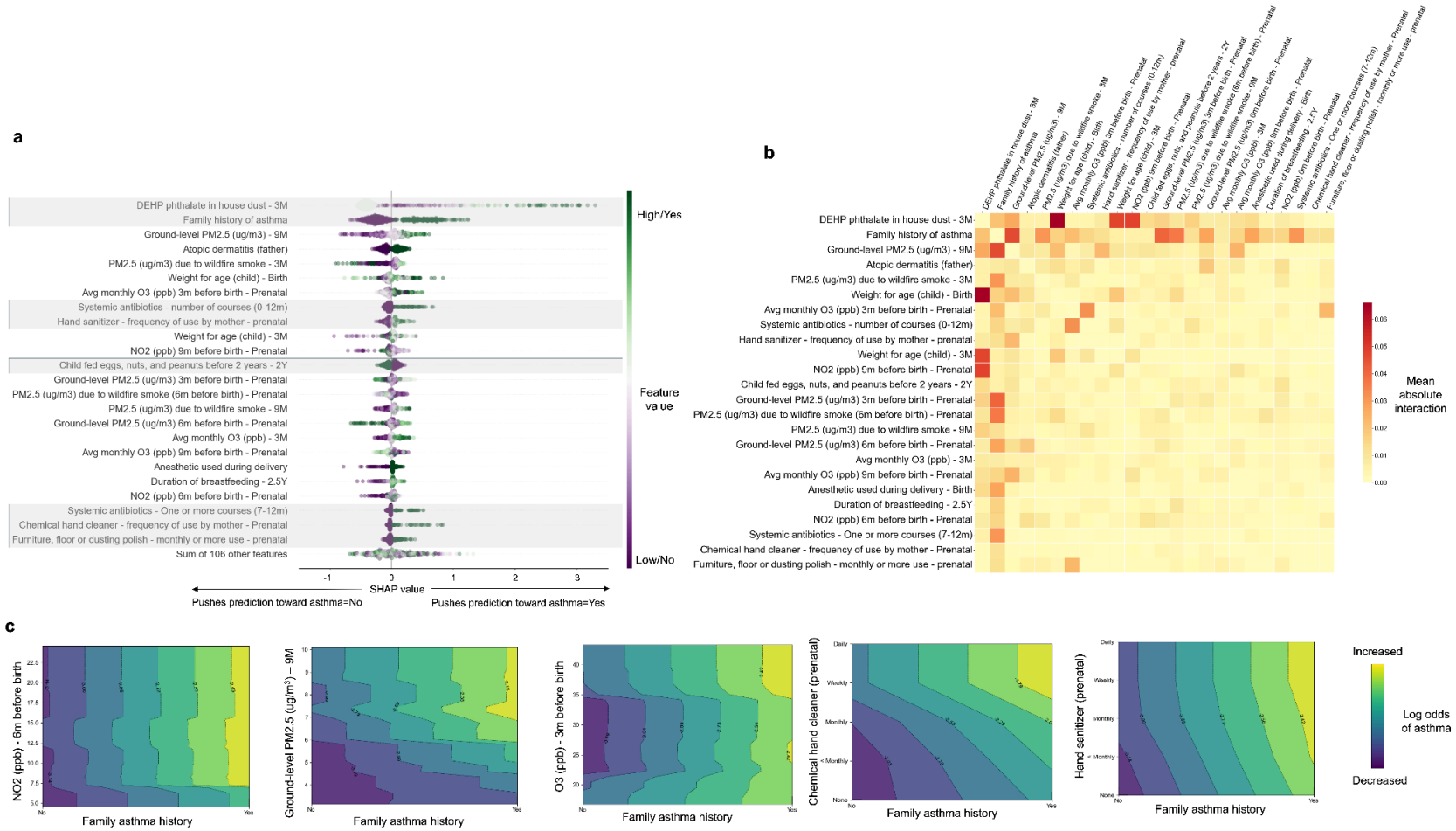
a
DEHP phthalate in house dust - 3M
Family history of asthma
Ground-level PM2.5 (ug/m3) - 9M
Atopic dermatitis (father)
PM2.5 (ug/m3) due to wildfire smoke - 3M
Weight for age (child) - Birth
Avg monthly O3 (ppb) 3m before birth - Prenatal
Systemic antibiotics - number of courses (0-12m)
Hand sanitizer - frequency of use by mother - prenatal
Weight for age (child) - 3M
NO2 (ppb) 9m before birth - Prenatal
Child fed eggs, nuts, and peanuts before 2 years - 2Y
Ground-level PM2.5 (ug/m3) 3m before birth - Prenatal
PM2.5 (ug/m3) due to wildfire smoke (6m before birth) - Prenatal
PM2.5 (ug/m3) due to wildfire smoke - 9M
Ground-level PM2.5 (ug/m3) 6m before birth - Prenatal
Avg monthly O3 (ppb) - 3M
Avg monthly O3 (ppb) 9m before birth - Prenatal
Anesthetic used during delivery
Duration of breastfeeding - 2.5Y
NO2 (ppb) 6m before birth - Prenatal
Systemic antibiotics - One or more courses (7-12m)
Chemical hand cleaner - frequency of use by mother - Prenatal
Furniture, floor or dusting polish - monthly or more use - prenatal
Sum of 106 other features
High/Yes
Feature value
Low/No
SHAP value
Pushes prediction toward asthma=No
Pushes prediction toward asthma=Yes
b
Mean absolute interaction
c
NO2 (ppb) - 6m before birth
Ground-level PM2.5 (ug/m³) – 9M
O3 (ppb) - 3m before birth
Chemical hand cleaner (prenatal)
Hand sanitizer (prenatal)
Family asthma history
Increased
Log odds of asthma
Decreased

