Supplementary Material

Epithelial 15-Lipoxygenase as a Convergent Node of Arachidonic Acid Dysregulation in Chronic Rhinosinusitis with Nasal Polyps: Integrated Genetic, Metabolomic and Single-Cell Evidence

Supplementary Methods

Study Design and Data Sources

This was an integrative analysis of public data that examined the arachidonic acid pathway in CRSwNP at five levels: genetics, local metabolism, tissue transcription, single cells and treatment response (Figure S1). Sources, sample sizes and uses of each dataset are given in Table S1. All data came from studies that had obtained ethical approval and released their data publicly; no new human samples were collected, so no additional ethical review was required.

Metabolite Mendelian Randomization Screen

Exposure data were taken from the GWAS of 1400 plasma metabolites by Chen et al. (8299 individuals of European ancestry) [1], and outcome data from the CRSwNP GWAS meta-analysis of FinnGen and UK Biobank by Saarentaus et al. (9626 cases, 685,602 controls; GWAS Catalog GCST90652531) [2]. We first matched the 8488 annotated features of the nasal secretion metabolome by name to the 871 named metabolites in the GWAS, which yielded 201 shared metabolites. Screening instruments were the significant associations recorded in the GWAS Catalog for each metabolite (with effect sizes and standard errors). After allele harmonization with the outcome data, variants were clumped with PLINK 2 against the 1000 Genomes European reference ($r^2 < 0.001$, 10 Mb window). Metabolites with a single instrument were analyzed with the Wald ratio and those with two or more with the IVW method. In total, 95 metabolites were tested, and multiple testing was controlled with the Benjamini–Hochberg procedure (FDR < 0.05).

Genome-Wide Instruments and Sensitivity Analyses

The four long-chain PUFAs with the strongest screening signals (arachidonate, adrenate, EPA and n6-DPA) were re-analyzed with instruments drawn from the full summary statistics: variants with $P < 1 \times 10^{-5}$ were clumped as above, and weak instruments with an

F statistic (β^2/SE^2) ≤ 10 were removed. Alleles were harmonized with harmonise_data in TwoSampleMR (action = 2; palindromic variants oriented by allele frequency). The IVW estimate was the primary analysis, supplemented by MR-Egger [3], weighted median [4], simple mode and weighted mode methods. Heterogeneity was assessed with Cochran's Q test and horizontal pleiotropy with the MR-Egger intercept and the MR-PRESSO global test (3000 simulations) [5]. Outlying instruments identified by MR-PRESSO were removed and the estimates recomputed, and leave-one-out analyses were performed. MR rests on three assumptions: the instruments are associated with the exposure (relevance), share no common cause with the outcome (independence), and affect the outcome only through the exposure (exclusion restriction). Weak instruments were removed to support relevance; the exposure GWAS (Canadian Longitudinal Study on Aging) and the outcome GWAS (FinnGen and UK Biobank) do not share participants; and pleiotropy-robust estimators, MR-PRESSO and colocalization were used to probe the exclusion restriction.

Colocalization at the FADS Locus

Within chr11:61.0–62.5 Mb (GRCh38), each of the four fatty acids, and linoleate as a negative control, was tested for colocalization with CRSwNP using approximate Bayes factors [6]. Prior probabilities were p1 = p2 = 1×10^{-4} and p12 = 1×10^{-5}; the prior standard deviation of effect sizes was 0.15 for continuous exposures and 0.2 for the binary outcome. We implemented the method from the Wakefield approximate Bayes factor and checked the implementation by colocalizing the outcome with itself (PP4 = 0.974). PP4 > 0.7 was taken to indicate a shared causal variant. The same approach was used for colocalization of CRSwNP with FADS1 and FADS2 expression in GTEx v8; because GTEx releases only significant eQTLs, this analysis was restricted to significant eQTL variants.

Gene-Level SMR and HEIDI Test

We prespecified 48 arachidonic acid pathway genes, comprising 32 synthetic and metabolic enzymes and 16 lipid-mediator receptors (Table S8). Significant cis-eQTLs of European ancestry from GTEx v8 [7] were converted to BESD format for six tissues (whole blood, lung, esophageal mucosa, spleen, minor salivary gland and EBV-transformed lymphocytes), with 1000 Genomes Europeans as the linkage disequilibrium reference. SMR v1.3.1 [8] was run with default settings (top cis-eQTL $P < 5 \times 10^{-8}$; HEIDI

test using 3–20 variants with P_eQTL < 1.57×10^{-3}). Significance was judged within each trait–tissue combination by Bonferroni correction (0.05 divided by the number of probes tested in that combination), together with HEIDI P ≥ 0.05, meaning no evidence that the association was driven by linkage. The asthma GWAS from the same study (GCST90652530) served as a comparison outcome. Linkage disequilibrium between variants was computed with PLINK 2 in 1000 Genomes Europeans.

