



















## ORIGINAL ARTICLE

## Asthma and Lower Airway Disease

# Epigenetic training of human bronchial epithelium cells by repeated rhinovirus infections

Marua Abu Risha<sup>1,2</sup>  | Karosham D. Reddy<sup>2,3,4</sup>  | Sai Sneha Priya Nemani<sup>2,3</sup>  |  
 Constanze Jakwerth<sup>5,6</sup>  | Carsten Schmidt-Weber<sup>5,6</sup>  | Thomas Bahmer<sup>2,7,8</sup>  |  
 Gesine Hansen<sup>9,10</sup>  | Erika von Mutius<sup>6,11,12</sup>  | Klaus F. Rabe<sup>5,6</sup>  |  
 Anna-Maria Dittrich<sup>9,10</sup>  | Ruth Grychtol<sup>9,10</sup>  | Nicole Maison<sup>6,11,12</sup>  |  
 Bianca Schaub<sup>6,12,13</sup>  | Matthias V. Kopp<sup>14</sup>  | Folke Brinkmann<sup>2,3</sup>  |  
 Silke Meiners<sup>2,15</sup>  | Uta Jappe<sup>1,2,16</sup>  | Markus Weckmann<sup>2,3,4</sup>  |  
 the ALLIANCE Study Group as part of the German Centre for Lung Research (DZL)

## Correspondence

Markus Weckmann, German Center for Lung Research (DZL), Airway Research Center North (ARCN), Borstel, Germany.  
 Email: [markus.weckmann@uksh.de](mailto:markus.weckmann@uksh.de)

## Funding information

Federal Ministry of Education and Science: German Center for Lung Research / Deutsches Zentrum für Lungenforschung (DZL) DZL 001C1

## Abstract

**Background:** Humans are subjected to various environmental stressors (bacteria, viruses, pollution) throughout life. As such, an inherent relationship exists between the effect of these exposures with age. The impact of these environmental stressors can manifest through DNA methylation (DNAm). However, whether these epigenetic effects selectively target genes, pathways, and biological regulatory mechanisms remains unclear. Due to the frequency of human rhinovirus (HRV) infections throughout life (particularly in early development), we propose the use of HRV under controlled conditions can model the effect of multiple exposures to environmental stressors.

**Methods:** We generated a prediction model by combining transcriptome and DNAm datasets from human epithelial cells after repeated HRV infections. We applied a novel experimental statistical design and method to systematically explore the multifaceted experimental space (number of infections, multiplicity of infections and duration). Our model included 35 samples, each characterized by the three parameters defining their infection status.

**Results:** Trainable genes were defined by a consistent linear directionality in DNAm and gene expression changes with successive infections. We identified 77 trainable genes which could be further explored in future studies. The identified methylation

**Abbreviations:** ATAC-seq, Assay for Transposase-Accessible Chromatin using sequencing; BAM, binary alignment map; DAMPs, damage-associated molecular pattern; DEGs, differentially expressed genes; DMPs, differentially methylated probes; DNAm, DNA methylation; DoE, Design of Experiments; DOI, duration of viral exposure; DOI, duration of viral exposure; ECM, Extracellular matrix; FDR, false discovery rate; GO, gene ontology; HRV, human rhinovirus; mock, uninfected cells from previous infected cells.; MOI, multiplicity of infection; NOI, number of repeated infections; noi0, uninfected cells; noi1, first infection; noi3, three repeated infections; noi5, five repeated infections; OFAT, One Factor at A Time (experimentation); PCA, principal component analysis; T.N.CpGs, trainable CpGs in non-coding regions; TG, trainable gene.

For affiliations refer to page 13.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

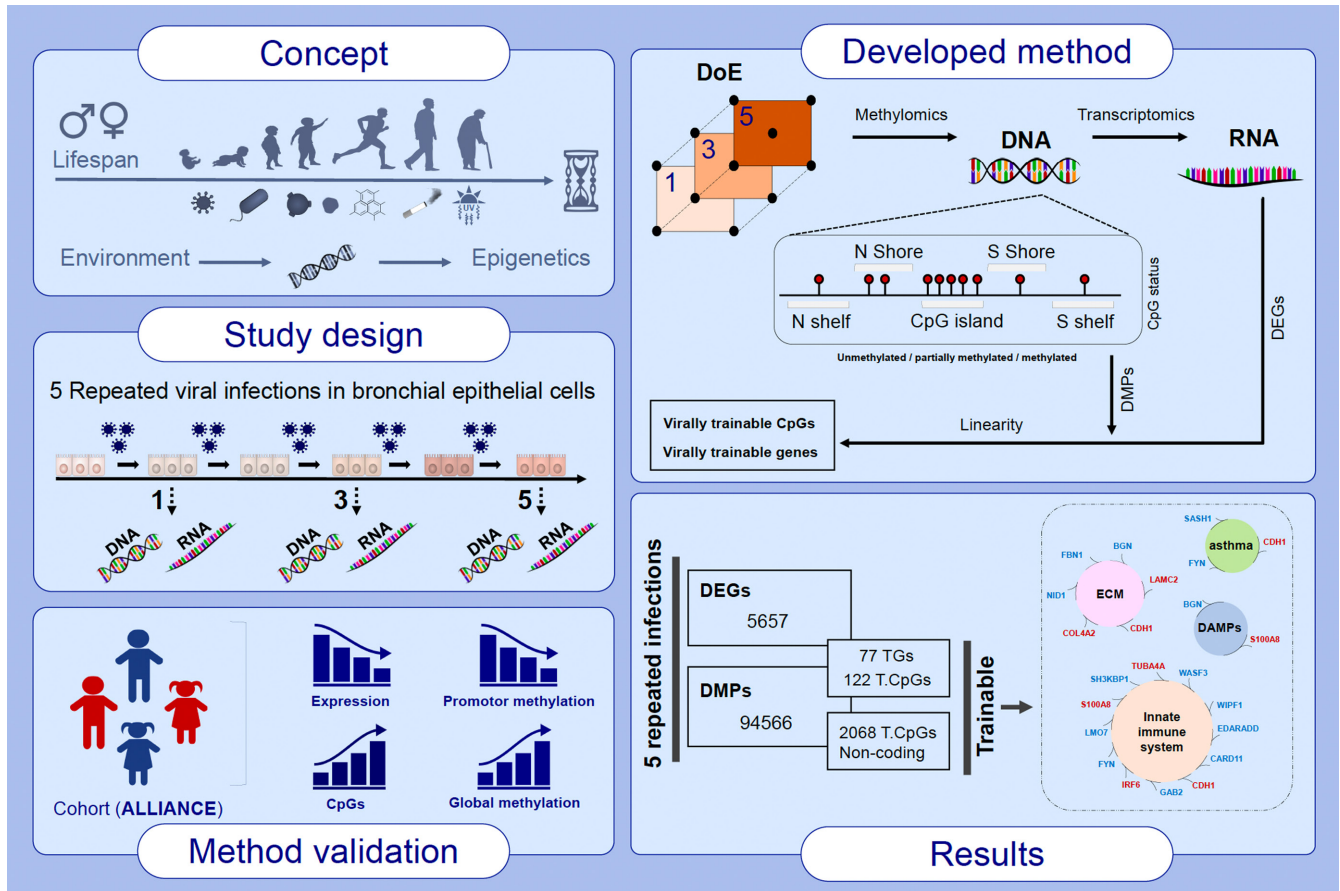
© 2024 The Author(s). *Allergy* published by European Academy of Allergy and Clinical Immunology and John Wiley & Sons Ltd.

sites were tracked within a pediatric cohort to determine the relative changes in candidate-trained sites with disease status and age.

**Conclusions:** Repeated viral infections induce an immune training response in bronchial epithelial cells. Training-sensitive DNAm sites indicate alternate divergent associations in asthma compared to healthy individuals. Our novel model presents a robust tool for identifying trainable genes, providing a foundation for future studies.

**KEYWORDS**

airway epithelium, asthma, human rhinovirus, trainable genes, trained immunity



**GRAPHICAL ABSTRACT**

This study aimed to identify trainable genes responsive to repeated HRV infections by examining changes in DNA methylation and gene expression in human bronchial epithelial cells (BEAS-2B). By using an innovative experimental design (DoE) and integrating transcriptome and methylation data, the developed method successfully served as a prediction model. Key findings included the identification of 77 trainable genes (TGs). The tracking of TGs methylation sites within an asthmatic pediatric cohort revealed differential associations with disease status and age. Additionally, distinct methylation sites associated with asthma and age highlight potential biomarkers for early diagnosis and future therapeutic targets. ALLIANCE, all-age asthma cohort; BEAS-2B, human bronchial epithelium cell line; DAMPS, damage-associated molecular pattern; DEGs, differentially expressed genes; DMPs, differentially methylated probes; DoE, design of experiments; ECM, extracellular matrix; HRV, human rhinovirus; TGs, trainable genes.

**1 | INTRODUCTION**

Trained immunity is a concept underlying the epigenetic and metabolic reprogramming of the innate immune system in response to various exposures. This concept, also known as inflammatory

memory, includes the long-term effect of inflammatory reprogramming on immune and structural cells, such as epithelial cells and fibroblasts.<sup>1</sup> In adaptive immune cells, immunological memory is determined by a set of genes transcribed in response to a specific pathogen.<sup>2</sup> On the other hand, trained innate immune cells modify

certain areas of DNA structures to enable varied gene transcription in response to the type of exposure and its duration.<sup>3–5</sup> The mechanisms underlying trained immunity have been investigated in-depth in the last decade.<sup>6–9</sup> The mechanisms are well characterized in immune cells such as macrophages, yet the role of trained immunity in the airway epithelium remains incompletely resolved. This is despite the crucial role of the epithelium as the first barrier against foreign particles and pathogens. Many studies show that innate immune cells can maintain long-term epigenetic memory of prior inflammation (reviewed in [10–12]), but it remains unclear if that long-term memory is selective for the exposure and involves specific gene sets. As such, we hypothesize that innate memory in epithelial cells is induced by genes that can be trained and that this training can be specific for particular triggers.