Figure 6 | Contribution of individual features, and their interactions for prediction of 5-year asthma using a gradient boosting classifier. **a.** Plot of top features contributing to prediction based on SHAP values (note features are similar to what were returned using a permutation approach, however, SHAP provides insight into direction of effect). Purple dots represent individuals with a lower value (or no response) while green represent individuals with a higher value (or yes response) to feature. Dots on the right of the baseline indicate that the value pushed the prediction toward asthma=yes, while dots on the left of baseline indicate value pushed prediction toward asthma=no. Grey shaded boxes indicate features that were important in the gradient boosting classifier model and also a significant factor identified by ExWAS. **b**. SHAP interaction heatmap showing mean absolute interaction between features with top SHAP values. **c**. Two-dimensional partial dependence plots of select interactions between family history of asthma, and air quality or prenatal cleaners, that increase the odds of developing asthma at age 5. From left to right: i. Average monthly NO2 (ppb) at 6m prior to birth, ground level PM2.5 (ug/m3) at 9 months post birth, iii. Average monthly O3 (ppb) at 3 months prior to birth, iv. Frequency of chemical hand cleaner use by mother during pregnancy (reported at 18th week), v. Frequency of hand sanitizer use by mother during pregnancy (reported at 18th week). Colormap represents predicted values on the log-odds scale, where higher values (yellow) indicate increased odds of asthma and lower values (purple) indicate decreased odds.

Figure 7

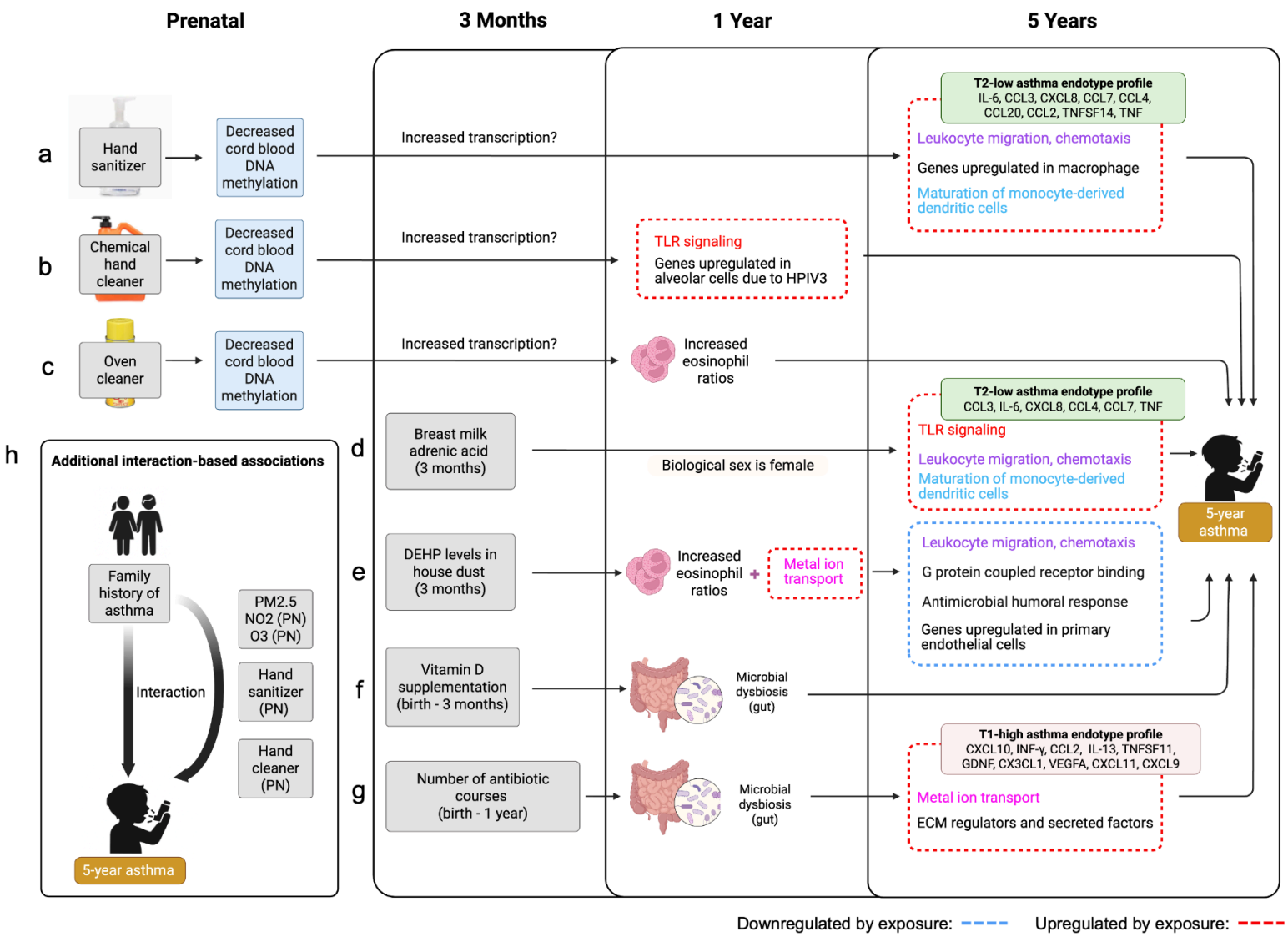
Prenatal
3 Months
1 Year
5 Years
a
Hand sanitizer
Decreased cord blood DNA methylation
Increased transcription?
b
Chemical hand cleaner
Decreased cord blood DNA methylation
Increased transcription?
TLR signaling
Genes upregulated in alveolar cells due to HPIV3
c
Oven cleaner
Decreased cord blood DNA methylation
Increased transcription?
Increased eosinophil ratios
T2-low asthma endotype profile
IL-6, CCL3, CXCL8, CCL7, CCL4, CCL20, CCL2, TNFSF14, TNF
Leukocyte migration, chemotaxis
Genes upregulated in macrophage
Maturation of monocyte-derived dendritic cells
h
Additional interaction-based associations
Family history of asthma
Interaction
PM2.5
NO2 (PN)
O3 (PN)
Hand sanitizer (PN)
Hand cleaner (PN)
5-year asthma
d
Breast milk adrenic acid (3 months)
Biological sex is female
T2-low asthma endotype profile
CCL3, IL-6, CXCL8, CCL4, CCL7, TNF
TLR signaling
Leukocyte migration, chemotaxis
Maturation of monocyte-derived dendritic cells
5-year asthma
e
DEHP levels in house dust (3 months)
Increased eosinophil ratios
+
Metal ion transport
Leukocyte migration, chemotaxis
G protein coupled receptor binding
Antimicrobial humoral response
Genes upregulated in primary endothelial cells
f
Vitamin D supplementation (birth - 3 months)
Microbial dysbiosis (gut)
g
Number of antibiotic courses (birth - 1 year)
Microbial dysbiosis (gut)
T1-high asthma endotype profile
CXCL10, INF-γ, CCL2, IL-13, TNFSF11, GDNF, CX3CL1, VEGFA, CXCL11, CXCL9
Metal ion transport
ECM regulators and secreted factors
Downregulated by exposure:
Upregulated by exposure:

