

# Epigenomic and proteomic analyses provide insights into early-life immune regulation and asthma development in infants

Received: 19 July 2024

Accepted: 17 February 2025

Published online: 15 April 2025

Yijun Li<sup>1</sup>, Zhaozhong Zhu<sup>2</sup>, Carlos A. Camargo Jr.<sup>2</sup>, Janice A. Espinola<sup>2</sup>, Kohei Hasegawa<sup>2,4</sup> & Liming Liang<sup>1,3,4</sup> ✉

Infants with severe bronchiolitis (i.e., bronchiolitis requiring hospitalization) face increased risks of respiratory diseases in childhood. We conduct epigenome-wide association studies in a multi-ethnic cohort of these infants. We identify 61 differentially methylated regions in infant blood (<1 year of age) associated with recurrent wheezing by age 3 (170 cases, 318 non-cases) and/or asthma by age 6 (112 cases, 394 non-cases). These differentially methylated regions are enriched in the enhancers of peripheral blood neutrophils. Several differentially methylated regions exhibit interaction with rhinovirus infection and/or specific blood cell types. In the same blood samples, circulating levels of 104 proteins correlate with the differentially methylated regions, and many proteins show phenotypic association with asthma. Through Mendelian randomization, we find causal evidence supporting a protective role of higher plasma ST2 (also known as IL1RL1) protein against asthma. DNA methylation is also associated with ST2 protein level in infant blood. Taken together, our findings suggest the contribution of DNA methylation to asthma development through regulating early-life systemic immune responses.

Bronchiolitis is an acute lower respiratory infection caused by respiratory viruses<sup>1</sup> and the leading cause of hospitalization in infants (age <1 year) in the United States<sup>2</sup>. Infants hospitalized for bronchiolitis (i.e., severe bronchiolitis), especially those with rhinovirus infection, have a substantially higher risk of developing recurrent wheezing and asthma in childhood<sup>3,4</sup>. However, the underlying biological mechanisms contributing to this elevated risk remains unclear.

DNA methylation (DNAm), a form of epigenetic modification, reflects the influence of genetics, environmental exposures, and gene-by-environment interactions<sup>5</sup>. DNAm has important regulatory functions in gene expression and contributes to the development of health outcomes<sup>5</sup>. Recent epigenome-wide association studies (EWAS) have linked DNAm level in children's blood to bronchiolitis severity<sup>6</sup>, recurrent wheezing<sup>7</sup>, and childhood asthma<sup>8,9</sup>. However, the association between infant DNAm and childhood respiratory outcomes

has not been investigated among infants with severe bronchiolitis. Such an investigation will provide important insights into the underlying mechanism of asthma development in this large, high-risk subpopulation.

Early-life immune regulation plays crucial roles in the development of childhood-onset asthma<sup>10</sup>. Proteins associated with immune functions, including cytokines and chemokines, contribute to the initiation and exacerbation of asthma<sup>11,12</sup>. In addition, large population-based proteomics studies have reported association between levels of DNAm and plasma proteins<sup>13,14</sup>. Therefore, integrating epigenome-wide DNA methylation data with immune-related proteomics data measured in infant blood offers promising avenues for exploring the impacts of early-life immune responses on the development of asthma.

To this end, we performed an epigenome-wide scan in a cohort of infants with severe bronchiolitis to examine the prospective

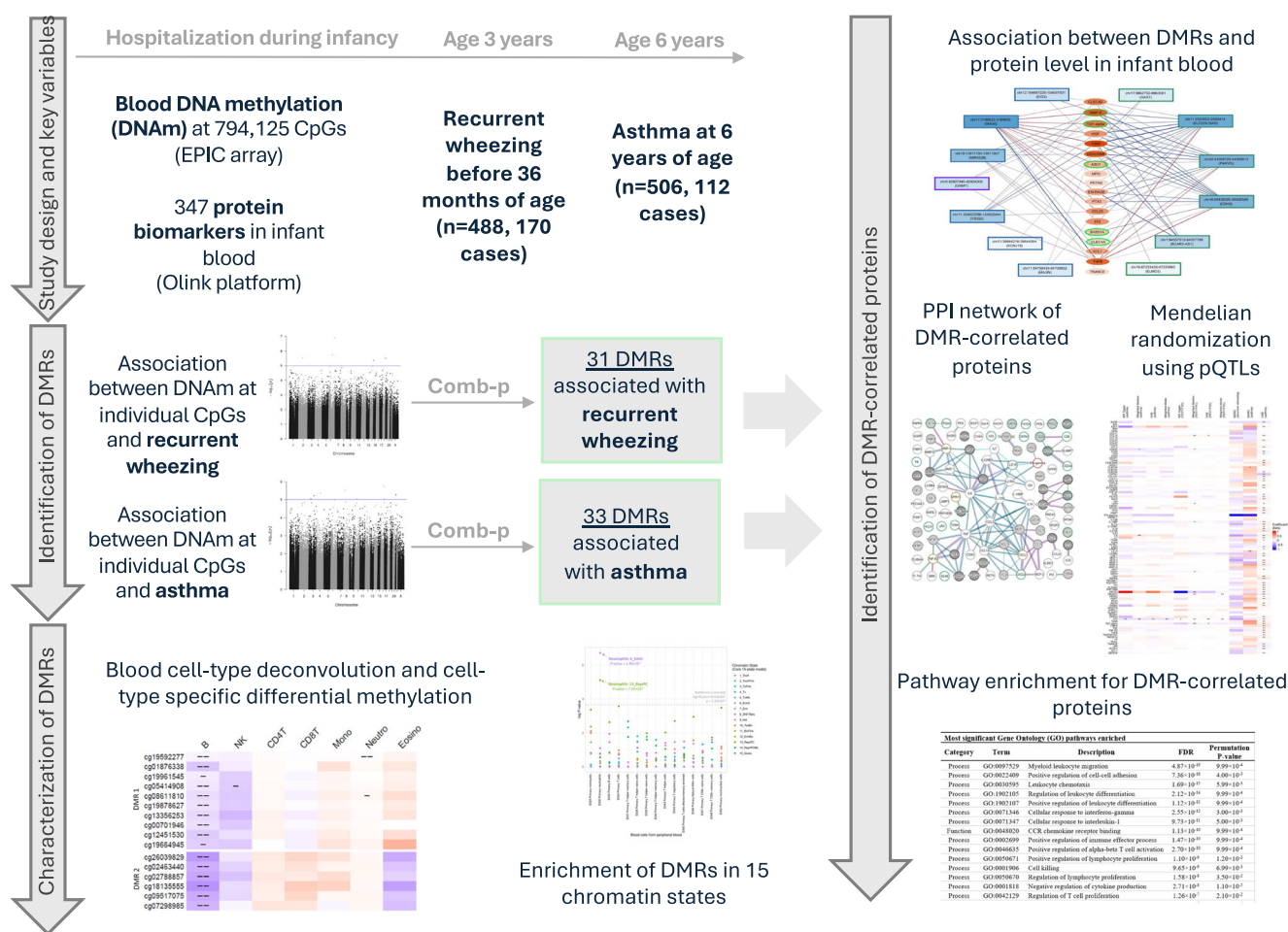
<sup>1</sup>Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA. <sup>2</sup>Department of Emergency Medicine, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA. <sup>3</sup>Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, USA. <sup>4</sup>These authors jointly supervised this work: Kohei Hasegawa, Liming Liang. ✉e-mail: [lliang@hsph.harvard.edu](mailto:lliang@hsph.harvard.edu)

association between DNAm in infant blood and development of recurrent wheezing by age 3 years and childhood asthma by age 6 years. Employing the state-of-the-art EPIC array, we examined the associations across almost 800,000 CpG sites throughout the genome, thereby nearly doubling the coverage of genome compared to previous studies<sup>8,9</sup>. Additionally, we utilized real-time polymerase chain reaction to determine rhinovirus infection status and accounted for its interaction effect in our EWAS discovery. Last, we integrated the EWAS results with data on 347 immune-related proteins measured in the same infant blood samples and identified proteins correlated with the DNAm differences. To prioritize potential intervention targets, we further evaluated the phenotypic associations and causal relationships of these proteins with asthma-related traits.

## Results

Among 625 infants who were enrolled in a prospective 17-center cohort study—the 35th Multicenter Airway Research Collaboration

(MARC-35) cohort—and had high-quality blood DNAm data measured at hospitalization before age 1 year, recurrent wheezing status before age 3 years was available for 488 infants, and asthma status by age 6 years was available for 506 infants (Fig. 1; Supplementary Fig. 1). These participants were included in the EWAS to identify differentially methylated regions (DMRs) associated with recurrent wheezing and asthma, respectively. Supplementary Data 1 shows the characteristics of these participants. Among participants with recurrent wheezing data, 170 (34.8%) developed recurrent wheezing by age 3 years; these infants were more likely to develop asthma by age 6 years compared to infants without recurrent wheezing (44.2% vs. 10.8%,  $p = 2.2 \times 10^{-16}$ ). A total of 112 (22.1%) participants developed asthma by age 6 years in our study sample. We observed notable differences in several baseline variables between recurrent wheezing/asthma cases and non-cases. These variables include age at hospitalization, race/ethnicity, number of previous breathing problems, history of eczema, rhinovirus infection, and maternal asthma.



**Fig. 1 | Study design and analysis workflow.** This study investigates the associations between epigenome-wide DNA methylation (794,125 high-quality CpG sites), measured using the Illumina Infinium MethylationEPIC array, in blood samples from infants with severe bronchiolitis, and the subsequent risk of recurrent wheezing (before 3 years of age,  $n = 448$ , 170 cases) and asthma (at 6 years of age,  $n = 506$ , 112 cases) during childhood. The analytical data were collected from participants of the 35th Multicenter Airway Research Collaboration (MARC-35) study. Thirty-one differentially methylated regions (DMRs) were associated with recurrent wheezing, and thirty-three DMRs were associated with asthma, taking into account potential interactions with rhinovirus infection during infancy. In silico blood cell-type deconvolution and chromatin state enrichment analyses were conducted to identify specific blood cell types likely driving the associations at these DMRs. The

chromatin state annotation was obtained from the Roadmap Epigenomics Consortium. We then explored the relationship between these DMRs and the levels of 347 proteins measured in the same bronchiolitis infant blood samples using the Olink multiplex platform. For proteins associated with the DMRs, we constructed a protein-protein interaction (PPI) network leveraging data from the STRING database, performed two-sample mendelian randomization analysis using protein quantitative trait loci (pQTLs, UK Biobank and deCODE) and GWAS summary statistics from published studies, and conducted pathway enrichment analysis using data from the Gene Ontology database. These analyses aim to provide additional biological insights into the early development of childhood respiratory diseases in infants with severe bronchiolitis underlying the DNA methylation differences.

### Epigenome-wide analysis identified novel DMRs associated with recurrent wheezing

We first examined the associations between DNAm and recurrent wheezing at 794,125 CpG sites that passed the quality control (Supplementary Fig. 2). No substantial genomic inflation was observed as indicated by the quantile-quantile plots (e.g.,  $\lambda_{\text{genomic control}} = 1.03$  for linear regression, Supplementary Fig. 3). We identified 31 DMRs associated with recurrent wheezing after accounting for multiple testing (Šidák  $p < 0.01$ ; Table 1, Supplementary Data 2, Supplementary Data 3). Of these, 21 DMRs were identified through the analysis based on linear regression, and additional 10 DMRs were identified through the analysis based on LRT-2df that leveraged potential interaction with rhinovirus infection to increase power. Ten DMRs exhibited different association patterns between participants who were rhinovirus-infected and those who were not (i.e., the nominal  $p$ -value  $< 0.05$  for the recurrent wheezing  $\times$  rhinovirus infection interaction at more than half of the CpG sites within a DMR).

The most significant recurrent wheezing DMR (chr18:1361190-13611807, Šidák  $p = 1.08 \times 10^{-10}$ ; Table 1) was annotated to the *LDLRAD4*. Across all 11 CpG sites in this region, we observed a lower DNAm level among the recurrent wheezing cases compared with the non-cases, and the effect sizes were more pronounced among participants who had rhinovirus infection during infancy (i.e., 7 out of 11 CpG site in the DMR had nominal  $p$ -value  $< 0.05$  for the interaction term).