Physiological and molecular changes in response to viral infections have long-term functional effects, enabling a fine-tuned response to subsequent pathogenic exposures. Human rhinoviruses (HRV) are among the most frequent viruses causing the common cold.<sup>13</sup> HRV is a single-stranded RNA (ssRNA) consisting of a single genome encoding 11 proteins.<sup>14</sup> HRV exists as multiple serotype species (HRV-A, HRV-B, and HRV-C), according to the International Committee on Taxonomy of Viruses (ICTV).<sup>15</sup> HRV often causes upper and lower respiratory infections in healthy individuals, although HRV instead triggers more severe and long-lasting lower respiratory infections in asthmatic children and adults.<sup>16–20</sup> The complex relationship between HRV infection and asthma development has been investigated in previous studies,<sup>13,21,22</sup> highlighting the central role of HRV infection in asthma exacerbations. Asthma is a chronic inflammatory respiratory disease characterized by airflow obstruction, wheezing, and airway hyperresponsiveness, causing episodic shortness of breath (exacerbations).<sup>23</sup> Multiple viral infections (particularly HRV) during childhood are a strong determinant of aberrant lung development and increased risk of asthma development.<sup>24</sup> Children often share common environmental exposures (exposure to microbes, viruses, smoking, and household pets) that influence the expression profile of immune-related genes mediated by epigenetic changes such as DNA methylation (DNAm).<sup>25</sup> The biological processes impacted by these environmental factors remain a key research focus. In particular, viral infections have gained significant attention over the past decade. Our recent studies identified changes to the airway epithelium that persist after viral clearance, which could aggravate asthma symptoms.<sup>11,12</sup> However, it remains unclear how DNAm is altered in response to these exposures and which biological pathways are affected.

We hypothesize that repeated infections with HRV cause DNAm changes that alter subsequent gene expression. As such, using a unique design of experiment (DoE) to infect bronchial epithelial cells multiple times with HRV, we aimed to identify trainable genes regulated by changes in DNA methylation. We sought to follow up on such candidate-trained DNA methylation sites in a pediatric asthma cohort.

## 2 | METHODS

### 2.1 | Design of experiments (DoE)

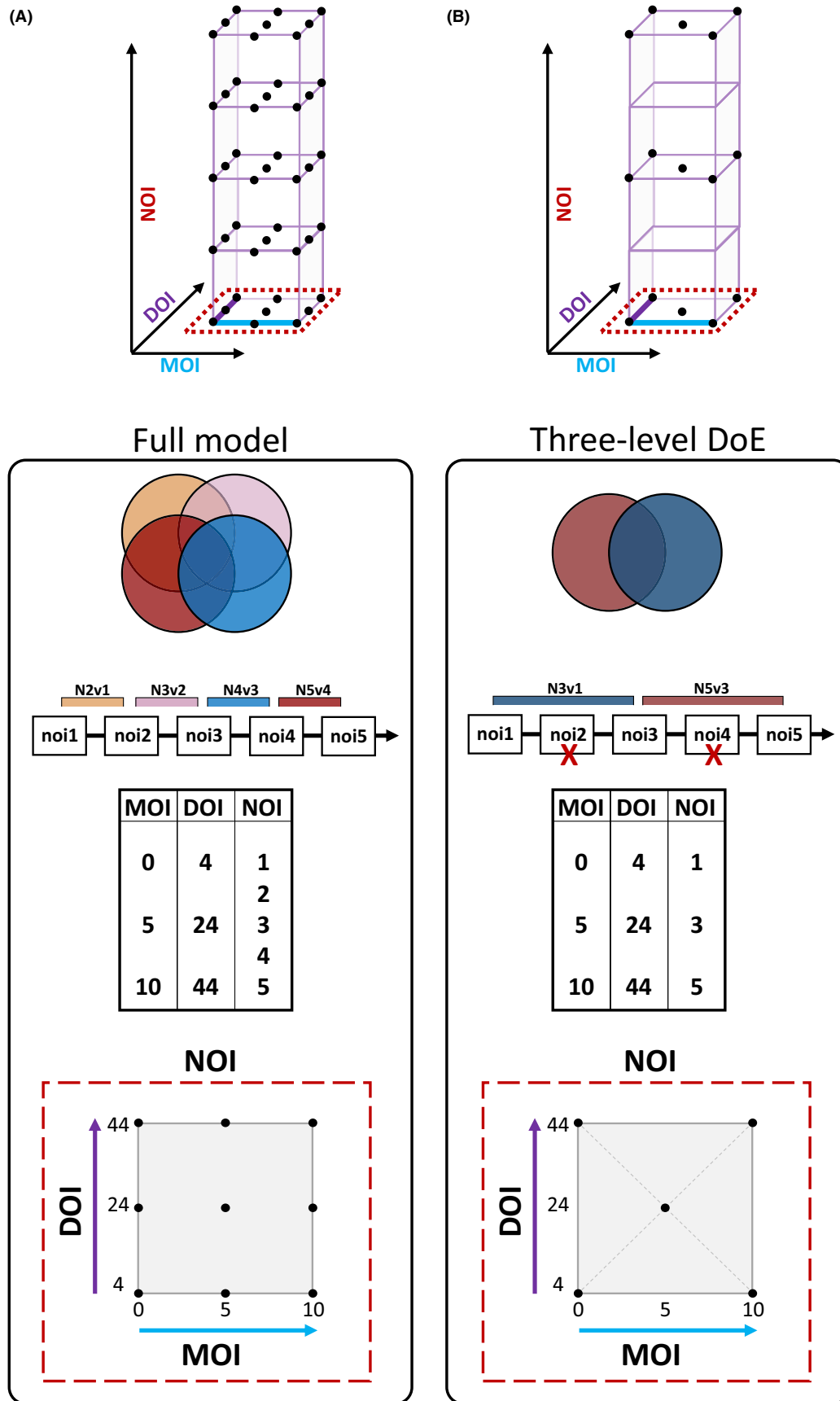
To analyse the effect of the duration of viral infection (DOI, hours of exposure to viral particles), the multiplicity of infection (MOI, number of viral particles per cell), and the number of repeated infections (NOI) demands a “One-Factor-at-A-Time” (OFAT) approach necessitating enormous resources. This would require a prohibitive number of samples, resources, and time, which would be hampered by batch effects due to the sheer size.<sup>26</sup> Especially when investigating methylation reprogramming after five repeated infections, the full model would have to include a huge number of samples (at least 135 samples). To overcome these limitations, we applied a statistical method to the experimental layout. Design of Experiments (DoE) is a mathematically supported approach to explore uncharted complex experimental spaces systematically. We devised a full-factorial, three-level DoE to leverage the continuous experimental space created by DOI, MOI, and NOI. These features were factorized: DOI: 4h, 24h, 44h; MOI: 0, 5, 10; and NOI: 1, 3, 5 (Figure S1). This statistical model was designed using the DoE full-factorial design tool of JMP (v12, SAS, USA) to deliver linear models between factorized variables and measurable biological outcomes (DNA methylation and gene expression). These regression models would further allow the imputation of values within the DoE not included by the factorized variables (Figure 1). We have applied this DoE previously.<sup>12</sup>

### 2.2 | Cell culture

As described previously,<sup>12</sup> human bronchial epithelial cells (BEAS-2B) were cultured in cell growth medium (BEGM) without gentamycin (GA)-1000 (Lonza, Switzerland) in pre-coated flasks (fibronectin, collagen type I, and bovine serum albumin). At 80% confluency, cells were infected with HRV-16 at MOI: 0, 5, and 10, and for DOI: 4h, 24h, and 44h. Infected cells were divided in two: (1) 80% of the cells were collected in RLT buffer (Qiagen, Hilden, Germany) for downstream DNA methylation (DNAm) and transcriptomic (RNAseq) analyses; (2) 20% of the cells were expanded to 80% confluency (approximately 72h) for re-infection (Figure S2). This method was repeated for up to five infections. A breakdown of treatment conditions is located in Table S1.

### 2.3 | Genome-wide methylation analysis and sequencing

As described previously,<sup>11</sup> DNA and RNA were isolated from each sample using the AllPrep DNA/RNA/Protein Mini Kit (Qiagen). RNA quality was evaluated via Agilent 2100 Bioanalyzer using Agilent RNA6000 Nano Chip (Agilent, USA). The concentrations of DNA



**FIGURE 1** Chart outlining the DoE design for the full (A) and three-level (B) factorial model. The three-level factorial design used in this study includes 35 different samples with three independent variables (MOI, DOI, and NOI) with three replicates. Each NOI (noi1, noi3, and noi5) includes five points that create two regression lines for analysis.

and RNA were measured by the FLUOstar Omega system (BMG Labtech, Ortenberg, Germany). DNAm analysis was performed via the HumanMethylation450 BeadChip Kit (Illumina, San Diego, USA), and RNA sequencing (RNAseq) was performed via the HiSeq2500 System (Illumina, San Diego, USA). DNAm and RNAseq were quantified by the Institute of Clinical Molecular Biology, Christian Albrechts University of Kiel, following the manufacturer's protocols.

## 2.4 | Pre-processing of RNA & DNA

For the RNAseq, raw data was quality-controlled using *fastqc* (v0.11.9). Data preprocessing was completed using *Trimmomatic-0.39*<sup>27</sup> to trim the repeated adapter-like sequences, low-quality reads (Q-score <20), and short reads (<35 bp). The remaining high-quality paired-end reads were aligned to the reference genome (*Homo sapiens*, GRCh38.107: Ensembl) using *bowtie2*.<sup>28</sup> The mapped reads were aligned to each gene with *SAMtools* to produce multiple binary alignment map (BAM) files.<sup>29</sup> Gene counts were extracted using *FeatureCounts*.<sup>30</sup>

For the Illumina 450K DNAm array, raw IDAT files were processed using the *minfi* pipeline (v1.43.0).<sup>31</sup> Processed data were normalized by quantile normalization for accurate DNA methylation estimation.<sup>32</sup> A quality control check based on mean detection *p*-value was carried out before and after normalization, and poor-quality samples (*p*-value >.05) were removed.

## 2.5 | Cohort data

Patient samples (*n*=120) analysed in this investigation were collected as part of the "All-age-asthma-cohort" (ALLIANCE).<sup>33</sup> The ALLIANCE study protocols were approved by the local medical ethics committee of the University of Lübeck (Vote 12-215; 18.12.2012), and all patients or their parent/guardian gave written informed consent. Nasal brushings were collected from patients with DNA isolated using the AllPrep DNA/RNA/Protein Mini Kit (Qiagen). DNA methylation (DNAm) was analysed using an Illumina EPIC BeadChip array by the Institute of Clinical Molecular Biology, Christian Albrechts University of Kiel. DNAm data quality and normalization were assessed using the R packages *minfi*, *MethylAid*, and *wateRmelon*. Normalization was completed using the *dasen* method in the *wateRmelon* package. Only pediatric study participants (age <18 years old) were included. Clinical characteristics were derived from patient medical history records, standardized questionnaires, structured interviews, and objective measurements. A summary of the basic demographics of the study population is included in Table S2.

## 2.6 | ALLIANCE cohort data sharing statement

The data used in this study are sensitive due to individual patient-level data, including but not limited to minors. It will not be made publicly available. Individual participant data that underlie the results (text,

tables, figures, including code/statistics) reported in this article can be released after de-identification and approved request. Investigators who seek access may contact the use-and-access committee (for each study). A breakdown of the cohort demographics is summarized in Table S2. All-Age-Asthma-Cohort: [alliance@imbs-luebeck.de](mailto:alliance@imbs-luebeck.de).