Fig. 7 | Schematic view of prenatal and early-life exposures, and a hypothesized model for different pathways to different asthma endotypes.

a. Daily use of hand sanitizer, **b.** weekly or more use of chemical hand cleaner, and **c.** weekly or more use of oven cleaner during the prenatal period corresponded with decreased cord blood DNA methylation, which may putatively increase transcription of genes contributing to lung inflammation and asthma. Prenatal hand sanitizer was also associated with enrichment of inflammatory pathways characterized by a leading edge subset of T2-low cytokines characteristic of a neutrophilic asthma endotype at 5 years. **d.** High breast milk fatty acid proportions, particularly adrenic acid (AdA), were associated with increased odds of asthma, while further analysis demonstrated that AdA proportions in the upper quintile were associated with 5- and 8-year asthma in females. AdA also exhibited increased leukocyte migration, maturation of monocyte-derived dendritic cells, and Toll-Like receptor signalling with a T2-low chemokine profile resembling prenatal daily hand sanitizer use. **e.** Exposure to DEHP concentrations greater than the top quartile in house dust was associated with higher levels of eosinophils and cytokines associated with metal ion transport at age 1. Cytokines associated with leukocyte migration and chemotaxis, G protein coupled receptor binding, antimicrobial humoral response and primary endothelial cell function were downregulated. **f.** Supplementation with vitamin D between birth and 3 months of age was associated with decreased alpha diversity (3 months and 1 year) and differences in bacterial taxa (3 months) but needs further investigation into associations with asthma that could account for **g.** Early-life antibiotic exposure was linked to asthma and accompanied by microbial dysbiosis in the gut. Children receiving 3 or more courses of antibiotics in the first year of life were enriched for pathways having increased levels of T1-high cytokines that may be associated with an emerging T2-low neutrophilic asthma subgroup. **h.** A family history of asthma or allergy represents a strong inherited risk factor that may also interact with other environmental exposures including PM2.5, NO2 and O3 to increase the risk of developing asthma. Figure generated using Biorender.

