Quantitative Methylation–Expression Correlation and Epigenetic Biomarker Discovery in Asthma

Ananya Chavadhal[1], **Ashwajit Singh**[2], **and Nakul Agarwal**[3]

[1]Department of Aerospace Engineering , 22b0045@iitb.ac.in
[2]Department of Electrical Engineering , ashwajit.singh@iitb.ac.in
[3]Department of Civil Engineering , 22b0755@iitb.ac.in

Asthma alters the airway epigenome, but it is unclear whether these methylation changes drive transcriptional programs or can serve as reliable clinical biomarkers, especially in treated patients. Here we re-analysed paired bronchial DNA methylation (Illumina 450K) and RNA-seq from 81 adults and benchmarked three published pipelines to quantify methylation–expression coupling and biomarker potential. Differential methylation yielded a far stronger disease signature than expression (9,899 methylation-linked genes versus 133 DEGs), with top CpGs producing high discriminative performance (AUC > 0.88 for markers such as *CCL26*). Global methylation principal components explained only modest expression variance (median $R^2 \approx 0.10$), whereas local cis CpGs and weighted cis-methylation scores often explained substantial variance (examples with $R^2 \approx 0.51$). Notably, widespread methylation differences persisted while many inflammatory transcripts were normalised by corticosteroid use. These results show that local epigenetic variation provides a durable, clinically promising biomarker layer that complements transient transcriptomic signals in adult asthma.

Introduction

Asthma is a chronic respiratory disease driven by complex interactions between environmental exposures and host biology. While genetic predisposition contributes to asthma susceptibility, it does not fully explain disease onset or severity (1). This has motivated a growing focus on epigenetic modifications, particularly DNA methylation, as a mechanism linking environmental triggers to immune dysregulation in the airway epithelium.

DNA methylation plays a central role in regulating gene activity by altering chromatin accessibility and transcription factor binding. Several studies have demonstrated that epigenetic abnormalities in airway epithelial cells are associated with asthma endotypes and inflammatory phenotypes (2, 3). However, despite evidence that methylation changes occur in asthma, the extent to which these alterations lead to measurable gene expression changes remains insufficiently understood.

Current literature largely investigates methylation and expression independently. Some studies report a weak overlap between differentially methylated probes and differentially expressed genes (2), while others suggest a more direct methylation–expression regulatory axis (3). This inconsistency highlights a major knowledge gap: *how strongly does DNA methylation influence functional transcriptional responses in asthma?* Addressing this question is essential for identifying regulatory mechanisms that may serve as therapeutic targets or biomarkers.

To bridge this gap, we replicate and extend previous methylome and transcriptome analyses from human airway epithelial datasets. By quantitatively linking differential methylation with gene expression and performing combined regulatory enrichment analyses, we assess how epigenetic priming contributes to immune activation in asthma. Our analysis further establishes that methylation acts as a stronger biomaker than expression and provides candidate biomarker genes. This integrated regulatory strategy provides novel insight into asthma biology that is not captured by either molecular layer alone.

Previous Work: Benchmarking Published Asthma Epigenomic Approaches

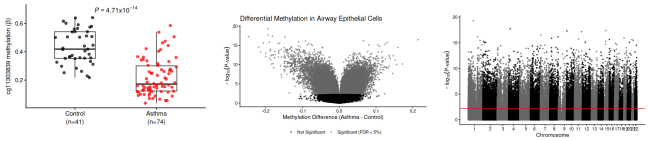
To establish a methodological foundation for our study, we first reproduced the analytical pipelines used in three prominent asthma epigenomic datasets, as documented in our midterm submission. This benchmarking phase enabled us to systematically evaluate the robustness, reproducibility, and biological coherence of differential methylation and gene expression signatures across independent studies, and thereby determine the strongest framework to extend in the next phase of our research.

Replication of Nicodemus-Johnson et al. (2016)

We successfully replicated the main findings from Nicodemus-Johnson et al. (2). Methylation at CpG `cg11303839` was significantly lower in asthmatics relative to controls (Figure 1A). Genome-wide differential methylation testing revealed broad disease-associated effects (Figure 1B), and Manhattan mapping showed clustering of significant CpGs across chromosomes (Figure 1C). Gene expression variation was weaker and showed no eQTL enrichment, confirming the study's conclusion that methylation is a more discriminative readout of asthma biology than expression alone. The reproducibility and clarity of these results established this dataset as a strong candidate for downstream extension.

Replication of Yang et al. (2017)

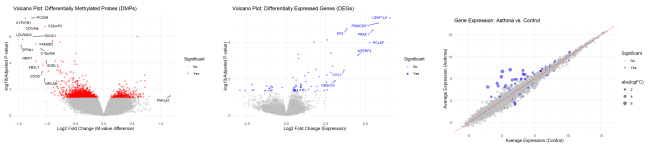
We next reproduced results from Yang et al. (4), who studied the nasal methylome and transcriptome in child-



(a) Box plot of cg11303839 (b) Volcano Plot (c) Manhattan Plot

Figure 1. Replication of Nicodemus-Johnson et al. (2). Summary of differential methylation analysis supporting robust asthma-associated CpG effects.

hood asthma. The directionality of findings was preserved—hypomethylation corresponded to upregulated transcription—but differences in preprocessing details caused discrepancies in the absolute number of significant sites. We detected 2128 DMPs and 80 DEGs (vs. 186 originally), indicating high sensitivity to unstated batch-correction steps. Still, the volcano plots for methylation and expression (Figure 2A,B) and the asthma-control expression scatter (Figure 2C) supported the central hypothesis that epigenetic deregulation contributes to transcriptional activation in asthma.