## 2.7 | Analysis & statistical methods

Principal component analysis (PCA) was conducted on both RNAseq/DNAm data to determine the effects of known variables and confounders. The *limma* package (v3.53.7) was used to analyse RNAseq and DNAm data.<sup>34</sup> The downstream analysis of the processed count data was carried out with *limma/R*, the mean-variance trend was converted into precision weights, then log-transformed by the *voom* function. A linear mixed model was used to analyse the entire experiment after the batch correction using the *sva* package.<sup>35</sup> For statistical inference, the empirical Bayes method estimated the likelihood distribution of the data. Using *ensemldb* (v2.22.0) differentially expressed genes (DEGs) were annotated to specific entries (e.g., protein coding, long non-coding RNA, non-coding genes).<sup>36</sup> Gene locations were obtained using the hg37 annotation using *biomaRT* (v2.54.1).<sup>37</sup> Functional analysis was performed with *ClusterProfiler* (v4.6.2).<sup>38</sup> Network graphic generated using Cytoscape (v3.10.1) with ClueGo and CluPedia.<sup>39-41</sup> *DMRcate* (v2.11.0) was used to determine, differentially methylated regions. This algorithm agglomerates CpG locations with an adjusted *p*-value below a false discovery rate (FDR) <.05. Specific parameters were used (Lambda=1000, C=2, minimum CpGs=5).<sup>42</sup> We also performed expression quantitative trait methylation (eQTM) analysis using *MatrixEQTL* (v2.3) to predict function associations between DNAm and gene expression.<sup>43</sup> MOI and DOI were entered as covariates in the model, and the correction for multiple testing was completed using the Benjamini-Hochberg method. Visualization of DNA co-methylation patterns was created with *coMET* (v1.30.0).<sup>44</sup> Data mining for the trainable genes and CpGs utilized different databases (e.g., Reactome, DAVID, PathCards, EWAS data Hub, UCSC Genome Browser Database).<sup>45-50</sup>

## 2.8 | Definition and methods to predict trainable genes

Genes demonstrating a persistent and mono-directional change in gene expression and DNAm were defined as "trainable." Those genes show a consistent and mono-directional change in expression and methylation with each subsequent infection. These changes must not be transient but rather maintain their altered state across subsequent infections and continue to change in the same direction. This method involves several stages, initially focusing on a set of genes that are differentially expressed in a linear fashion after each infection. The identification of functionally relevant differential methylation was inferred by changes in methylation status from

unmethylated (beta value  $\leq .2$ ) to partially methylated (beta value between  $.2$  and  $.6$ ) or to fully methylated (beta value  $\geq .6$ ). For this, differentially methylated sites between baseline (noi0) and after five consecutive infections (noi5) were compared, selecting the CpG sites exhibiting a shift between the above statuses. Next, to identify the CpGs changing with each infection, we extracted significant differentially methylated probes (DMPs). Only CpGs located near protein-coding genes were included (i.e., gene body, 5'UTR, 3'UTR, 200 and 1500 base pairs upstream of the transcriptional start site). Similarly, the expression results were analysed by selecting only genes that meet the regression criteria (FDR  $< 0.05$ ).

By merging RNAseq and DNAm results, we compiled a gene list that demonstrates a linear gene expression change associated with DMPs in response to multiple HRV infections. These are termed "trainable genes" (TGs) (Figure 2). A separate analysis was completed

for non-protein-coding gene regions that were included within our developed method (Table S21), as these CpGs also indicate an adherence to our defined training criteria. Methods for predicting the untrainable genes are described in Appendix S1 and the untrainable genes listed in Table S23. The grouping concept for each developed model (trainable/untrainable) is shown in Figure S3.

### 3 | RESULTS

#### 3.1 | RNA expression and DNA methylation profiles differ after each infection

This study was designed to identify trainable genes responding to repeated viral infections. RNAseq and DNAm data were

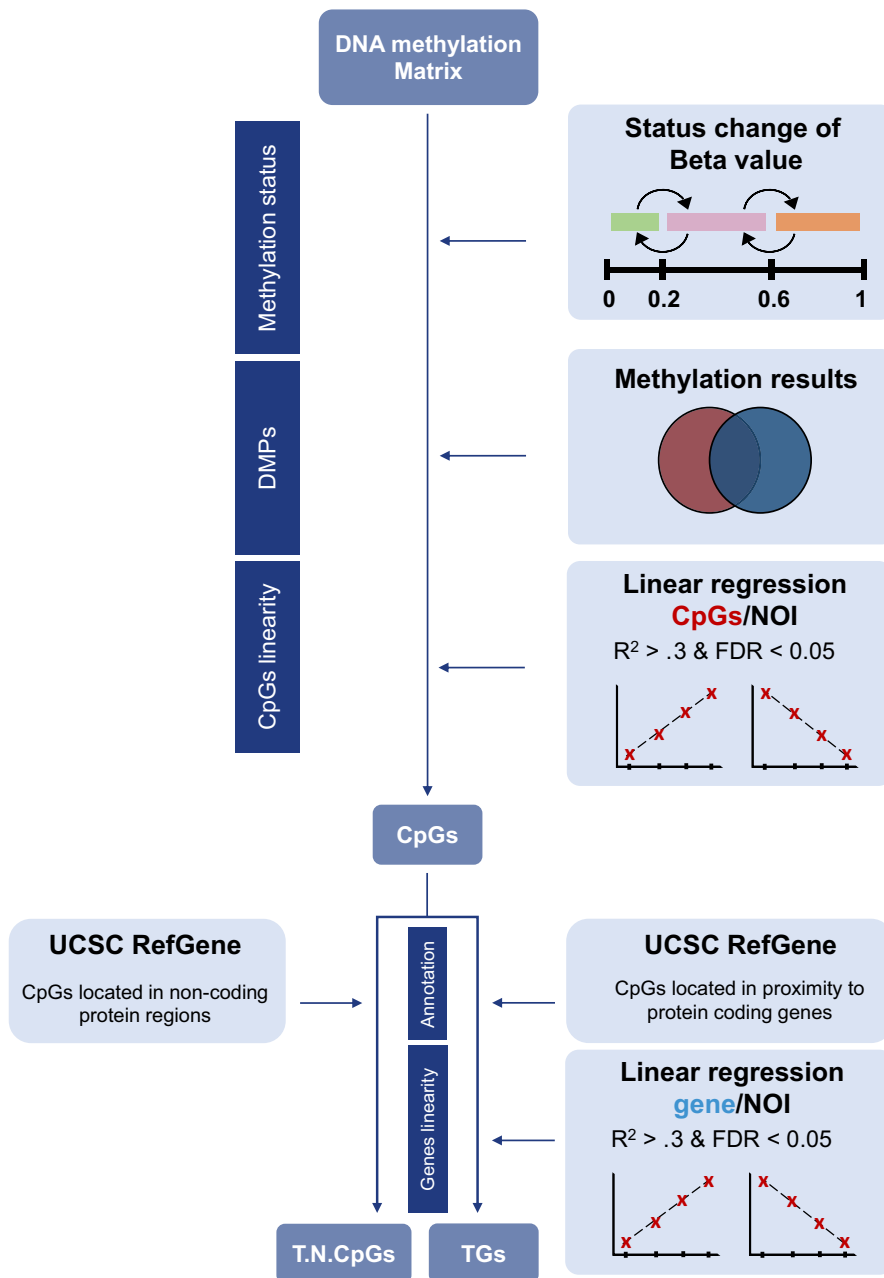


FIGURE 2 Flow chart of the method to obtain trainable genes. (Right) The method used to obtain the trainable genes (TGs). (Left) The method used to obtain the trainable CpGs related to non-coding regions (T.N.CpGs: Trainable CpGs in non-coding regions).

sampled following the first (N1), third (N3), and fifth (N5) infection (Figure 3A), and non-infected "mock" BEAS-2B cells served as a control. Out of 35 libraries, 80% of reads were mapped to the genome, resulting in a total of 61,806 transcripts. After pre-processing, we retained 17,740 transcripts for differential gene expression analysis. To minimize false positives, differentially expressed genes (DEGs) were determined at a false discovery rate (FDR) <0.05. This analysis was completed across three comparisons: N3vs1, N5vs3, and N5vs1 (Tables S4–S6). For N3vN1, 476 DEGs were identified (196 upregulated and 280 downregulated), and for N5vN3 173 genes (100 upregulated and 73 downregulated) (Figure 3B). The pathway analysis of these DEGs (N3vN1 and N5vN3) indicated shared biological processes such as the Wnt signaling pathway, cell-substrate adhesion, and epithelial-to-mesenchymal transition (Figure 3C) (Table S7).

For verification, we investigated an additional group (N5vN1) to compare the cumulative effect of all infections to the stepwise approach from above. We found 2834 upregulated and 2823 downregulated genes. Gene ontology (GO) analysis after five repeated infections revealed GO terms associated with cell cycle, DNA repair, cholesterol biosynthesis, and mitochondrial translation (Table S8). We also observed the regulation of the apoptotic signaling pathway, indicating that these DEGs are part of damage-associated molecular patterns (DAMPs). Subsequently, we identified 36 genes encoding DAMPs (5 downregulated and 31 upregulated) (Table S9).

Comparing the differences between repeated infections, we observed 18 DEGs (Figure 3A) that showed significant variations. Examining these patterns, we found genes that maintained their regression direction (e.g., MRC2, ATRNL1, MAPKAPK3, WIPF1, NID1, MAPKAPK3, TTL7, ABCA8, TRPC1, CNTNAP3B, ZNF83, EVC2, and GPR137C), while others were inconsistent and changed their regression direction, such as GCNT2, PGF, SPECC1, VGLL4, TCF7L2, and VPS33B (Table S10).

A similar approach was conducted for DNAm analysis. Differentially methylated probes (DMPs) were determined at an FDR <0.05. Overall, 54,035 DMPs were identified for N3v1 (28,595 up and 25,440 down) and 5286 DMPs (2351 up and 2935 down) in N5v3. For N5v1, we found 94,566 DMPs (46,382 up and 48,184 down) (Tables S12–S14). To define DNAm associated with five repeated infections, differentially methylated regions (DMR) between noi5 and noi1 were identified (Table S15). The DMPs identified from these results were then used to filter the trainable genes. Additional results for the effect of the first infection (Noi1 vs. Noi0) on both expression and methylation are listed in Tables S3 and S11.