Re-Quantification of the Nasal Secretion Metabolome

MTBLS13192 contains UPLC-MS data of nasal secretions from 28 patients described in the source study as having type 2 chronic rhinosinusitis with nasal polyps assessed on postoperative pathology and 12 healthy controls, with 40 samples and 3 pooled quality-control samples in each of positive and negative ion modes. As the public metabolite table did not contain per-sample abundances, we converted the raw .raw files to mzML with ThermoRawFileParser and performed targeted integration of every annotated feature (m/z and retention time) in the submitters' metabolite table: the apex was located on the extracted ion chromatogram (±10 ppm) within ±0.3 min of the annotated retention time, and the peak was integrated by the trapezoidal rule within ±0.15 min of the apex.

Sample numbers in the public files carried no group labels. By number, samples fell into two groups (1–12 and 13–40), which we checked against the direction of the differential metabolites reported by the submitters: six of seven measurable metabolites agreed, identifying samples 1–12 as controls. Acquisition timestamps showed that the two groups had been injected in alternation (rank-sum test of acquisition order, P = 0.69), so injection order was not confounded with group.

Features were kept if detected in at least 80% of samples in both groups. Data were normalized first to total area and then by probabilistic quotient normalization (PQN) [9] against the median control spectrum, and features with a relative standard deviation ≥ 30% in quality-control samples were removed, leaving 4890 features. Data were $\log_2$-transformed and missing values replaced by the minimum of each feature. Groups were compared with the Mann–Whitney U test with Benjamini–Hochberg correction; fold changes were ratios of group means, and variable importance in projection (VIP) came from a two-component PLS-DA model of autoscaled data. Annotation confidence was graded according to the Metabolomics Standards Initiative (MSI) [10]: an annotation was

level 2 (MS2 support) if a data-dependent acquisition (DDA) MS2 spectrum in any sample matched both its precursor m/z (±10 ppm) and retention time (±0.3 min), and level 3/4 otherwise (MS1 and retention time only). Metabolite-level analyses were limited to comparisons prespecified by the genetic results.

Targeted Identification and Quantification of Lipoxygenase Products

To look for ALOX15 products, we retrieved all negative-mode MS2 spectra whose precursor matched a monohydroxy fatty acid $[M-H]^-$ (±10 ppm): HETE (m/z 319.2279), HODE (295.2279), HDHA (343.2279) and HEPE (317.2122). Because positional isomers share the same precursor mass, isomers were assigned from published diagnostic fragments (±0.01 Da): m/z 219.1391 for 15-HETE, 179.1078 for 12-HETE, 115.0401 for 5-HETE, 195.1391 for 13-HODE, 171.1027 for 9-HODE, 245.1547 for 17-HDHA, 205.1234 for 14-HDHA and 219.1391 for 15-HEPE. A diagnostic ion had to reach at least 5% of the base peak. Spectra were grouped into 0.15 min retention-time bins, and an isomer was accepted only when supported by at least three spectra.

Quantification used MS1 extracted ion chromatograms (±10 ppm). Retention times were first aligned on the apex of the main 13-HODE peak in each sample (offsets −0.014 to 0.012 min), and each isomer was then integrated within ±0.045 min (HETE, HODE) or ±0.035 min (HDHA, HEPE) of its apex and normalized with the PQN factors from the global analysis. All six targeted products had positive integrated peak areas in all 40 samples; therefore, there were no missing or zero values and no imputation was performed. Groups were compared using two-sided Mann–Whitney U tests, with Benjamini–Hochberg correction across the six products. The within-sample 15-HETE/12-HETE ratio was used as a relative lipid-product measure intended to reduce common dilution effects, rather than as a direct assay of enzyme activity. The ratio was finite and positive in all 40 samples, with no missing values or zero denominators. Quantitative repeatability was assessed from the relative standard deviation of each peak in pooled quality-control samples. No authentic standards were used, so these identifications correspond to MSI level 2.

