

Ancestry-Associated Pathway Enrichment and Network Architecture in Severe Asthma: Airway Remodelling, Epithelial Barrier and Sex-Adjusted Analysis

Adekunle O. Adeluwoye

adekunle.a@urekabiotec.org

UREKA Biotec/OmicsLogic Africa

Watafua Miriam

UREKA Biotec/OmicsLogic Africa

Lawal Jamiu

UREKA Biotec/OmicsLogic Africa

Katherine O. David

UREKA Biotec/OmicsLogic Africa

Ayanfeoluwa Alabetutu

UREKA Biotec/OmicsLogic Africa

Research Article

Keywords: severe asthma, pathway enrichment, GSEA, ssGSEA, ancestry, airway remodelling, epithelial barrier, steroid response, co-expression network, transcriptomics

Posted Date: October 15th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11370104/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Ancestry-Associated Pathway Enrichment and Network Architecture in Severe Asthma: Airway Remodelling, Epithelial Barrier and Sex-Adjusted Analysis

Adekunle O. Adeluwoye^{1,2*}; Watafua Miriam^{1,3,4}; Lawal Jamiu^{1,5}; Katherine O. David^{1,6}; Ayanfeoluwa Alabetutu^{1,7}

1 Bioinformatics and Big Data Analysis Programmes and Training, UREKA Biotech/OmicsLogic Africa, Ibadan, Nigeria

2 Department of Medical Laboratory Science, KolaDaisi University, Ibadan, Nigeria

3 School of Engineering and Physical Sciences, Heriot-Watt University, Edinburgh, United Kingdom

4 Department of Biochemistry, Faculty of Life Science, University of Maiduguri, Borno State, Nigeria

5 Department of Public Health, MIVA Open University, Abuja, Nigeria

6 Department of Biomedical Technology, Federal University of Technology, Akure, Nigeria

7 Research Institute of Virology, Federal Research Center for Fundamental and Translational Medicine, Novosibirsk, Russia

*Corresponding author: Adekunle O. Adeluwoye; adekunle.a@urekabiotech.org; Department of Medical Laboratory Science, KolaDaisi University, Ibadan, Nigeria

Running title: Pathway and network analysis in severe asthma

Abstract

Background: Systems-level transcriptomic analyses may identify coordinated biological programmes that are not apparent from single-gene testing, but pathway and co-expression results from whole blood are sensitive to sex, cell composition and sample size.

Objective: To characterise ancestry-associated pathway enrichment, sample-level pathway scores and curated-gene co-expression structure in severe asthma.

Methods: We analysed RMA-normalised whole-blood microarray profiles from 24 participants with severe asthma in GSE69683: 12 recorded as black_african and 12 as white_caucasian, with eight females and four males in each group. Gene-level expression was derived from 21,653 unambiguously mapped genes. Pre-ranked GSEA used the sex-adjusted ancestry t statistic across ten curated asthma pathways; ssGSEA scores were compared using sex-adjusted ordinary least squares. Spearman networks among 128 measurable curated genes used a pre-specified $|\rho| \geq 0.60$ and $FDR < 0.05$ threshold.

Results: Pre-ranked GSEA identified Epithelial Barrier Function (NES -1.871, FDR 0.0152) and Airway Remodelling (NES -1.766, FDR 0.0193), both enriched towards the European-ancestry group. Adjusted ssGSEA identified Airway Remodelling (FDR 0.0001), Bronchial Hyperresponsiveness, Innate Immunity/TLR Signalling and Steroid Response (each FDR 0.0348), again with higher mean scores in the European-ancestry group. The full network retained 173 of 8,128 tested pairs and connected 81 genes. COL1A1 and ARRB2 had the highest degree (13 each), while SPDEF had the highest betweenness; subgroup networks were too sparse for reliable comparative network structure claims.

Conclusions: Airway Remodelling was the only pathway supported by both enrichment approaches, while epithelial-barrier, bronchial hyperresponsiveness, innate/TLR and steroid-response findings depended on the analytical question being asked. The co-expression results identify descriptive network-central candidates, but sparse subgroup graphs preclude reliable comparative network inference. Whole-blood cell composition, unavailable treatment outcomes and the small selected cohort constrain interpretation; independent replication in larger ancestry-diverse cohorts with cell-composition and richer clinical data is required.

Keywords: severe asthma; pathway enrichment; GSEA; ssGSEA; ancestry; airway remodelling; epithelial barrier; steroid response; co-expression network; transcriptomics

1. Introduction

Severe asthma is not a single molecular state. Clinical phenotypes overlap with distinct inflammatory and structural programmes, and transcriptomic studies have shown that disease severity, type 2 inflammation, neutrophilic inflammation and corticosteroid response can be represented differently across blood and airway compartments [1-10]. The U-BIOPRED programme was developed in part to address this heterogeneity through deeply characterised severe-asthma cohorts, and both peripheral-blood and airway-tissue analyses have demonstrated that clinically similar patients may carry different molecular patterns [4-8]. These observations make pathway-level analysis attractive because a biological process may shift coherently even when no single transcript dominates the signal.

Ancestry-related differences add another layer of complexity. Asthma morbidity and mortality remain unequally distributed across populations, and African ancestry has been associated with asthma risk and immunological traits in admixed cohorts [11-15]. Race/ethnicity categories recorded in clinical datasets are not identical to genotype-derived ancestry measurements, and gene expression is shaped by genetic regulation, environmental exposure, social conditions and tissue composition [13,16,17]. Systems analyses that compare population groups therefore require careful language: the recorded grouping may be useful for detecting patterns, but it should not be treated as a direct genetic mechanism or as a biological explanation for health disparities.

Gene set approaches ask questions that differ from conventional differential expression. Pre-ranked gene set enrichment analysis (GSEA) tests whether members of a defined pathway accumulate toward one end of a ranked genome-wide statistic, whereas single-sample approaches such as ssGSEA or GSVA-like scoring summarise pathway-associated expression within individual samples [18-20]. These approaches are complementary rather than interchangeable, and disagreement between them can occur because one evaluates population-level rank concentration while the other produces sample-level scores that can be modelled against covariates [18-22]. This distinction is particularly important in small cohorts, where pathway scores can stabilise weakly distributed effects but can also amplify assumptions embedded in the chosen gene sets.

Several asthma programmes are well suited to this type of analysis. Airway remodelling involves extracellular-matrix turnover, transforming growth factor signalling, metalloproteinases and their inhibitors, while epithelial-barrier biology links tight junctions, mucus differentiation and airway responsiveness; MMP9, TGFB1, TIMP2, collagen-related genes, SPDEF and claudin biology are all represented in these processes [23-30]. Although these mechanisms are best established in airway tissues, coordinated expression of related genes in whole blood may still reflect systemic inflammatory states or circulating-cell programmes that track with airway disease.

The curated pathways used here also include genes with direct relevance to bronchial responsiveness and treatment biology. ALOX5 and LTC4S participate in leukotriene synthesis and have been studied in relation to leukotriene-modifier response, while CHRM3 encodes the M3 muscarinic receptor involved in airway smooth-muscle contraction [31-33]. ADRB2 and its regulator ARRB2 are relevant to beta2-agonist signalling, and experimental work shows that beta-arrestin 2 can influence both allergic inflammation and beta2-adrenergic receptor desensitisation [34-37]. Glucocorticoid-response genes and

pathways are similarly heterogeneous: GLCCI1, FKBP5, DUSP1 and related signalling components have been linked to variation in corticosteroid response or steroid-insensitive inflammatory signalling [38-43]. Innate and neutrophilic pathways provide a further interpretive context because CXCL8-rich, neutrophil-associated severe asthma can involve extracellular traps and inflammasome activity [44]. These findings provide biological context for pathway-level associations, but cross-sectional expression data cannot demonstrate a pharmacological response phenotype.

Co-expression analysis provides a further systems-level view by describing genes that vary together across samples. Correlation-network methods have been used to identify asthma modules, exacerbation programmes and candidate regulators in airway and nasal transcriptomes [45-48]. Sex also matters for network and pathway interpretation because peripheral-blood expression contains strong sex-chromosome signals, including XIST, RPS4Y1, DDX3Y and KDM5D, and sex-specific expression effects have been documented in asthma and across human tissues [49-53]. Finally, whole blood is a mixture of leukocyte populations. The Duffy-null ACKR1 variant, common in people of African descent, is associated with lower neutrophil counts, and computational deconvolution studies show how bulk-blood expression can be materially altered by cell proportions [54-57].