References

1. Hersh, A., von Scheven, E. & Yelin, E. Adult outcomes of childhood-onset rheumatic diseases. *Nat Rev Rheumatol* **7**, 290–295 (2011).
2. Organization, W. H. Global health risks : mortality and burden of disease attributable to selected major risks. vi, 62 p. (2009).
3. Armocida, B. *et al.* Burden of non-communicable diseases among adolescents aged 10–24 years in the EU, 1990–2019: a systematic analysis of the Global Burden of Diseases Study 2019. *Lancet Child Adolesc Health* **6**, 367–383 (2022).
4. Di Cesare, M. *et al.* The epidemiological burden of obesity in childhood: a worldwide epidemic requiring urgent action. *BMC Med* **17**, 212 (2019).
5. Boat, T. F. & Whitsett, J. A. How Can the Pediatric Community Enhance Funding for Child Health Research? *JAMA Pediatrics* **175**, 1212–1214 (2021).
6. Salam, R. A. *et al.* An evidence gap map of interventions for noncommunicable diseases and risk factors among children and adolescents. *Nat Med* **30**, 290–301 (2024).
7. Garner, R. & Kohen, D. Changes in the prevalence of asthma among Canadian children. *Health Reports* **19**, 7 (2008).
8. To, T., Dell, S., Dick, P. & Cicutto, L. The Burden of Illness Experienced by Young Children Associated with Asthma: A Population-Based Cohort Study. *Journal of Asthma* **45**, 45–49 (2008).
9. Gupta, R. *et al.* The Economic Impact of Childhood Food Allergy in the United States. *JAMA Pediatrics* **167**, 1026–1031 (2013).
10. Foppiano, F. & Schaub, B. Childhood asthma phenotypes and endotypes: a glance into the mosaic. *Molecular and Cellular Pediatrics* **10**, 9 (2023).
11. Kuruvilla, M. E., Lee, F. E.-H. & Lee, G. B. Understanding Asthma Phenotypes, Endotypes, and Mechanisms of Disease. *Clin Rev Allergy Immunol* **56**, 219–233 (2019).

12. Dharmage, S. C., Perret, J. L. & Custovic, A. Epidemiology of Asthma in Children and Adults. *Front Pediatr* **7**, 246 (2019).
13. Pinot De Moira, A., Pearce, N., Pedersen, M. & Nybo Andersen, A.-M. The influence of early-life animal exposure on the risk of childhood atopic dermatitis, asthma and allergic rhinoconjunctivitis: findings from the Danish National Birth Cohort. *International Journal of Epidemiology* dyad040 (2023) doi:10.1093/ije/dyad040.
14. Nwaru, B. I. *et al.* Timing of infant feeding in relation to childhood asthma and allergic diseases. *Journal of Allergy and Clinical Immunology* **131**, 78–86 (2013).
15. Burke, H. *et al.* Prenatal and Passive Smoke Exposure and Incidence of Asthma and Wheeze: Systematic Review and Meta-analysis. *Pediatrics* **129**, 735–744 (2012).
16. Ball, T. M. *et al.* Siblings, Day-Care Attendance, and the Risk of Asthma and Wheezing during Childhood. *New England Journal of Medicine* **343**, 538–543 (2000).
17. Patrick, D. M. *et al.* Decreasing antibiotic use, the gut microbiota, and asthma incidence in children: evidence from population-based and prospective cohort studies. *The Lancet Respiratory Medicine* **8**, 1094–1105 (2020).
18. Asthma and the social determinants of health. *Annals of Allergy, Asthma & Immunology* **128**, 5–11 (2022).
19. Lavigne, É. *et al.* Fine particulate matter concentration and composition and the incidence of childhood asthma. *Environment International* **152**, 106486 (2021).
20. Sbihi, H. *et al.* Thinking bigger: How early-life environmental exposures shape the gut microbiome and influence the development of asthma and allergic disease. *Allergy* **74**, 2103–2115 (2019).
21. Gold, D. R. *et al.* NIAID, NIEHS, NHLBI, and MCAN Workshop Report: The indoor environment and childhood asthma—implications for home environmental intervention in asthma prevention and management. *Journal of Allergy and Clinical Immunology* **140**, 933–949 (2017).

22. Christensen, S. *et al.* Prenatal environmental tobacco smoke exposure increases allergic asthma risk with methylation changes in mice. *Environmental and Molecular Mutagenesis* **58**, 423–433 (2017).
23. Rosas-Salazar, C. & Hartert, T. V. Prenatal exposures and the development of childhood wheezing illnesses. *Current Opinion in Allergy and Clinical Immunology* **17**, 110 (2017).
24. Georas, S. N. & Khurana, S. Update on Asthma Biology. *Journal of Allergy and Clinical Immunology* https://doi.org/10.1016/j.jaci.2024.01.024 (2024) doi:10.1016/j.jaci.2024.01.024.
25. Guillien, A. *et al.* Exposome Profiles and Asthma among French Adults. *Am J Respir Crit Care Med* **206**, 1208–1219 (2022).
26. Khalili, R. *et al.* Multiple prenatal exposures and acute-care clinical encounters for asthma among children born to mothers living near a Superfund site. *Am J Epidemiol* **193**, 1088–1096 (2024).
27. Savouré, M., Eminson, K., Sese, L., Dumas, O. & Cai, Y. S. The exposome in respiratory diseases: multiple preventable risk factors from early life to adulthood. *Breathe* **19**, (2023).
28. Vrijheid, M. *et al.* Advancing tools for human early lifecourse exposome research and translation (ATHLETE). *Environ Epidemiol* **5**, e166 (2021).
29. Agier, L. *et al.* Early-life exposome and lung function in children in Europe: an analysis of data from the longitudinal, population-based HELIX cohort. *The Lancet Planetary Health* **3**, e81–e92 (2019).
30. Miller, G. W. & Banbury Exposomics Consortium. Integrating exposomics into biomedicine. *Science* **388**, 356–358 (2025).
31. Wild, C. P. Complementing the Genome with an “Exposome”: The Outstanding Challenge of Environmental Exposure Measurement in Molecular Epidemiology. *Cancer Epidemiol Biomarkers Prev* **14**, 1847–1850 (2005).
32. Subbarao, P. *et al.* The Canadian Healthy Infant Longitudinal Development (CHILD) Study: examining developmental origins of allergy and asthma. *Thorax* **70**, 998–1000 (2015).