### Epigenome-wide analysis identified novel DMRs associated with childhood asthma

We then examined the associations between DNAm at 794,125 CpG sites and childhood asthma. No substantial genomic inflation was observed as indicated by the quantile-quantile plots (e.g.,  $\lambda_{\text{genomic control}} = 1.05$  for linear regression, Supplementary Fig. 4). We identified 33 DMRs associated with asthma after accounting for multiple testing (Šidák  $p < 0.01$ ; Table 2, Supplementary Data 4, Supplementary Data 5). Of these, 20 DMRs were identified through the analysis based on linear regression, and an additional 13 DMRs were identified through the analysis based on LRT-2df. Seven DMRs exhibited different association patterns between participants who were rhinovirus-infected and those who were not (i.e., the nominal  $p$ -value  $< 0.05$  for the asthma  $\times$  rhinovirus infection interaction at more than half of the CpG sites within a DMR).

The most significant asthma DMR (chr17:6899085-6899758, Šidák  $p = 1.31 \times 10^{-13}$ ; Table 2) overlapped the promoter region of *ALOX12*. We observed a higher DNAm level among the asthma cases compared to the non-cases across all 12 CpG sites in this region. This asthma DMR overlapped with a recurrent wheezing DMR near *ALOX12* (chr17:6898821-6899577, Šidák  $p = 3.15 \times 10^{-8}$ ; Table 1). In addition, two other asthma DMRs were also associated with recurrent wheezing, with consistent direction of effects: chr6:28129313-28129656 (*ZKSCAN8PI*; higher DNAm in cases) and chr6:42927940-42928200 (*GNMT*; lower DNAm in cases).

### Cell-type specific associations at the infant blood DMRs

The proportions of seven blood cell types (B cells, NK cells, CD4 T cells, CD8 T cells, monocytes, neutrophils, and eosinophils) were inferred based on the epigenome-wide DNAm data. Utilizing the CellDMC method<sup>15</sup>, which tests the interactions between the cell type fractions in each sample and the outcome of interest, we identified blood cell types that may drive the observed associations between the DMRs and the childhood respiratory outcomes.

Supplementary Fig. 5 illustrates the cell-type specific associations of DNAm with recurrent wheezing at selected DMRs (DMRs displaying nominal statistical significance, i.e.,  $p < 0.05$ , for cell-type specific associations are shown). At the most statistically significant DMR for recurrent wheezing (chr18:1361190-13611807; *LDLRAD4*), the association appeared to be driven by effects in eosinophils, monocytes, and

neutrophils. Other recurrent wheezing DMRs, such as chr2:54086854-54087343 (*GPR75*, *ASB3*; B cells), chr8:22132926-22133076 (*PIWIL2*; B cells), chr11:124622096-124622444 (*VSIG2*; neutrophils, eosinophils), and chr17:46681111-46681401 (*HOXB6*; CD4 T cells, CD8 T cells, eosinophils), also appeared to be influenced by cell-type-specific DNAm effects.

Supplementary Fig. 6 illustrates the cell-type specific associations of DNAm with asthma at selected DMRs with nominal statistical significance ( $p < 0.05$ ) for cell-type specific associations. Asthma DMRs with evidence for cell-type specific DNAm effects include chr6:28058724-28058973 (*ZSCAN12PI*; CD4 T cells), chr7:95025855-95026248 (*PON3*; NK cells), chr16:66400320-66400599 (*CDH5*; monocytes, neutrophils, eosinophils), and chr16:67233432-67233983 (*ELMO3*; CD4 T cells).

### DMRs for childhood respiratory outcomes were enriched in neutrophil-specific chromatin states

We conducted enrichment analysis to examine whether the 61 DMRs for recurrent wheezing and/or asthma were enriched with specific chromatin states in human peripheral blood cells from the Roadmap Epigenomics Project<sup>16</sup>. Our analysis (Fig. 2, Supplementary Data 6) revealed statistically significant enrichment of these DMRs in the Genic Enhancer (6\_EnhG,  $p = 1.90 \times 10^{-7}$ ) and Repressed Polycomb (13\_ReprPC,  $p = 7.65 \times 10^{-6}$ ) regions in primary neutrophils (Accession: E30), highlighting the potential roles of epigenetic regulation in neutrophils during early asthma development.

### DMRs for childhood respiratory outcomes correlated with circulating levels of immune proteins in infant blood

We examined the correlations between DNAm level at the DMRs and 347 immune-related proteins in infant blood, controlling for age, sex, race/ethnicity, and other key covariates. A total of 38 DMRs showed correlation with at least one protein level in infant blood (FDR  $< 0.05$ ; Supplementary Data 1). Among them, the 4 DMRs correlated with the greatest number of proteins, namely, chr17:2169272-2169852 (*SMG6*, 51 proteins), chr22:44568725-44568913 (*PARVG*, 42 proteins), chr16:66400320-66400599 (*CDH5*, 41 proteins), and chr11:2920052-2920414 (*SLC22A18AS*, 37 proteins), shared correlation with the same set of 30 proteins (Supplementary Fig. 7a). These 30 proteins were also correlated with a large number of other DMRs. Some examples of these proteins are CEACAM8 (20 DMRs), MMP-9 (19 DMRs), TNFB (17 DMRs), AZU1 (17 DMRs), OSM (17 DMRs), HGF (17 DMRs), TGF-alpha (15 DMRs), and CLEC4D (15 DMRs) (Supplementary Fig. 7b).

To visually represent the DMR-protein correlations in infant blood, we created Fig. 3 to highlight the 18 proteins with the most statistically significant correlations with DMRs (FDR  $< 0.0005$ ).

### Phenotypic associations between DMR-correlated proteins and asthma

A total of 104 protein markers were statistically significantly (FDR  $< 0.05$ ) correlated with the DMRs for childhood respiratory outcomes. Among these DMR-correlated proteins, levels of LIF-R and CHI3L1 in infant blood were prospectively associated with childhood asthma in a small subset of MARC-35 participants with available data (50 asthma cases, 59 non-cases) after correcting for multiple testing (FDR  $< 0.05$ ; Fig. 4). Supplementary Data 8 provides detailed results for the phenotypic associations between protein levels and the respiratory outcomes. Using this data, we also found 6 and 24 proteins associated with later recurrent wheezing and asthma outcomes, respectively, with nominal significance ( $p < 0.05$ ). Of them, IFN- $\gamma$  was nominally associated with both recurrent wheezing ( $p = 0.006$ ) and asthma ( $p = 0.032$ ). We further leveraged recent findings from the UK Biobank (UKB)<sup>17</sup> where associations between plasma proteomics and prevalent asthma were assessed in a larger population ( $N = 46,595$ ; mean

**Table 1 | Differentially methylated regions (DMRs, Šidák p-value < 0.01) associated with recurrent wheezing before 3 years among infants hospitalized for bronchiolitis (n = 488)**

| DMRs (chromosomeposition) | Width (bp) | N of CpGs | Šidák p-value          | Nearest genes      | Model <sup>a</sup> | Direction of effect <sup>b</sup> | Overlap with promoter | Distance to TSS (bp) <sup>c</sup> | Evidence for open chromatin <sup>d</sup> | Statistically significant interaction by rhinovirus infection <sup>e</sup> |
|---------------------------|------------|-----------|------------------------|--------------------|--------------------|----------------------------------|-----------------------|-----------------------------------|------------------------------------------|----------------------------------------------------------------------------|
| chr1:156616554-156617074  | 521        | 6         | $2.71 \times 10^{-8}$  | BCAN               | LRT-2df            | -                                | No                    | 4814                              | No                                       | Yes                                                                        |
| chr2:54086854-54087343    | 490        | 10        | $2.66 \times 10^{-4}$  | GPR75; ASB3        | both               | -                                | Yes                   | 0                                 | Yes                                      | No                                                                         |
| chr2:186603233-186603639  | 407        | 9         | $4.81 \times 10^{-4}$  | FSIP2              | LRT-2df            | +                                | Yes                   | 0                                 | Yes                                      | Yes                                                                        |
| chr4:187422005-187422343  | 339        | 7         | $2.72 \times 10^{-8}$  | F11-AS1            | linear             | -                                | Yes                   | 0                                 | No                                       | n/a                                                                        |
| chr5:78985425-78985900    | 476        | 9         | $2.87 \times 10^{-4}$  | CMYA5              | LRT-2df            | +                                | Yes                   | 0                                 | No                                       | Yes                                                                        |
| chr6:28129313-28129656    | 344        | 9         | $4.06 \times 10^{-5}$  | ZKSCAN8P1          | linear             | +                                | Yes                   | 0                                 | Yes                                      | n/a                                                                        |
| chr6:42927940-42928303    | 364        | 12        | $5.47 \times 10^{-9}$  | GNMT               | both               | -                                | Yes                   | -197                              | Yes                                      | No                                                                         |
| chr6:142623433-142623767  | 335        | 5         | $6.75 \times 10^{-4}$  | ADGRG6             | both               | -                                | No                    | 377                               | Yes                                      | No                                                                         |
| chr7:4911022-4911224      | 203        | 5         | $1.38 \times 10^{-7}$  | RADIL              | both               | -                                | No                    | 12,111                            | No                                       | No                                                                         |
| chr8:1765217-1765679      | 463        | 9         | $1.21 \times 10^{-5}$  | MIR596             | linear             | -                                | Yes                   | 0                                 | Yes                                      | n/a                                                                        |
| chr8:21232926-22133076    | 151        | 6         | $2.85 \times 10^{-4}$  | PIWIL2             | LRT-2df            | +                                | No                    | 116                               | No                                       | Yes                                                                        |
| chr8:2371171-23712052     | 342        | 5         | $1.21 \times 10^{-5}$  | STC1               | both               | +                                | No                    | 268                               | No                                       | No                                                                         |
| chr11:64739436-64739852   | 417        | 9         | $3.98 \times 10^{-5}$  | MAJIN              | linear             | +                                | Yes                   | 0                                 | Yes                                      | n/a                                                                        |
| chr11:124622096-124622444 | 349        | 5         | $1.09 \times 10^{-5}$  | VSIG2              | LRT-2df            | -                                | Yes                   | 0                                 | No                                       | Yes                                                                        |
| chr12:52281482-52282204   | 723        | 12        | $7.63 \times 10^{-10}$ | ANKRD33            | both               | +                                | Yes                   | 0                                 | Yes                                      | No                                                                         |
| chr12:104697220-104697631 | 412        | 11        | $1.08 \times 10^{-5}$  | EID3; TXNRP1       | LRT-2df            | -                                | Yes                   | 0                                 | Yes                                      | Yes                                                                        |
| chr12:122018897-122019117 | 221        | 7         | $2.34 \times 10^{-4}$  | KDM2B              | LRT-2df            | +                                | Yes                   | 0                                 | No                                       | Yes                                                                        |
| chr12:130822256-130822605 | 350        | 7         | $1.42 \times 10^{-4}$  | PIWIL1             | LRT-2df            | -                                | Yes                   | 0                                 | No                                       | No                                                                         |
| chr15:101085028-101085341 | 314        | 5         | $5.35 \times 10^{-3}$  | CERS3              | linear             | +                                | Yes                   | -103                              | Yes                                      | n/a                                                                        |
| chr16:66955988-66956242   | 255        | 5         | $3.97 \times 10^{-3}$  | RRAD               | linear             | -                                | No                    | 3197                              | No                                       | n/a                                                                        |
| chr17:2169272-2169852     | 581        | 7         | $8.33 \times 10^{-6}$  | SMG6               | both               | -                                | Yes                   | 0                                 | Yes                                      | No                                                                         |
| chr17:6898821-6899577     | 757        | 12        | $3.15 \times 10^{-8}$  | ALOX12             | both               | +                                | Yes                   | 0                                 | No                                       | No                                                                         |
| chr17:7253308-7253720     | 413        | 7         | $2.45 \times 10^{-4}$  | ACAP1              | linear             | +                                | No                    | 13,460                            | Yes                                      | n/a                                                                        |
| chr17:36997420-36997740   | 321        | 8         | $5.43 \times 10^{-6}$  | C17orf98           | both               | -                                | Yes                   | 0                                 | Yes                                      | No                                                                         |
| chr17:46681111-46681401   | 291        | 6         | $4.07 \times 10^{-3}$  | HOXB6              | linear             | +                                | No                    | 933                               | No                                       | n/a                                                                        |
| chr17:47209546-47209827   | 282        | 6         | $3.62 \times 10^{-5}$  | B4GALNT2           | both               | -                                | Yes                   | 0                                 | Yes                                      | No                                                                         |
| chr18:13611190-13611807   | 618        | 11        | $1.08 \times 10^{-10}$ | LDLRAD4            | both               | -                                | Yes                   | 77                                | Yes                                      | Yes                                                                        |
| chr19:18496756-18497201   | 446        | 7         | $2.75 \times 10^{-5}$  | GDF15              | both               | +                                | Yes                   | 0                                 | Yes                                      | No                                                                         |
| chr21:39644216-39644354   | 139        | 6         | $4.95 \times 10^{-4}$  | KCNJ15             | LRT-2df            | -                                | Yes                   | 0                                 | No                                       | Yes                                                                        |
| chrX:48684811-48685182    | 372        | 5         | $1.91 \times 10^{-4}$  | ERAS               | linear             | -                                | Yes                   | 0                                 | No                                       | n/a                                                                        |
| chrX:70712403-70712810    | 408        | 7         | $2.74 \times 10^{-8}$  | INGX; BCYRN1; TAF1 | LRT-2df            | +                                | Yes                   | 0                                 | Yes                                      | Yes                                                                        |