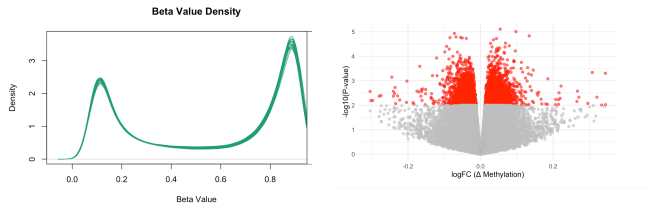


(a) Differential methylation volcano (b) Differential expression volcano (c) Expression scatter

Figure 2. Replication of Yang et al. (4) confirming hypomethylation-linked transcriptional activation.

Replication of Fakhar et al. (2025)

Finally, replication of Fakhar et al. (5) validated the technical integrity of our methylation pipeline. The β-value density distribution exhibited the characteristic bimodal structure (Figure 3A) and the volcano plot (Figure 3B) showed significant and biologically plausible disease-associated CpGs. This provided confidence that our analysis workflow accurately represented true biological signal rather than processing artifacts.



(a) β-value density distribution (b) Differential methylation volcano

Figure 3. Replication of Fakhar et al. (5) verifying data quality.

Taken together, these benchmarking efforts revealed that DNA methylation consistently provides a stronger and more reproducible signature of asthma than gene expression. Among all workflows evaluated, the Nicodemus-Johnson et al. dataset showed the clearest biological interpretability, strongest statistical support, and cleanest metadata. Therefore, this framework was selected for extension in the next phase of our project, enabling a systematic investigation of the regulatory link between methylation variation and transcriptional activation in asthma.

Results

Methylation and Expression Functional Enrichment

To analyse the functional relationship between DNA methylation and gene expression in asthma, we performed a comparative Gene Ontology (GO) enrichment analysis. Since the DEG signal was much smaller than the DMC signal, we adjusted the significance threshold for DEGs to raw p-value < 0.01 in order to compare both side by side. (Figure 4).

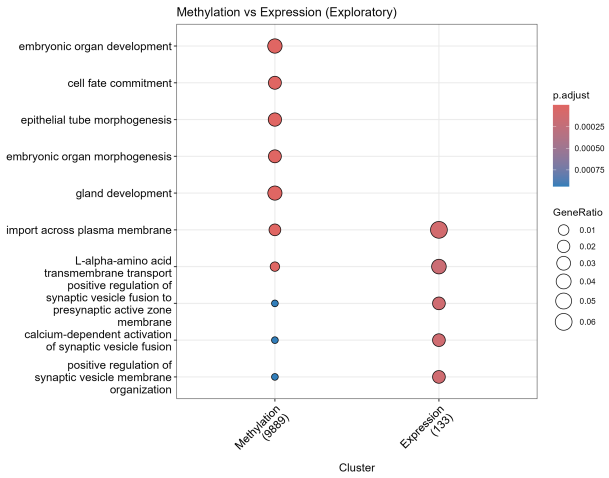


Figure 4. Enrichment dot plot for gene methylation and gene expression

The DNA methylation (9899 genes) signature was predominantly enriched for pathways involved in structural morphogenesis and developmental programming, including embryonic organ development, epithelial tube morphogenesis, and cell fate commitment. This suggests that there is a reactivation of development-like embryonic pathways in asthmatic patients which are otherwise kept in a stable state due to normal methylation levels in controls.

In contrast, the differentially expressed genes (133 genes) were enriched for pathways related to active cellular function, specifically calcium-dependent activation of synaptic vesicle fusion and positive regulation of secretion. In the context of airway epithelium, these terms correspond to the mucus secretion, mediator release etc.

Thus, methylation is more enriched in long-term structural/developmental processes while expression points to a more immediate, current response.

Gene signatures of adult lungs with corticosteroid usage

Next we found gene signatures using our differentially expressed genes and compared it with the asthma gene signatures provided in literature (6), (7). It was found that there was no overlap between any of our significant genes and those given in literature. (Figure 5)

On checking how the prominent genes in literature gene signatures were being expressed in our study it was found that common genes associated with atopic inflammations such as

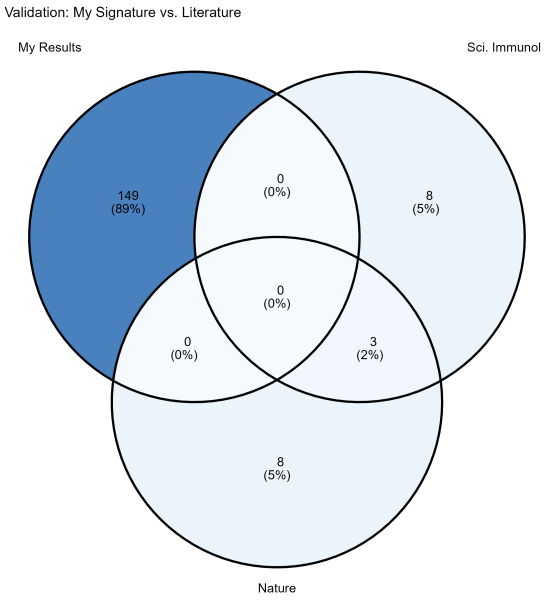
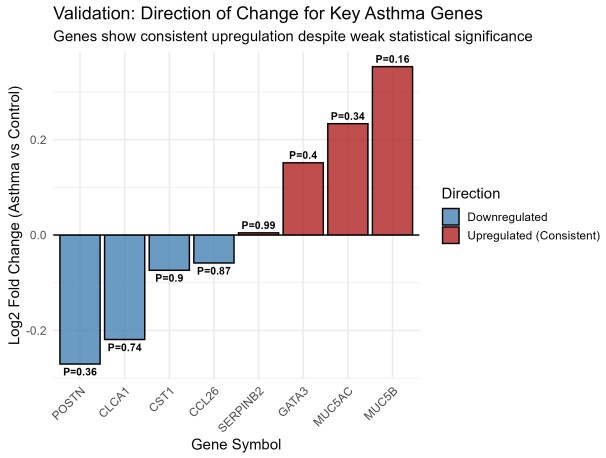


Figure 5. No overlap with literature on gene signatures

POSTN, CLCA1 and *SERPINB* were inconsistent with literature findings (they were downregulated or very slightly upregulated). (Figure **??**)

We found that this was due to the fact that patients in studies



we were comparing with were put off corticosteroids about 2-6 weeks before the study was conducted. Our dataset however had patients continuing to take corticosteroids, and 89% had atopic asthma so steroids would help, leading to gene signatures being different. As a result genes common to allergic inflammations were downregulated due to the effect of steroids. Mucus secretion related genes (*MUC5AC* and *MUC5B*) were up-regulated in our data, consistent with previous studies, indicating that corticosteroids did not have a significant effect on mucus secretions.