33. Hoskinson, C. *et al.* Delayed gut microbiota maturation in the first year of life is a hallmark of pediatric allergic disease. *Nat Commun* **14**, 4785 (2023).
34. Arrieta, M.-C. *et al.* Early infancy microbial and metabolic alterations affect risk of childhood asthma. *Science Translational Medicine* **7**, 307ra152-307ra152 (2015).
35. Miliku, K. *et al.* Human milk oligosaccharide profiles and food sensitization among infants in the CHILD Study. *Allergy* **73**, 2070–2073 (2018).
36. Sbihi, H. *et al.* Prenatal exposure to traffic-related air pollution, the gestational epigenetic clock, and risk of early-life allergic sensitization. *Journal of Allergy and Clinical Immunology* **144**, 1729-1731.e5 (2019).
37. Baluch, N., Gallant, M. & Ellis, A. K. Exposomal research in the context of birth cohorts: What have they taught us? *Annals of Allergy, Asthma & Immunology* **125**, 639–645 (2020).
38. Azeez, R. *et al.* Investigating SARS-CoV-2 infection and the health and psychosocial impact of the COVID-19 pandemic in the Canadian CHILD Cohort. *Epidemiol Health* e2023091 (1697122801) doi:10.4178/epih.e2023091.
39. Brook, J. R. *et al.* The Canadian Urban Environmental Health Research Consortium – a protocol for building a national environmental exposure data platform for integrated analyses of urban form and health. *BMC Public Health* **18**, 114 (2018).
40. Dai, R. *et al.* Wheeze trajectories: Determinants and outcomes in the CHILD Cohort Study. *Journal of Allergy and Clinical Immunology* **149**, 2153–2165 (2022).
41. He, P. *et al.* Early prediction of pediatric asthma in the Canadian Healthy Infant Longitudinal Development (CHILD) birth cohort using machine learning. *Pediatr Res* 1–8 (2024) doi:10.1038/s41390-023-02988-2.
42. Miliku, K. *et al.* Sex-specific associations of human milk long-chain polyunsaturated fatty acids and infant allergic conditions. *Pediatric Allergy and Immunology* **32**, 1173–1182 (2021).
43. Kooijman, M. N. *et al.* The Generation R Study: design and cohort update 2017. *Eur J Epidemiol* **31**, 1243–1264 (2016).

44. Environmental Influences on Child Health Outcomes (ECHO)-Wide Cohort | American Journal of Epidemiology | Oxford Academic. https://academic.oup.com/aje/article/192/8/1249/7085738.
45. Maitre, L. *et al.* Multi-omics signatures of the human early life exposome. *Nat Commun* **13**, 7024 (2022).
46. Yang, H.-J. *et al.* The Cohort for Childhood Origin of Asthma and allergic diseases (COCOA) study: design, rationale and methods. *BMC Pulmonary Medicine* **14**, 109 (2014).
47. Lee, E. *et al.* Insights from the COCOA birth cohort: The origins of childhood allergic diseases and future perspectives. *Allergology International* **73**, 3–12 (2024).
48. Bisgaard, H. *et al.* 25 Years of translational research in the Copenhagen Prospective Studies on Asthma in Childhood (COPSAC). *Journal of Allergy and Clinical Immunology* https://doi.org/10.1016/j.jaci.2022.11.022 (2023) doi:10.1016/j.jaci.2022.11.022.
49. Golding, J. The Avon Longitudinal Study of Parents and Children (ALSPAC)--study design and collaborative opportunities. *European Journal of Endocrinology* **151**, U119–U123 (2004).
50. Hohmann, C. *et al.* Sex-specific incidence of asthma, rhinitis and respiratory multimorbidity before and after puberty onset: individual participant meta-analysis of five birth cohorts collaborating in MeDALL. *BMJ Open Respir Res* **6**, e000460 (2019).
51. Almqvist, C. *et al.* Season of birth, childhood asthma and allergy in a nationwide cohort—Mediation through lower respiratory infections. *Clinical & Experimental Allergy* **50**, 222–230 (2020).
52. Burke, W., Fesinmeyer, M., Reed, K., Hampson, L. & Carlsten, C. Family history as a predictor of asthma risk. *American Journal of Preventive Medicine* **24**, 160–169 (2003).
53. Cardet, J. C. *et al.* Income is an independent risk factor for worse asthma outcomes. *Journal of Allergy and Clinical Immunology* **141**, 754-760.e3 (2018).

54. Bener, A., Ehlayel, M. S., Alsowaidi, S. & Sabbah, A. Role of breast feeding in primary prevention of asthma and allergic diseases in a traditional society. *Eur Ann Allergy Clin Immunol* **39**, 337–343 (2007).