Tissue Transcriptomes

The discovery set comprised raw counts from 42 nasal polyps and 28 healthy inferior turbinates in GSE136825 [11]. Genes with counts ≥ 10 in at least 10 samples were kept, and differential expression was tested with DESeq2 [12] with group as the only covariate;

differentially expressed genes were defined by FDR < 0.05 and $|\log_2$ fold change$| > 1$. Pathway genes were taken from KEGG arachidonic acid metabolism (hsa00590) and biosynthesis of unsaturated fatty acids (hsa01040), 90 genes in total. Overall pathway enrichment was tested with the hypergeometric test, and a preranked competitive test was performed with limma cameraPR [13] using DESeq2 Wald statistics. Functional enrichment of upregulated and downregulated genes was assessed separately by over-representation analysis against GO biological process and KEGG gene sets from msigdbr.

In the external cohorts, counts from GSE179265 [14] (RNA-seq) underwent variance-stabilizing transformation, and series-matrix values from GSE36830 [15] (GPL570 array) were $\log_2$-transformed, keeping the probe with the highest mean expression for each gene. Key enzymes were compared between groups in each of the three cohorts with the Wilcoxon rank-sum test.

Pathway Scores, Eosinophil Activation and Immune Infiltration

A gene-set score was the mean z-score of its member genes. Scores for fatty acid-related GO biological processes (arachidonic acid, eicosanoid, leukotriene and prostanoid metabolism, the lipoxygenase pathway, the epoxygenase P450 pathway and others) were compared between groups with the Wilcoxon test and checked with cameraPR. An eosinophil activation score was built from 14 eosinophil genes (CLC, CCR3, IL5RA, SIGLEC8, EPX, PRG2, PRG3, RNASE2, RNASE3, ADORA3, HRH4, ADGRE1, CEBPE and SMPD3), and its relation to key enzymes was assessed by Spearman correlation within polyp samples only. In GSE198950, chronic rhinosinusitis samples were divided by the pathology report into < 10 and ≥ 10 eosinophils per high-power field, and key enzymes were compared with the Wilcoxon test. The column order of this expression matrix did not match the sample table, so samples were matched through the identifiers in the sample description field rather than by column position.

Immune cell composition was estimated with an R implementation of CIBERSORT (e1071 nu-SVR) [16]: with LM22 as the signature matrix, each sample was fitted by linear ν-support vector regression (ν chosen from 0.25, 0.5 and 0.75 by the lowest root-mean-square error), negative coefficients were set to zero and the remainder normalized to proportions. Groups were compared with the Wilcoxon test and Benjamini–Hochberg

correction, and relations between key enzymes and cell fractions were assessed by Spearman correlation.

Single-Cell Transcriptome Analysis

Raw sequencing data for GSE196169 [17] are not public, so we used the per-sample raw count matrices provided by the authors (19 nasal polyps; 9 profiled with 10x Chromium v2 and 10 with v3; immune cells enriched by flow sorting). Quality control kept cells with 200–6000 detected genes and < 20% mitochondrial reads, and genes detected in at least 10 cells. Data were normalized as in Seurat LogNormalize (counts scaled to 10^4 per cell, then log1p). Highly variable genes were ranked by standardized dispersion separately within the v2 and v3 chemistries, and the 2000 genes with the highest mean rank were selected. To reduce batch offsets between samples and chemistries, highly variable genes were centered within each sample, then scaled and clipped at ±10; 30 principal components were extracted by truncated singular value decomposition, and cells were grouped into 40 clusters by k-means.

Cell types were annotated from canonical marker genes. Each cluster was scored by the mean z-score of each marker set and the labels were then adjusted manually with reference to the cluster's top genes; for example, clusters high in GATA3 and IL13 were labelled ILC2/Th2. One cluster dominated by mitochondrial genes, MALAT1 and stress genes was judged to be low quality and removed, leaving 36,991 cells of 19 cell types. t-SNE (perplexity 30) on the first 20 principal components was used for visualization. The localization of each gene was described by mean expression and the fraction of expressing cells. Robustness was examined per sample: among samples with at least 20 cells of a cell type, we counted the samples in which that cell type showed the highest expression. For the relation between ALOX15 and CYP2J2 within epithelial cells, Spearman coefficients were computed within each sample and their median was tested against zero with the Wilcoxon signed-rank test.