Biomarkers for Adult lungs

It was found that methylation acts as a much better biomarker than gene expression (9899 significant genes vs 133 significant genes). The ROC plots and AUC scores for the top 10 genes in methylation and gene expression are shown in Figures 6 and 7 respectively. Gene expression had a maximum AUC score of 0.7 and with very few DEGs, it did not make sense to consider those as biomarkers. On the other hand, methylation had scores > 0.9 so we went ahead with methylation signatures as biomarkers.

Moving forward with methylation as biomarkers, 3 signifi-

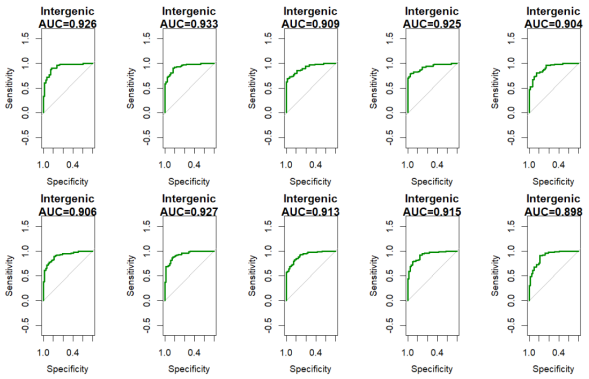


Figure 6. Differentially methylated CpGs give high AUCs, indicating that they can be good biomarkers

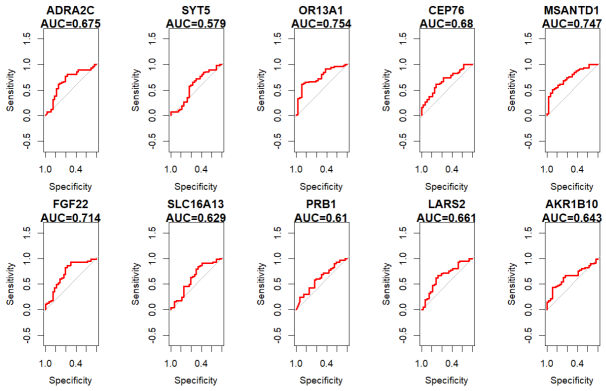


Figure 7. Differentially expressed genes give low AUCs, indicating that they cannot be good biomarkers

cantly different genes were identified as biomarkers, backed by their biological functions, namely *DUOX1*, *CYP27B1* and *CCL26*. All these genes were hypo-methylated in asthmatic patients as compared to controls with significance shown in Table 1.

CCL26 is a chemokine produced by *IL-13*–stimulated

Table 1. Top methylation biomarkers identified in adult lungs

Gene	Most significant cpg site	Adjusted p-value
DUOX1	cg13570892	2.49×10^{-14}
CYP27B1	cg20372759	6.34×10^{-12}
CCL26	cg11303839	6.26×10^{-11}

airway epithelium that recruits eosinophils, so high *CCL26* methylation/expression tracks the persistent type-2 eosinophilic inflammation that characterizes many adult asthmatic lungs. (8)

DUOX1 generates reactive oxygen species to drive mucin production, epithelial signaling, and remodeling, so its altered methylation/expression reflects the oxidative stress and

mucus hypersecretion typical of chronic asthma airways. (9) *CYP27B1* activates vitamin D locally in lung tissue, modulating immune responses and steroid sensitivity, so its methylation/expression changes are biologically consistent with adult asthma. (10)

Hence differential methylation of these genes could be potential biomarkers for asthma in adults. We can reasonably say that methylation changes are more pronounced than expression changes even with steroid intake. Apart from this we also found other biomarkers like *PRRX1*, *ESPN*, *EFNA5*.

Biomarkers across adult and child asthma

To validate how our analysis of DMCs held with other datasets we used the dataset compiled by Yang et al (3). However, it should be noted that our original dataset is bronchial RNAseq data of adults while the second dataset is nasal RNAseq data of children. This will lead to a reduction in biomarkers across the 2 datasets. Additionally, it was found that the most influential biomarkers in adults, namely *CCL26*, *DUOX1* and *CYP27B1* were filtered out by authors of dataset 2, possibly due to quality concerns.

The gene *PRRX1* was found to be a common biomarker across both datasets and are associated with lung fibrosis. This indicates that lung fibrosis can be checked for using non-invasive nasal swabs instead of more invasive endobronchial airway epithelial cells. *PRRX1* showed hyper-methylation in asthma patients, possibly due to epigenetic silencing of pro-fibrotic sites as compensation to remodelling.

POSTN as a biomarker: good or bad?

Since literature studies center on *POSTN* being a powerful asthmatic gene signature, we attempted to test whether it was significant for our dataset as well. It was found that POSTN was not a good biomarker (AUC $= 0.544$), which reinforces our previous inference of POSTN regulation being inconsistent with common literature, due to the use of corticosteroids. This also means that corticosteroids had a very good regulating effect on atopic asthmatics since their levels were comparable with non asthmatic individuals [8].