55. Almqvist, C., Cnattingius, S., Lichtenstein, P. & Lundholm, C. The impact of birth mode of delivery on childhood asthma and allergic diseases–a sibling study. *Clinical & Experimental Allergy* **42**, 1369–1376 (2012).

56. Zacharasiewicz, A. Maternal smoking in pregnancy and its influence on childhood asthma. *ERJ Open Res* **2**, 00042–02016 (2016).

57. Condella, A. *et al.* Multicenter Study of Albuterol Use Among Infants Hospitalized with Bronchiolitis. *West J Emerg Med* **19**, 475–483 (2018).

58. Hill, D. A., Grundmeier, R. W., Ram, G. & Spergel, J. M. The epidemiologic characteristics of healthcare provider-diagnosed eczema, asthma, allergic rhinitis, and food allergy in children: a retrospective cohort study. *BMC Pediatr* **16**, 1–8 (2016).

59. Nordling, E. *et al.* Traffic-Related Air Pollution and Childhood Respiratory Symptoms, Function and Allergies. *Epidemiology* **19**, 401 (2008).

60. Brunst, K. J. *et al.* Timing and Duration of Traffic-related Air Pollution Exposure and the Risk for Childhood Wheeze and Asthma. *Am J Respir Crit Care Med* **192**, 421–427 (2015).

61. Martin, J., Townshend, J. & Brodlie, M. Diagnosis and management of asthma in children. *bmjpo* **6**, (2022).

62. Zhang, B. *et al.* Maternal smoking during pregnancy and cord blood DNA methylation: new insight on sex differences and effect modification by maternal folate levels. *Epigenetics* **13**, 505–518 (2018).

63. Miyake, K. *et al.* Association between DNA methylation in cord blood and maternal smoking: The Hokkaido Study on Environment and Children's Health. *Sci Rep* **8**, 5654 (2018).

64. Hartl, D. *et al.* Acidic Mammalian Chitinase Is Secreted via an ADAM17/Epidermal Growth Factor Receptor-dependent Pathway and Stimulates Chemokine Production by Pulmonary Epithelial Cells*. *Journal of Biological Chemistry* **283**, 33472–33482 (2008).

65. Macke, E. L. *et al.* Loss of Chondroitin Sulfate Modification Causes Inflammation and Neurodegeneration in skt Mice. *Genetics* **214**, 121–134 (2020).

66. Nair, S., Bist, P., Dikshit, N. & Krishnan, M. N. Global functional profiling of human ubiquitome identifies E3 ubiquitin ligase DCST1 as a novel negative regulator of Type-I interferon signaling. *Sci Rep* **6**, 36179 (2016).

67. Sedlyarov, V. *et al.* The Bicarbonate Transporter SLC4A7 Plays a Key Role in Macrophage Phagosome Acidification. *Cell Host & Microbe* **23**, 766-774.e5 (2018).

68. Heuvel, M. M. van den, Vanhee, D. D. C., Postmus, P. E., Hoefsmit, E. C. M. & Beelen, R. H. J. Functional and phenotypic differences of monocyte-derived dendritic cells from allergic and nonallergic patients. *Journal of Allergy and Clinical Immunology* **101**, 90–95 (1998).

69. Branchett, W. J., Saraiva, M. & O'Garra, A. Regulation of inflammation by Interleukin-10 in the intestinal and respiratory mucosa. *Current Opinion in Immunology* **91**, 102495 (2024).

70. Hinks, T. S. C., Levine, S. J. & Brusselle, G. G. Treatment options in type-2 low asthma. *European Respiratory Journal* **57**, (2021).

71. Enweasor, C., Flayer, C. H. & Haczku, A. Ozone-Induced Oxidative Stress, Neutrophilic Airway Inflammation, and Glucocorticoid Resistance in Asthma. *Front. Immunol.* **12**, (2021).

72. Branicka, O., Gawlik, R. & Glück, J. Eosinophil to lymphocyte ratio may predict OCS reduction and change in quality of life (AQLQ) resulting from asthma biological treatment. *Immunopharmacology and Immunotoxicology* **46**, 212–217 (2024).

73. Zhang, X.-Y. *et al.* Full blood count parameters for the detection of asthma inflammatory phenotypes. *Clinical & Experimental Allergy* **44**, 1137–1145 (2014).

74. Aldakheel, F. M. *et al.* Immune cell ratio and coagulation markers in assessing prognosis of asthma: a cross-sectional study from Saudi Arabia. *Front Immunol* **14**, 1206636 (2023).

75. Sherriff, A., Farrow, A., Golding, J. & Henderson, J. Frequent use of chemical household products is associated with persistent wheezing in pre-school age children. *Thorax* **60**, 45–49 (2005).

76. Henderson, J., Sherriff, A., Farrow, A. & Ayres, J. G. Household chemicals, persistent wheezing and lung function: effect modification by atopy? *European Respiratory Journal* **31**, 547–554 (2008).

77. Kojima, R. *et al.* Prenatal occupational disinfectant exposure and childhood allergies: the Japan Environment and Children's study. *Occup Environ Med* **79**, 521–526 (2022).