### 3.2 | Trainable genes exhibit diversity

We have identified genes that exhibit linear expression changes with each subsequent infection paired with a linear change in proximal DMPs. As such, we have identified 77 potential virally

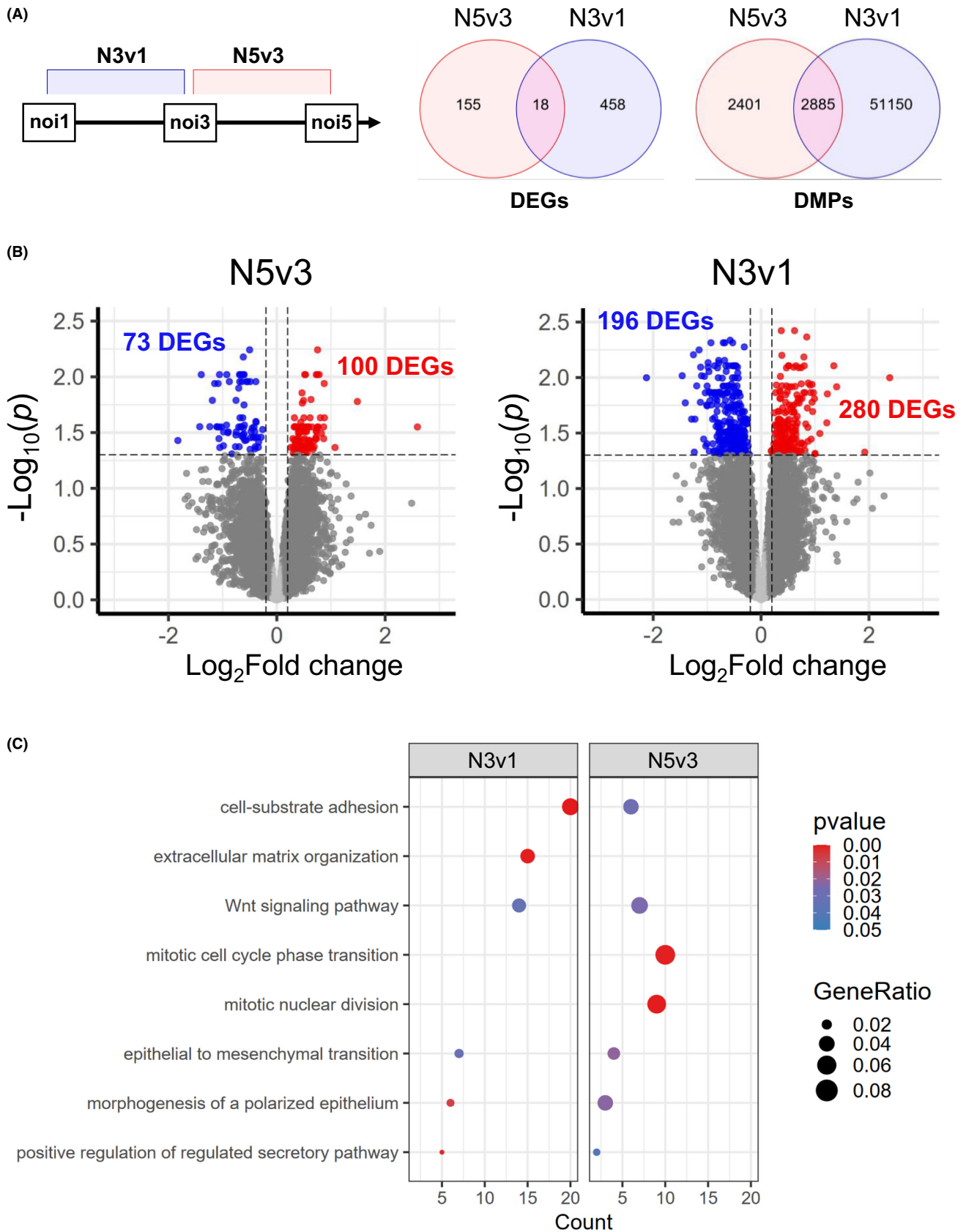
trained genes (TGs) (Table S16). eQTM analysis was conducted on this subset of genes to statistically determine whether a cis-regulatory relationship exists between CpG sites and gene expression. We identified a total of 787 cis-eQTMs (Table S17). Figure 4A displays heatmaps representing the expression profile of the four different expression response patterns across multiple infections. The expression of the TGs is displayed in the Figures S4–S7. Pathway enrichment analysis of the TGs is presented in Figure 4B as a network. We observe that key functional processes of cell-cell communication, cellular signaling, and immune responses are enriched within our identified TGs. The functional analysis results (reactome pathways) for TGs are listed in Table S18. These results show that the trainable genes are diverse and are involved in many important processes.

In Table S19, the trainable genes were distinguished based on their function and expression level. The top 20 trainable genes with a high difference in their expression level across each infection were ABCA8, ARHGAP20, ARL4C, BGN, BMPR1B, CA2, EVC2, IRF6, KIF5C, KLF12, LAMC2, MECOM, MED12L, MEGF6, PABPC1L, SDK1, WFDC2, WIPF1, ZNF347, and ZSCAN18 with an FDR <1 × 10<sup>-5</sup>. Other trainable genes showed enrichment for extracellular matrix organization (BGN, CDH1, COL4A2, FBN1, LAMC2, and NID1), cytokine signaling (EDARADD, FYN, GAB2, and IRF6), NF-κB pathway (BMPR1B and MECOM), asthma (SASH1, FYN, and CDH1), DAMPs (BGN and S100A8), transcription regulation (KLF12, MECOM, SIM1, IRF6, MED12L, ZEB2, ZSCAN18, ZNF117, ZNF347, ZNF518B, and ZNF718), and the innate immune system (CARD11, CDH1, EDARADD, FYN, GAB2, IRF6, LMO7, S100A8, SH3KBP1, TUBA4A, WASF3, and WIPF1).

The primary focus of this model is to identify trainable genes associated with DNA methylation sites (CpGs) (Table S20). These CpGs are located in protein-coding genes. However, not all the CpGs are located in such protein-coding regions. Consequently, our model also found 2068 trainable CpGs in non-protein coding regions (T.N.CpGs) (Table S21). Using data mining and methylation studies from the EWAS Data Hub, we report 51 asthma-associated CpGs located in non-coding regions (Table S22). Further filtering based on the 54 CpGs that are enhancer elements (FDR <1 × 10<sup>-3</sup>) highlights the regulation by non-coding RNAs. We identified 10 hypomethylated CpGs (cg16418183, cg04098829, cg15812586, cg21915998, cg25054816, cg03002688, cg22542373, cg10764891, cg12940886) and 14 hypermethylated CpGs (cg07044938, cg05318186, cg22903457, cg12352210, cg23375169, cg09371059, cg25381253, cg10954469, cg07352127, cg22542222, cg08599249, cg18780021, cg17130544, cg05484949).

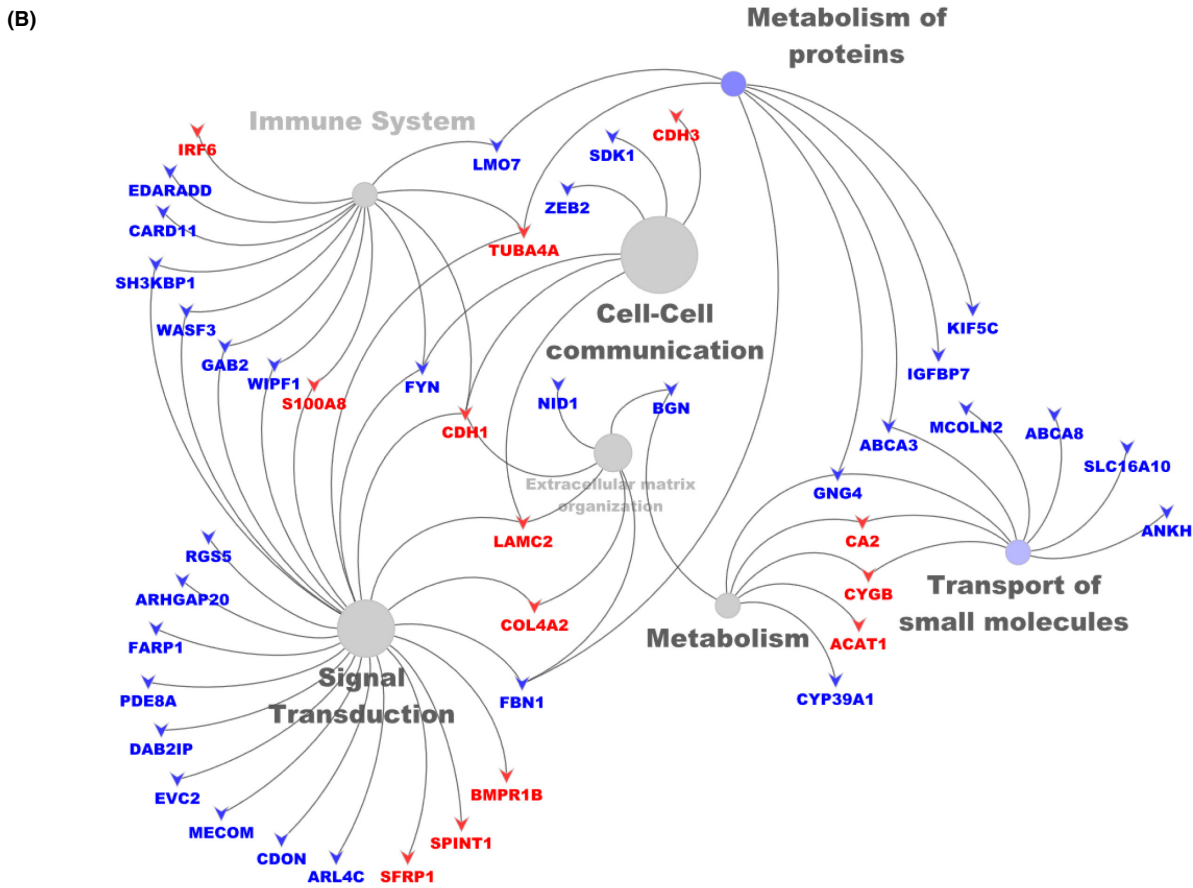
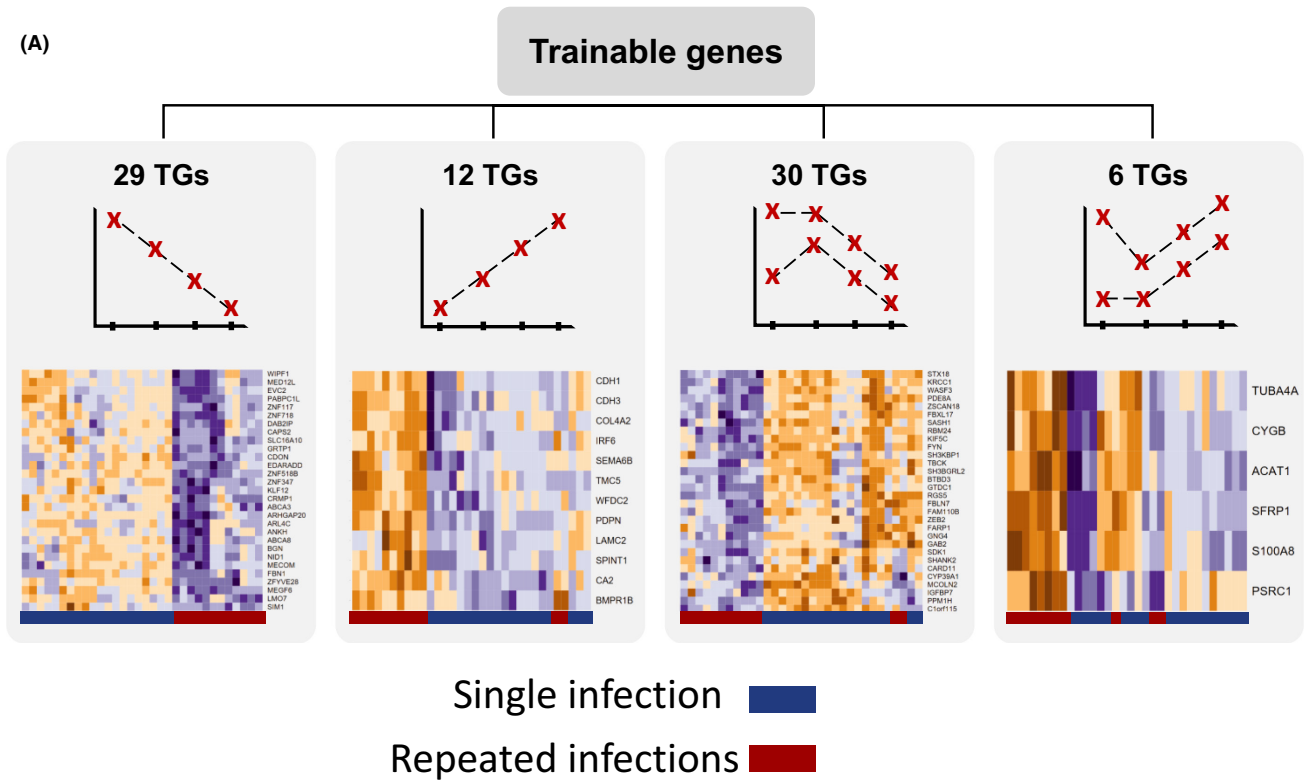
### 3.3 | Comparative analysis of methylation patterns of the trainable gene S100A8