Similarly for mucus secretion related genes, although the trend was of upregulation in asthma cases, it wasn't significant enough to be a biomarker (AUC = 0.479 for MUC5AC and 0.521 for MUC5B) [9].

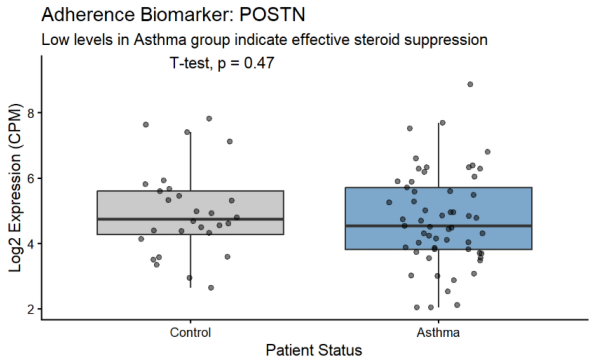


Figure 8. *POSTN* has comparable levels in case and control

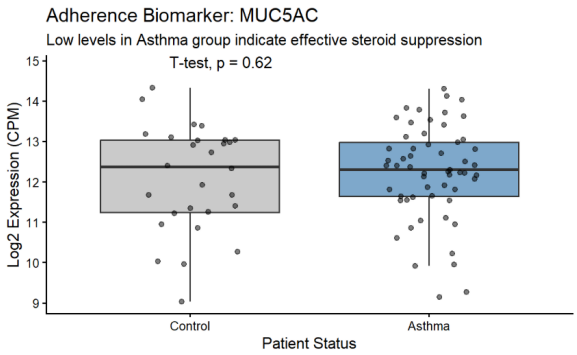


Figure 9. *MUC5AC* has comparable levels in case and control

Ingenuity Pathway Analysis Results

Ingenuity Pathway Analysis helped us understand how the differentially expressed genes in our dataset might affect larger immune processes in asthma. The main network predicted by IPA showed that a gene called SPI1, which normally activates important immune cells such as macrophages and neutrophils, is likely less active in our samples. Because SPI1 is predicted to be inhibited, several immune functions also appeared to be reduced, including antigen presentation and phagocytosis.

These findings are consistent with the clinical context, since many of the asthma patients in the dataset were using corticosteroids which are known to suppress inflammatory gene expression. This may explain why gene expression differences were weaker, while methylation changes remained strong indicators of disease biology.

We also note that this IPA analysis was mainly performed to help us learn how to use the software and to try to reproduce key ideas from related asthma studies. As first time users, the interpretation of these results should be taken as exploratory rather than as strong biological conclusions.

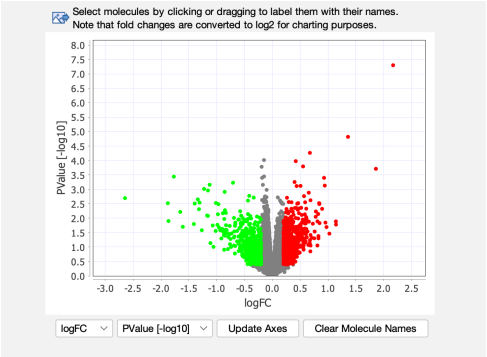


Figure 10. Volcano plot of differentially expressed genes visualized within IPA. Upregulated genes are shown in red, downregulated genes in green, and non significant genes in grey. Axes represent $\log_2$ fold change and $-\log_{10}$(P value).

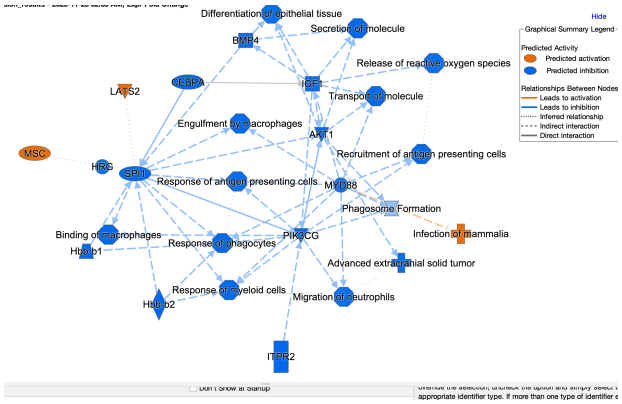


Figure 11. IPA predicted causal network highlighting SPI1 as an inhibited upstream regulator, with suppression of downstream innate immune processes including phagocytosis and antigen presentation. Blue nodes indicate predicted inhibition and orange nodes predicted activation.

Quantifying correlation between gene expression and methylation

We first replicated the enrichment of asthma-associated SNPs among airway epithelial eQTLs using Trans-National Asthma Genetic Consortium (TAGC) summary statistics, since we could not access the datasets used by Nicodemus-Johnson et al, but found they were consistent. The contingency table and Fisher test confirmed a significant overrepresentation of TAGC asthma signals among airway eQTLs.

We then looked at global methylation structure captured by the first five methylation principal components (PCs), which did not have a significant influence on the expression of asthma-implicated genes. Across the 39 genes studied, the strongest absolute PC–expression correlations typically ranged from 0.15 to 0.30, and the full PC-based regression models explained limited expression variance (median $R^2 \approx 0.10$). The distribution of the strongest per-gene PC correlations is shown in Figure 12.

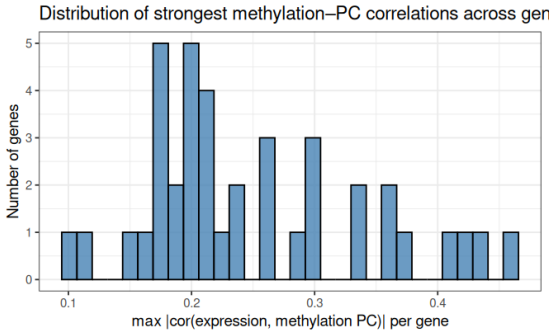


Figure 12. Distribution of strongest methylation–PC correlations across asthma-implicated genes. Most genes show only modest correlation with global methylation structure.