Genome build: GRCh37/hg19. The DMRs were identified by the comb-p method based on the results of linear regression or LRT-2df. A region was identified as a DMR if it had a Šidák p-value < 0.01 and at least 5 CpG sites.  
bp base pair. DMR differentially methylated regions, TSS transcription starting site, LRT-2df likelihood ratio test with 2 degree of freedom.  
“Statistical model in which the DMR was discovered: “linear” represents linear regression without accounting for interaction between recurrent wheezing and rhinovirus infection; “LRT-2df” represents 2 degree of freedom likelihood ratio test accounting for interaction between recurrent wheezing and rhinovirus infection; “both” represents that the DMR was identified by both models (in such case, results with more CpG sites were presented).  
“Distance to transcription starting site: a negative number means the DMR is upstream from the TSS; a positive number means the DMR is downstream from the TSS.  
“Direction of effect was obtained from linear regression or LRT-2df for the main effect of recurrent wheezing. A + represents higher methylation in recurrent wheezing cases compared to non-cases among infants without rhinovirus infection. In the case when the direction of effect was inconsistent across CpG sites within a DMR, we took the direction of the average of coefficients across the CpG sites.  
“Evidence for open chromatin was obtained by single cell ATAC-seq peak in human peripheral blood mononuclear cells (PBMCs) from a healthy donor.  
“For DMRs identified by LRT-2df, we obtained the nominal p-value for the interaction term recurrent wheezing/rhinovirus infection. A “Yes” represents that interaction-p < 0.05 for more than half of the CpG sites in that DMR. Gene names are italicized.

**Table 2 | Differentially methylated regions (DMRs, Šidák  $p$ -value < 0.01) associated with asthma at 6 years among infants hospitalized for bronchiolitis ( $n = 506$ )**

| DMRs (chromosome:position) | Width (bp) | N of CpGs | Šidák $p$ -value       | Nearest genes | Model <sup>a</sup> | Direction of effect <sup>b</sup> | Overlap with promoter | Distance to TSS (bp) <sup>c</sup> | Evidence for open chromatin <sup>d</sup> | Statistically significant interaction by rhinovirus infection <sup>e</sup> |
|----------------------------|------------|-----------|------------------------|---------------|--------------------|----------------------------------|-----------------------|-----------------------------------|------------------------------------------|----------------------------------------------------------------------------|
| chr1:19969931-19970337     | 407        | 10        | $1.61 \times 10^{-5}$  | NBL1          | both               | +                                | No                    | 28                                | No                                       | No                                                                         |
| chr1:46668513-46668840     | 328        | 5         | $7.10 \times 10^{-4}$  | POMGN17       | LRT-2df            | +                                | No                    | 17,137                            | Yes                                      | No                                                                         |
| chr1:94057512-94057789     | 278        | 7         | $4.24 \times 10^{-3}$  | BCAR3-AS1     | linear             | -                                | Yes                   | 0                                 | Yes                                      | n/a                                                                        |
| chr1:247681297-247681766   | 470        | 6         | $4.27 \times 10^{-3}$  | GCSAML        | linear             | +                                | No                    | 10,903                            | Yes                                      | n/a                                                                        |
| chr3:48632568-48633059     | 492        | 8         | $4.23 \times 10^{-6}$  | COL7A1        | both               | -                                | Yes                   | 0                                 | No                                       | No                                                                         |
| chr5:78365647-78365830     | 184        | 7         | $1.75 \times 10^{-2}$  | BHMT2         | LRT-2df            | -                                | No                    | 100                               | No                                       | No                                                                         |
| chr5:135416029-135416613   | 585        | 9         | $2.76 \times 10^{-5}$  | VTRNA2-1      | linear             | -                                | Yes                   | 0                                 | Yes                                      | n/a                                                                        |
| chr6:26225246-26225407     | 162        | 7         | $1.97 \times 10^{-5}$  | HISTH3E       | linear             | -                                | Yes                   | 0                                 | No                                       | n/a                                                                        |
| chr6:28058724-28058973     | 250        | 8         | $1.17 \times 10^{-9}$  | ZSCAN12P1     | LRT-2df            | -                                | No                    | 139                               | Yes                                      | Yes                                                                        |
| chr6:28129313-28129656     | 344        | 9         | $8.32 \times 10^{-5}$  | ZKSCAN8P1     | both               | +                                | Yes                   | 0                                 | Yes                                      | No                                                                         |
| chr6:42927940-42928200     | 261        | 9         | $7.22 \times 10^{-8}$  | GNMT          | both               | -                                | Yes                   | -300                              | Yes                                      | Yes                                                                        |
| chr7:95025855-95026248     | 394        | 14        | $2.20 \times 10^{-4}$  | PON3          | both               | +                                | Yes                   | -168                              | Yes                                      | No                                                                         |
| chr7:150644459-150644866   | 408        | 5         | $1.10 \times 10^{-3}$  | KCNH2         | both               | +                                | No                    | 8051                              | No                                       | No                                                                         |
| chr8:144631768-144631921   | 154        | 5         | $8.49 \times 10^{-4}$  | GSDMD         | LRT-2df            | +                                | No                    | -3636                             | Yes                                      | Yes                                                                        |
| chr10:131989303-131989849  | 547        | 6         | $7.38 \times 10^{-6}$  | GLRX3         | linear             | -                                | No                    | 54,664                            | No                                       | n/a                                                                        |
| chr11:2890258-2890725      | 468        | 24        | $6.86 \times 10^{-7}$  | KCNQ1DN       | linear             | -                                | Yes                   | -538                              | No                                       | n/a                                                                        |
| chr11:2920052-2920414      | 363        | 7         | $1.52 \times 10^{-3}$  | SLC22A18AS    | LRT-2df            | +                                | No                    | 4761                              | Yes                                      | Yes                                                                        |
| chr12:44152565-44152940    | 376        | 9         | $3.46 \times 10^{-5}$  | IRAK4; PUS7L  | linear             | +                                | Yes                   | 0                                 | Yes                                      | n/a                                                                        |
| chr13:11403452-11403796    | 345        | 7         | $2.41 \times 10^{-5}$  | ADPHL1        | both               | -                                | Yes                   | 0                                 | Yes                                      | No                                                                         |
| chr15:23810334-23810861    | 528        | 7         | $1.59 \times 10^{-4}$  | MKRN3         | LRT-2df            | -                                | Yes                   | 0                                 | Yes                                      | No                                                                         |
| chr15:93652578-93653048    | 471        | 5         | $4.20 \times 10^{-7}$  | RGMA          | linear             | +                                | No                    | -20,135                           | No                                       | n/a                                                                        |
| chr15:101389974-101390350  | 377        | 5         | $3.99 \times 10^{-4}$  | ALDH1A3       | LRT-2df            | -                                | No                    | -29,659                           | Yes                                      | Yes                                                                        |
| chr16:66400320-66400599    | 280        | 8         | $5.17 \times 10^{-5}$  | CDH5          | both               | +                                | Yes                   | 0                                 | No                                       | No                                                                         |
| chr16:67233432-67233983    | 552        | 7         | $1.32 \times 10^{-5}$  | ELMO3         | LRT-2df            | -                                | No                    | 20                                | No                                       | No                                                                         |
| chr17:6899085-6899758      | 674        | 12        | $1.31 \times 10^{-13}$ | ALOX12        | both               | +                                | Yes                   | 0                                 | No                                       | No                                                                         |
| chr17:9862752-9863081      | 330        | 10        | $9.06 \times 10^{-3}$  | GAS7          | LRT-2df            | -                                | Yes                   | 0                                 | Yes                                      | Yes                                                                        |
| chr17:38935747-38936092    | 346        | 5         | $2.50 \times 10^{-4}$  | KRT27         | LRT-2df            | +                                | No                    | 2694                              | Yes                                      | Yes                                                                        |
| chr17:47571835-47572132    | 298        | 5         | $8.26 \times 10^{-3}$  | NGFR          | both               | +                                | Yes                   | -523                              | Yes                                      | No                                                                         |
| chr18:77623282-77623598    | 317        | 7         | $6.08 \times 10^{-3}$  | KCNKG2        | linear             | +                                | Yes                   | -70                               | No                                       | n/a                                                                        |
| chr19:45504242-45504516    | 275        | 6         | $2.69 \times 10^{-3}$  | RELB          | LRT-2df            | -                                | Yes                   | -191                              | Yes                                      | No                                                                         |
| chr19:49198900-49199234    | 335        | 5         | $1.71 \times 10^{-4}$  | FUT2          | LRT-2df            | +                                | Yes                   | 0                                 | No                                       | No                                                                         |
| chr21:37915044-37915391    | 348        | 10        | $8.64 \times 10^{-4}$  | CLDN14        | both               | +                                | Yes                   | 0                                 | No                                       | No                                                                         |
| chr22:44568725-44568913    | 189        | 6         | $7.20 \times 10^{-3}$  | PARVG         | LRT-2df            | +                                | Yes                   | 0                                 | Yes                                      | No                                                                         |

Genome build: GRCh37/hg19. The DMRs were identified by the comb-p method based on the results of linear regression or LRT-2df. A region was identified as a DMR if it had a Šidák  $p$ -value < 0.01 and at least 5 CpG sites. bp base pair. DMR differentially methylated regions, n/a not applied, TSS transcription starting site, LRT-2df likelihood ratio test with 2 degree of freedom.