78. Casas, L. *et al.* The use of household cleaning products during pregnancy and lower respiratory tract infections and wheezing during early life. *Int J Public Health* **58**, 757–764 (2013).

79. Parks, J. *et al.* Association of use of cleaning products with respiratory health in a Canadian birth cohort. *CMAJ* **192**, E154–E161 (2020).

80. Sheikhpour, M., Maleki, M., Ebrahimi Vargoorani, M. & Amiri, V. A review of epigenetic changes in asthma: methylation and acetylation. *Clinical Epigenetics* **13**, 65 (2021).

81. Tse, T. J. *et al.* Toxicology of alcohol-based hand rubs formulated with technical-grade ethanol. *Toxicology Reports* **8**, 785–792 (2021).

82. Fu, J. *et al.* Sex-Specific and Trimester-Specific Associations of Prenatal Exposure to Bisphenols, Parabens, and Triclosan with Neonatal Birth Size and Gestational Age. *Environ. Sci. Technol.* **58**, 13687–13696 (2024).

83. Vetrano, A. M. *et al.* Inflammatory Effects of Phthalates in Neonatal Neutrophils. *Pediatr Res* **68**, 134–139 (2010).

84. Miyamoto, M. & Sasakawa, S. Effects of Plasticizers and Plastic Bags on Granulocyte Function during Storage. *Vox Sanguinis* **53**, 19–22 (1987).

85. Song, A.-Y. *et al.* Integrating epidemiology and adverse outcome pathway analysis to assess the risk of childhood asthma associated with di(2-ethylhexyl) phthalate exposure. *Journal of Hazardous Materials Advances* **21**, 101058 (2026).

86. Lu, C. *et al.* Interaction effect of prenatal and postnatal exposure to ambient air pollution and temperature on childhood asthma. *Environment International* **167**, 107456 (2022).

87. Huang, L. *et al.* Airborne phthalates in indoor environment: Partition state and influential built environmental conditions. *Chemosphere* **254**, 126782 (2020).

88. Tseng, H.-H. *et al.* Di-(2-ethylhexyl) Phthalate Promotes Allergic Lung Inflammation by Modulating CD8α+ Dendritic Cell Differentiation via Metabolite MEHP-PPARγ Axis. *Front Immunol* **13**, 581854 (2022).

89. Suen, J.-L. *et al.* Environmental Factor-Mediated Transgenerational Inheritance of Igf2r Hypomethylation and Pulmonary Allergic Response via Targeting Dendritic Cells. *Front. Immunol.* **11**, (2020).

90. Dutta, S., Haggerty, D. K., Rappolee, D. A. & Ruden, D. M. Phthalate Exposure and Long-Term Epigenomic Consequences: A Review. *Front. Genet.* **11**, (2020).

91. Wang, I.-J., Karmaus, W. J., Chen, S.-L., Holloway, J. W. & Ewart, S. Effects of phthalate exposure on asthma may be mediated through alterations in DNA methylation. *Clin Epigenet* **7**, 27 (2015).

92. Navaranjan, G. *et al.* Early life exposure to phthalates and the development of childhood asthma among Canadian children. *Environmental Research* **197**, 110981 (2021).

93. Government of Canada, P. S. and P. C. Vitamin D supplementation for breastfeeding : 2004 Health Canada recommendation : H44-74/2004E-PDF - Government of Canada Publications - Canada.ca. https://publications.gc.ca/site/eng/9.687087/publication.html (2002).

94. Ellsworth, L. *et al.* Impact of maternal overweight and obesity on milk composition and infant growth. *Maternal & Child Nutrition* **16**, e12979 (2020).

95. Weiss, S. T. & Gold, D. R. Gender differences in asthma. *Pediatric Pulmonology* **19**, 153–155 (1995).

96. Soto-Ramírez, N. *et al.* Fatty Acids in Breast Milk Associated with Asthma-Like Symptoms and Atopy in Infancy: A Longitudinal Study. *Journal of Asthma* **49**, 926–934 (2012).

97. Cho, Y., Osgood, R. S., Bell, L. N., Karoly, E. D. & Shore, S. A. Ozone-induced changes in the serum metabolome: Role of the microbiome. *PLOS ONE* **14**, e0221633 (2019).

98. Friedman, J. H. Greedy function approximation: A gradient boosting machine. *The Annals of Statistics* **29**, 1189–1232 (2001).

99. O'Connor, K. M., Das, A. B., Winterbourn, C. C. & Hampton, M. B. Inhibition of DNA methylation in proliferating human lymphoma cells by immune cell oxidants. *J Biol Chem* **295**, 7839–7848 (2020).

100. Klibaner-Schiff, E. *et al.* Environmental exposures influence multigenerational epigenetic transmission. *Clin Epigenet* **16**, 145 (2024).

101. Development of a Symptom-Based Tool for Screening of Children at High Risk of Preschool Asthma | Pediatrics | JAMA Network Open | JAMA Network. https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2797046.

102. Cruz-Hernandez, C., Goeuriot, S., Giuffrida, F., Thakkar, S. K. & Destaillats, F. Direct quantification of fatty acids in human milk by gas chromatography. *Journal of Chromatography A* **1284**, 174–179 (2013).