We therefore used the GSE69683 severe-asthma cohort to ask two linked questions. First, which of ten curated asthma pathways showed ancestry-associated enrichment by pre-ranked GSEA and which showed sex-adjusted differences in per-sample ssGSEA scores? Second, what descriptive co-expression structure was present among the 128 measurable curated genes, and were the subgroup networks sufficiently dense to justify ancestry- or sex-specific network claims? The primary genome-wide differential-expression counts and curated-gene associations from this cohort are reported in the companion gene-level analysis. The present systems-level analysis uses the same cohort and covariate definitions and focuses on pathway enrichment and network architecture.

2. Materials and Methods

2.1 Study design, data source and cohort

This cross-sectional secondary analysis used publicly available whole-blood expression data from GEO accession GSE69683 generated within U-BIOPRED [4,5]. The selected cohort comprised 24 participants with severe asthma: 12 recorded in GEO as race: black_african and 12 as race: white_caucasian, with eight females and four males in each group. Group labels were parsed programmatically from the GEO sample-characteristics fields and verified against a version-controlled sample manifest. The concise terms African-ancestry group and European-ancestry group are used for these recorded source categories; genetic ancestry was not independently re-estimated in this secondary analysis.

2.2 Expression preprocessing and probe-to-gene mapping

The series was generated on the Affymetrix HT HG-U133+ PM Array Plate (GPL13158; 54,715 probe sets). RMA-normalised, log₂-transformed values were used as supplied; RMA is an established framework for background adjustment, quantile normalisation and probe-set summarisation of Affymetrix data [58]. GPL13158 shares probe-set identifiers with GPL570 but is a distinct array design. For annotation only, the GPL13158 '_PM_' infix was removed and identifiers were cross-mapped to GPL570 annotation. Unambiguous probes were averaged per gene, producing 21,653 gene-level expression profiles. Ambiguous and unmapped probes were retained in audit products.

2.3 Curated pathway definitions

Ten literature-curated pathways represented Immune Response, Airway Remodelling, Oxidative Stress, Bronchial Hyperresponsiveness, Epithelial Barrier Function, Innate Immunity/TLR Signalling, Type 17/Neutrophilic Inflammation, Hypoxia/Metabolic Response, Steroid Response and Drug Metabolism. The pathway source contained 140 pathway-gene memberships, 129 unique genes and 128 genes measurable in the collapsed expression matrix. A single tab-separated source generated both the pathway table and GMT file so that the gene-set definitions used for statistical analysis and reporting remained identical. The curated sets were intentionally compact and hypothesis-directed rather than intended as a comprehensive asthma interactome. The pathway set was not expanded post hoc to a broad pathway library after inspection of the results; such an analysis would test a substantially larger hypothesis space and is better treated as a separate sensitivity analysis.

2.4 Sex-adjusted ancestry ranking and pre-ranked GSEA

At gene level, expression was modelled as $\text{expression} \sim \text{ancestry} + \text{sex}$ by ordinary least squares. The ancestry coefficient represented African minus European expression with sex held constant. Genes were ranked by the ancestry t statistic from this adjusted model. Pre-ranked GSEA was performed with GSEAPy using 1,000 permutations, minimum gene-set size 5, maximum size 500 and random seed 42, following the rank-based enrichment framework of Subramanian and colleagues [18]. Positive normalised enrichment scores (NES) indicated enrichment towards the African-ancestry group and negative scores towards the European-ancestry group. $\text{FDR} < 0.05$ was considered significant.

2.5 ssGSEA and pathway-level modelling

ssGSEA generated one rank-based pathway score per sample using the same curated pathway definitions [19,20]. The primary pathway comparison used $\text{score} \sim \text{ancestry} + \text{sex}$, again reporting the African-minus-European coefficient. Because ssGSEA scores are enrichment statistics rather than expression measurements, between-group coefficients were reported as score differences and not as fold changes. Benjamini-Hochberg correction was applied across the ten pathways [59]. An unadjusted Wilcoxon rank-sum comparison was retained as a secondary sensitivity analysis because it cannot incorporate the sex covariate.

2.6 Co-expression network analysis

Spearman correlations were calculated across all 8,128 unique pairs among the 128 measurable curated genes after removal of zero-variance features. Edges were retained only when $|\text{rho}| \geq 0.60$ and $\text{FDR} < 0.05$. The same pre-specified threshold was applied to the full cohort and to ancestry- and sex-defined subgroups. Degree and betweenness centrality were exported for every node, allowing network prominence to be determined from machine-readable metrics rather than visual appearance. Because subgroup sizes ranged from eight to 16 and the ancestry groups each contained 12 participants, subgroup networks were treated as exploratory and were not used for reliable comparative network structure claims. The interpretation follows the general principle that correlation networks describe co-variation and require additional evidence before causal or regulatory language is justified [45].

2.7 Reproducibility and analysis provenance

All numerical results, figures and summary counts reported here derive from a single fully executed notebook

Cross_Ancestry_Asthma_Transcriptomics_PUBLICATION_MASTER_EXECUTED_v1.ipynb,

whose recorded execution began on 30 July 2026 at 14:27:12 UTC. The run used GSE69683 on GPL13158, random seed 42, GSEA minimum set size 5 and 1,000 permutations; it indexed 72 generated files and passed all programmed validation checks. Machine-readable GSEA, ssGSEA, network-pair, edge, node, centrality and checksum outputs were written to persistent repository directories.

3. Results

3.1 Pre-ranked GSEA identifies epithelial-barrier and airway-remodelling enrichment

Two of the ten curated pathways met $FDR < 0.05$ in the sex-adjusted pre-ranked analysis (Table 1). Epithelial Barrier Function showed the strongest enrichment (NES -1.871, FDR 0.0152), with CLDN18, SPDEF, MUC5AC, FLG and CLDN1 contributing to the leading edge. Airway Remodelling was also enriched towards the European-ancestry group (NES -1.766, FDR 0.0193), with MMP9, TIMP2, TGFB1, PDGFB, COL1A1, TGFBR1, TGFBR2, COL3A1, TIMP1, PDGFA, MMP12 and FGF2 represented in the leading edge. The remaining eight pathways did not reach the FDR threshold. Bronchial Hyperresponsiveness, Steroid Response and Innate Immunity/TLR Signalling nevertheless showed negative NES values, whereas Hypoxia/Metabolic Response and Immune Response showed positive NES values, but these directions were not statistically significant.

Table 1. Pre-ranked GSEA across ten curated asthma pathways.

Pathway	NES	FDR	Leading-edge genes	Enrichment direction
Epithelial barrier function	-1.871	0.0152	CLDN18; SPDEF; MUC5AC; FLG; CLDN1	European
Airway remodelling	-1.766	0.0193	MMP9; TIMP2; TGFB1; PDGFB; COL1A1; TGFBR1; TGFBR2; COL3A1; TIMP1; PDGFA; MMP12; FGF2	European
Hypoxia/metabolic response	+1.407	0.2854	LDHA; SLC2A1	African
Bronchial hyperresponsiveness	-1.377	0.2476	ALOX5; LTC4S; CHRM3; EDNRA	European
Steroid response	-1.338	0.2219	ZBTB16; DUSP1	European
Innate immunity/TLR signalling	-1.306	0.2078	CXCL8; TLR2; MYD88	European
Immune response	+1.071	0.5561	CD28; GATA3; CTLA4; ICOS; STAT4	African
Drug metabolism	-0.898	0.6986	ARRB2; NAT2; DUSP1; HSD17B1; SLCO1B1; CYP2C19; others	European
Oxidative stress	+0.708	0.8533	PRDX1; GCLC; NOX1; GCLM	African
Type 17/neutrophilic inflammation	-0.647	0.9035	CXCL1	European