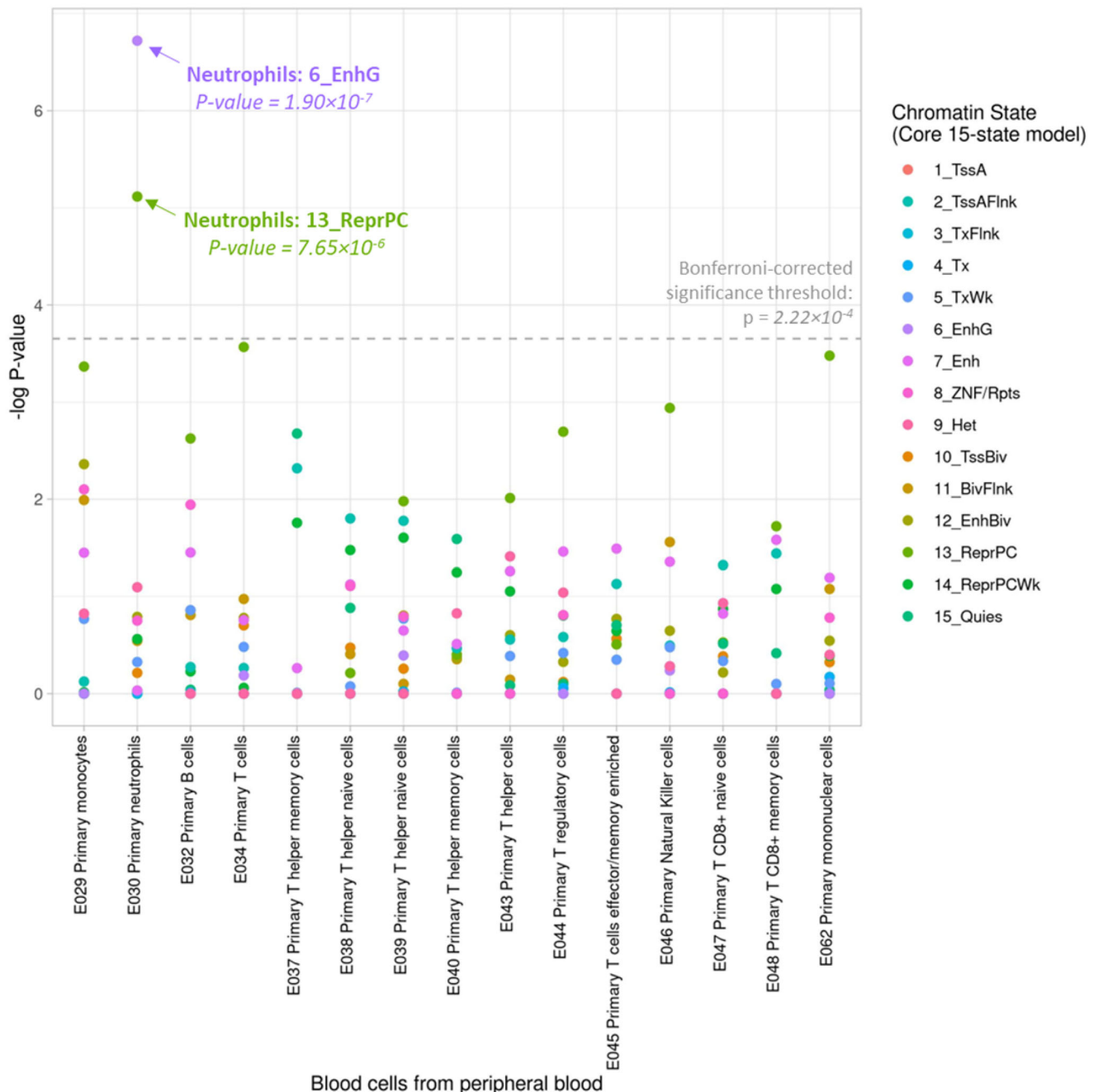
<sup>a</sup>Statistical model in which the DMR was discovered: “linear” represents linear regression without accounting for interaction between asthma and rhinovirus infection; “LRT-2df” represents 2 degree of freedom likelihood ratio test accounting for interaction between asthma and rhinovirus infection; “both” represents that the DMR was identified by both models (in such case, results with more CpG sites were presented).

<sup>b</sup>Distance to transcription starting site: a negative number means the DMR is upstream from the TSS; a positive number means the DMR is downstream from the TSS.

<sup>c</sup>Direction of effect was obtained from linear regression or LRT-2df for the main effect of asthma. A + represents higher methylation in asthma cases compared to non-cases among infants without rhinovirus infection. In the case when the direction of effect was inconsistent across CpG sites within a DMR, we took the direction of the average of coefficients across the CpG sites.

<sup>d</sup>Evidence for open chromatin was obtained by single cell ATAC-seq peak in human peripheral blood mononuclear cells (PBMCs) from a healthy donor.

<sup>e</sup>For DMRs identified by LRT-2df, we obtained the nominal  $p$ -value for the interaction term asthma\*rhinovirus infection. A “Yes” represents that interaction- $p$  < 0.05 for more than half of the CpG sites in that DMR. Gene names are italicized.



**Fig. 2 | Enrichment of DMRs in 15 chromatin states in peripheral blood cells from the NIH Roadmap Epigenomics projects.** The 15 chromatin states were consolidated from the NIH Roadmap Epigenomics project based on five histone modification marks (H3K4me3, H3K4me1, H3K36me3, H3K9me3, H3K27me3) in human cells. Chromatin states are 1\_TssA (Active TSS), 2\_TssAFlnk (Flanking Active TSS), 3\_TxFlnk (Transcr. at gene 5' and 3'), 4\_Tx (Strong transcription), 5\_TxWk (Weak transcription), 6\_EnhG (Genic enhancers), 7\_Enh (Enhancers), 8\_ZNF/Rpts

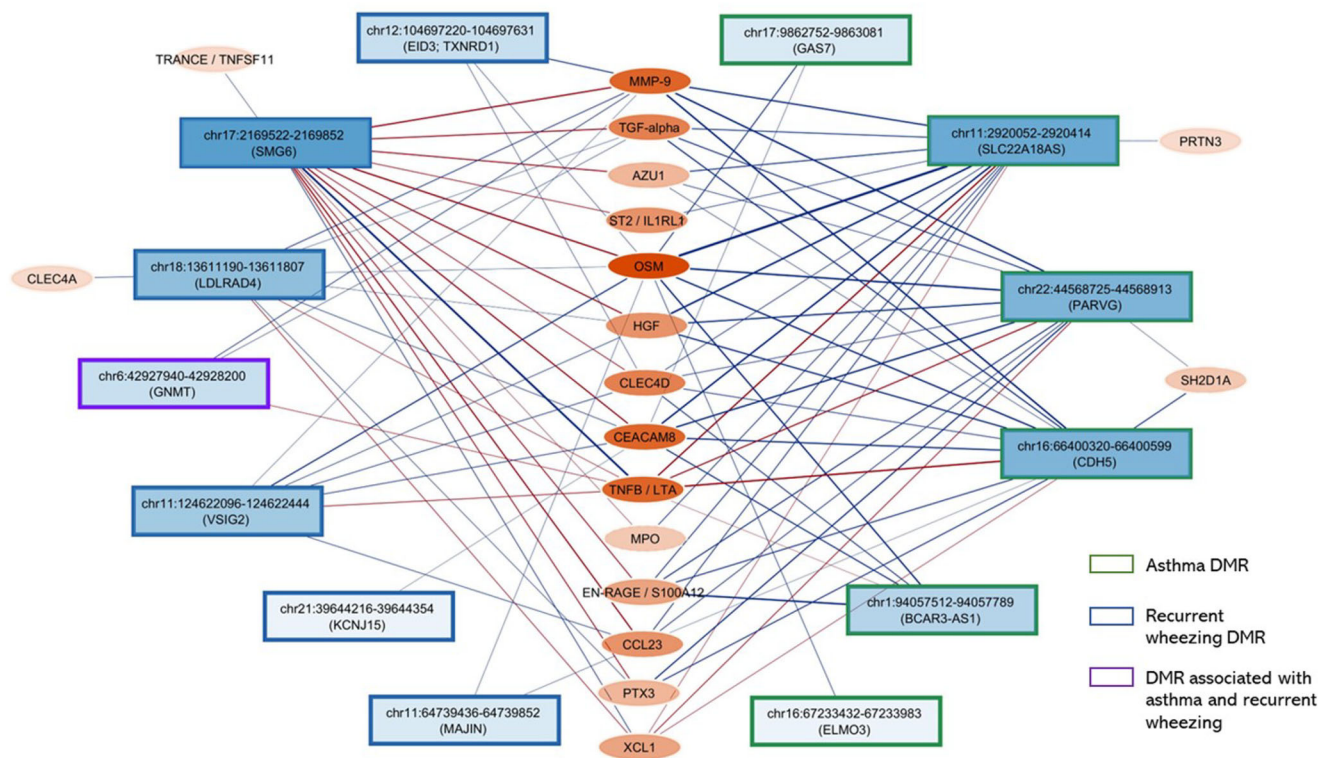
(ZNF gene & repeats), 9\_Het (Heterochromatin), 10\_TssBiv (Bivalent/Poised TSS), 11\_BivFlnk (Flanking Bivalent TSS/Enh), 12\_EnhBiv (Bivalent Enhancer), 13\_ReprPC (Repressed PolyComb), 14\_ReprPCWk (Weak Repressed PolyComb), 15\_Quies (Quiescent/Low). The enrichment *p*-value was obtained based on two-sided binomial test in 1000 permuted samples. To address multiple testing, we employed a Bonferroni-corrected significant threshold of  $p = 2.22 \times 10^{-4}$  (the horizontal dashed line).

age: 57 years). Among all DMR-correlated proteins we identified, 63.5% ( $n = 66$ ) also showed were associated with prevalent asthma in UKB (FDR < 0.05; Fig. 4, Supplementary Data 8).

### Protein–protein interactions and pathway enrichments of the DMR-correlated proteins

We derived protein-protein interaction (PPI) network from the STRING database<sup>18</sup> based on previous experimental evidence and co-expression data (Supplementary Fig. 8). This network suggested that

many DMR-correlated proteins may interact with each other. For instance, CEACAM8, AZU1, PRTN3, and MPO were significantly associated with DNAm at the DMRs (FDR < 0.0005). These proteins were also interconnected with each other through co-expression evidence, suggesting shared biological functions. Similarly, MMP-9, OSM, and HGF were interconnected with each other and also demonstrated evidence for interacting with other cytokines (e.g., IL-6, IL12RB1, IL18, TNF) and chemokines (e.g., CCL3, MCP-1/CCL2) through the PPI network. Additionally, we identified several pairs of known ligand-



**Fig. 3 | Association between DNAm at DMRs and circulating levels of proteins (FDR < 0.0005) among infants hospitalized for bronchiolitis ( $n = 112$ ).** Rectangular nodes are DMRs, with the color shade corresponding to number of proteins the DMRs associated with. Oval nodes are proteins, with the color shade corresponding to number of DMRs the protein associated with. Edges represent associations between DMRs and proteins with a Benjamini-Hochberg FDR below 0.0005

after adjusting for covariates. The thickness of the edges corresponds to the statistical significance of the association (obtained from linear regression, two-sided test) calculated as average  $-\log_{10}(P)$  of the DNAm-protein correlation across all CpG sites within a DMR. Red edges show positive association. Blue edges show negative association.

receptor interactions among the DMR-correlated proteins, including IL-18 and IL-18BP, IL-33 and ST2, IL-12B and IL12RB1, TNF and TNF-R2/TNFRSF1B, and TRAIL/TNFSF10 and TNFSFR10C.

The Gene Ontology (GO) pathway analyses for the 104 DMR-correlated proteins revealed multiple significantly enriched biological and molecular pathways (FDR < 0.05; Supplementary Data 9). Among the most significantly enriched pathways, several were related to the migration, differentiation, and proliferation of leukocytes or T cells. Other top enriched pathways included positive regulation of cell-cell adhesion (FDR =  $7.36 \times 10^{-16}$ ), cellular response to interferon-gamma (FDR =  $2.55 \times 10^{-12}$ ), cellular response to interleukin-1 (FDR =  $9.73 \times 10^{-11}$ ), regulation of tumor necrosis factor production (FDR =  $8.22 \times 10^{-6}$ ), and positive regulation of NIK/NF- $\kappa$ B signaling (FDR =  $1.95 \times 10^{-5}$ ).

### Mendelian randomization for the DMR-correlated proteins and asthma-related outcomes

Utilizing protein quantitative trait loci (pQTLs) mapped in a recent study in the Icelandic deCODE population<sup>19</sup>, we conducted a two-sample mendelian randomization (MR) analysis to examine the causal relationship between the DMR-correlated proteins and asthma and lung function (FEV1/FVC) (Fig. 4, Supplementary Data 1,0). Notably, we found evidence supporting putative causal effects (FDR < 0.05) of plasma ST2 level on both asthma and FEV1/FVC through all MR methods we applied. Our results revealed that a higher plasma ST2 level may decrease the risk of asthma and increase lung function. These results remained robust in the sensitivity analyses when pQTLs derived from a recent UKB study<sup>17</sup> were used as the genetic instruments (Supplementary Fig. 9, Supplementary Data 1,0), as well as when an alternative asthma genome-wide association study (GWAS)

summary statistics were used in the MR analyses (Supplementary Data 11).