103. Azad, M. B. *et al.* 'Human Milk Oligosaccharide Concentrations Are Associated with Multiple Fixed and Modifiable Maternal Characteristics, Environmental Factors, and Feeding Practices. *The Journal of Nutrition* **148**, 1733–1742 (2018).

104. Seabold, S. & Perktold, J. Statsmodels: Econometric and Statistical Modeling with Python. *scipy* https://doi.org/10.25080/Majora-92bf1922-011 (2010) doi:10.25080/Majora-92bf1922-011.

105. PATEL, C. J. & MANRAI, A. K. DEVELOPMENT OF EXPOSOME CORRELATION GLOBES TO MAP OUT ENVIRONMENT-WIDE ASSOCIATIONS. *Pac Symp Biocomput* **20**, 231–242 (2015).

106. Krzywinski, M. *et al.* Circos: An information aesthetic for comparative genomics. *Genome Res.* **19**, 1639–1645 (2009).

107. Nevola (kathy-nevola), K. *et al.* OlinkAnalyze: Facilitate Analysis of Proteomic Data from Olink. (2024).

108. Moraes, T. J. *et al.* The Canadian Healthy Infant Longitudinal Development Birth Cohort Study: Biological Samples and Biobanking. *Paediatric and Perinatal Epidemiology* **29**, 84–92 (2015).

109. Dai, D. L. Y. *et al.* Breastfeeding enrichment of B. longum subsp. infantis mitigates the effect of antibiotics on the microbiota and childhood asthma risk. *Med (N Y)* **4**, 92-112.e5 (2023).

110. Beghini, F. *et al.* Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3. *eLife* **10**, e65088.

111. Lahti, L. & Shetty, S. microbiome R package. https://doi.org/10.18129/B9.bioc.microbiome (2017) doi:10.18129/B9.bioc.microbiome.

112. McMurdie, P. J. & Holmes, S. phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLOS ONE* **8**, e61217 (2013).

113. Mallick, H. *et al.* Multivariable association discovery in population-scale meta-omics studies. *PLOS Computational Biology* **17**, e1009442 (2021).

114. Lu, Z. *et al.* Early-preschool wheeze trajectories are predominantly nonallergic with distinct biologic and microbiome traits. *Journal of Allergy and Clinical Immunology* https://doi.org/10.1016/j.jaci.2025.07.034 (2025) doi:10.1016/j.jaci.2025.07.034.

115. Boogaart, K. G. van den, Tolosana-Delgado, R. & Bren, M. compositions: Compositional Data Analysis. (2024).

116. Nickols, W. A. *et al.* MaAsLin 3: Refining and extending generalized multivariable linear models for meta-omic association discovery. *bioRxiv* 2024.12.13.628459 (2024) doi:10.1101/2024.12.13.628459.

117. Gu, Z., Eils, R. & Schlesner, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847–2849 (2016).

118. Merrill, S. M. *et al.* Associations of peripheral blood DNA methylation and estimated monocyte proportion differences during infancy with toddler attachment style. *Attach Hum Dev* **25**, 132–161 (2023).

119. Konwar, C. *et al.* Risk-focused differences in molecular processes implicated in SARS-CoV-2 infection: corollaries in DNA methylation and gene expression. *Epigenetics & Chromatin* **14**, 1–18 (2021).

120. Ewastools: Infinium Human Methylation BeadChip pipeline for population epigenetics integrated into Galaxy - PMC. https://pmc-ncbi-nlm-nih-gov.proxy.lib.sfu.ca/articles/PMC7219210/.

121. Heiss, J. A. & Just, A. C. Identifying mislabeled and contaminated DNA methylation microarray data: an extended quality control toolset with examples from GEO. *Clin Epigenetics* **10**, 73 (2018).

122. Aryee, M. J. *et al.* Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* **30**, 1363–1369 (2014).

123. Du, P., Kibbe, W. A. & Lin, S. M. lumi: a pipeline for processing Illumina microarray. *Bioinformatics* **24**, 1547–1548 (2008).

124. Pidsley, R. *et al.* Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling. *Genome Biol* **17**, 208 (2016).

125. Pidsley, R. *et al.* A data-driven approach to preprocessing Illumina 450K methylation array data. *BMC Genomics* **14**, 293 (2013).

126. Johnson, W. E., Li, C. & Rabinovic, A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8**, 118–127 (2007).

127. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)* **57**, 289–300 (1995).

128. Korthauer, K. *et al.* A practical guide to methods controlling false discoveries in computational biology. *Genome Biol* **20**, 118 (2019).

129. Venables, B. & Ripley, B. Modern applied statistics with S. in *Springer* (2002). doi:10.1007/b97626.

130. Jones, M. J., Moore, S. R. & Kobor, M. S. Principles and Challenges of Applying Epigenetic Epidemiology to Psychology. *Annu Rev Psychol* **69**, 459–485 (2018).

131. Cohen, S., Kamarck, T. & Mermelstein, R. A global measure of perceived stress. *J Health Soc Behav* **24**, 385–396 (1983).

132. Radloff, L. S. The CES-D scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement* **1**, 385–401 (1977).

133. Valvi, D. *et al.* Variability and predictors of urinary phthalate metabolites in Spanish pregnant women. *International Journal of Hygiene and Environmental Health* **218**, 220–231 (2015).