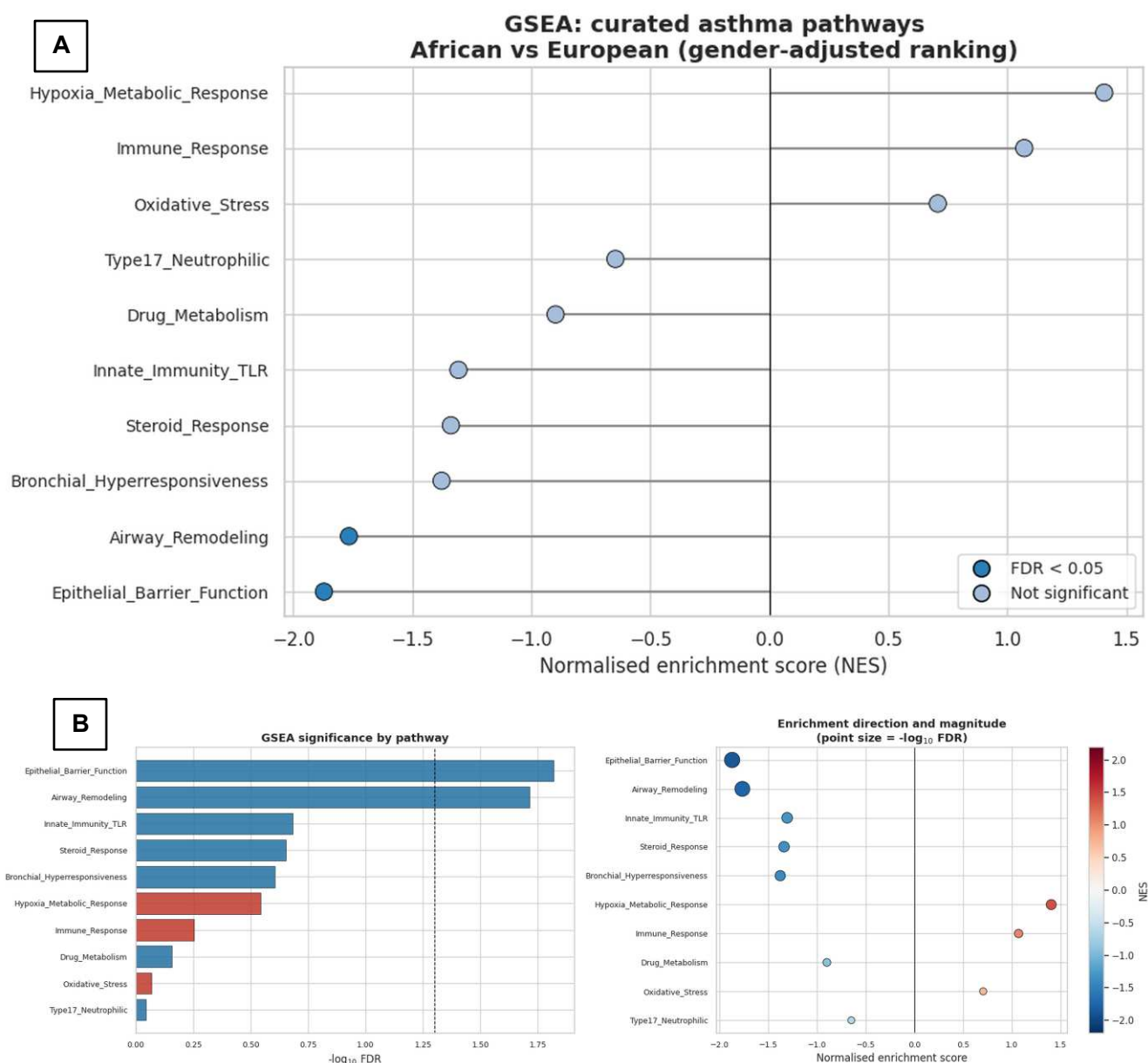


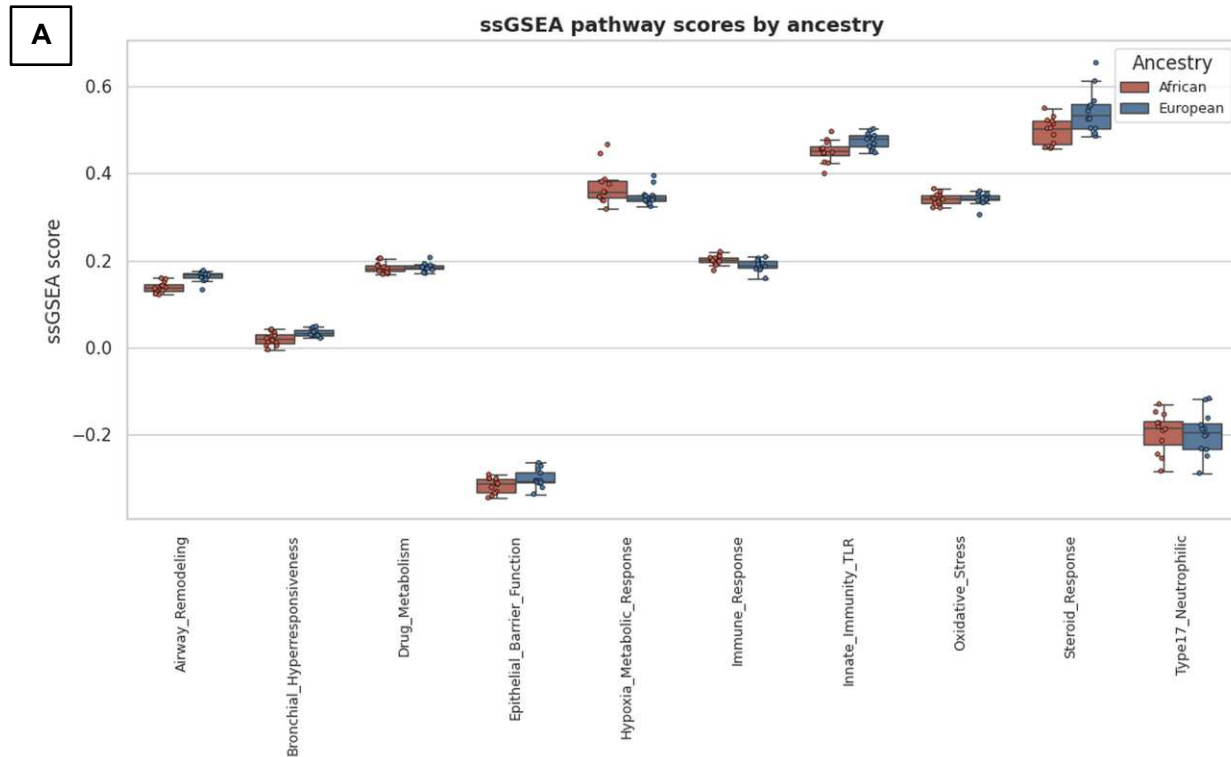
Figure 1. Pre-ranked GSEA across ten curated asthma pathways. (A) Lollipop plot of normalised enrichment scores from the sex-adjusted ancestry ranking. (B) Significance display of enrichment direction and FDR. Negative NES indicates enrichment towards the European-ancestry group.

3.2 Sex-adjusted ssGSEA identifies four pathway-score differences

The sample-level analysis produced a different but overlapping result (Table 2). Airway Remodelling showed the largest adjusted difference, with a lower mean score in the African-ancestry group (African minus European -0.0253 ; FDR 0.0001). Bronchial Hyperresponsiveness (-0.0145), Innate Immunity/TLR Signalling (-0.0250) and Steroid Response (-0.0450) were also lower in the African-ancestry group and each remained significant after correction (FDR 0.0348). Immune Response and Epithelial Barrier Function were nominally associated but did not survive the ten-pathway correction (FDR 0.0530 each). Airway Remodelling was therefore the only pathway supported by both pre-ranked GSEA and adjusted ssGSEA.

Table 2. Sex-adjusted ssGSEA pathway-score differences.

Pathway	African - European score difference	P value	FDR	Significant	Higher score
Airway remodelling	-0.0253	<0.0001	0.0001	Yes	European
Bronchial hyperresponsiveness	-0.0145	0.0090	0.0348	Yes	European
Innate immunity/TLR signalling	-0.0250	0.0139	0.0348	Yes	European
Steroid response	-0.0450	0.0138	0.0348	Yes	European
Immune response	+0.0118	0.0318	0.0530	No	African
Epithelial barrier function	-0.0158	0.0311	0.0530	No	European
Hypoxia/metabolic response	+0.0232	0.1001	0.1430	No	African
Drug metabolism	-0.0015	0.7437	0.9266	No	European
Oxidative stress	-0.0007	0.9051	0.9266	No	European
Type 17/neutrophilic inflammation	+0.0019	0.9266	0.9266	No	African