Independent Replication of Single-Cell Localization

Because the immune enrichment and lack of stromal cells in GSE196169 bias its cell composition, we replicated the localization in GSE276503 [18], in which tissue was enzymatically dissociated and sequenced without sorting; only the per-sample count matrices of its 15 nasal polyps were used. Quality control, normalization, highly variable

genes, principal components and clustering followed exactly the steps and parameters above. Clusters were first scored automatically from marker genes and then corrected manually after inspecting the fraction of cells expressing each marker: clusters positive for CD1C, CLEC10A and FCER1A were labelled dendritic cells, clusters positive for FCN1 and S100A8 but negative for FCGR3B were labelled monocytes, and one KRT8+/KRT5+ cluster with a median of only 660 detected genes was kept separately as low-complexity epithelium. Per-sample robustness was assessed as above, and the share of ALOX15 expression contributed by each compartment was computed as the sum of normalized expression across its cells.

Fresh-tissue droplet single-cell sequencing captures eosinophils poorly, and GSE276503 contained no recognizable eosinophil population. We therefore also analyzed GSE299751 [19], single-cell data from nasal lavage of children with asthma generated with 10x Flex fixed-RNA probes, in which epithelial cells and eosinophils are present in the same samples. Barcodes with ≥ 200 UMIs were kept from the raw matrices and then filtered to cells with ≥ 200 detected genes, ≥ 500 UMIs and < 20% mitochondrial reads; the remaining steps were as above (30 clusters). Eosinophils were defined as CLC+, CCR3+ and SIGLEC8+ but HDC−, and basophils as HDC+ and GATA2+. Because eosinophils were sequenced far less deeply than epithelium (median 590 vs. 12,379 UMIs), detection rates cannot be compared directly. We therefore binomially down-sampled the counts of all other cells to depths drawn at random from the eosinophil UMI distribution, repeated this 200 times, and compared ALOX15 detection at equal depth together with ALOX15 counts per 10,000 UMIs. This analysis compared transcript detection and relative transcript abundance in nasal lavage from children with asthma, rather than directly measuring cellular contributions in CRSwNP tissue. Lower ALOX15 detection in eosinophils does not exclude a substantial functional contribution, which may depend on per-cell protein expression and enzyme activity.

Inference of Lipid-Mediator Communication

For seven lipid mediators (LTB4, cysteinyl leukotrienes, PGD2, PGE2, TXA2, PGI2 and PGF2α), synthetic steps and receptors were defined. For each synthetic step the most highly expressed isoenzyme was used; the sender score was the geometric mean of the mean expression of the synthetic steps, the receiver score was the mean expression of the receptor, and their product was the communication score for a pair of cell types. Following

CellPhoneDB [20], empirical P values were derived from a null distribution built by permuting cell-type labels 1000 times. For the top-scoring pair of each mediator, we checked sample by sample whether the sender and receiver cell types agreed, including only cell types with at least 20 cells.

Restoration after Dupilumab

GSE267331 contains Affymetrix HG-U133 Plus 2.0 array data from nasal brushings in the SINUS-52 trial [21]. We used the authors' processed series-matrix values ($\log_2$ scale), keeping the probe with the highest mean expression for each gene. Participants with paired baseline and week-24 samples were included: 47 receiving dupilumab and 19 receiving placebo. Disease-signature genes were the differentially expressed genes of GSE136825, of which 1747 were measurable on the array. The array background was 2.31, and only the 356 genes with a median $\log_2$ expression > 6 in brushings were retained.

For each gene, the within-participant change (week 24 minus baseline) was compared between arms with Welch's t test. Restoration was signed so that movement opposite to the disease effect was positive: restoration = −sign(disease effect) × (mean change with dupilumab − mean change with placebo). The restored fraction was restoration divided by the absolute disease effect. P values were corrected by the Benjamini–Hochberg procedure, and genes with FDR < 0.05 and positive restoration were classed as restored. A non-significant difference does not show that a gene was not restored, so "not restored" was judged with two one-sided tests of equivalence (TOST) [22], with a margin of 50% of the absolute disease effect: TOST P < 0.05 was classed as equivalent to no restoration and the rest as indeterminate. Key enzymes with a median expression ≤ 6, close to array background, were not interpreted.