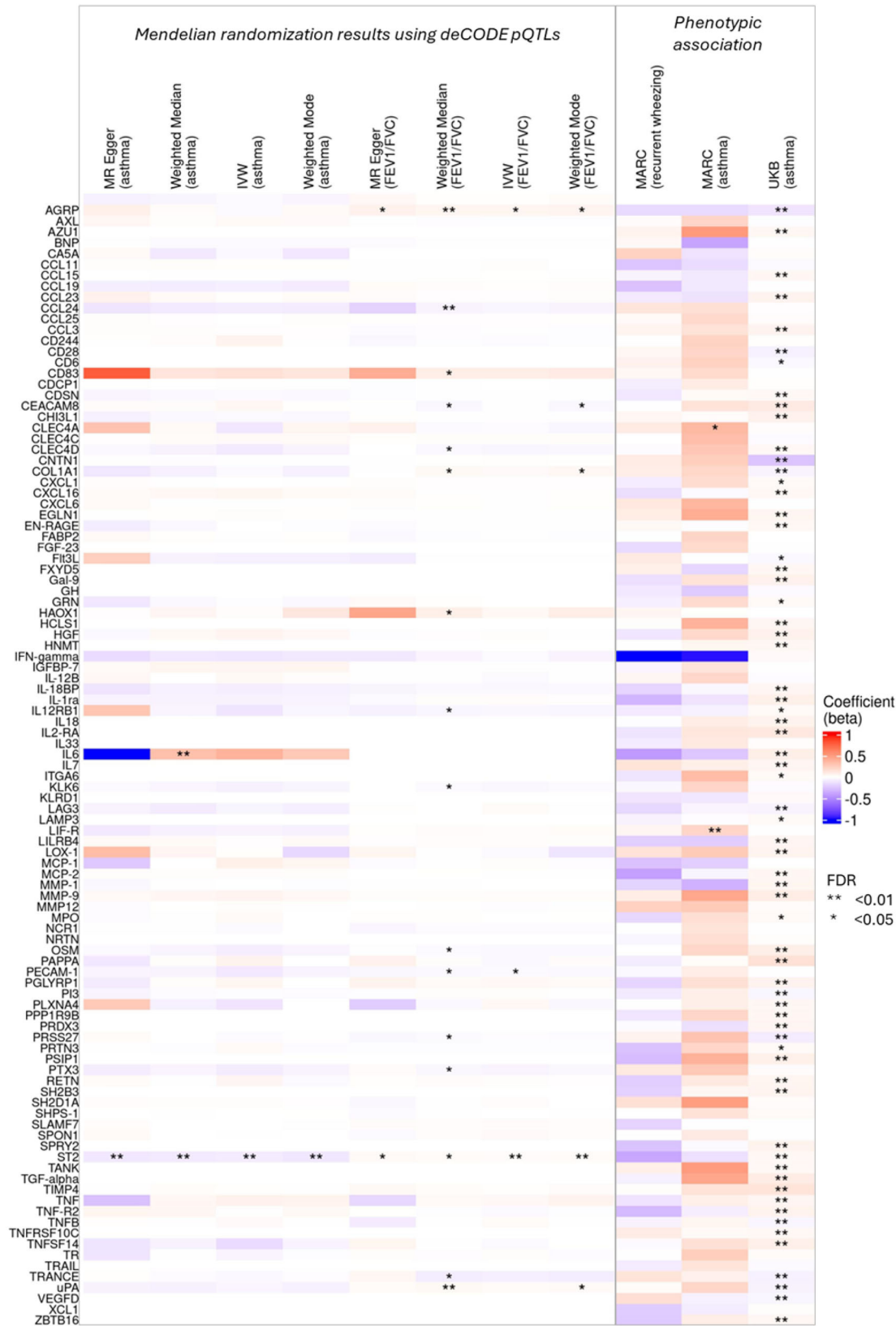
We further investigated whether ST2 may have causal effects on other asthma-related outcomes through MR. We found evidence supporting causal relationships between plasma ST2 level and allergic conditions, including allergic rhinitis and atopic dermatitis (Supplementary Data 11).

Additionally, our MR analyses suggested potential causal effects of plasma AGRP and CHI3L1 levels on FEV1/FVC when the deCODE pQTLs were used as genetic instruments; however, we did not find same results in the analysis using the UKB pQTLs (Supplementary Data 10).

### Associations of DNAm with ST2 protein in infant blood

Given the consistent MR evidence suggesting potential causal effects of plasma ST2 protein level on asthma and related outcomes, we further evaluated the association of DNAm at the *ST2* (also known as *IL1RL1*) locus with ST2 protein level in infant blood ( $n = 112$ , Supplementary Data 12). Among 19 CpG sites at this locus included on the EPIC array, DNAm at 6 CpG sites were significantly associated ST2 protein level after adjusting for key covariates (FDR < 0.05). When further adjusting for estimated blood cell types, the positive association between DNAm at cg17738684 (at distal promoter of *ST2*) and the ST2 protein remained statistically significant (FDR =  $3.06 \times 10^{-3}$ ). Here we note that DNAm at this CpG site was not associated with childhood asthma or recurrent wheezing outcomes in our study sample (Supplementary Data 12).

Additionally, at 8 DMRs associated with recurrent wheezing and/or asthma, we observed correlations between DNAm and the ST2 protein level in infant blood (FDR < 0.05, Supplementary Data 7).



**Fig. 4 | Putative causal relationships and phenotypic associations between DMR-correlated proteins and asthma-related outcomes.** Evidence for the putative causal relationship was obtained using two-sample mendelian randomization (MR Egger, weighted median, inverse variance weighted (IVW), weighted mode) with pQTLs from deCODE as genetic instruments. Phenotypic associations

were obtained using linear regression. The analyses to identify phenotypic association in the MARC-35 cohort were conducted in a small subset of participants with available data (recurrent wheezing: 57 cases, 50 non-cases; asthma: 50 cases, 59 non-cases).

These DMRs are chr1:94057512-94057789 (*BCAR3-ASI*), chr11:2920052-2920414 (*SLC22A18AS*), chr11:124622096-124622444 (*VSIG2*), chr16:66400320-66400599 (*CDH5*), chr17:2169522-2169852 (*SMG6*), chr17:6899085-6899758 (*ALOX12*), chr18:13611190-13611807 (*LDLRAD4*), and chr22:44568725-44568913 (*PARVG*).

## Discussion

In a multi-ethnic cohort of infants with severe bronchiolitis, we conducted EWAS and identified 61 differentially methylated regions (DMRs) where DNAm levels in infant blood were associated with recurrent wheezing by age 3 years and/or asthma by age 6 years. These DMRs were enriched in the genic enhancer of neutrophils in peripheral blood, and many of them demonstrated cell-type-specific associations in deconvolution analysis. Furthermore, we identified 104 proteins correlated with the DMRs in the same infant blood samples; among these proteins, many also had phenotypic associations with respiratory outcomes. Leveraging pQTLs resources from two recent large proteomics GWASs, we conducted mendelian randomization for the DMR-correlated proteins. We reported putative causal effects of plasma ST2 level on asthma, FEV1/FVC, and allergic conditions that often comorbid with asthma. Lastly, we investigated the potential epigenetic regulation at the *ST2* locus and found significant association between *ST2* promoter DNAm and *ST2* protein level in infant blood.

In this study, we focused on childhood recurrent wheezing and asthma – two respiratory outcomes with substantially higher risk in infants with severe bronchiolitis compared to the general pediatric population<sup>3</sup>. While these two conditions are different, they share similar symptoms and potentially overlapping pathobiology since early-age recurrent wheezing often elevates the risk of later chronic asthma<sup>20</sup>. Consistent with prior research in broader pediatric populations<sup>21</sup>, we identified a DMR spanning the promoter region of *ALOX12* that exhibited differential methylation by both recurrent wheezing and childhood asthma in infants with severe bronchiolitis. Additionally, we identified two novel DMRs associated with both recurrent wheezing and asthma, annotated to *ZKSCAN8PI* and *GNMT*, respectively. These three DMRs may offer valuable insights into the shared etiological factors contributing to recurrent wheezing and asthma in children. Of note, the DMR at *ZKSCAN8PI* also harbors a genetic variant associated with asthma in prior work<sup>22</sup>.

The most statistically significant recurrent wheezing DMR was chr18:13611190-13611807 (*LDLRAD4*; Šidák  $p = 1.08 \times 10^{-10}$ ), where a lower DNAm was associated with an elevated risk, and the effect size was larger among those with rhinovirus infection (Supplementary Data 3; Supplementary Fig. 11). A previous study has linked the gene expression of *LDLRAD4*, a known negative regulator of TGF- $\beta$  signaling, to asthma severity<sup>23</sup>. In fact, existing evidence has suggested that various molecular mechanisms, such as miRNA<sup>24</sup> and IFN- $\gamma$  signaling<sup>25</sup>, regulate the TGF- $\beta$  response during asthma development. In the current study, we also observed correlation between IFN- $\gamma$  and later respiratory outcomes, as well as between IFN- $\gamma$  and *LDLRAD4* DNAm in infant blood. It is possible that TGF- $\beta$  response is the shared biological pathway underlying these observed associations, linking the molecular features to subsequent recurrent wheezing outcome. Future investigations are warranted to elucidate the complex regulatory functions across DNAm, IFN- $\gamma$ , and TGF- $\beta$  signaling during infancy that may contribute to later wheezing symptoms, especially among high-risk individuals with rhinovirus bronchiolitis.

We leveraged potential interaction effect of rhinovirus infection in the EWAS and uncovered additional DMRs, including chr8:22132926-22133076 (*PIWIL2*) and chr12:130822256-130822605 (*PIWIL1*). Specifically, the association of hypomethylation at *PIWIL2* with recurrent wheezing was observed exclusively among those with rhinovirus infection (Supplementary Data 3; Supplementary Fig. 10); this association was likely driven by B cell-specific effects (Supplementary Fig. 5). In contrast, the association between *PIWIL1* DNAm and

recurrent wheezing seems to have opposite direction of effects between rhinovirus infection status (Supplementary Data 3). Piwi-interacting RNA (piRNA) is a class of small non-coding RNAs expressed in human germline and somatic cells<sup>26</sup>. These piRNAs interact with PIWI proteins to form piRNA/PIWI complex, exerting regulatory functions in gene silencing through transcriptional and post-transcriptional mechanisms<sup>26</sup>. Recent studies have linked piRNA to various respiratory diseases<sup>26</sup>. An experimental study found that *PIWIL2* downregulates TGF- $\beta$  signal transduction and reduces lung fibrosis<sup>27</sup>; another study reported change in piRNA expression in airway epithelial cells after respiratory virus infection<sup>28</sup>. In concordance with the existing literature, our results further suggest that DNAm of piRNAs may play a role in host response to rhinovirus infection during infancy, contributing to the development of recurrent wheezing in childhood.

Protein serve as the fundamental unit that executes various biological functions for health and disease development. In this study, we leveraged the correlation between DNAm and circulating protein levels to enhance the interpretation of our EWAS results. A recurrent wheezing DMR (chr17:2169272-2169852) at the promoter region of *SMG6* correlated with 51 proteins in infant blood (Supplementary Fig. 7; Supplementary Data 7). Additionally, an asthma DMR at *CDH5* (chr16:66400320-66400599) also displayed correlations with a large number of proteins in infant blood (Supplementary Fig. 7; Supplementary Data 7). We note here that the number of unique protein correlations for each DMR did not correlate with the number of CpG sites within the DMR (Pearson's correlation  $r = 0.09$ ). *SMG6* is an endonuclease in the nonsense-mediated mRNA decay (NMD) pathway, which can degrade erroneous transcripts and perform post-transcriptional regulation<sup>29</sup>. Cadherin 5 (*CDH5*) is a calcium-dependent endothelial adhesion molecule with essential functions in initiating Th2 inflammation and mediating airway remodeling<sup>30</sup>. Evidence has linked *CDH5* to endothelial cell autophagy which keeps neutrophil trafficking under control<sup>31</sup>. Notably, among 37 proteins shared correlations with both the DMR at *CDH5* and the DMR at *SMG6*, many have known functions in neutrophil migration and degranulation. These include Carcinoembryonic Antigen-Related Cell Adhesion Molecule 8 (CEACAM8), Azurocidin (AZU1), Myeloperoxidase (MPO), Myeloblastin (PRTN3), S100 Calcium-binding Protein A12 (S100A12/EN-RAGE), and C-C Motif Chemokine Ligand 23 (CCL23)<sup>32–34</sup>. Consistently, the DMRs we discovered were significantly enriched in the enhancer region of neutrophils in peripheral blood (Fig. 2, Supplementary Data 6). Altogether, these findings highlighted that neutrophil activity during infancy may contribute to asthma etiology and this process may be regulated by epigenetic mechanisms.