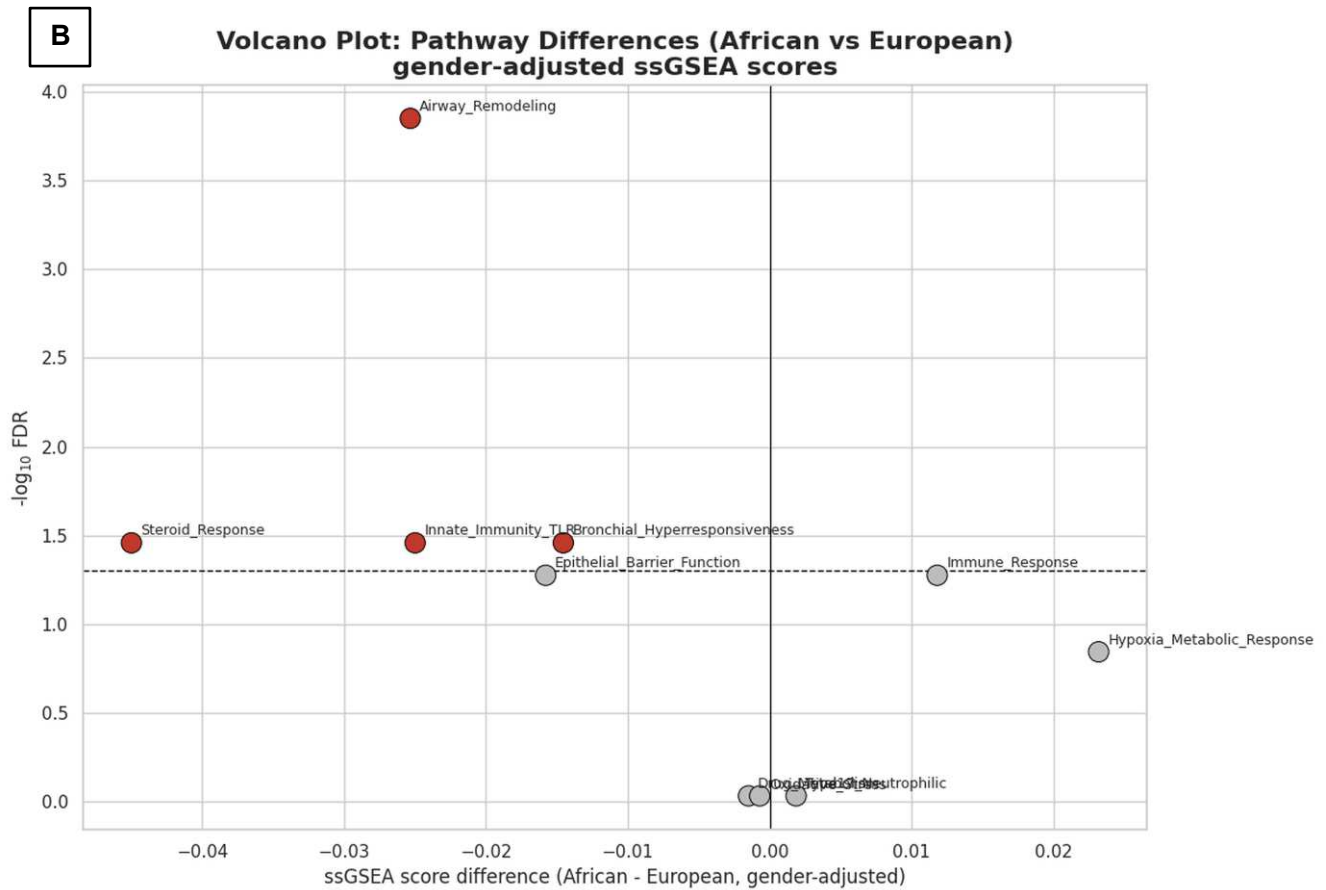
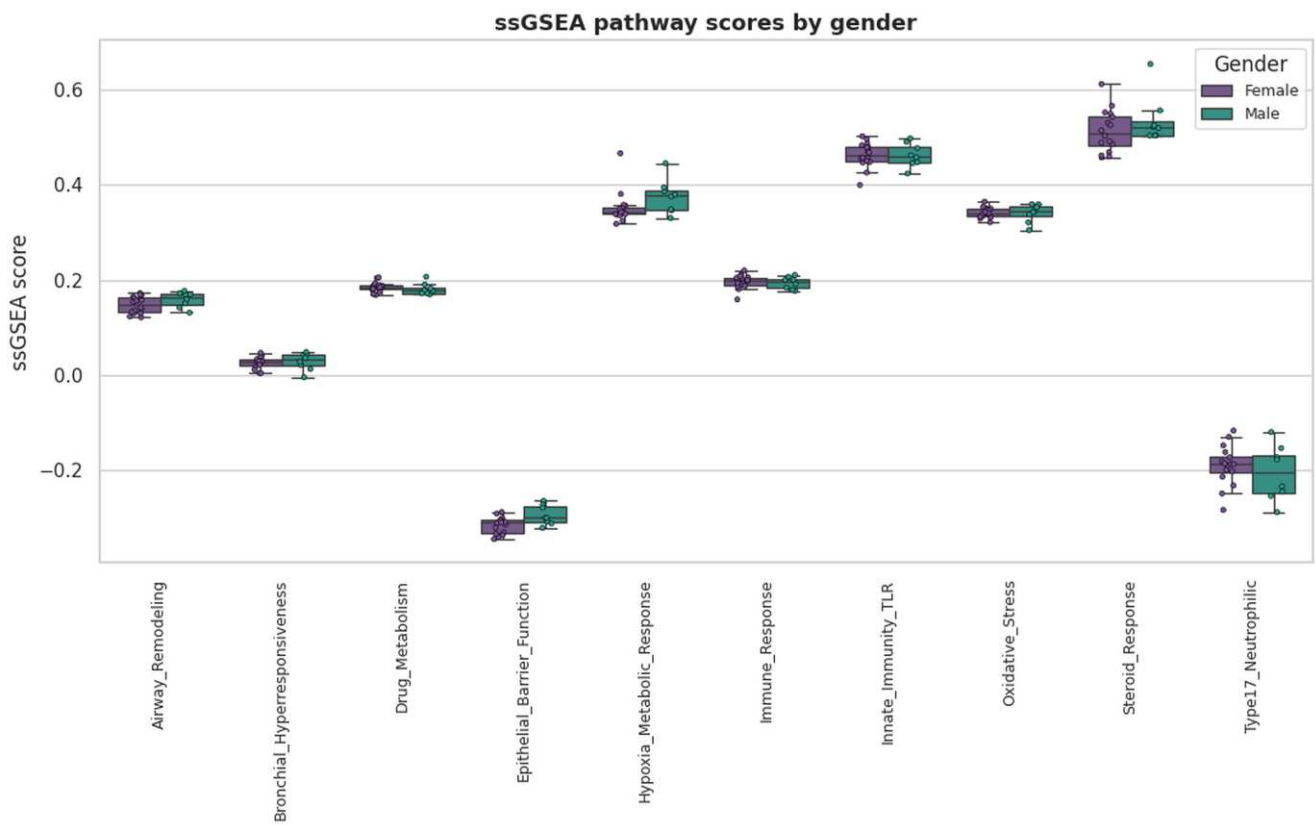


Figure 2. Per-sample pathway activity. (A) ssGSEA score distributions by ancestry and sex. (B) Volcano plot of adjusted African-minus-European ssGSEA score differences. The x-axis represents a pathway-score difference, not a fold change.

3.3 The full-cohort network contains informative centrality structure, whereas subgroup networks are sparse

The full-cohort network retained 173 of 8,128 tested gene pairs and connected 81 of 128 curated genes (Table 3). The strongest degree centrality was shared by COL1A1 and ARRB2 (degree 13 each), followed by SPDEF (12), CD28 and LDHA (11 each), and ALOX5 (10). Betweenness centrality was highest for SPDEF (0.0471), followed by COL1A1 (0.0338), CD28 (0.0320), SMAD2 (0.0316) and SMAD4 (0.0300); MMP9 and CXCL1 also ranked within the top ten by betweenness. ARRB2 was tied for first by degree and ranked ninth by betweenness, while ALOX5 ranked sixth by degree and twelfth by betweenness. These measures describe network position rather than causal regulatory control.

Table 3. Co-expression network summary under the fixed threshold.

Group	Samples	Genes	Pairs tested	Edges retained	Connected nodes / interpretation
All samples	24	128	8,128	173	81
African ancestry	12	128	8,128	2	2-4 (exploratory)
European ancestry	12	128	8,128	0	0
Female	16	128	8,128	1	Exploratory
Male	8	128	8,128	8	Exploratory