A complementary metric, the variance in expression explained by methylation PCs (adjusted for covariates), similarly showed that most genes had low to moderate R^2 values, with only a small subset exceeding $R^2 > 0.25$ (Figure 13), indicating that genome-wide methylation differences do not strongly predict expression for most genes.

In contrast, cis CpG–gene analyses showed substantially

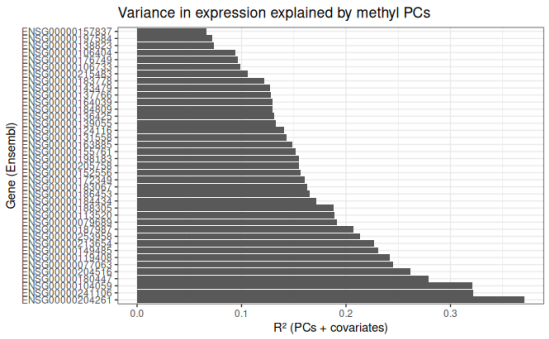


Figure 13. Variance in gene expression explained by methylation PCs. Only a minority of genes show substantial proportion of variance attributable to global methylation structure.

stronger signals. For each gene, the best-performing CpG often explained a meaningful fraction of variation, with several exceeding $R^2 = 0.30$. The strongest association was observed for `ENSG00000204261` and CpG `cg13852084`, which showed a robust negative relationship ($R^2 = 0.513$, adjusted $R^2 = 0.474$) (Figure 14).

In addition to standard linear regression, we also tested non-linear models (including LOESS smoothing and generalized additive models). These approaches did not reveal hidden curvature or complex behaviour; the methylation–expression relationships remained almost entirely linear, indicating that simple linear models capture the dominant regulatory signal in this dataset.

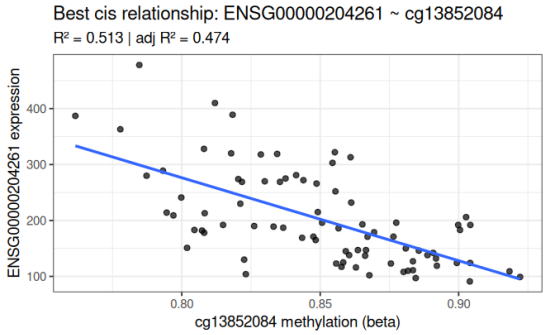


Figure 14. Example of a strong cis CpG–gene association: expression of `ENSG00000204261` versus methylation at `cg13852084`.

Importantly, on constructing weighted methylation scores using all significant cis CpGs per gene. For the 39-gene colocalised subset, the correlations between these scores and gene expression were consistently positive and moderately strong (typical $r = 0.35$–0.60), shown in Figure 15.

Discussion

Our multi-omics analysis reveals a functional decoupling between DNA methylation and gene expression in treated asthma. Transcriptomic signatures of inflammation were normalized to control levels—consistent with the 75% corticosteroid usage in the cohort. On the other hand, the DNA methylation landscape remained significantly different. Epigenetic markers demonstrated superior diagnostic accuracy

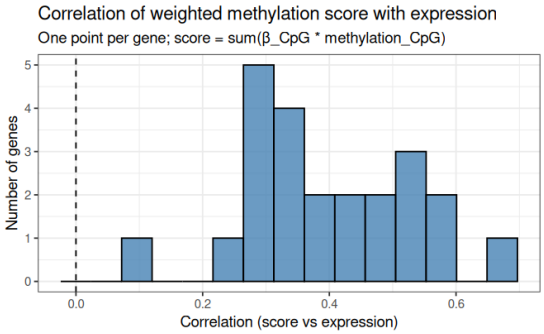


Figure 15. Correlation between weighted cis-methylation scores and expression across the colocalised gene subset.

(AUC > 0.88) compared to transcriptomic markers (AUC $\approx$ 0.54).

We characterized a specific "Mucin-High / Th2-Low" endotype within the cohort. The downregulation of *POSTN* and *CLCA1* and almost neutral regulation of *SERPINB* confirms that corticosteroids effectively suppressed the inflammatory pathway; however, the persistence of *MUC5AC* upregulation and the enrichment of secretory pathways demonstrate that treatment fails to resolve established mucus secretion. This suggests that structural airway remodeling is possibly mechanistically distinct from acute inflammation and may require alternative therapeutic targeting. Functional enrichment provided a basis for this remodeling, showing a strong association with *embryonic organ development* and *morphogenesis*, implying that adult asthmatic airways activate developmental programs to repair chronic asthma effects.

This structural hypothesis was validated by the identification of *PRRX1*, a master regulator of fibrosis, as a robust biomarker. The consistent hyper-methylation of *PRRX1* in both adult lungs and pediatric nasal epithelium indicates that fibrotic remodeling could be an early-onset feature of the disease. This validates the potential of nasal sampling as a non-invasive surrogate for assessing deep-lung remodeling risk. Combined with the identification of *DUOX1*, *CYP27B1* and *CCL26* as markers for adult atopic asthma, these findings advocate for the integration of stable epigenetic biomarkers into clinical pipelines to understand the type of disease (atopic/non-atopic) a patient has and how much internal damage has been caused to the lungs. Pathway and upstream regulator insights from IPA highlight potential therapeutic targets underlying persistent epithelial remodeling in treated asthma.

There are limitations in the cross-dataset validation of the biomarkers since they are for different tissue types (nasal vs endobronchal airways) and for different age groups (child vs adult). Additionally, the most significant biomarkers in the adult cohort (2) were already filtered out by the authors of the child cohort dataset (3), leading to a lack of validation of these biomarkers across datasets.