Leveraging the pQTLs resources from two largest plasma proteomics GWAS to date (deCODE:  $N = 35,559$ ; UKB:  $N = 46,595$ )<sup>17,19</sup>, we performed MR analyses to explore the potential causal links between the DMR-correlated proteins and asthma-related outcomes. Our investigation, employing various MR methods, consistently indicated that elevated level of circulating ST2 (also called soluble ST2 or interleukin-1 receptor-like 1a, IL1RL1-a) likely reduces the risks of asthma and allergic diseases and improves lung function. We note that in the main MR analyses where pQTLs from the deCODE study were used as instruments, there were no overlapping participants between the outcomes GWASs and the pQTLs GWAS. In the sensitivity analysis where pQTLs from the UKB study were used as instruments, less than 10% of the study population in the outcome GWAS were included in the pQTLs GWAS. However, given this small proportion, the bias in the two-sample MR results due to overlapping UKB participants should be minimal (<1%)<sup>35</sup>. ST2 is a member of the interleukin-1 superfamily and plays crucial roles in immune responses<sup>36</sup>. Its soluble isoform binds to extracellular interleukin-33 (IL-33), lowering the concentration of IL-33 available for interacting with the corresponding transmembrane receptor (ST2L, also called IL1RL1-b), thereby attenuating the pro-

inflammatory cascade<sup>37</sup>. Previous research based on a smaller pQTLs GWAS discovery ( $N = 21,758$ ) reported MR evidence for the protective effects of plasma ST2 against asthma and allergic rhinitis<sup>38</sup>. Our results provided compelling evidence to corroborate these findings. We emphasize that the MR findings should be interpreted with caution, as this statistical method relies on several assumptions. Downstream validation experiments are necessary to confirm the putative causal relationships suggested by the MR analyses.

Previous genetic studies have identified single nucleotide polymorphisms (SNPs) at the *IL1RL1*(*ST2*) locus associated with childhood wheezing phenotypes<sup>39</sup>, as well as correlations between these SNPs and *IL1RL1* mRNA and protein expression in the airway epithelium of asthma patients<sup>40</sup>. The present analyses further revealed novel associations between DNAm at multiple DMRs in relation to childhood respiratory outcomes and circulating ST2 level in infants with severe bronchiolitis. In addition, we identified a CpG site (cg17738684) in the distant promoter of *ST2*, where DNAm was associated with ST2 protein level in infant blood, independent of key covariates and blood cell type composition. This contrasts with a previous study reporting no such association in children's blood<sup>41</sup>. This discrepancy may be explained by differences in population characteristics. Specifically, in our study, all participants were infants experiencing acute bronchiolitis at the time of blood draw, whereas in the prior study, the blood samples were collected from children aged 4–5 years without acute infection<sup>41</sup>. In fact, we previously reported that DNAm changes related to infant bronchiolitis severity were enriched in interleukin-1 mediated signaling pathway<sup>6</sup>. Continued research is imperative to elucidate the mechanism and timing through which epigenetics influences the cellular responses related to ST2 and interleukin-1 signaling, and the subsequent effects on asthma development. While genetic data analysis is beyond the scope of the current study, future investigation into whether the differential methylation at ST2 reflects genetic susceptibility to childhood asthma would offer valuable insights into the underlying biological mechanisms. With ongoing clinical trials investigating the efficacy of drugs targeting the ST2-IL33 pathway in treating adult respiratory diseases<sup>37</sup>, these insights are valuable for identifying and prioritizing high-risk children for preventive and therapeutic interventions in forthcoming studies.

In this study, we adopted a regional-based approach to detect differential methylation, recognizing that methylation levels are generally highly spatially correlated along the DNA sequence as a result of the processivity of DNA methyltransferases and other enzymes modifying the epigenome<sup>42</sup>. Comparing to analysis detecting differential methylation at individual CpG sites, the regional analysis provided advantages such as reducing the burden of multiple testing by removing spatial redundancy and enhancing the robustness and functional relevance of the results<sup>42</sup>. In addition, we applied the *SmartSVA* method<sup>43</sup> to address cell type heterogeneity and batch effects. This method has been shown to adequately account for cell type heterogeneity in EWAS, reducing false positive while preserving statistical power<sup>43</sup>. We conducted additional sensitivity analysis by adjusting for the estimated cell type proportions as covariates and found that the majority of the DMRs remained statistically significant (Supplementary Data 13). This suggests that the *SmartSVA* approach used in our primary analysis effectively accounted for potential confounding by cell type composition. Future studies should aim to collect direct measurements of blood cell metrics in infants alongside DNAm profiling to better capture cell type heterogeneity across samples, as the DNAm-based blood cell composition estimates are only approximations of the true values. Last, we used a relatively stringent significance threshold (Šidák  $p$ -value  $< 0.01$ ) for calling DMRs to control type I error. Indeed, among the DMRs identified in our study, many are proximal to genes previously associated with asthma etiology, reinforcing the biological relevance of our findings. We performed an ad hoc permutation analysis by randomly shuffling the case/non-case

labels for the outcomes. While the number of positive findings in the permuted datasets (6–8 DMRs) was significantly lower than that in our main analysis (20–24 DMRs from linear regression or LRT), it is still not negligible. This suggests that our current statistical models may not yet be optimally calibrated. Therefore, it is important to emphasize that our findings warrant replication in future studies.

This study has potential limitations. First, despite controlling key covariates and cell type heterogeneity in the EWAS, confounding due to unmeasured variables was possible given the observational design. However, since DNAm was measured many years before the respiratory outcomes, our results were less likely to be confounded by exposures after infancy. Second, while several DMRs in our study showed interaction with rhinovirus infection, we were unable to determine if DNAm contributed to infants' susceptibility in response to rhinovirus exposure, or if DNAm and rhinovirus infection had synergistic effects on asthma risk through independent biological pathways. Third, we were unable to establish temporal relationships between DNAm and protein levels. It is possible that the observed DMR-protein correlations reflect some underlying disease processes that lead to changes in both DNAm and protein levels. This may also explain why only a few DMR-associated proteins had support from MR evidence for their effects on asthma-related outcomes. Since the DMRs associated with proteins were not in the genetic regions coding for the proteins, it is also possible that the DNAm reflects a downstream effect of changes in protein levels. Future studies should collect appropriate data to elucidate these relationships. Last, all our study participants were infants with severe bronchiolitis, which was not representative of the general pediatric population. Moreover, data to distinguish different phenotypes/endotypes of recurrent wheezing and asthma cases was unavailable. Future investigations should evaluate the generalizability of our results to other populations and specific disease subtypes.

In summary, in a multi-ethnic cohort of infants with severe bronchiolitis, we discovered multiple novel genomic regions where DNAm in infant blood was prospectively associated with recurrent wheezing and/or asthma in childhood. By integrating epigenome-wide methylation, proteomics, and existing multi-omics resources, our analyses highlighted biological pathways (e.g., TGF- $\beta$  signaling) and effector cell types (e.g., neutrophils) that likely play critical roles during the early stages of asthma development. Furthermore, we provided consistent evidence to support potential causal effects of plasma ST2 level on asthma and related traits and reported DNAm changes in infant blood relevant to this emerging therapeutic target. Overall, our findings advance current understanding of asthma etiology and may pave the way for the development of early interventions to prevent asthma.

## Methods

### Study design, setting, and participants

The institutional review board at each participating hospital approved the study, with written informed consent obtained from the parent or guardian. Details of the participating hospitals can be found in Supplementary Data 14.

The study design and analysis workflow are summarized in Fig. 1. We conducted analysis using data from the 35th Multicenter Airway Research Collaboration (MARC-35) study<sup>44</sup>. Detailed information regarding the study design, setting, and participants can be found in the Supplementary Information. Briefly, MARC-35 is a multicenter prospective cohort study enrolling infants (age  $< 1$  year) hospitalized for bronchiolitis (severe bronchiolitis) during three consecutive bronchiolitis seasons, spanning from 2011 to 2014, across 17 medical centers located in 14 U.S. states (Supplementary Data 14). The diagnosis of bronchiolitis was made by attending physicians and defined as an acute respiratory illness characterized by a combination of rhinitis, cough, tachypnea, wheezing, crackles, or retraction, according to the American Academy of Pediatrics bronchiolitis guidelines<sup>45</sup>. Infants with

preexisting heart or lung disease, immunodeficiency, immunosuppression, or a gestational age of <32 weeks were excluded. All enrolled participants received treatment at the discretion of their attending physicians.

A total of 921 infants were enrolled in the longitudinal MARC-35 cohort. For the current analysis, we limited our study population to participants who had data on DNAm in infant blood and at least one of the following outcomes: recurrent wheezing by age 3 years ( $n = 488$ , 170 cases, 318 non-cases) or asthma by age 6 years ( $n = 506$ , 112 cases, 394 non-cases) (Supplementary Fig. 1).

### Data and sample collection

Participant's demographic data, including sex assigned at birth, age, and parent-reported race and ethnicity, as well as clinical data, including family, environmental, and medical history, and details of the acute illness, were collected via structured interview and medical chart reviews using standardized protocol<sup>4,45</sup>. Following the initial hospitalization at enrollment, trained study personnel initiated telephone interviews with parents/legal guardians at six-month intervals, complemented by medical record evaluations conducted by physicians. All collected data were reviewed at the Emergency Medicine Network (EMNet) Coordinating Center at Massachusetts General Hospital (Boston, MA, USA)<sup>4</sup>. Whole blood samples and nasopharyngeal airway samples were collected within 24 h of hospitalization by trained study personnel at each site following standardized protocols established in previous studies<sup>4,6</sup>. Details regarding the data and sample collection can be found in the Supplementary Information.

### DNA methylation profiling and quality control

Epigenome-wide DNAm was profiled in the blood samples using the Illumina Infinium MethylationEPIC BeadChip (Illumina, San Diego, CA). To ensure the quality of the DNAm data, we applied multiple sample-level and probe-level quality control (QC) filters (Supplementary Fig. 1, Supplementary Fig. 2) following the established data preprocessing pipeline in the R/Bioconductor package minfi<sup>46</sup>. Details of DNA extraction and methylation QC are described in the Supplementary Information and our recent study<sup>6</sup>. Following the QC steps, we applied the single-sample normal-exponential using out-of-band probes (ssNoob) procedure to correct background and dye bias<sup>47</sup>. The current analysis was restricted to probes on chromosomes 1–22 and X.

### Proteomics data measurement and quality control

Levels of circulating proteins were profiled in infant blood samples using the Olink multiplex platform (Olink Bioscience, Uppsala, Sweden). We added internal and external control samples and conducted data QC using the OlinkAnalyze R package<sup>48</sup>. The expression value of each protein was normalized to a unit on the log<sub>2</sub> scale (Normalized Protein eXpression, NPX value), which is proportional to the protein concentration. Details of proteomics profiling and QC are described in the Supplementary Information and our previous study<sup>49</sup>. The current analysis included 345 unique proteins measured on four Olink panels: Immune Response, Inflammation, Cardiovascular II, Cardiovascular III. These proteins are relevant to inflammation or immune functions and are considered most relevant to asthma development.

### Outcomes

The clinical outcomes of interest are (i) recurrent wheezing by age 3 years and (ii) asthma by age 6 years. Specifically, recurrent wheezing was defined as having at least 2 corticosteroid-requiring exacerbations in 6 months or having at least 4 wheezing episodes in 1 year that last at least 1 day and affect sleep, according to the 2007 National Institutes of Health (NIH) asthma guidelines<sup>50</sup>. Asthma was defined based on a commonly used epidemiologic definition<sup>51</sup>: physician diagnosis of asthma by age 6 years with asthma medication use (e.g., inhaled

corticosteroids) or asthma-related symptoms (e.g., wheezing, nocturnal cough) in the preceding year.

### Identification of Rhinovirus infection

The frozen nasopharyngeal samples were first shipped to Baylor College of Medicine (Houston, TX, USA) for identification of rhinovirus viruses. Complementary DNA was generated using a rhinovirus-specific primer, and then a two-step real-time polymerase chain reaction (PCR) was performed. The details of the primer and probe have been described elsewhere<sup>52</sup>.