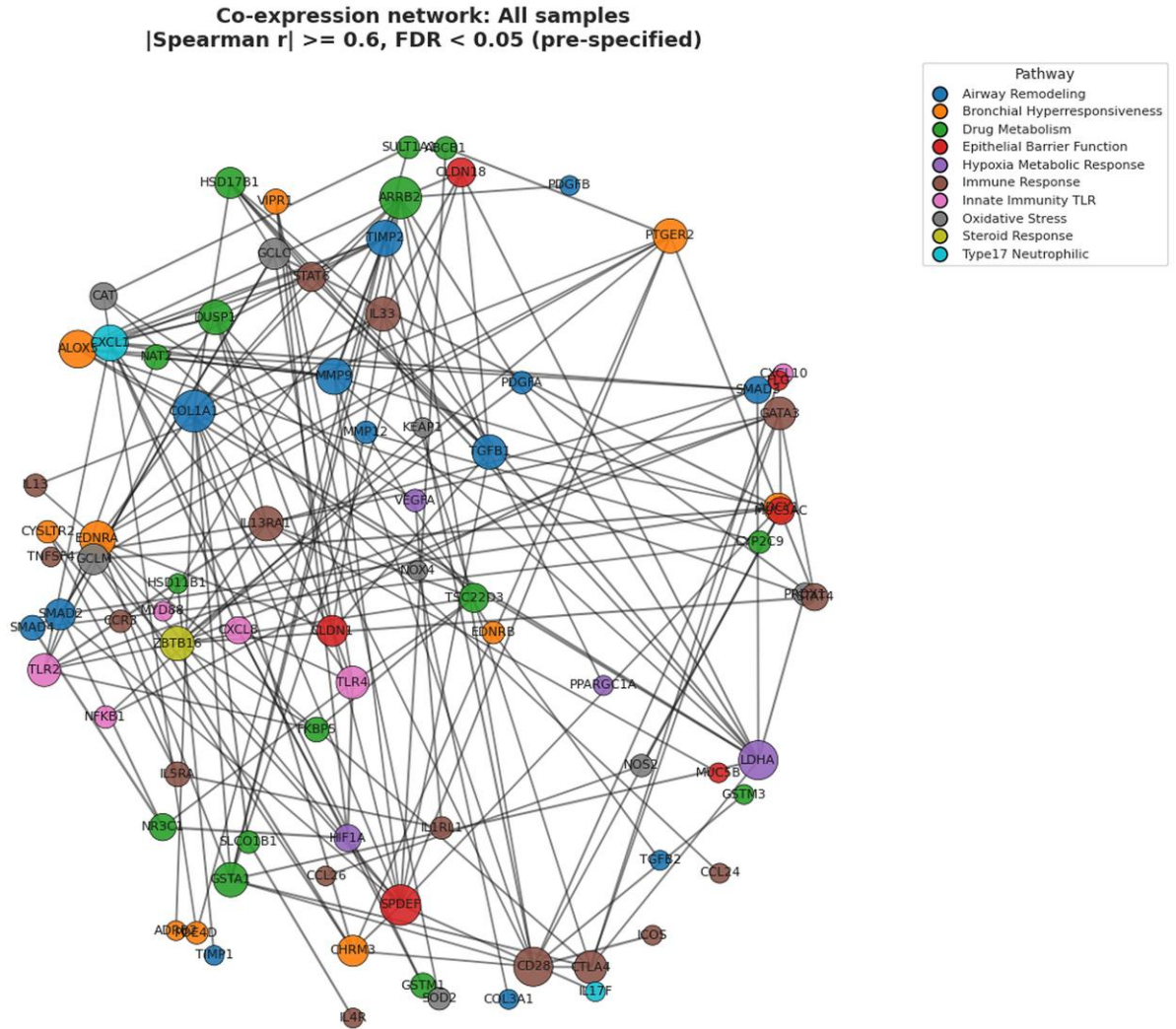


Figure 3. Full-cohort curated-gene co-expression network. Nodes represent the 128 measurable curated genes and edges represent Spearman correlations retained at $|\rho| \geq 0.60$ and $FDR < 0.05$. Node colour indicates curated pathway membership and node size indicates degree. Hub and bridge statements are based on the machine-readable degree and betweenness tables rather than visual prominence.

3.4 Subgroup networks do not support comparative network claims

Applying the same fixed threshold to smaller subgroups yielded only two retained edges in the African-ancestry group, none in the European-ancestry group, one among females and eight among males. The difference in edge counts was therefore dominated by very sparse graphs and small sample sizes rather than by a sufficiently stable network structure for biological comparison. Subgroup networks are retained in Supplementary Figure S1 for transparency, but no ancestry-specific or sex-specific rewiring claim is made.

4. Discussion

4.1 Convergent evidence for airway remodelling

The clearest systems-level finding is not that every pathway moved in the same direction, but that Airway Remodelling was detected by two analytically different approaches. The GSEA result reflects concentration of remodelling genes toward the European end of the sex-adjusted genome-wide ranking, whereas the ssGSEA result reflects lower sample-level remodelling scores in the African-ancestry group. Convergence across these questions gives the remodelling signal more interpretive weight than pathways detected by only one method, although it remains a blood-derived association rather than direct evidence of airway structural change. The leading-edge genes are biologically credible: MMP9, TIMP2 and TGFB1 participate in extracellular-matrix turnover and fibrotic signalling, and collagen and growth-factor genes have established roles in airway remodelling [23-25,29,30]. The finding is also compatible with prior severe-asthma transcriptomic work showing that molecular heterogeneity extends beyond canonical type 2 inflammation [5,6,8].

4.2 Epithelial-barrier enrichment and the GSEA-ssGSEA tension

The epithelial-barrier result illustrates why the two enrichment methods should not be treated as interchangeable. Epithelial Barrier Function reached FDR significance in pre-ranked GSEA but not in adjusted ssGSEA, even though its ssGSEA coefficient was in the same European-higher direction and was nominally associated before correction. This is not a contradiction: GSEA is sensitive to coordinated positioning of pathway genes in a ranked statistic, while ssGSEA converts each sample into a pathway score and then asks whether those scores differ after covariate adjustment [18-22]. The leading-edge genes, including CLDN18, SPDEF, MUC5AC, FLG and CLDN1, connect plausibly to epithelial integrity, mucus differentiation and barrier function [26-28]. The important limitation is anatomical: these transcripts were measured in whole blood, so the result may represent circulating correlates of epithelial disease rather than epithelial-cell expression itself.

4.3 Bronchial hyperresponsiveness, innate/TLR and steroid-response associations

The three additional ssGSEA associations warrant cautious attention because each captures a clinically relevant programme but none reached significance by pre-ranked GSEA. Bronchial Hyperresponsiveness includes ALOX5, LTC4S, CHRM3 and EDNRA, linking leukotriene synthesis and cholinergic smooth-muscle biology to the observed score difference [31-33]. Innate Immunity/TLR Signalling includes CXCL8, TLR2 and MYD88; a lower mean score in the African-ancestry group is therefore not equivalent to 'less innate inflammation', particularly in bulk blood where neutrophil abundance can strongly influence the measured signal. Severe neutrophilic asthma has been associated with CXCL8-related inflammatory activity, extracellular DNA, neutrophil extracellular traps and inflammasome activation [9,10,42,44]. The Steroid Response pathway likewise contains genes such as ZBTB16, DUSP1, NR3C1 and FKBP5, but a pathway-score difference cannot be interpreted as corticosteroid sensitivity because actual dose, adherence, longitudinal response and pharmacodynamic measurements were unavailable [38,41-43].

4.4 Potential pharmacogenomic leads should remain distinct from treatment-response claims

Several genes occupying these pathways are pharmacologically relevant, but the present data support prioritisation rather than clinical inference. Genetic studies have linked ALOX5 and LTC4S variation to leukotriene-modifier response and ADRB2 variation to heterogeneity in beta2-agonist outcomes, while

mechanistic studies place beta-arrestin 2 within both allergic inflammation and beta2-adrenergic receptor regulation [31,32,34-37,39,40]. In the companion gene-level analysis, ARRB2, ALOX5 and LTC4S differed between the ancestry groups; in the present network, ARRB2 was tied for the highest degree and ALOX5 was among the more highly connected nodes. These observations make ARRB2 and ALOX5 reasonable candidates for treatment-linked follow-up, but they do not demonstrate ancestry-specific drug response. This distinction is reinforced by the absence of significant GSEA or ssGSEA enrichment for the broad Drug Metabolism pathway and by the lack of treatment-outcome data.

4.5 Network centrality is informative at full-cohort scale but not within small subgroups

The full-cohort graph supports a descriptive view of inter-pathway co-variation. COL1A1 and ARRB2 had the highest degree, while SPDEF had the highest betweenness, followed by COL1A1, CD28, SMAD2 and SMAD4; MMP9 and CXCL1 also ranked highly by betweenness. These rankings are consistent with the broader use of co-expression networks to identify connected molecular programmes and candidate regulators in asthma [45-48], but centrality does not establish that a gene is a causal driver. The subgroup results make this distinction especially important. With only two retained edges in the African-ancestry network and none in the European-ancestry network, visual differences between subgroup plots are not sufficient for reliable network structure claims; larger samples, bootstrap or resampling-based stability analysis, and external validation would be needed before comparing network organisation across groups.