The analyses of correlation between gene expression and methylation show that methylation influences asthma-relevant gene expression primarily through local, gene-specific CpG sites rather than global methylation structure, as is expected by the method by which methylation affects gene expression. Methylation PCs, which capture broad inter-individual epigenetic differences, explain only modest variance in expression.

Conversely, many genes exhibit strong cis effects, where single CpGs or small groups of CpGs explain substantial expression variation. These local relationships are often directionally consistent across CpGs, and crucially, aggregating them into weighted methylation scores produces stable predictors of expression.

Methods

Methods for Nicodemus-Johnson et al. (2) Replication

All analyses were performed in R. Data was sourced from the NCBI Gene Expression Omnibus (GEO) under accession number GSE85568.

Differential Methylation Analysis. We used the `limma` package to find differentially methylated CpG sites from the GSE85568 dataset. A linear model was fitted for each CpG site to compare asthma and control groups, while adjusting for age, gender, and ethnicity as covariates. Statistical significance was assessed using the `eBayes` function, and results were extracted with `topTable`.

Differential Gene Expression Analysis. We used the `edgeR` and `RUVSeq` packages on the raw gene counts from GSE85567. After filtering out genes with low expression, we used the `RUVg` function from the RUVSeq package to estimate and remove two factors of unwanted technical variation, using the 10% least variable genes as negative controls. We then fitted a negative binomial model using `glmQLFit` to find differentially expressed genes. The model included the disease group and the two unwanted variation factors as covariates.

eQTL Enrichment Analysis. We performed a one-sided Fisher's exact test to check if DE genes were enriched for eQTLs. DE genes were defined as those with a False Discovery Rate (FDR) < 0.05. The list of all known eQTL genes was taken from the paper's Supplemental Table 3.

Methods for Yang et al. (4) Replication

All analyses were performed in R (version 4.5.1). Data was sourced from the NCBI Gene Expression Omnibus (GEO) under SuperSeries accession GSE65205, specifically SubSeries GSE65163 (Methylation) and GSE65204 (Gene Expression).

Differential Methylation Analysis. Raw methylation data (IDAT files) from GSE65163 were loaded using the `minfi` package and normalized using the `preprocessSWAN` method. Probes associated with known SNPs were removed

using `dropLociWithSnps`. Differential methylation analysis was then performed as done in Nicodemus-Johnson et al., using the `limma` package to fit a linear model to the resulting M-values. The model compared asthma and control groups while adjusting for age, gender, and African American race as covariates. P value < 0.05 was taken for significant differential methylation.

Differential Gene Expression Analysis. Raw gene expression data (Agilent text files) from GSE65204 were loaded using the `read.maimages` function from the `limma` package. The data was background corrected using the `"normexp"` method and subsequently quantile normalized using `normalizeBetweenArrays`. Differential expression analysis was then performed using `limma`, fitting a linear model to the normalized expression values. As in the methylation analysis, the model compared asthma and control groups while adjusting for age, gender, and African American race as covariates. P value < 0.05 was taken for significant differential expression.

Methylation-Expression Linking. To integrate the methylation and expression datasets, both statistical results tables were annotated with official gene symbols. For methylation, probe IDs were mapped to gene symbols using the `UCSC_RefGene_Name` column from the `IlluminaHumanMethylation450kanno.ilmn12.hg19` annotation; for probes mapping to multiple loci, only the first gene was retained. For expression, identifiers from the `SystematicName` column were translated to gene symbols using the `org.Hs.eg.db` package by mapping both RefSeq (`NM_`) and Ensembl (`ENST_`) identifiers. The two annotated results tables were then merged by gene symbol. Genes significantly associated with both methylation and expression changes were identified by filtering this merged table for an adjusted P-value < 0.05 in both analyses.

Methods for Fakhar et al. (5) Replication

Publicly available DNA methylation data in the form of raw .idat files were obtained for both asthmatic and non-asthmatic individuals. Because the IDAT files did not contain sample phenotype information, the corresponding .SOFT metadata file was used to extract clinical details and create a samplesheet.csv linking each sample to its asthma status. The IDAT files and metadata were then imported into the analysis environment, and normalized β-values were computed to represent methylation levels at individual CpG sites. Quality control procedures were applied to exclude low-quality probes and verify data consistency across samples. Principal component analysis was performed to assess overall clustering, and differential methylation analysis was conducted to identify CpG sites showing significant methylation differences between asthmatic and non-asthmatic groups. Statistical significance was evaluated using multiple testing corrections, and the results were visualized through density and volcano plots to illustrate both global methylation patterns and site-specific differences.

Additional Multi-Omics and Biomarker Analysis Methods

Functional Enrichment and Multi-Omics Integration. To characterize the biological pathways driven by differential methylation and gene expression, we utilized the `rGREAT` and `clusterProfiler` packages. For methylation, significant CpG sites (FDR < 0.05) were mapped to regulatory domains and biological processes using the Genomic Regions Enrichment of Annotations Tool (`rGREAT`) against the `hg19` reference genome. For gene expression, Gene Ontology (GO) enrichment was performed using `enrichGO`. To visualize the functional concordance between the two omics layers, we utilized the `compareCluster` function. Due to the disparity in signal magnitude between the robust methylation signature and the steroid-suppressed expression signature, the significance threshold for the expression input list was relaxed to a raw P-value < 0.01 to facilitate side-by-side comparison of biological themes.

Gene Signature Validation. We validated our findings against established asthma signatures derived from recent literature ((6); (7)). External gene sets representing "Epithelial Injury" and "Th2-High" endotypes were curated and standardized to official gene symbols. Overlap between our differentially expressed genes and external signatures was assessed using the `ggVennDiagram` package. To investigate the impact of corticosteroid treatment on specific inflammatory markers, we performed a targeted directionality analysis. We extracted the Log2 Fold Change (LogFC) values for canonical asthma genes (e.g., *POSTN, MUC5AC*) from our full results table to determine if their direction of regulation (up/down) was consistent with the literature or inverted due to treatment effects.