Throughout development, individuals are exposed to multiple and varied pathological stimuli, which can lead to worse outcomes in



**FIGURE 3** Results of expression profiles of human bronchial epithelium after repeated infections. (A) Schematic diagram of the comparison groups and Venn diagram showing the number of significant overlapping DEGs and DMPs between N3v1 and N5v3. (B) Volcano plot showing expression differences between repeated infection groups ( $n=35$ ). Red indicates DEGs that differ at  $FDR < 0.05$  and have a positive foldchange; blue indicates DEGs that differ at  $FDR \leq 0.05$  and have a negative foldchange. (C) top gene ontology terms.





**FIGURE 4** Expression response patterns for trainable genes. (A) The TGs are in four divided groups. These groups are based on expression linearity. (B) Shared pathways between TGs, genes shown in blue, indicate a downregulation, but genes shown in red indicate upregulation among repeated infections. Pathways are shown as grey circles.

diseases such as asthma. These exposures can similarly induce a training response that may be altered in the context of disease. As such, the clinical relevance of our identified virus-trainable genes was analysed across an individual's chronological age. We analysed the global methylation level and promoter region methylation. Our candidate in vitro TGs were compared to patient-derived data from the ALLIANCE pediatric cohort (All-age-asthma-cohort). *KLF12* is an important transcription factor that demonstrates the high difference in its expression level in response to multiple infections (Appendix S1: Figure SC2). *S100A8* has a robust involvement in the innate immune system and DAMPs. Figure 5 presents *S100A8* with its trainable CpG site (cg20335425). The CpG methylation level decreased with repeated infections, although interestingly the same CpG showed increased methylation in vivo (ALLIANCE). Further, both the global and promoter methylations of *S100A8* followed a different pattern in both datasets. Importantly, while healthy controls experienced an age-dependent increase of methylations in the *S100A8* promoter region and globally, asthmatic patients maintained methylation levels.

#### 4 | DISCUSSION

To our knowledge, this study is the first of its kind to identify virally trained genes (TGs) and associated CpGs, in respiratory epithelium. The unique design of this study facilitated a robust approach that helped identify TGs using a sample-reducing Design of Experiments (DoE). The study comprised 35 samples consisting of three different measurements for each factor (infection duration, multiplicity of infection, and number of infections). This model combined gene expression and DNA methylation profiles of each sample, enabling the identification of the training-associated changes induced by each round of infection.

Our definition of TGs includes genes that show persistent, mono-directional changes in their expression and DNA methylation. We found 77 genes and 122 CpGs trainable by multiple HRV infections. eQTM analysis estimated the functional impact of these CpGs on gene expression, where 105 CpGs indicated an association with the expression of 64 genes. Of these, 11 genes were involved in transcription regulation (*KLF12*, *MECOM*, *SIM1*, *IRF6*, *MED12L*, *ZEB2*, *ZSCAN18*, *ZNF117*, *ZNF347*, *ZNF518B*, *ZNF718*), while the remaining TGs indicated enrichment for signal transduction, cell-cell communication, developmental biology, epithelial-to-mesenchymal transition, and ECM organization pathways. These results suggest that virally-trained genes play a central role in cell-environment interactions. As such, a dysregulation of these genes could lead to functional and structural dysfunction in response to further pathological exposures.

Epithelial barrier dysfunction is a central feature of asthma (and allergy), which is consistently reported to contribute to disease progression and chronicity.<sup>51</sup> Remarkably, viral infections may cause aberrant mucosal barrier function, implicated by a higher susceptibility to further viral infections and symptom persistence.<sup>52</sup> Highlighting

the connection between our identified training responses and disease, some of the identified trainable CpG sites have been reported in asthma EWAS studies. These CpGs are not proximal to specific genes but are instead located within non-coding regions.<sup>53,54</sup> This likely indicates a more complex epigenetic regulation of training response mechanisms arising from viral infections, which lies beyond the current investigation's scope.

Our data revealed that the regulation of 36 DAMP-associated genes significantly changed after five HRV infections (31 upregulated, 5 downregulated). Our method identified only two of these DAMPs as trainable (*BGN* and *S100A8*). Interestingly, *S100A8* was recently associated with *S100A9* in multiple COVID-19 studies, with their protein complex mediating host proinflammatory responses during infection.<sup>55-57</sup> Furthermore, the *S100A8/S100A9* complex was characterized as a potential biomarker for neutrophilic asthma<sup>58,59</sup> and hypothesized to play a part in airway remodeling in asthma.<sup>60</sup> Despite this, further and more detailed investigations are needed to fully understand the *S100A8* function. Hypomethylation of cg20335425 in the *S100A8* intronic region shows an opposing methylation regression pattern with age compared to repeated infections.

The exact influence of intronic methylation sites on gene expression is not fully understood, as introns contain complex regulatory elements (enhancers and silencers) potentially influencing chromatin accessibility for transcription factors.<sup>61-63</sup> The significance of intronic regulation has been highlighted in cancer research.<sup>64</sup> For instance, we previously reported an intronic epigenetic regulation enhancer function of *ZNF263* being diminished through virus-induced increased methylation.<sup>12</sup> Whilst the methylation level for cg20335425 was significantly changed in our in vitro experiments, an alternate response to environmental conditions may exist in the human respiratory mucosa. Our results suggest that the *S100A8* gene is trainable upon viral exposure with promoter and global gene methylation levels showing age-related increase in healthy children. However, this is not apparent in asthmatic patients. This indicates that asthma might cause an altered regulation of *S100A8*. The implications of this observation require elucidation within the context of exacerbation risk and response to pathogenic stimuli. Our results suggest that *S100A8* might be exposome-sensitive and potentially influenced by the accumulation of infections over a person's lifetime.

A recent study that focused on single 48-h exposure to HRV 16 in the mucosal epithelium of patients with chronic rhinosinusitis identified genes that represent overlapping results with our own findings.<sup>65</sup> We identified 10 genes in the referenced study that are identical with genes we found to be trainable in our investigation (overlap). These 10 genes (*FYN*, *GAB2*, *IGFBP7*, *S100A8*, *DAB2IP*, *BMP1B*, *CDON*, *FBN1*, *FBXL17*, *SFRP1*) are found in specific modules (black, red, yellow) that may be associated with cellular processes, structural integrity, immune response, inflammation, signaling pathways, and regulatory mechanisms. This overlap suggests that these genes play a crucial role in the epithelial response to rhinovirus

infection across different types of epithelial cells and conditions. These findings underscore the importance of these genes in the context of chronic infections. The repeated infections in our study and the chronic condition in the other study both highlight how persistent viral exposure can lead to sustained epigenetic and transcriptional changes, potentially contributing to disease progression and severity.

Our study has greatly enriched the current understanding of trained immunity. As we observed, repeated HRV infections certainly increase or hinder gene expression. However, this influence differs with each infection (i.e., the set of genes changes with each infection), complicating the ability to highlight the precise impact of repeated HRV infections. However, our methodology enables the identification of CpGs across each subsequent infection to filter for the associated genes that may not be significantly affected after fewer infections but become measurable with later infections. This ability to pair DNA methylation with gene expression is a key advantage of the current study, revealing a connection between viral infection, DNA methylation, and subsequent gene expression changes.