4.6 Whole-blood composition and interpretation of the recorded ancestry groups

Whole-blood transcriptomics cannot separate cell-intrinsic regulation from differences in circulating cell abundance without additional information. This is particularly relevant here because the Duffy-null ACKR1 variant is common in people of African descent and is strongly associated with lower neutrophil counts [54,55]. Bulk-tissue deconvolution methods demonstrate that variation in cell proportions can materially alter gene-expression summaries and downstream association signals [56,57]. Consequently, ancestry-associated pathway scores involving CXCL8, MMP9, ALOX5 or other myeloid-linked genes may partly reflect composition rather than per-cell transcriptional regulation. The grouping variable follows the race/ethnicity categories recorded in the source GEO metadata, and genetic ancestry was not independently re-estimated in this secondary analysis. The observed associations should therefore be interpreted as differences between the recorded study groups rather than direct genetic effects of ancestry.

4.7 Strengths and limitations

A strength of this analysis is the use of one version-controlled cohort, one sex-adjusted statistical framework, programmatically linked pathway definitions, fixed network thresholds and machine-readable output manifests across GSEA, ssGSEA and network analyses. The principal limitation is the small selected sample of 24 participants, compounded by unmeasured whole-blood cell composition, incomplete clinical covariates and the absence of independent genetic-ancestry re-estimation in this secondary analysis, which limits power and prevents separation of cell abundance, environmental or social exposure, treatment and cell-intrinsic expression effects. The pathway analysis was deliberately targeted rather than exhaustive, treatment-response measurements were unavailable, and the eight- to 16-sample subgroup networks were too sparse for reliable comparative network inference; broader gene-

set sensitivity analysis and independent replication would strengthen future work, while the present pre-specified pathway analysis is interpreted within its stated scope.

4.8 Implications and next studies

The next study should test whether the remodelling signal persists after accounting for leukocyte composition, richer clinical treatment variables and independent ancestry characterisation where available in a substantially larger cohort. Replication should also include airway-derived material, because whole-blood and airway transcriptomes capture different biological compartments, and should pre-specify whether the objective is pathway discovery, clinical stratification or treatment-response prediction. Asthma phenotypes are further modified by clinical factors such as obesity and age of onset [60], reinforcing the need for richer covariate data rather than ancestry-only comparisons. Current international guidance also emphasises the heterogeneity of asthma and the need to relate biological findings to clinically meaningful phenotypes and treatment response [61]. A broader MSigDB/Reactome-style sensitivity analysis is technically feasible from the existing sex-adjusted 21,653-gene ranking, but it would introduce a new post hoc hypothesis space and is better treated as a separate analysis. For the network analysis specifically, resampling-based stability estimates and an external cohort should precede any claim that centrality or module structure differs by ancestry or sex.

5. Conclusion

In this severe-asthma whole-blood cohort, the most consistent systems-level difference between the recorded ancestry groups involved Airway Remodelling, which was supported by both pre-ranked GSEA and sex-adjusted ssGSEA. Epithelial-barrier, bronchial-hyperresponsiveness, innate/TLR and steroid-response findings depended on the analytical question and should be interpreted as complementary exploratory evidence rather than interchangeable confirmation. The full-cohort co-expression network identified central genes including COL1A1, ARRB2 and SPDEF, but sparse subgroup graphs precluded reliable ancestry- or sex-specific network comparisons. Whole-blood composition and incomplete clinical covariates remain important sources of uncertainty, and these findings require independent replication in larger ancestry-diverse, cell-resolved and treatment-linked cohorts before mechanistic or clinical-response claims are made.

Declarations

Data availability. The expression data analysed in this study are publicly available from NCBI Gene Expression Omnibus under accession GSE69683 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE69683>). The version-controlled sample manifest, analysis notebook, curated pathway definitions, machine-readable GSEA and ssGSEA results, co-expression network tables, figures, validation files and checksums will be made available at <https://github.com/UREKA-Biotec-OmicsLogic-Africa/cross-ancestry-asthma-transcriptomics-public> and archived at <https://doi.org/10.5281/zenodo.23227701>.

Ethical approval. This study is a secondary analysis of publicly available de-identified data and required no additional ethical approval. The original U-BIOPRED study obtained ethical approvals at participating sites.

Informed consent. Informed consent was obtained by the original U-BIOPRED investigators from study participants in accordance with the approved study protocols. No new participant contact or data collection occurred in the present secondary analysis.

Funding. No dedicated funding was received for this analysis. The analyses were conducted using the publicly available GSE69683 dataset generated by the U-BIOPRED study. The U-BIOPRED consortium was supported by its own funding sources, as detailed in the original publications. No additional financial support was obtained for this secondary analysis.

Competing interests. The authors declare no competing interests.

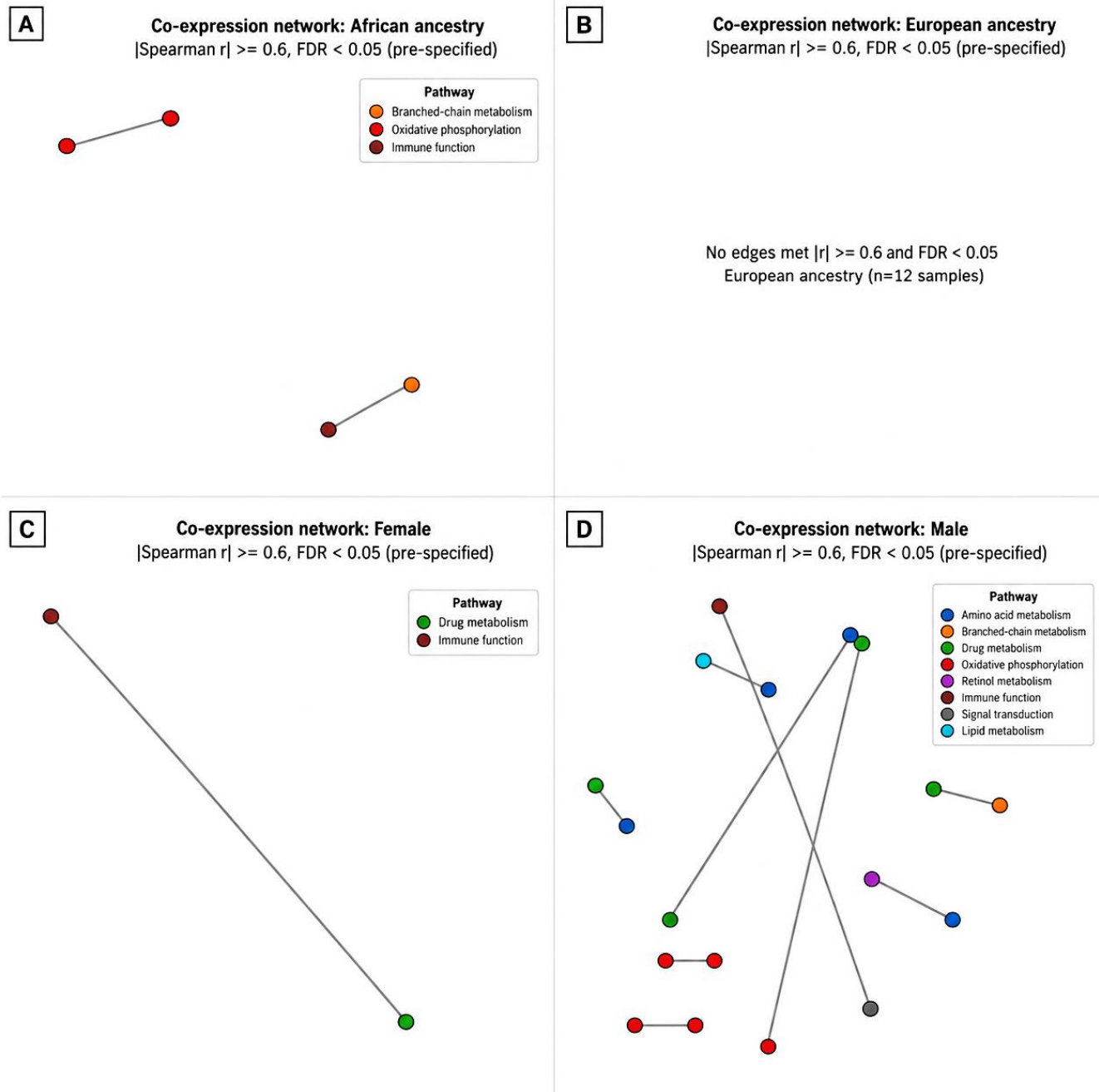
Author contributions. A.O.A: Conceptualisation, Methodology, Supervision, Project administration, Formal analysis, Interpretation, Writing - original draft, Writing - review and editing. W.M: Investigation, Data curation, Formal analysis, Visualisation, Writing - original draft, Writing - review and editing. K.D: Validation, Literature curation, Visualisation, Writing - review and editing. L.J: Validation, Reproducibility checks, Writing - review and editing. A.A: Methodology, Software, Validation, Formal analysis, Writing - review and editing. All authors read the final manuscript and approved it before submission.