The disease-signature score was the mean z-score of upregulated genes minus that of downregulated genes, and the change in score was compared between arms with Welch's t test. To see whether restoration differed by cell of origin, we computed an epithelial specificity for each gene from the single-cell data, $\log_2$[(mean epithelial expression + 0.01)/(mean expression in the highest immune cell type + 0.01)]. Genes with specificity > 1 and expression in ≥ 10% of cells of some cell type were classed as epithelium-specific and those < −1 as immune-specific, and restored fractions of the two groups were compared with the Mann–Whitney U test.

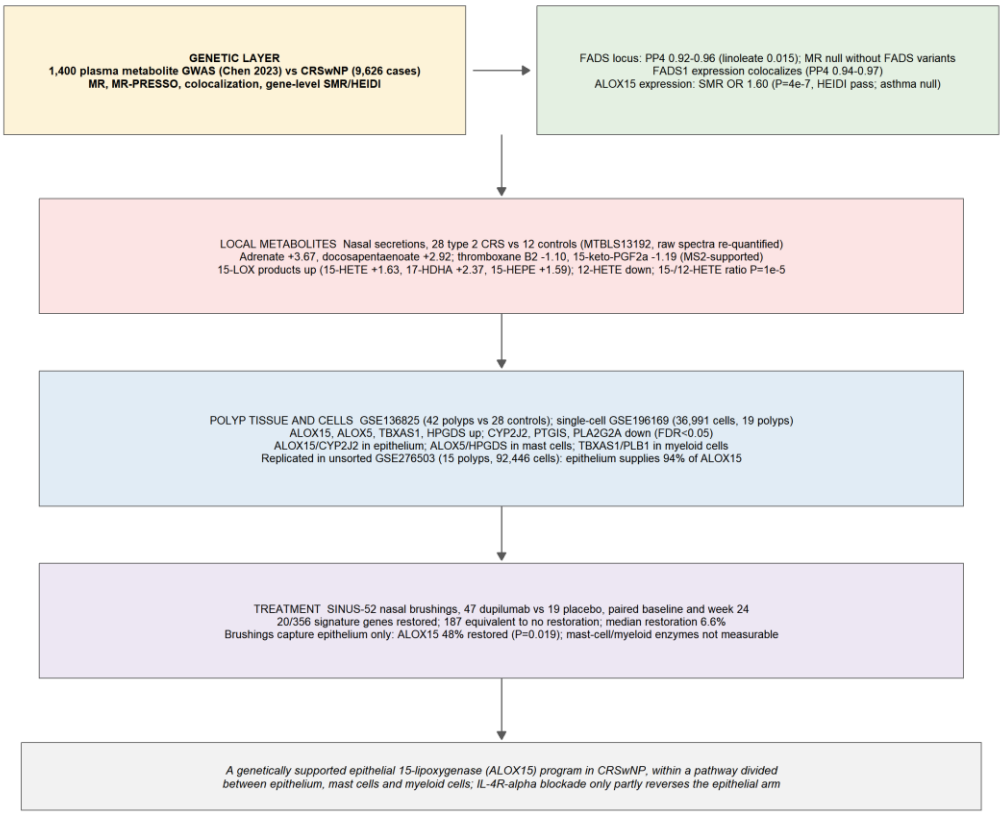
Diagnostic Model (Secondary Analysis)

The diagnostic model was a secondary analysis. With GSE136825 as the training set, pathway genes that were differentially expressed and measurable in all three cohorts were z-standardized within each cohort. Five feature selection methods (all genes, LASSO, elastic net, random forest and stepwise logistic regression) were crossed with 10 classifiers (LASSO, ridge regression, elastic net with two α values, random forest, XGBoost, radial-kernel SVM, naive Bayes, LDA and gradient boosting machine), giving 50 combinations whose AUCs were evaluated in GSE179265 and GSE36830. The training AUC of the final four-gene logistic regression model was estimated by tenfold cross-validation repeated 20 times. To test whether the model outperformed chance, we drew 1000 random four-gene sets from the 1750 differentially expressed genes measurable in all three cohorts and put them through the same procedure; 867 sets were fitted successfully. The empirical P value was (number of random sets with AUC not lower than observed + 1)/(number of valid sets + 1).