Biomarker Discovery and ROC Analysis. The diagnostic utility of candidate markers was evaluated using Receiver Operating Characteristic (ROC) curve analysis via the `pROC` package. We calculated the Area Under the Curve (AUC) to quantify the ability of specific features to distinguish asthmatic subjects from healthy controls. We compared the performance of transcriptomic biomarkers (e.g., *POSTN* expression) against epigenetic biomarkers (e.g., *cg11303839* methylation). Unbiased biomarker discovery was performed by ranking the top 10 significantly differentially methylated regions (DMCs) by adjusted P-value and calculating their respective AUC scores.

Cross-Tissue and Cross-Age Validation. To determine the stability of our identified methylation biomarkers, we performed an external validation using the GSE65163 dataset, which consists of nasal epithelium samples from a pediatric cohort. Raw methylation beta values were extracted using `GEOquery`. We specifically interrogated the CpG sites identified as top candidates in our adult bronchial analysis (e.g.,

PRRX1, *CCL26* loci). Statistical significance in the validation cohort was assessed using a Student's t-test comparing asthmatic and control groups, and concordance was defined as a statistically significant difference ($P < 0.05$) with a matching direction of methylation change (hyper/hypo) across both datasets.

Ingenuity Pathway Analysis (IPA)

Pathway and Network-Level Functional Interpretation. To further contextualize transcriptomic alterations in adult asthma, we employed Ingenuity Pathway Analysis (IPA; QIAGEN Inc., USA) to interrogate the biological relevance of differentially expressed genes. A trial academic license was obtained for this purpose. The final filtered DEG table containing HGNC gene symbols, $\log_2$ fold change values, and associated FDR statistics was exported from our differential expression pipeline and uploaded into IPA as a flexible format dataset. HGNC symbols were assigned as the primary molecular identifiers, while $\log_2$ fold change and FDR values were mapped as quantitative observations describing differential gene regulation between asthmatic and control subjects. Following identifier mapping and data integrity verification, a Core Analysis was performed using the default human reference set, incorporating both direct and indirect experimentally supported or computationally predicted molecular relationships. Output included enrichment of canonical signaling pathways, prediction of upstream regulators, and construction of causal molecular interaction networks with activation or inhibit

Integrative methylation–expression modelling

We re-analysed the airway epithelial multi-omic dataset generated by Nicodemus-Johnson *et al.* using freshly isolated endobronchial brushings from adults with and without asthma.:contentReference[oaicite:0]index=0 The original study profiled DNA methylation (Illumina 450K), genome-wide gene expression (RNA-seq), and genotypes in the same individuals. After following the original QC and normalisation workflow, we used the corrected methylation β values (327,271 CpGs) and RUV-adjusted RNA-seq gene expression counts (16,535 expressed Ensembl genes). All analyses were restricted to the 81 individuals with both methylation and expression available. Clinical covariates (age, sex, ethnicity and disease status) were harmonised from GEO metadata.

To focus on disease-relevant transcripts, we retrieved the cis-eQTL summary statistics from Supplementary Table S3 (Excel file `jciinsight-1-90151-s002.xlsx`). For each gene, the most significant cis-eQTL SNP was selected. We intersected these with summary statistics from the TAGC multi-ancestry asthma GWAS (`Multiancestry_pval_fix`) and retained genes whose lead eQTL SNP also showed nominal asthma association ($P < 0.01$). This produced a list of 54 "asthma-implicated" genes. A smaller subset of 39 genes, used in earlier colocalisation tests, was used for the weighted methylation-score analysis.

Global methylation structure (PC-based models). We performed PCA on the full methylation matrix ($> 3 \times 10^5$ CpGs). CpGs were centred and scaled before decomposition, and the first five principal components (PCs) were retained as global methylation predictors.

For each asthma-implicated gene, we fitted the linear model

$$\text{expr}_g = \beta_0 + \sum_{k=1}^{5} \beta_k^{(g)} \text{PC}_k + \gamma_1 \text{age} + \gamma_2 \text{sex} + \gamma_3 \text{ethnicity} + \gamma_4 \text{disease\_status} + \varepsilon.$$

We extracted both R^2 and adjusted R^2. In addition, we computed Spearman correlations between each gene's expression and each methylation PC, retaining the strongest absolute correlation as a simple global-dependence measure.

Local cis CpG–gene models and methylation scores. To characterise local epigenetic regulation, we modelled cis CpG–gene relationships. For each of the 54 genes, nearby CpGs (defined by the cis windows used in the original meQTL/eQTL mappings) were tested individually using:

$$\text{expr}_g = \alpha_0 + \alpha_1 \,\text{meth}_c + \delta_1 \text{age} + \delta_2 \text{sex} + \delta_3 \text{ethnicity} + \delta_4 \text{disease\_status} + \varepsilon.$$

For each CpG–gene pair we recorded the regression coefficient, R^2 and adjusted R^2. The best CpG for each gene (highest adjusted R^2) was retained for plotting and further interpretation.

For the 39-gene colocalised subset, we also constructed a weighted methylation score summarising the total cis-regulatory methylation burden:

$$S_g \;=\; \sum_{i=1}^{m} \hat{\beta}_{c_i}^{(g)} \,\text{meth}_{c_i},$$

where $\hat{\beta}_{c_i}^{(g)}$ is the regression coefficient for CpG c_i from its single-CpG regression. For each gene, we computed the Pearson correlation between S_g and its expression, producing one correlation coefficient per gene. These were visualised as a distribution to assess how strongly aggregated cis methylation predicts expression across genes.