Acknowledgements. We thank the U-BIOPRED participants and investigators and the original study authors for making the data publicly available.

AI use disclosure. Generative artificial intelligence tools were employed solely to assist with language editing, grammar refinement, and structural suggestions during manuscript preparation. All data analyses, numerical results, and scientific interpretations were conducted independently by the authors. The authors affirm that no AI system was used to generate, manipulate, or alter primary data, statistical analyses, or scientific conclusions. Responsibility for the accuracy, integrity, and final content of the manuscript rests entirely with the authors.

Supplementary material

- Supplementary Table S1: complete pre-ranked GSEA results, leading-edge genes and ranked-gene output.
- Supplementary Table S2: sample-by-pathway ssGSEA score matrix, sex-adjusted OLS results and Wilcoxon comparison.
- Supplementary Table S3: full-cohort and subgroup correlation-pair, edge, node-centrality and network-summary files, including ranked degree and betweenness results.



- Supplementary Figure S1: subgroup co-expression network grid for African-ancestry, European-ancestry, female and male groups. The same fixed threshold was applied in every subgroup; these graphs are descriptive and are not interpreted as evidence of ancestry- or sex-specific network organisation.

References

1. Wenzel SE. Asthma phenotypes: the evolution from clinical to molecular approaches. *Nat Med.* 2012;18(5):716-725. doi:10.1038/nm.2678.
2. Lambrecht BN, Hammad H. The immunology of asthma. *Nat Immunol.* 2015;16(1):45-56. doi:10.1038/ni.3049.

3. Israel E, Reddel HK. Severe and difficult-to-treat asthma in adults. *N Engl J Med*. 2017;377(10):965-976. doi:10.1056/NEJMr1608969.
4. Shaw DE, Sousa AR, Fowler SJ, Fleming LJ, Roberts G, Corfield J, et al. Clinical and inflammatory characteristics of the European U-BIOPRED adult severe asthma cohort. *Eur Respir J*. 2015;46(5):1308-1321. doi:10.1183/13993003.00779-2015.
5. Bigler J, Boedigheimer M, Schofield JPR, Skipp PJ, Corfield J, Rowe A, et al. A severe asthma disease signature from gene expression profiling of peripheral blood from U-BIOPRED cohorts. *Am J Respir Crit Care Med*. 2017;195(10):1311-1320. doi:10.1164/rccm.201604-0866OC.
6. Kuo CHS, Pavlidis S, Loza M, Baribaud F, Rowe A, Pandis I, et al. A transcriptome-driven analysis of epithelial brushings and bronchial biopsies to define asthma phenotypes in U-BIOPRED. *Am J Respir Crit Care Med*. 2017;195(4):443-455. doi:10.1164/rccm.201512-2452OC.
7. Woodruff PG, Modrek B, Choy DF, Jia G, Abbas AR, Ellwanger A, et al. T-helper type 2-driven inflammation defines major subphenotypes of asthma. *Am J Respir Crit Care Med*. 2009;180(5):388-395. doi:10.1164/rccm.200903-0392OC.
8. Modena BD, Bleecker ER, Busse WW, Erzurum SC, Gaston BM, Jarjour NN, et al. Gene expression correlated with severe asthma characteristics reveals heterogeneous mechanisms of severe disease. *Am J Respir Crit Care Med*. 2017;195(11):1449-1463. doi:10.1164/rccm.201607-1407OC.
9. Hinks TSC, Levine SJ, Brusselle GG. Treatment options in type-2 low asthma. *Eur Respir J*. 2021;57(1):2000528. doi:10.1183/13993003.00528-2020.
10. Simpson JL, Scott R, Boyle MJ, Gibson PG. Inflammatory subtypes in asthma: assessment and identification using induced sputum. *Respirology*. 2006;11(1):54-61. doi:10.1111/j.1440-1843.2006.00784.x.
11. Akinbami LJ, Moorman JE, Bailey C, Zahran HS, King M, Johnson CA, et al. Trends in asthma prevalence, health care use, and mortality in the United States, 2001-2010. *NCHS Data Brief*. 2012;(94):1-8. PMID:22617340.
12. Celedon JC, Roman J, Schraufnagel DE, Thomas A, Samet J. Respiratory health equality in the United States: the American Thoracic Society perspective. *Ann Am Thorac Soc*. 2014;11(4):473-479. doi:10.1513/AnnalsATS.201402-059PS.
13. Oh SS, Galanter J, Thakur N, Pino-Yanes M, Barcelo NE, White MJ, et al. Diversity in clinical and biomedical research: a promise yet to be fulfilled. *PLoS Med*. 2015;12(12):e1001918. doi:10.1371/journal.pmed.1001918.
14. Flores C, Ma SF, Pino-Yanes M, Wade MS, Perez-Mendez L, Kittles RA, et al. African ancestry is associated with asthma risk in African Americans. *PLoS One*. 2012;7(1):e26807. doi:10.1371/journal.pone.0026807.
15. Vergara C, Murray T, Rafaels N, Lewis R, Campbell M, Foster C, et al. African ancestry is a risk factor for asthma and high total IgE levels in African admixed populations. *Genet Epidemiol*. 2013;37(4):393-401. doi:10.1002/gepi.21702.
16. Gamazon ER, Segre AV, van de Bunt M, Wen X, Xi HS, Hormozdiari F, et al. Using an atlas of gene regulation across 44 human tissues to inform complex disease- and trait-associated variation. *Nat Genet*. 2018;50(7):956-967. doi:10.1038/s41588-018-0154-4.
17. Van Driest SL, Abul-Husn NS, Glessner JT, Bastarache L, Nirenberg S, Schildcrout JS, et al. Association between a common, benign genotype and unnecessary bone marrow biopsies among African American patients. *JAMA Intern Med*. 2021;181(8):1100-1105. doi:10.1001/jamainternmed.2021.3108.
18. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. 2005;102(43):15545-15550. doi:10.1073/pnas.0506580102.

19. Barbie DA, Tamayo P, Boehm JS, Kim SY, Moody SE, Dunn IF, et al. Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature*. 2009;462(7269):108-112. doi:10.1038/nature08460.
20. Hanzelmann S, Castelo R, Guinney J. GSEA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics*. 2013;14:7. doi:10.1186/1471-2105-14-7.
21. Reimand J, Isserlin R, Voisin V, Kucera M, Tannus-Lopes C, Rostamianfar A, et al. Pathway enrichment analysis and visualization of omics data using g:Profiler, GSEA, Cytoscape and EnrichmentMap. *Nat Protoc*. 2019;14(2):482-517. doi:10.1038/s41596-018-0103-9.
22. Liberzon A, Birger C, Thorvaldsdottir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database hallmark gene set collection. *Cell Syst*. 2015;1(6):417-425. doi:10.1016/j.cels.2015.12.004.
23. Ojiaku CA, Yoo EJ, Panettieri RA Jr. Transforming growth factor beta-1 function in airway remodelling and hyperresponsiveness: the missing link? *Am J Respir Cell Mol Biol*. 2017;56(4):432-442. doi:10.1165/rcmb.2016-0307TR.
24. Ohbayashi H, Shimokata K. Matrix metalloproteinase-9 and airway remodeling in asthma. *Curr Drug Targets Inflamm Allergy*. 2005;4(2):177-181. doi:10.2174/1568010053586246.
25. Cho JY, Miller M, McElwain K, McElwain S, Shim JY, Raz E, et al. Remodeling associated expression of matrix metalloproteinase 9 but not tissue inhibitor of metalloproteinase 1 in airway epithelium: modulation by immunostimulatory DNA. *J Allergy Clin Immunol*. 2006;117(3):618-625. doi:10.1016/j.jaci.2005.12.1324.
26. Lambrecht BN, Hammad H. The airway epithelium in asthma. *Nat Med*. 2012;18(5):684-692. doi:10.1038/nm.2737.
27. Park KS, Korfhagen TR, Bruno MD, Kitzmiller JA, Wan H, Wert SE, et al. SPDEF regulates goblet cell hyperplasia in the airway epithelium. *J Clin Invest*. 2007;117(4):978-988. doi:10.1172/JCI29176.
28. Sweerus K, Lachowicz-Scroggins M, Gordon E, LaFemina M, Huang X, Parikh M, et al. Claudin-18 deficiency is associated with airway epithelial barrier dysfunction and asthma. *J Allergy Clin Immunol*. 2017;139(1):72-81.e1. doi:10.1016/j.jaci.2016.02.035.
29. Hough KP, Curtiss ML, Blain TJ, Liu RM, Trevor J, Deshane JS, et al. Airway remodeling in asthma. *Front Med (Lausanne)*. 2020;7:191. doi:10.3389/fmed.2020.00191.
30. Boxall C, Holgate ST, Davies DE. The contribution of transforming growth factor-beta and epidermal growth factor signalling to airway remodelling in chronic asthma. *Eur Respir J*. 2006;27(1):208-229. doi:10.1183/09031936.06.00130004.
31. Drazen JM, Yandava CN, Dube L, Szczerback N, Hippensteel R, Pillari A, et al. Pharmacogenetic association between ALOX5 promoter genotype and the response to anti-asthma treatment. *Nat Genet*. 1999;22(2):168-170. doi:10.1038/9680.
32. Lima JJ, Zhang S, Grant A, Shao L, Tantisira KG, Allayee H, et al. Influence of leukotriene pathway polymorphisms on response to montelukast in asthma. *Am J Respir Crit Care Med*. 2006;173(4):379-385. doi:10.1164/rccm.200509-1412OC.
33. Donfack J, Kogut P, Forsythe S, Solway J, Ober C. Sequence variation in the promoter region of the cholinergic receptor muscarinic 3 gene and asthma and atopy. *J Allergy Clin Immunol*. 2003;111(3):527-532. doi:10.1067/mai.2003.71.
34. Bleecker ER, Postma DS, Lawrance RM, Meyers DA, Ambrose HJ, Goldman M. Effect of ADRB2 polymorphisms on response to longacting beta2-agonist therapy: a pharmacogenetic analysis of two randomised studies. *Lancet*. 2007;370(9605):2118-2125. doi:10.1016/S0140-6736(07)61906-0.
35. Wechsler ME, Kunselman SJ, Chinchilli VM, Bleecker E, Boushey HA, Calhoun WJ, et al. Effect of beta2-adrenergic receptor polymorphism on response to longacting beta2 agonist in asthma