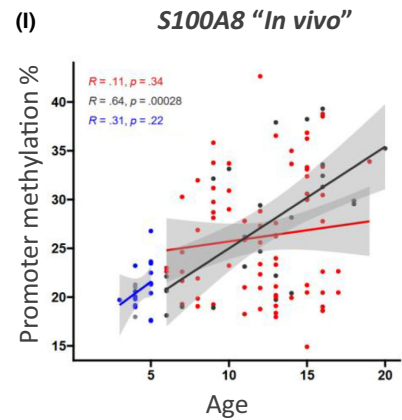
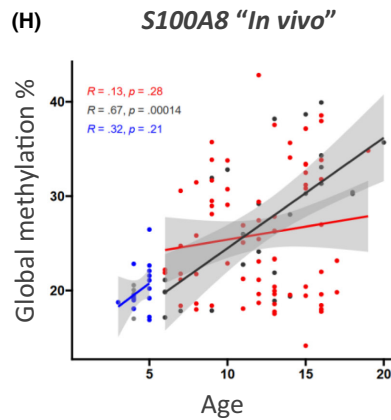
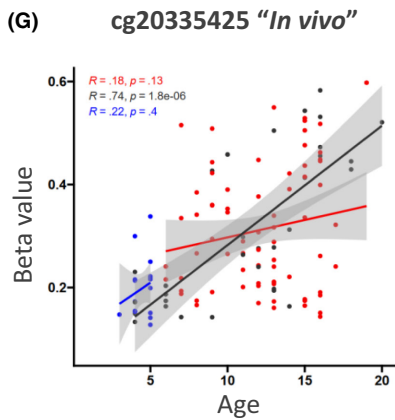
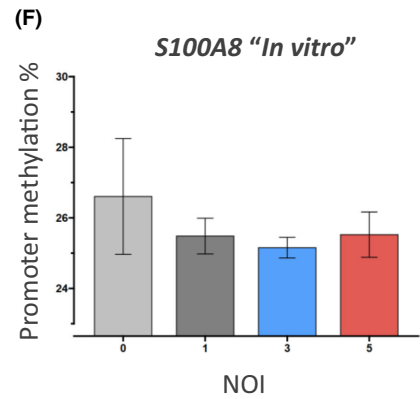
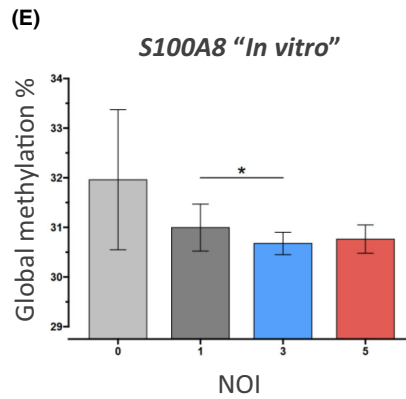
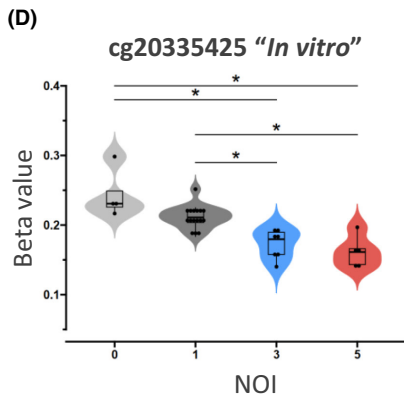
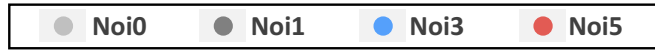
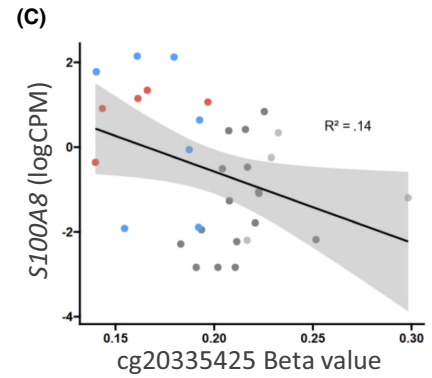
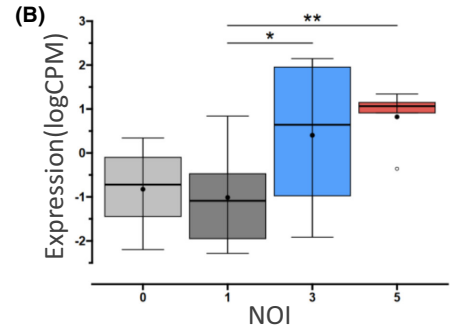
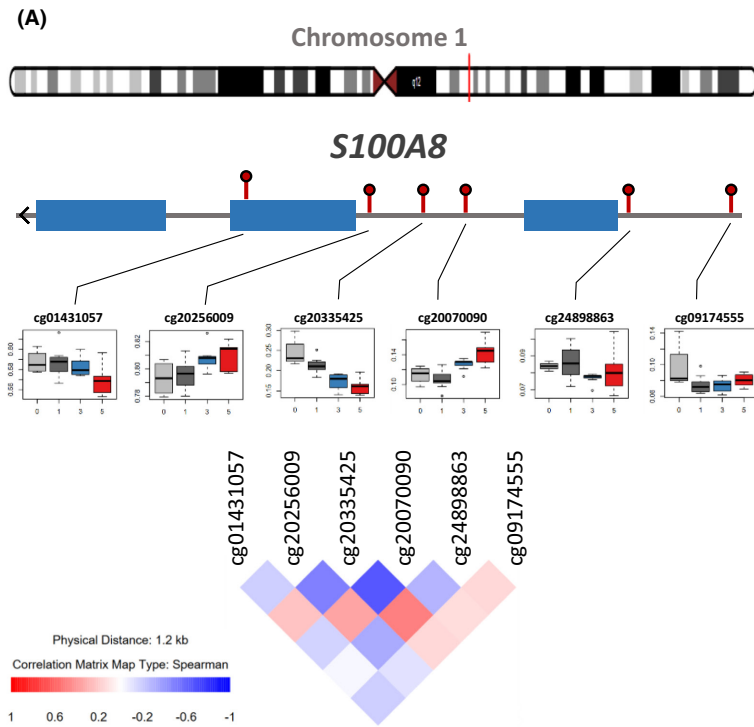
Our DoE has advantages, but it comes with some limitations. As an advantage, implementing the 3 factorial DoE effectively reduced the number of samples required for analysis, decreasing both cost and time. However, using the 3-factorial DoE may result in the loss of information from the infected cells (noi2 and noi4) that were not part of the 3-factorial design. And in these cells, biologically relevant changes may have occurred. Not including these cells in the design may have increased the number of identified trainable genes and CpGs. This was shown in the cohort data, where we found some of the trained CpGs (*SASH1*; cg06282952, *KLF12*; cg09677330, *WIPF1*; cg10037068, *KIF5C*; cg11664251, *RBM24*; cg14466942, *SDK1*; cg19297245/cg19369262, *S100A8*; cg20335425) to demonstrate a linear relationship with chronological age. Determining why not all identified TGs significantly correlate with age is a complex matter that could be attributed to a multitude of factors such as complex regulatory mechanisms influenced by environmental and lifestyle challenges faced by the individuals participating in the cohort study (environmental smoke exposure, diet, physical activity, work environment). Another reason could be related to the type of cells investigated in both the cell culture (bronchial epithelial cell line (BEAS-2B)) and the cohort study (nasal epithelial cells). Furthermore, our method was designed to assess trainable genes associated with DNAm; other trainable genes modulated by different regulation mechanisms (i.e., LncRNA, miRNA, histone modification) were beyond the scope of the current analysis. Despite these limitations, we are still able to demonstrate an ability to project the results from our DoE viral infection model onto a patient cohort. Through this analysis, we identify virally-trained genes that present divergent regulatory patterns (via DNA methylation) between healthy and asthma-diagnosed individuals. This underscores the potential utility of our cost-effective and time-efficient DoE in contributing to future

projects exploring complex interactions between environmental exposures and molecular mechanisms. By choosing this approach of using repeated viral infections in cell culture and comparing it with the methylation levels across different age groups (healthy individuals and asthmatics), we were able to identify specific sites that may be associated with asthma or cellular aging. This approach effectively helped overcome the ethical challenges of experimentally inducing recurrent infections in humans.

Despite all the limitations we faced, our developed method remains a strong foundation for future development. By enhancing the related factors (lab workflows, design, cell types, techniques, and analysis), we can significantly improve its accuracy and expand its applications. The model presented herein demonstrates a robust method to investigate virally trained genes. However, this model is limited by the necessity to use a cell line rather than primary human epithelial cells. The use of primary cells is restricted by their relatively limited cell culture life.<sup>66</sup> Future studies that are able to integrate and model the effects of multiple viral infections in primary cells would provide deeper and robust insight into the biologically relevant pathways that are trainable. This work would effectively build upon and provide confirmation of the results from our current work. Another way to validate the results effectively is by having both cell culture and cohort data derived from the same cell types. In our case, we only had access to human nasal epithelial cells for comparison with our bronchial epithelial cell culture experiments. Another way to improve the work accuracy is by incorporating additional techniques that will yield better insights into regulatory processes related to time effects. This includes using single-cell analysis, histone modifications, and ATAC-seq. And again, this all can be changed based on the type of statistical method used. In our case, we aimed to investigate the training aspect of HRV where we assume linearity, so picking the linear regression method using the Limma package was optimal in our case. Another valid method is the generalized linear model (GLM) via edgeR/R<sup>67,68</sup> which can give a better understanding of the training threshold due to its ability to deal with non-normal distributions of expression or methylation data.

## 5 | CONCLUSION

The presented investigation introduces a novel study design that enables the identification of trainable genes and their associated CpGs in the context of repeated viral infections. Our unique approach revealed that virally-trained genes such as DAMPs (BGN and S100A8) may serve as genes that respond to the accumulation of infections throughout a person's lifetime. Notably, we observe that some TGs indicate distinct DNAm patterns with age between healthy and asthmatic individuals. Despite the limitations of our model, it presents a significant step forward in the field of trained immunity, providing a strong platform for the application of future studies to investigate this complex biological mechanism. Such findings will have valuable clinical applications to determine the



**FIGURE 5** *S100A8* and cg20335425. (A) Regional plot and co-methylation patterns at the *S100A8* gene in BEAS-2B. (B) Gene expression of *S100A8* and (C) its correlation with the cg20335425. (D) In vitro results of cg20335425. (E) In vitro results of global methylation level of all *S100A8* CpGs. (F) In vitro results of promoter methylation of *S100A8*-related CpGs. (G) In vivo results of cg20335425. (H) In vivo results of global methylation level of all *S100A8* CpGs. (I) In vivo results of promoter methylation of *S100A8*-related CpGs. Asterisks (\* or \*\*) denote significant differences between groups, respectively, at  $p < 5 \times 10^{-2}$  and  $1 \times 10^{-3}$ .

contribution of early life training responses and the relationship between environmental exposures, epigenetic regulation, asthma, and aging.

## AUTHOR CONTRIBUTIONS

MAR, UJ, and MW conceived the original idea. MAR, KDR, and SSPN contributed to and performed the data analysis. MAR, KDR, UJ, and MW interpreted the data, discussed, and drafted the original manuscript composition. All authors have discussed, contributed, and approved the final version of the manuscript. MAR: data analysis (DoE results for both RNA expression and DNA methylation data), established new methods (obtaining trainable genes, untrainable genes), optimized method through data mining, wrote original draft, created all figures and all tables, edited manuscript. UJ: Obtained the funding; discussed the data; supervised the analysis; contributed to drafting the original manuscript and critically revised it; approved the final version. KDR: data analysis (ALLIANCE cohort), drafting of the manuscript. MW: Conception of the idea, data analysis, drafting of the manuscript. SM: untrainable genes original idea, edited manuscript.