### Statistical analysis

**Identifying differentially methylated regions (DMRs).** The analytical workflow is summarized in Fig. 1. We followed a two-step approach to identify DMRs for recurrent wheezing and asthma outcomes, respectively. For each outcome, we first examined the associations between the outcome and individual CpG sites using linear regression model and an additional likelihood ratio test with 2 degree-of-freedom (LRT-2df) that leveraged interaction effect by rhinovirus infection to increase power. In the linear regression model, we regressed the DNAm  $M$ -value (a logit-transformation of the %DNAm level) on the outcome (recurrent wheezing or asthma), adjusting for covariates and surrogate variables. We applied the empirical Bayes approach<sup>53</sup> to linear regression to obtain robust estimates of standard error and  $p$ -value. The LRT-2df model assessed the overall association of DNAm with the outcome ( $H_0$ : outcome main effect and outcome  $\times$  rhinovirus infection interaction are both zero). In both the linear regression and the LRT-2df models, we adjusted the following covariates: participants' sex, age at hospitalization, race/ethnicity, insurance type, maternal age, maternal asthma, and maternal smoking during pregnancy. Additionally, surrogate variables were estimated using the R package "SmartSVA"<sup>43</sup> and were adjusted in both models. SmartSVA conducts robust reference-free adjustment of cell mixture and correction for technical batches in EWAS and can control the type I error well while preserving power<sup>43</sup>. In the second step, we applied the comb-p method<sup>54</sup> to combine spatially correlated  $p$ -values and identify genomic regions showing consistent evidence for association with the respiratory outcome across multiple CpG sites (details see Supplementary Information). We provided the  $p$ -values at individual CpG sites from the linear regression and the LRT-2df models to the comb-p algorithm, respectively. DMRs were determined as those with a Šidák  $p < 0.01$  (corrected for multiple testing) and at least 5 CpG sites. For DMRs identified by the LRT-2df model, we obtained the nominal  $p$ -value for the outcome  $\times$  rhinovirus interaction at each CpG site.

Annotations for the DMRs were retrieved from the EPIC manifest file and UCSC genome browser (GRCh37/hg19). Evidence for chromatin accessibility was obtained by identifying the overlap between the DMRs and an ATAC-seq peak reference data measured in human peripheral blood mononuclear cells (PBMCs) from a healthy donor<sup>55</sup>.

**Cell-type specific effect through deconvolution.** To examine if the associations at the DMRs were driven by specific cell types, we applied the CellDMC algorithm<sup>15</sup>, which uses a deconvolution approach to detect cell type specific differential methylation. We first inferred the proportions of seven blood cell types (B cells, NK cells, CD4 T cells, CD8 T cells, monocytes, neutrophils, eosinophils) using the DNAm data and an established reference-based algorithm<sup>56</sup>, and then estimated the DNAm differences by the outcome status that were specific to each blood cell type. These analyses were conducted using the R package "EpiDISH"<sup>57</sup> for 474 CpG sites within the DMRs.

**Cell-type specific enrichment in chromatin states.** We further investigated the enrichment of the DMRs in specific chromatin states across 15 human peripheral blood cell populations using eFORGE 2.0<sup>58</sup>. For each cell type, we tested the enrichment in 15 chromatin states

inferred from histone modification marks and consolidated by the Roadmap Epigenomics Consortium<sup>16</sup>. In this analysis, we applied a 1-kb proximity filter to the CpG sites within the DMRs and adjusted for the background CpG sites included on the EPIC array (details see Supplementary Information). To address multiple testing, we employed a Bonferroni-corrected significance threshold ( $p = 2.22 \times 10^{-4}$ ).

**Linking circulating proteins levels to the DNAm in infant blood.** In a subset of our study participants with available proteomics data ( $n = 112$ ), we used linear regression models to examine the association between DNAm and levels of circulating proteins, controlling the same covariates as above. This analysis was run for each CpG-protein pair from the 474 CpG sites (within DMRs) and 347 proteins. False discovery rate (FDR) was calculated using the Benjamini-Hochberg procedure. Proteins correlated with any CpG site with an FDR < 0.05 were considered as DMR-correlated protein and included in downstream analyses.

For these proteins, we used adjusted linear regression to investigate whether the protein levels in infant blood were associated with subsequent recurrent wheezing ( $n = 107$ ) and asthma ( $n = 109$ ) outcomes within the MARC-35 cohort. In addition, we compared our findings to a recent publication<sup>17</sup> from UKB to explore the relationship between these proteins and prevalent asthma in a larger population.

In addition, we focused on 19 CpG sites at the *ST2* locus (chr2:102926502-10296819) on the EPIC array and evaluated the association of DNAm with *ST2* protein level in infant blood (details see Supplementary Information).

**Protein-proteins interaction (PPI) network and pathway enrichment analysis.** We leveraged the publicly available database STRING<sup>18</sup> to construct a PPI network for the DMR-correlated proteins based on the following types of existing evidence: (1) PPI from curated databases, (2) experimentally determined PPI, (3) Protein co-expression, and (4) Protein homology.

Furthermore, we performed Gene Ontology (GO) pathway enrichment analysis for DMR-correlated proteins using R/Bioconductor package “STRINGdb”<sup>18</sup>. We calculated Benjamini-Hochberg FDR within each category. Given the Olink protein panel was designed for immune-related functions, we calculated permutation  $p$ -values for the enriched pathways to mitigate the impact of proteins selection on the pathway enrichment results. Details of the permutation test are described in Supplementary Information. We restricted pathways with 30–300 background proteins to eliminate pathways with either too specific or too broad functional annotations.

**Mendelian randomization (MR) for DMR-correlated proteins.** We conducted two-sample MR analysis to investigate the potential causal relationship of all DMR-correlated proteins on asthma<sup>59</sup> and FEV1/FVC<sup>60</sup> using GWAS summary statistics. Leveraging recent large-scale GWAS for plasma proteomics from deCODE (SomaScan platform)<sup>19</sup> and UKB (Olink platform)<sup>17</sup>, we obtained a list of genome-wide significant cis- and trans-pQTLs as candidates for the genetic instruments. After clumping and selection of instruments (details see Supplementary Information), we applied four MR methods (MR-Egger, inverse variance weighted, weighed median, weighted mode) to proteins with at least 3 pQTLs using R package “TwoSampleMR”<sup>61</sup>. Benjamini-Hochberg FDR was calculated to correct for multiple testing. Additionally, we conducted two-sample MR analysis to further examine the potential causal relationship of *ST2* protein level on other allergy-related traits (allergic rhinitis, atopic dermatitis, serum IgE level)<sup>62–66</sup>.

All statistical tests included in the current study are two-sided.

## Reporting summary

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

## Data availability

The raw data generated in this study have been deposited to the NIH/NIAID ImmPort database under Accession ID [SDY2306](https://www.ncbi.nlm.nih.gov/immport/SDY2306). The raw data is available under restricted access to ensure compliance with the informed consent forms of the MARC-35 study and the genomic data sharing plan. Requests for access can be made by directly contacting the corresponding author. The expected timeframe for response and the duration of access is subject to the data sharing policy of the NIH/NIAID ImmPort database. The EWAS summary statistics are available at <http://lianglab.rc.fas.harvard.edu/AsthmaWheezingEWAS/>. All other data are publicly available through the original studies' website. The EPIC array manifest file is available at <https://support.illumina.com/downloads/infinium-methylationepic-v1-0-product-files.html>. Annotation information from UCSC genome browser is available at <https://genome.ucsc.edu/>. The ATAC-seq peak reference data in human PBMCs can be downloaded from <https://www.10xgenomics.com/datasets/pbmc-from-a-healthy-donor-no-cell-sorting-10-k-1-standard-2-0-0>. The chromatin states data from the NIH Roadmap Epigenomics Project is available at [https://egg2.wustl.edu/roadmap/web\\_portal/index.html](https://egg2.wustl.edu/roadmap/web_portal/index.html). The STRING protein-protein-interaction data is available at <https://string-db.org/>. The deCODE pQTLs GWAS summary statistics is available at <https://www.decode.com/summarydata/>. The UKB pQTLs GWAS summary statistics is available at <https://metabolomips.org/ukbbpgwas/>. The GWAS summary statistics for asthma, FEV1/FVC, and other outcomes can be downloaded from <https://gwas.mrcieu.ac.uk/> (accession numbers: GCST90014325, GCST007431, GCST90018795, GCST005038, GCST90038664, GCST90027161, GCST90027161).

## Code availability

The DNA methylation data pre-processing was conducted using the R/Bioconductor package *minfi* v1.44.0 (<https://bioconductor.org/packages/release/bioc/html/minfi.html>). The epigenome-wide linear regression models were implemented using the R package *meffil* v1.3.6 (<https://github.com/perishky/meffil>). The epigenome-wide likelihood ratio tests were conducted using the R package *lmtest* v0.9.40 (<https://cran.r-project.org/web/packages/lmtest/index.html>). The cell type and batch control in EWAS was conducted by the R package *SmartSVA* v0.1.3 (<https://cran.r-project.org/web/packages/SmartSVA/index.html>). The region-based analyses were performed using *comb-p* v0.50.6 (<https://github.com/brentp/combined-p-values>). The blood cell type deconvolution analysis was performed using the R/Bioconductor package *EpiDISH* v2.14.1 (<https://github.com/sjczheng/EpiDISH/tree/master>). The DMR chromate states enrichment analysis was performed using *eFORGE* v2.0 (<https://eforge.altiusinstitute.org/>). The proteomics data pre-processing was conducted using R package *OlinkAnalyze* v3.3.1 (<https://github.com/Olink-Proteomics/OlinkRPackage>). The protein GO pathway enrichment analysis was performed using R/Bioconductor package *STRINGdb* v2.10.1 (<https://www.bioconductor.org/packages/release/bioc/html/STRINGdb.html>). The two-sample mendelian randomization was conducted using R package *TwoSampleMR* v0.5.10 (<https://mrcieu.github.io/TwoSampleMR/>).

## References

- Hasegawa, K., Dumas, O., Hartert, T. V. & Camargo, C. A. Advancing our understanding of infant bronchiolitis through phenotyping and endotyping: clinical and molecular approaches. *Expert Rev. Respir. Med.* **10**, 891–899 (2016).
- Fujiogi, M. et al. Trends in bronchiolitis hospitalizations in the United States: 2000–2016. *Pediatrics* **144**, e20192614 (2019).
- Hasegawa, K., Mansbach, J. M. & Camargo, C. A. Infectious pathogens and bronchiolitis outcomes. *Expert Rev. Anti Infect. Ther.* **12**, 817–828 (2014).
- Hasegawa, K. et al. Association of Rhinovirus C bronchiolitis and immunoglobulin E sensitization during infancy with development of recurrent wheeze. *JAMA Pediatr.* **173**, 544–552 (2019).