- (LARGE trial): a genotype-stratified, randomised, placebo-controlled, crossover trial. *Lancet*. 2009;374(9703):1754-1764. doi:10.1016/S0140-6736(09)61492-6.
36. Walker JKL, Fong AM, Lawson BL, Savov JD, Patel DD, Schwartz DA, et al. Beta-arrestin-2 regulates the development of allergic asthma. *J Clin Invest*. 2003;112(4):566-574. doi:10.1172/JCI17265.
 37. Deshpande DA, Theriot BS, Penn RB, Walker JKL. Beta-arrestins specifically constrain beta2-adrenergic receptor signaling and function in airway smooth muscle. *FASEB J*. 2008;22(7):2134-2141. doi:10.1096/fj.07-102459.
 38. Tantisira KG, Lasky-Su J, Harada M, Murphy A, Litonjua AA, Himes BE, et al. Genomewide association between GLCCI1 and response to glucocorticoid therapy in asthma. *N Engl J Med*. 2011;365(13):1173-1183. doi:10.1056/NEJMoa0911353.
 39. Lima JJ, Blake KV, Tantisira KG, Weiss ST. Pharmacogenetics of asthma. *Curr Opin Pulm Med*. 2009;15(1):57-62. doi:10.1097/MCP.0b013e32831da8be.
 40. Garcia-Menaya JM, Cordobes-Duran C, Garcia-Martin E, Agundez JAG. Pharmacogenetic factors affecting asthma treatment response: potential implications for drug therapy. *Front Pharmacol*. 2019;10:520. doi:10.3389/fphar.2019.00520.
 41. Alsaffar SF, Alqaraghuli HA, Yenzeel JH, Ghazi HF. The association of FKBP5 polymorphism with asthma susceptibility in asthmatic patients. *J Basic Clin Physiol Pharmacol*. 2021;32(4):479-484. doi:10.1515/jbcpp-2020-0450.
 42. Robins S, Roussel L, Schachter A, Risse PA, Mogas AK, Olivenstein R, et al. Steroid-insensitive ERK1/2 activity drives CXCL8 synthesis and neutrophilia by airway smooth muscle. *Am J Respir Cell Mol Biol*. 2011;45(5):984-990. doi:10.1165/rcmb.2010-0450OC.
 43. Thomson NC. Addressing corticosteroid insensitivity in adults with asthma. *Expert Rev Respir Med*. 2016;10(2):137-156. doi:10.1586/17476348.2016.1133304.
 44. Lachowicz-Scroggins ME, Dunican EM, Charbit AR, Raymond W, Looney MR, Peters MC, et al. Extracellular DNA, neutrophil extracellular traps, and inflammasome activation in severe asthma. *Am J Respir Crit Care Med*. 2019;199(9):1076-1085. doi:10.1164/rccm.201810-1869OC.
 45. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*. 2008;9:559. doi:10.1186/1471-2105-9-559.
 46. Altman MC, Gill MA, Whalen E, Babineau DC, Shao B, Liu AH, et al. Transcriptome networks identify mechanisms of viral and nonviral asthma exacerbations in children. *Nat Immunol*. 2019;20(5):637-651. doi:10.1038/s41590-019-0347-8.
 47. Do AN, Chun Y, Grishina G, Grishin A, Rogers AJ, Raby BA, et al. Network study of nasal transcriptome profiles reveals master regulator genes of asthma. *J Allergy Clin Immunol*. 2021;147(3):879-893. doi:10.1016/j.jaci.2020.07.006.
 48. Zhang Z, Wang J, Chen O. Identification of biomarkers and pathogenesis in severe asthma by coexpression network analysis. *BMC Med Genomics*. 2021;14:51. doi:10.1186/s12920-021-00892-4.
 49. Jansen R, Batista S, Brooks AI, Tischfield JA, Willemssen G, van Grootheest G, et al. Sex differences in the human peripheral blood transcriptome. *BMC Genomics*. 2014;15:33. doi:10.1186/1471-2164-15-33.
 50. Staedtler F, Hartmann N, Letzkus M, Bongiovanni S, Scherer A, Marc P, et al. Robust and tissue-independent gender-specific transcript biomarkers. *Biomarkers*. 2013;18(5):436-445. doi:10.3109/1354750X.2013.811538.
 51. Oliva M, Munoz-Aguirre M, Kim-Hellmuth S, Wucher V, Gewirtz ADH, Cotter DJ, et al. The impact of sex on gene expression across human tissues. *Science*. 2020;369(6509):eaba3066. doi:10.1126/science.aba3066.

52. Mersha TB, Martin LJ, Biagini Myers JM, Butsch Kovacic M, He H, Lindsey M, et al. Genomic architecture of asthma differs by sex. *Genomics*. 2015;106(1):15-22. doi:10.1016/j.ygeno.2015.03.003.
53. Gautam Y, Afanador Y, Abebe T, Lopez JE, Mersha TB. Genome-wide analysis revealed sex-specific gene expression in asthmatics. *Hum Mol Genet*. 2019;28(15):2600-2614. doi:10.1093/hmg/ddz074.
54. Reich D, Nalls MA, Kao WHL, Akylbekova EL, Tandon A, Patterson N, et al. Reduced neutrophil count in people of African descent is due to a regulatory variant in the Duffy antigen receptor for chemokines gene. *PLoS Genet*. 2009;5(1):e1000360. doi:10.1371/journal.pgen.1000360.
55. Merz LE, Story CM, Osei MA, Jolley K, Ren S, Park HS, et al. Absolute neutrophil count by Duffy status among healthy Black and African American adults. *Blood Adv*. 2023;7(3):317-320. doi:10.1182/bloodadvances.2022007679.
56. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods*. 2015;12(5):453-457. doi:10.1038/nmeth.3337.
57. Aguirre-Gamboa R, de Klein N, di Tommaso J, Claringbould A, van der Wijst MGP, de Vries D, et al. Deconvolution of bulk blood eQTL effects into immune cell subpopulations. *BMC Bioinformatics*. 2020;21(1):243. doi:10.1186/s12859-020-03576-5.
58. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics*. 2003;4(2):249-264. doi:10.1093/biostatistics/4.2.249.
59. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Series B Stat Methodol*. 1995;57(1):289-300. doi:10.1111/j.2517-6161.1995.tb02031.x.
60. Holguin F, Bleeker ER, Busse WW, Calhoun WJ, Castro M, Erzurum SC, et al. Obesity and asthma: an association modified by age of asthma onset. *J Allergy Clin Immunol*. 2011;127(6):1486-1493.e2. doi:10.1016/j.jaci.2011.03.036.
61. Global Initiative for Asthma. Global Strategy for Asthma Management and Prevention, 2024 update. Updated May 2024. Available from: <https://ginasthma.org/>.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [MS2Table2ssGSEAadjusted.tsv](#)
- [MS2Table3networksummary.tsv](#)
- [MS2Table1GSEA.tsv](#)
- [Additionalfile2.zip](#)
- [Additionalfile3.zip](#)
- [Additionalfile1.zip](#)
- [Additionalfile4.png](#)