## AFFILIATIONS

<sup>1</sup>Division of Clinical and Molecular Allergology, PA Chronic Lung Diseases, Research Center Borstel, Borstel, Germany

<sup>2</sup>German Center for Lung Research (DZL), Airway Research Center North (ARCN), Borstel, Germany

<sup>3</sup>Department of Paediatric Pneumology & Allergology, University Clinic Schleswig-Holstein (UKSH), Lübeck, Germany

<sup>4</sup>Division of Epigenetics of Chronic Lung Diseases, Priority Area Chronic Lung Diseases, Research Center Borstel – Leibniz Lung Center, Borstel, Germany

<sup>5</sup>Centre of Allergy and Environment (ZAUM), Technische Universität and Helmholtz Centre Munich, Munich, Germany

<sup>6</sup>Comprehensive Pneumology Centre Munich (CPC-M), Member of The German Centre for Lung Research (DZL), Munich, Germany

<sup>7</sup>Department of Pneumology, LungenClinic Grosshansdorf, Grosshansdorf, Germany

<sup>8</sup>University Hospital Schleswig-Holstein Campus Kiel, Department for Internal Medicine I, Kiel, Germany

<sup>9</sup>Department of Paediatric Pneumology, Allergology and Neonatology, Hannover Medical School, Hannover, Germany

<sup>10</sup>Biomedical Research in End stage and Obstructive Lung Disease Hannover (BREATH), Member of the German Centre of Lung Research (DZL), Munich, Germany

<sup>11</sup>Institute for Asthma and Allergy Prevention, Helmholtz Centre Munich, German Research Centre for Environmental Health, Neuherberg, Germany

<sup>12</sup>Department of Pulmonary and Allergy, Dr von Hauner Children's Hospital, University Children's Hospital, Ludwig Maximilian's University, Munich, Germany

<sup>13</sup>German Center for Child and Adolescent Health (DZKJ), Dr von Hauner Children's Hospital, LMU Munich, Munich, Germany

<sup>14</sup>Department of Paediatric Respiratory Medicine, Inselspital, University

Children's Hospital of Bern, University of Bern, Bern, Switzerland

<sup>15</sup>Immunology and Cell Biology, PA Chronic Lung Diseases, Research Center Borstel, Borstel, Germany

<sup>16</sup>Interdisciplinary Allergy Outpatient Clinic, Department of Pneumology, University of Lübeck, Lübeck, Germany

## ACKNOWLEDGEMENTS

The funding from the Federal Ministry of Education and Science (Deutsches Zentrum für Lungenforschung, DZL) enabled this project, for which we are very grateful. The authors would also like to acknowledge the patients and clinical staff who contributed to the generation of cohort samples analysed in this work.

## FUNDING INFORMATION

Federal Ministry of Education and Science: German Center for Lung Research/Deutsches Zentrum für Lungenforschung (DZL) DZL001C1.

## CONFLICT OF INTEREST STATEMENT

MAR, KDR, SSPN, CJ, CSW, TB, GH, EVM, AMD, RG, NM, FB, SM, and UJ declare no conflict of interest regarding the content of this manuscript. MW reports grants from the COVID-19 Research Initiative Schleswig-Holstein, the Follow-Up of Respiratory Infections in Schleswig-Holstein (FRISH), the German Center of Lung Research (DZL, Funding No. 82DZL001B6), intramural funding of the Christian-Albrechts-University Kiel, the University of Lübeck, and the Leibniz Lung Center, Research Center Borstel. Funding institutions did not participate in the design and conduct of this study. MVK reports grant from the BMBF for the Deutsches Zentrum für Lungenforschung (DZL) and received consultant fees from Sanofi Aventis GmbH, Chiesi GmbH, Allergopharma GmbH along with payments or honoraria for lectures, presentations, speakers' bureaus, manuscript writing, or educational events from Sanofi Aventis GmbH, Infectopharm GmbH, Allergopharma GmbH. KFR received personal payments or honoraria from AstraZeneca, Boehringer Ingelheim, Chiesi Pharmaceuticals, CSL Behring, Sanofi & Regeneron, GlaxoSmithKline, Berlin Chemie, and Menarini; K.F. Rabe also discloses participation on data safety monitoring boards/advisory boards for AstraZeneca, Boehringer Ingelheim, and Sanofi Regeneron, and leadership or fiduciary role in the German Center for Lung Research (DZL), German Chest Society (DGP), and American Thoracic Society (ATS). BS reports grants from the BMBF (German center for lung research, CPC-Munich, DZL 82DZL033C2, Combat Lung diseases FP4), the German Center for Child and Adolescent Health (DZKJ; LMU/LMU Klinikum: 01GL2406A), from DFG (DFG-SCHA 997/8–1 (BS); DFG-SCHA 997/9–1, DFG-SCHA-997/10–1, DFG-SCHA-997/11–1). BS reports consulting fees from GlaxoSmithKline, Novartis, and Sanofi;

payment/honoraria and participation on a Data Safety Monitoring Board or Advisory Board from Sanofi.

## DATA AVAILABILITY STATEMENT

Raw data for BEAS-2B cells are available upon request.

## ORCID

Marua Abu Risha  <https://orcid.org/0009-0005-6614-9136>  
 Karosham D. Reddy  <https://orcid.org/0000-0001-9002-6930>  
 Sai Sneha Priya Nemani  <https://orcid.org/0009-0004-3625-3373>  
 Constanze Jakwerth  <https://orcid.org/0000-0002-2902-0393>  
 Carsten Schmidt-Weber  <https://orcid.org/0000-0002-3203-8084>  
 Thomas Bahmer  <https://orcid.org/0000-0002-7677-7057>  
 Gesine Hansen  <https://orcid.org/0000-0001-5185-9454>  
 Erika von Mutius  <https://orcid.org/0000-0002-8893-4515>  
 Klaus F. Rabe  <https://orcid.org/0000-0002-7020-1401>  
 Anna-Maria Dittrich  <https://orcid.org/0000-0002-9582-1025>  
 Ruth Grychtol  <https://orcid.org/0000-0002-8084-104X>  
 Nicole Maison  <https://orcid.org/0000-0002-8731-4218>  
 Bianca Schaub  <https://orcid.org/0000-0003-1652-8873>  
 Matthias V. Kopp  <https://orcid.org/0000-0003-1989-5492>  
 Folke Brinkmann  <https://orcid.org/0000-0002-4307-4232>  
 Silke Meiners  <https://orcid.org/0000-0003-3678-7995>  
 Uta Jappe  <https://orcid.org/0000-0001-5602-3672>  
 Markus Weckmann  <https://orcid.org/0000-0001-5342-979X>

## REFERENCES

- Ordoas-Montanes J, Beyaz S, Rakoff-Nahoum S, Shalek AK. Distribution and storage of inflammatory memory in barrier tissues. *Nat Rev Immunol*. 2020;20(5):308-320. doi:10.1038/s41577-019-0263-z
- Dunn J, McCuaig R, Tu WJ, Hardy K, Rao S. Multi-layered epigenetic mechanisms contribute to transcriptional memory in T lymphocytes. *BMC Immunol*. 2015;16(1):27. doi:10.1186/s12865-015-0089-9
- Hajishengallis G, Netea MG, Chavakis T. Innate immune memory, trained immunity and nomenclature clarification. *Nat Immunol*. 2023;24(9):1393-1394. doi:10.1038/s41590-023-01595-x
- Acevedo OA, Berrios RV, Rodríguez-Guilarte L, Lillo-Dapremont B, Kalergis AM. Molecular and cellular mechanisms modulating trained immunity by various cell types in response to pathogen encounter. *Front Immunol*. 2021;12:745332. doi:10.3389/fimmu.2021.745332
- Fanucchi S, Mhlanga MM. Lnc-ing trained immunity to chromatin architecture. *Front Cell Dev Biol*. 2019;7:2. doi:10.3389/fcell.2019.00002
- Domínguez-Andrés J, Dos Santos JC, Bekkering S, et al. Trained immunity: adaptation within innate immune mechanisms. *Physiol Rev*. 2023;103(1):313-346. doi:10.1152/physrev.00031.2021
- Al B, Suen TK, Placek K, Netea MG. Innate (learned) memory. *J Allergy Clin Immunol*. 2023;152(3):551-566. doi:10.1016/j.jaci.2023.06.014
- Netea MG. Immunological memory in innate immunity. *J Innate Immun*. 2014;6(2):117-118. doi:10.1159/000357283
- Netea MG, Domínguez-Andrés J, Barreiro LB, et al. Defining trained immunity and its role in health and disease. *Nat Rev Immunol*. 2020;20(6):375-388. doi:10.1038/s41577-020-0285-6
- Larsen SB, Cowley CJ, Sajjath SM, et al. Establishment, maintenance, and recall of inflammatory memory. *Cell Stem Cell*. 2021;28(10):1758-1774.e8. doi:10.1016/j.stem.2021.07.001
- Pech M, Weckmann M, König IR, et al. Rhinovirus infections change DNA methylation and mRNA expression in children with asthma. *PLoS One*. 2018;13(11):e0205275. doi:10.1371/journal.pone.0205275
- Nemani SSP, Vermeulen CJ, Pech M, et al. COL4A3 expression in asthmatic epithelium depends on intronic methylation and ZNF263 binding. *ERJ Open Research*. 2021;7(2):00802-2020. doi:10.1183/23120541.00802-2020
- Gern JE. The ABCs of rhinoviruses, wheezing, and asthma. *J Virol*. 2010;84(15):7418-7426. doi:10.1128/jvi.02290-09
- Palmenberg AC, Rathe JA, Liggett SB. Analysis of the complete genome sequences of human rhinovirus. *J Allergy Clin Immunol*. 2010;125(6):1190-1199; quiz 1200-1. doi:10.1016/j.jaci.2010.04.010
- Walker PJ, Siddell SG, Lefkowitz EJ, et al. Recent changes to virus taxonomy ratified by the international committee on taxonomy of viruses. *Arch Virol*. 2022;167(11):2429-2440. doi:10.1007/s00705-022-05516-5
- Kennedy JL, Turner RB, Braciale T, Heymann PW, Borish L. Pathogenesis of rhinovirus infection. *Curr Opin Virol*. 2012;2(3):287-293. doi:10.1016/j.coviro.2012.03.008
- Cox DW, Khoo SK, Zhang G, et al. Rhinovirus is the most common virus and rhinovirus-C is the most common species in paediatric intensive care respiratory admissions. *Eur Respir J*. 2018;52(2):1800207. doi:10.1183/13993003.00207-2018
- Arruda E, Boyle TR, Winther B, Pevear DC, Gwaltney JM Jr, Hayden FG. Localization of human rhinovirus replication in the upper respiratory tract by in situ hybridization. *J Infect Dis*. 1995;171(5):1329-1333. doi:10.1093/infdis/171.5.1329
- Papadopoulos NG, Bates PJ, Bardin PG, et al. Rhinoviruses infect the lower airways. *J Infect Dis*. 2000;181(6):1875-1884. doi:10.1086/315513
- Yang Z, Mitländer H, Vuorinen T, Finotto S. Mechanism of rhinovirus immunity and asthma. *Front Immunol*. 2021;12:731846. doi:10.3389/fimmu.2021.731846
- Ortega H, Nickle D, Carter L. Rhinovirus and asthma: challenges and opportunities. *Rev Med Virol*. 2021;31(4):e2193. doi:10.1002/rmv.2193
- Heinonen S, Jartti T, Garcia C, et al. Rhinovirus detection in symptomatic and asymptomatic children: value of host transcriptome analysis. *Am J Respir Crit Care Med*. 2016;193(7):772-782. doi:10.1164/rccm.201504-0749OC
- Paul Chowdhury B, Gorska MM. Modeling asthma in mice using common aeroallergens. *Methods Mol Biol (Clifton, NJ)*. 2022;2506:1-18. doi:10.1007/978-1-0716-2364-0\_1
- Nissen G, Hinsenbrock S, Rausch TK, et al. Lung function of preterm children parsed by a polygenic risk score for adult COPD. *NEJM Evid*. 2023;2:EVIDoa2200279.
- van Breugel M, Qi C, Xu Z, et al. Nasal DNA methylation at three CpG sites predicts childhood allergic disease. *Nat Commun*. 2022;13(1):7415. doi:10.1038/s41467-022-35088-6
- Gilman J, Walls L, Bandiera L, Menolascina F. Statistical Design of Experiments for synthetic biology. *ACS Synth Biol*. 2021;10(1):1-18. doi:10.1021/acssynbio.0c00385
- Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics (Oxford, England)*. 2014;30(15):2114-2120. doi:10.1093/bioinformatics/btu170
- Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012;9(4):357-359. doi:10.1038/nmeth.1923
- Li H, Handsaker B, Wysoker A, et al. The sequence alignment/map format and SAMtools. *Bioinformatics (Oxford, England)*. 2009;25(16):2078-2079. doi:10.1093/bioinformatics/btp352