5. Cavalli, G. & Heard, E. Advances in epigenetics link genetics to the environment and disease. *Nature* **571**, 489–499 (2019).
6. Zhu, Z. et al. Epigenome-wide association analysis of infant bronchiolitis severity: a multicenter prospective cohort study. *Nat. Commun.* **14**, 5495 (2023).
7. Arathimos, R. et al. Epigenome-wide association study of asthma and wheeze in childhood and adolescence. *Clin. Epigenetics* **9**, 112 (2017).
8. Reese, S. E. et al. Epigenome-wide meta-analysis of DNA methylation and childhood asthma. *J. Allergy Clin. Immunol.* **143**, 2062–2074 (2019).
9. Xu, C.-J. et al. DNA methylation in childhood asthma: an epigenome-wide meta-analysis. *Lancet* **379**. [https://doi.org/10.1016/S2213-2600\(18\)30052-3](https://doi.org/10.1016/S2213-2600(18)30052-3) (2018).
10. Krusche, J., Basse, S. & Schaub, B. Role of early life immune regulation in asthma development. *Semin. Immunopathol.* **42**, 29–42 (2020).
11. Lambrecht, B. N., Hammad, H. & Fahy, J. V. The cytokines of asthma. *Immunity* **50**, 975–991 (2019).
12. Lukacs, N. W. Role of chemokines in the pathogenesis of asthma. *Nat. Rev. Immunol.* **1**, 108–116 (2001).
13. Zaghlool, S. B. et al. Epigenetics meets proteomics in an epigenome-wide association study with circulating blood plasma protein traits. *Nat. Commun.* **11**, 15 (2020).
14. Myte, R., Sundkvist, A., Van Guelpen, B. & Harlid, S. Circulating levels of inflammatory markers and DNA methylation, an analysis of repeated samples from a population based cohort. *Epigenetics* **14**, 649–659 (2019).
15. Zheng, S. C., Breeze, C. E., Beck, S. & Teschendorff, A. E. Identification of differentially methylated cell types in epigenome-wide association studies. *Nat. Methods* **15**, 1059–1066 (2018).
16. Kundaje, A. et al. Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).
17. Sun, B. B. et al. Plasma proteomic associations with genetics and health in the UK Biobank. *Nature* **622**, 329–338 (2023).
18. Szklarczyk, D. et al. The STRING database in 2021: customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res.* **49**, D605–D612 (2020).
19. Eldjarn, G. H. et al. Large-scale plasma proteomics comparisons through genetics and disease associations. *Nature* **622**, 348–358 (2023).
20. Tenero, L., Piazza, M. & Piacentini, G. Recurrent wheezing in children. *Transl. Pediatr.* **5**, 31–36 (2016).
21. Morales, E. et al. DNA hypomethylation at ALOX12 is associated with persistent wheezing in childhood. *Am. J. Respir. Crit. Care Med.* **185**, 937–943 (2012).
22. Zhu, Z. et al. Shared genetics of asthma and mental health disorders: a large-scale genome-wide cross-trait analysis. *Eur. Respir. J.* **54**, 1901507 (2019).
23. Weathington, N. et al. BAL cell gene expression in severe asthma reveals mechanisms of severe disease and influences of medications. *Am. J. Respir. Crit. Care Med.* **200**, 837–856 (2019).
24. Zhu, Z. et al. Nasal airway microRNA profiling of infants with severe bronchiolitis and risk of childhood asthma: a multicentre prospective study. *Eur. Respir. J.* **62**, 2300502 (2023).
25. Zhu, Z. et al. Integrated nasopharyngeal airway metagenome and asthma genetic risk endotyping of severe bronchiolitis in infancy and risk of childhood asthma. *Eur Respir J.* <https://doi.org/10.1183/13993003.01130-2024> (2024).
26. Yao, Y. et al. The emerging role of the piRNA/PIWI complex in respiratory tract diseases. *Respir. Res.* **24**, 76 (2023).
27. Zou, G.-L. et al. The role of Nrf2/PIWIL2/purine metabolism axis in controlling radiation-induced lung fibrosis. *Am. J. Cancer Res.* **10**, 2752–2767 (2020).
28. Corsello, T., Kudlicki, A. S., Liu, T. & Casola, A. Respiratory syncytial virus infection changes the piwi-interacting RNA content of airway epithelial cells. *Front. Mol. Biosci.* **9**, 931354 (2022).
29. Sun, B. & Chen, L. Mapping genetic variants for nonsense-mediated mRNA decay regulation across human tissues. *Genome Biol.* **24**, 164 (2023).
30. Asosingh, K. et al. Nascent endothelium initiates Th2 polarization of asthma. *J. Immunol.* **190**, 3458–3465 (2013).
31. Reglero-Real, N., Pérez-Gutiérrez, L. & Nourshargh, S. Endothelial cell autophagy keeps neutrophil trafficking under control. *Autophagy* **17**, 4509–4511 (2021).
32. Lacy, P. Mechanisms of degranulation in neutrophils. *Allergy Asthma Clin. Immunol.* **2**, 98 (2006).
33. Akgun, E. et al. Proteins associated with neutrophil degranulation are upregulated in nasopharyngeal swabs from SARS-CoV-2 patients. *PLoS ONE* **15**, e0240012 (2020).
34. Rouleau, P. et al. the calcium-binding protein S100A12 induces neutrophil adhesion, migration, and release from bone marrow in mouse at concentrations similar to those found in human inflammatory arthritis. *Clin. Immunol.* **107**, 46–54 (2003).
35. Burgess, S., Davies, N. M. & Thompson, S. G. Bias due to participant overlap in two-sample Mendelian randomization. *Genet. Epidemiol.* **40**, 597–608 (2016).
36. Kakkar, R. & Lee, R. T. The IL-33/ST2 pathway: therapeutic target and novel biomarker. *Nat. Rev. Drug Discov.* **7**, 827–840 (2008).
37. Riera-Martínez, L., Cànoves-Gómez, L., Iglesias, A., Martín-Medina, A. & Cosío, B. G. The role of IL-33/ST2 in COPD and its future as an antibody therapy. *Int. J. Mol. Sci.* **24**, 8702 (2023).
38. Folkersen, L. et al. Genomic and drug target evaluation of 90 cardiovascular proteins in 30,931 individuals. *Nat. Metab.* **2**, 1135–1148 (2020).
39. Savenije, O. E. et al. Association of IL33-IL-1 receptor-like 1 (IL1RL1) pathway polymorphisms with wheezing phenotypes and asthma in childhood. *J. Allergy Clin. Immunol.* **134**, 170–177 (2014).
40. Portelli, M. A. et al. Phenotypic and functional translation of IL1RL1 locus polymorphisms in lung tissue and asthmatic airway epithelium. *JCI Insight* **5**, e132446 (2020).
41. Dijk, F. N. et al. Genetic regulation of IL1RL1 methylation and IL1RL1-a protein levels in asthma. *Eur. Respir. J.* **51**, 1701377 (2018).
42. Teschendorff, A. E. & Relton, C. L. Statistical and integrative system-level analysis of DNA methylation data. *Nat. Rev. Genet.* **19**, 129–147 (2018).
43. Chen, J. et al. Fast and robust adjustment of cell mixtures in epigenome-wide association studies with SmartSVA. *BMC Genom.* **18**, 413 (2017).
44. Hasegawa, K. et al. Association of nasopharyngeal microbiota profiles with bronchiolitis severity in infants hospitalised for bronchiolitis. *Eur. Respir. J.* **48**, 1329–1339 (2016).
45. Ralston, S. L. et al. Clinical practice guideline: the diagnosis, management, and prevention of bronchiolitis. *Pediatrics* **134**, e1474–e1502 (2014).
46. 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).
47. Fortin, J. P., Triche, T. J. & Hansen, K. D. Preprocessing, normalization and integration of the Illumina HumanMethylationEPIC array with minfi. *Bioinformatics* **33**, 558–560 (2017).
48. Nevola, K. et al. OlinkAnalyze: facilitate analysis of proteomic data from Olink. R package version 3.3.1. <https://CRAN.R-project.org/package=OlinkAnalyze> (2023).
49. Ooka, T. et al. Proteomics endotyping of infants with severe bronchiolitis and risk of childhood asthma. *Allergy* **77**, 3350–3361 (2022).
50. National Asthma Education and Prevention Program. Expert panel report 3 (EPR-3): guidelines for the diagnosis and management of

- asthma-summary report 2007. *J. Allergy Clin. Immunol.* **120**, S94–S138 (2007).
51. Raita, Y. et al. Nasopharyngeal metatranscriptome profiles of infants with bronchiolitis and risk of childhood asthma: a multi-centre prospective study. *Eur. Respir. J.* **60**, 2102293 (2022).
  52. Lu, X. et al. Real-time reverse transcription-PCR assay for comprehensive detection of human rhinoviruses. *J. Clin. Microbiol.* **46**, 533–539 (2008).
  53. Smyth, G. K. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.* **3**, Article3 (2004).
  54. Pedersen, B. S., Schwartz, D. A., Yang, I. V. & Kechris, K. J. Comb-p: software for combining, analyzing, grouping and correcting spatially correlated *P*-values. *Bioinformatics* **28**, 2986–2988 (2012).
  55. PBMC from a Healthy Donor - No Cell Sorting (10k). *10x Genomics* <https://www.10xgenomics.com/datasets/pbmc-from-a-healthy-donor-no-cell-sorting-10-k-1-standard-2-0-0>.
  56. Teschendorff, A. E., Breeze, C. E., Zheng, S. C. & Beck, S. A comparison of reference-based algorithms for correcting cell-type heterogeneity in Epigenome-Wide Association Studies. *BMC Bioinforma.* **18**, 105 (2017).
  57. Zheng, S. C. et al. EpiDISH web server: epigenetic dissection of intra-sample-heterogeneity with online GUI. *Bioinformatics* **36**, 1950–1951 (2020).
  58. Breeze, C. E. et al. eFORGE v2.0: updated analysis of cell type-specific signal in epigenomic data. *Bioinformatics* **35**, 4767–4769 (2019).
  59. Valette, K. et al. Prioritization of candidate causal genes for asthma in susceptibility loci derived from UK Biobank. *Commun. Biol.* **4**, 1–15 (2021).
  60. Shrine, N. et al. New genetic signals for lung function highlight pathways and chronic obstructive pulmonary disease associations across multiple ancestries. *Nat. Genet.* **51**, 481–493 (2019).
  61. Hemani, G., Tilling, K. & Smith, G. D. Orienting the causal relationship between imprecisely measured traits using GWAS summary data. *PLOS Genet.* **13**, e1007081 (2017).
  62. Ferreira, M. A. et al. Shared genetic origin of asthma, hay fever and eczema elucidates allergic disease biology. *Nat. Genet.* **49**, 1752–1757 (2017).
  63. Dönertaş, H. M., Fabian, D. K., Fuentealba, M., Partridge, L. & Thornton, J. M. Common genetic associations between age-related diseases. *Nat. Aging* **1**, 400–412 (2021).
  64. Sliz, E. et al. Uniting biobank resources reveals novel genetic pathways modulating susceptibility for atopic dermatitis. *J. Allergy Clin. Immunol.* **149**, 1105–1112.e9 (2022).
  65. Scepanovic, P. et al. Human genetic variants and age are the strongest predictors of humoral immune responses to common pathogens and vaccines. *Genome Med.* **10**, 59 (2018).
  66. Sakaue, S. et al. A cross-population atlas of genetic associations for 220 human phenotypes. *Nat. Genet.* **53**, 1415–1424 (2021).

## Acknowledgements

This study was supported by grants from the National Institutes of Health (Bethesda, MD): R01 AI-148338 (L.L. and K.H.), K01 AI-153558 (Z.Z.), U01 AI-087881 (C.A.C.), R01 AI-114552 (C.A.C.), R01 AI-127507 (C.A.C.), R01 AI-134940 (K.H.), R01 AI-137091 (K.H.), and UG3/UH3 OD-023253 (C.A.C.); Massachusetts General Hospital Department of Emergency Medicine Fellowship/Eleanor and Miles Shore Faculty Development Awards Program (Z.Z.); and the Harvard University William F. Milton Fund (Z.Z.). The content of this manuscript is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. We would like to thank the participants and researchers from the Multicenter Airway Research Collaboration (MARC) who significantly contributed or collected data.

## Author contributions

Y.L. conceptualized and designed the study, performed data quality control and statistical analysis, drafted the initial manuscript, and approved the final manuscript. Z.Z. critically reviewed and revised the initial manuscript, and approved the final manuscript. C.A.C. conceptualized and designed the parent study, obtained funding, collected the data, supervised the conduct of parent study, critically reviewed and revised the initial manuscript, and approved the final manuscript. J.A.E. performed data quality control, reviewed the manuscript, and approved the final manuscript. K.H. conceptualized and designed the study, obtained funding, supervised the conduct of study, reviewed the manuscript, and approved the final manuscript. L.L. conceptualized and designed the study, obtained funding, collected the data, supervised the conduct of study, critically reviewed and revised the initial manuscript, and approved the final manuscript.

## Competing interests

The authors declare the following competing interests. Z.Z. reports grants from National Institutes of Health during the conduct of the study. C.A.C. reports grants from National Institutes of Health during the conduct of the study. K.H. reports grants from National Institutes of Health during the conduct of the study; grants from Novartis, outside the submitted work. L.L. reports grants from National Institutes of Health during the conduct of the study. All other authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-025-57288-6>.

**Correspondence** and requests for materials should be addressed to Liming Liang.

**Peer review information** *Nature Communications* thanks John Holloway and Klaus Bønnelykke for their contribution to the peer review of this work. A peer review file is available.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

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

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

© The Author(s) 2025