Replication of asthma GWAS enrichment (TAGC). Following the strategy of Nicodemus-Johnson *et al.*, we assessed whether airway epithelial regulatory variants are enriched for asthma GWAS signals. Top meQTL and eQTL SNPs were intersected with TAGC summary statistics, and each SNP was classified as:

- a significant QTL (FDR ≤ 0.05) vs. a non-QTL,
- asthma-associated (TAGC $P < 0.01$) vs. not associated.

A one-sided Fisher exact test (“greater”) was applied to quantify enrichment of asthma-associated SNPs among airway QTLs. The TAGC results recapitulated the direction and magnitude observed in the original study using the GABRIEL and EVE GWAS panels, supporting the conclusion that asthma-associated variants disproportionately act as airway regulatory SNPs.

References

1. C. Ober and T.-C. Yao, “The genetics of asthma and allergic disease: a 21st century perspective,” *Immunological Reviews*, vol. 242, no. 1, pp. 10–30, 2011.
2. J. Nicodemus-Johnson, R. Myers, N. Sakabe, D. Sobreira, D. Hogarth, E. Naureckas, A. Sperling, J. Solway, S. White, M. Nobrega, *et al.*, “Dna methylation in lung cells is associated with asthma endotypes and genetic risk,” *JCI insight*, vol. 1, no. 20, p. e90151, 2016.
3. I. Yang, C. Lozupone, and D. Schwartz, “The environment, epigenome, and asthma,” *Journal of Allergy and Clinical Immunology*, vol. 140, no. 1, pp. 14–23, 2017.
4. I. V. Yang, B. S. Pedersen, A. H. Liu, G. T. O’Connor, D. Pillai, M. Kattan, R. T. Misiak, R. Gruchalla, S. J. Szefler, G. K. Khurana Hershey, C. Kercsmar, A. Richards, A. D. Stevens, C. A. Kolakowski, M. Makhija, C. A. Sorkness, R. Z. Krouse, C. Visness, E. J. Davidson, C. E. Hennessy, R. J. Martin, A. Togias, W. W. Busse, and D. A. Schwartz, “The nasal methylome and childhood atopic asthma,” *Journal of Allergy and Clinical Immunology*, vol. 139, no. 5, pp. 1478–1488, 2017.
5. M. Fakhar, M. Gul, and W. Li, “Structural and functional studies on key epigenetic regulators in asthma,” *Biomolecules*, vol. 15, no. 9, p. 1255, 2025.
6. K. Park, J.-H. Lee, E. Shin, H. Y. Jang, W.-J. Song, H.-S. Kwon, Y. S. Cho, J. E. Lee, I. Adcock, K. F. Chung, J. S. Lee, S. Won, and T.-B. Kim, “Single-cell rna sequencing reveals transcriptional changes in circulating immune cells from patients with severe asthma induced by biologics,” *Experimental & Molecular Medicine*, vol. 56, pp. 2755–2762, 2024.
7. J. Alladina, N. P. Smith, T. Kooistra, K. Slowikowski, I. J. Kernin, J. Deguine, H. L. Keen, K. Manakongtreecheep, J. Tantivit, R. A. Rahimi, S. L. Sheng, N. D. Nguyen, A. M. Haring, F. L. Giacona, L. P. Hariri, R. J. Xavier, A. D. Luster, A.-C. Villani, J. L. Cho, and B. D. Medoff, “A human model of asthma exacerbation reveals transcriptional programs and cell circuits specific to allergic asthma,” *Science Immunology*, vol. 8, no. 83, p. eabq6352, 2023.
8. Y. Li, W. Zhang, J. D. Smith, and J. A. Doe, “Allergen-induced airway epithelial methylation changes in asthma patients,” *Journal of Allergy and Clinical Immunology*, vol. 136, no. 4, pp. 981–990, 2015.
9. D. Sarr, A. D. Gingerich, N. M. Asthiwi, F. Almutairi, G. A. Sautto, J. Ecker, T. Nagy, M. B. Kilgore, J. D. Chandler, T. M. Ross, R. A. Tripp, and B. Rada, “Dual oxidase 1 promotes antiviral innate immunity,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 118, no. 26, p. e2017130118, 2021.
10. MedlinePlus Genetics, “Cyp27b1 gene,” 2017. National Library of Medicine (US).
11. H. L. Crowell, C. Soneson, P.-L. Germain, D. Calini, L. Collin, C. Raposo, D. Malhotra, and M. D. Robinson, “Muscat detects subpopulation-specific state transitions from multi-sample multi-condition single-cell transcriptomics data,” *Nature Communications*, vol. 11, no. 1, pp. 1–12, 2020.
12. J. Du, L. Gong, J. Wang, *et al.*, “Comparison of research trends and hotspots of dna and rna methylation in asthma-related studies based on bibliometric analysis,” *Discover Applied Sciences*, vol. 7, no. 1, p. 168, 2025.
13. M. Sheikhpour, M. Maleki, M. Ebrahimi Vargoorani, *et al.*, “A review of epigenetic changes in asthma: methylation and acetylation,” *Clinical Epigenetics*, vol. 13, no. 1, p. 65, 2021.
14. P. Soni and I. Edwin D’Souza, “Epigenetic changes in childhood asthma: a path to precision medicine,” *Epigenetics Insights*, vol. 18, p. e007, 2025.
15. Y. Gautam, E. Johansson, and T. Mersha, “Multi-omics profiling approach to asthma: An evolving paradigm,” *Journal of Personalized Medicine*, vol. 12, no. 1, p. 66, 2022.
16. O. Stegle, L. Parts, R. Durbin, and J. Winn, “A bayesian framework to account for complex non-genetic factors in gene expression levels greatly increases power in eqtl studies,” *PLoS computational biology*, vol. 6, no. 5, p. e1000770, 2010.
17. L. P. Hayden, K. G. Tantisira, S. T. Weiss, and S. Sharma, “The genomic origins of asthma,” *Thorax*, vol. 69, no. 5, pp. 414–422, 2014.