30. Liao Y, Smyth GK, Shi W. FeatureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics (Oxford, England)*. 2014;30(7):923-930. doi:10.1093/bioinformatics/btt656
31. Aryee MJ, Jaffe AE, Corrada-Bravo H, et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics (Oxford, England)*. 2014;30(10):1363-1369. doi:10.1093/bioinformatics/btu049
32. Touleimat N, Tost J. Complete pipeline for Infinium® human methylation 450K BeadChip data processing using subset quantile normalization for accurate DNA methylation estimation. *Epigenomics*. 2012;4(3):325-341. doi:10.2217/epi.12.21
33. Fuchs O, Bahmer T, Weckmann M, et al. The all age asthma cohort (ALLIANCE) – from early beginnings to chronic disease: a longitudinal cohort study. *BMC Pulm Med*. 2018;18(1):140. doi:10.1186/s12890-018-0705-6
34. Ritchie ME, Phipson B, Wu D, et al. Limma powers differential expression analyses for RNA-seq sequencing and microarray studies. *Nucleic Acids Res*. 2015;43(7):e47. doi:10.1093/nar/gkv007
35. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics (Oxford, England)*. 2012;28(6):882-883. doi:10.1093/bioinformatics/bts034
36. Rainer J, Gatto L, Weichenberger CX. EnsemblDb: an R package to create and use Ensembl-based annotation resources. *Bioinformatics (Oxford, England)*. 2019;35(17):3151-3153. doi:10.1093/bioinformatics/btz031
37. Durinck S, Spellman PT, Birney E, Huber W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat Protoc*. 2009;4(8):1184-1191. doi:10.1038/nprot.2009.97
38. Wu T, Hu E, Xu S, et al. ClusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation (Cambridge (Mass))*. 2021;2(3):100141. doi:10.1016/j.xinn.2021.100141
39. Shannon P, Markiel A, Ozier O, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. 2003;13(11):2498-2504. doi:10.1101/gr.1239303
40. Bindea G, Mlecnik B, Hackl H, et al. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics (Oxford, England)*. 2009;25(8):1091-1093. doi:10.1093/bioinformatics/btp101
41. Bindea G, Galon J, Mlecnik B. CluePedia Cytoscape plugin: pathway insights using integrated experimental and in silico data. *Bioinformatics (Oxford, England)*. 2013;29(5):661-663. doi:10.1093/bioinformatics/btt019
42. Peters TJ, Buckley MJ, Statham AL, et al. De novo identification of differentially methylated regions in the human genome. *Epigenetics Chromatin*. 2015;8(1):6. doi:10.1186/1756-8935-8-6
43. Shabalina AA. Matrix eQTL: Ultra Fast eQTL Analysis Via Large Matrix Operations. *Bioinformatics (Oxford, England)*. 2012;28(10):1353-1358. doi:10.1093/bioinformatics/bts163
44. Martin TC, Yet I, Tsai PC, Bell JT. coMET: visualisation of regional epigenome-wide association scan results and DNA co-methylation patterns. *BMC Bioinformatics*. 2015;16(1):131. doi:10.1186/s12859-015-0568-2
45. Milacic M, Beavers D, Conley P, et al. The Reactome pathway knowledgebase 2024. *Nucleic Acids Res*. 2024;52(D1):D672-d678. doi:10.1093/nar/gkac1025
46. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*. 2009;4(1):44-57. doi:10.1038/nprot.2008.211
47. Sherman BT, Hao M, Qiu J, et al. DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). *Nucleic Acids Res*. 2022;50(W1):W216-w221. doi:10.1093/nar/gkac194
48. Belinky F, Nativ N, Stelzer G, et al. PathCards: multi-source consolidation of human biological pathways. *Database*. 2015;2015:bav006. doi:10.1093/database/bav006
49. Xiong Z, Li M, Yang F, et al. EWAS data hub: a resource of DNA methylation array data and metadata. *Nucleic Acids Res*. 2020;48(D1):D890-D895. doi:10.1093/nar/gkz840
50. Nassar LR, Barber GP, Benet-Pagès A, et al. The UCSC genome browser database: 2023 update. *Nucleic Acids Res*. 2023;51(D1):D1188-d1195. doi:10.1093/nar/gkac1072
51. Carlier FM, de Fays C, Pilette C. Epithelial barrier dysfunction in chronic respiratory diseases. *Front Physiol*. 2021;12:691227. doi:10.3389/fphys.2021.691227
52. Fähnrich A, Reddy KD, Ott F, et al. scRNA-Seq Reveals Persistent Aberrant Differentiation of Nasal Epithelium Driven by TNF $\alpha$  and TGF $\beta$  in Post-COVID Syndrome. *bioRxiv*. 2024. doi:10.1101/2024.01.10.57480
53. Nicodemus-Johnson J, Myers RA, Sakabe NJ, et al. DNA methylation in lung cells is associated with asthma endotypes and genetic risk. *JCI Insight*. 2016;1(20):e90151. doi:10.1172/jci.insight.90151
54. Arathimos R, Suderman M, Sharp GC, et al. Epigenome-wide association study of asthma and wheeze in childhood and adolescence. *Clin Epigenetics*. 2017;9:112. doi:10.1186/s13148-017-0414-7
55. Oguariri RM, Brann TW, Adelsberger JW, et al. Short communication: S100A8 and S100A9, biomarkers of SARS-CoV-2 infection and other diseases, suppress HIV replication in primary macrophages. *AIDS Res Hum Retrovir*. 2022;38(5):401-405. doi:10.1089/aid.2021.0193
56. Deguchi A, Yamamoto T, Shibata N, Maru Y. S100A8 may govern hyper-inflammation in severe COVID-19. *FASEB J*. 2021;35(9):e21798. doi:10.1096/fj.202101013
57. Mellett L, Khader SA. S100A8/A9 in COVID-19 pathogenesis: impact on clinical outcomes. *Cytokine Growth Factor Rev*. 2022;63:90-97. doi:10.1016/j.cytogfr.2021.10.004
58. Decaestecker T, Bos S, Lorent N, et al. Elevated serum calprotectin (S100A8/A9) in patients with severe asthma. *J Asthma*. 2022;59(6):1110-1115. doi:10.1080/02770903.2021.1914649
59. Kim DH, Gu A, Lee JS, et al. Suppressive effects of S100A8 and S100A9 on neutrophil apoptosis by cytokine release of human bronchial epithelial cells in asthma. *Int J Med Sci*. 2020;17(4):498-509. doi:10.7150/ijms.37833
60. Halayko AJ, Ghavami S. S100A8/A9: a mediator of severe asthma pathogenesis and morbidity? *Can J Physiol Pharmacol*. 2009;87(10):743-755. doi:10.1139/y09-054
61. Bulger M, Groudine M. Functional and mechanistic diversity of distal transcription enhancers. *Cell*. 2011;144(3):327-339. doi:10.1016/j.cell.2011.01.024
62. Rose AB. Introns as gene regulators: a brick on the accelerator. *Front Genet*. 2018;9:672. doi:10.3389/fgene.2018.00672
63. Klemm SL, Shipony Z, Greenleaf WJ. Chromatin accessibility and the regulatory epigenome. *Nat Rev Genet*. 2019;20(4):207-220. doi:10.1038/s41576-018-0089-8
64. Ehrlich M, Lacey M. DNA Hypomethylation and Hemimethylation in cancer. In: Karpf AR, ed. *Epigenetic Alterations in Oncogenesis*. New York; 2013:31-56.
65. Soliai MM, Kato A, Naughton KA, et al. Epigenetic responses to rhinovirus exposure in airway epithelial cells are correlated with key transcriptional pathways in chronic rhinosinusitis. *Allergy*. 2023;78(10):2698-2711. doi:10.1111/all.15837
66. Reeves SR, Barrow KA, White MP, Rich LM, Naushab M, Debley JS. Stability of gene expression by primary bronchial epithelial cells over increasing passage number. *BMC Pulm Med*. 2018;18(1):91. doi:10.1186/s12890-018-0652-2

67. Chen Y, Chen L, Lun ATL, Baldoni PL, Smyth GK. edgeR 4.0: powerful differential analysis of sequencing data with expanded functionality and improved support for small counts and larger datasets. *bioRxiv*. 2024. doi:[10.1101/2024.01.21.576131](https://doi.org/10.1101/2024.01.21.576131)
68. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics (Oxford, England)*. 2010;26(1):139-140. doi:[10.1093/bioinformatics/btp616](https://doi.org/10.1093/bioinformatics/btp616)

**How to cite this article:** Risha MA, Reddy KD, Nemani SSP, et al. Epigenetic training of human bronchial epithelium cells by repeated rhinovirus infections. *Allergy*. 2024;00:1-16. doi:[10.1111/all.16388](https://doi.org/10.1111/all.16388)

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.