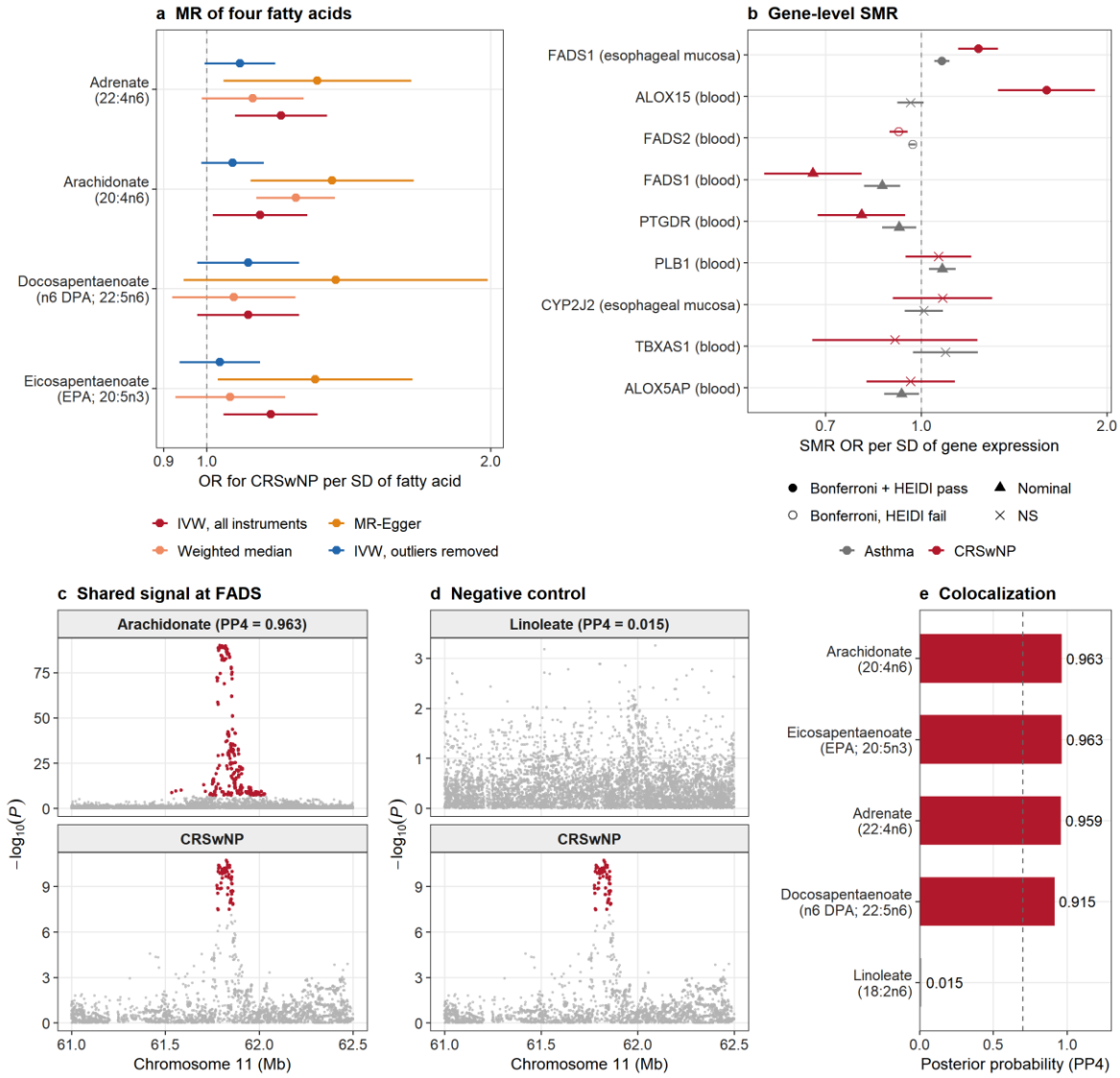
Statistical Analysis and Software

Unless stated otherwise, tests were two-sided with $P < 0.05$ considered significant, and multiple testing was controlled by the Benjamini–Hochberg FDR. Analyses used R 4.5.3 (TwoSampleMR 0.7.5, MRPRESSO 1.0, DESeq2 1.50.2, limma 3.66.0, glmnet 4.1.10, e1071 1.7.17, randomForest 4.7.1.2, xgboost 3.2.1.1 and msigdbr 26.1.0), Python 3.12.10 (NumPy 2.5.0, SciPy 1.18.0, scikit-learn 1.9.0 and pandas 3.0.3), PLINK 2, SMR 1.3.1 and ThermoRawFileParser.

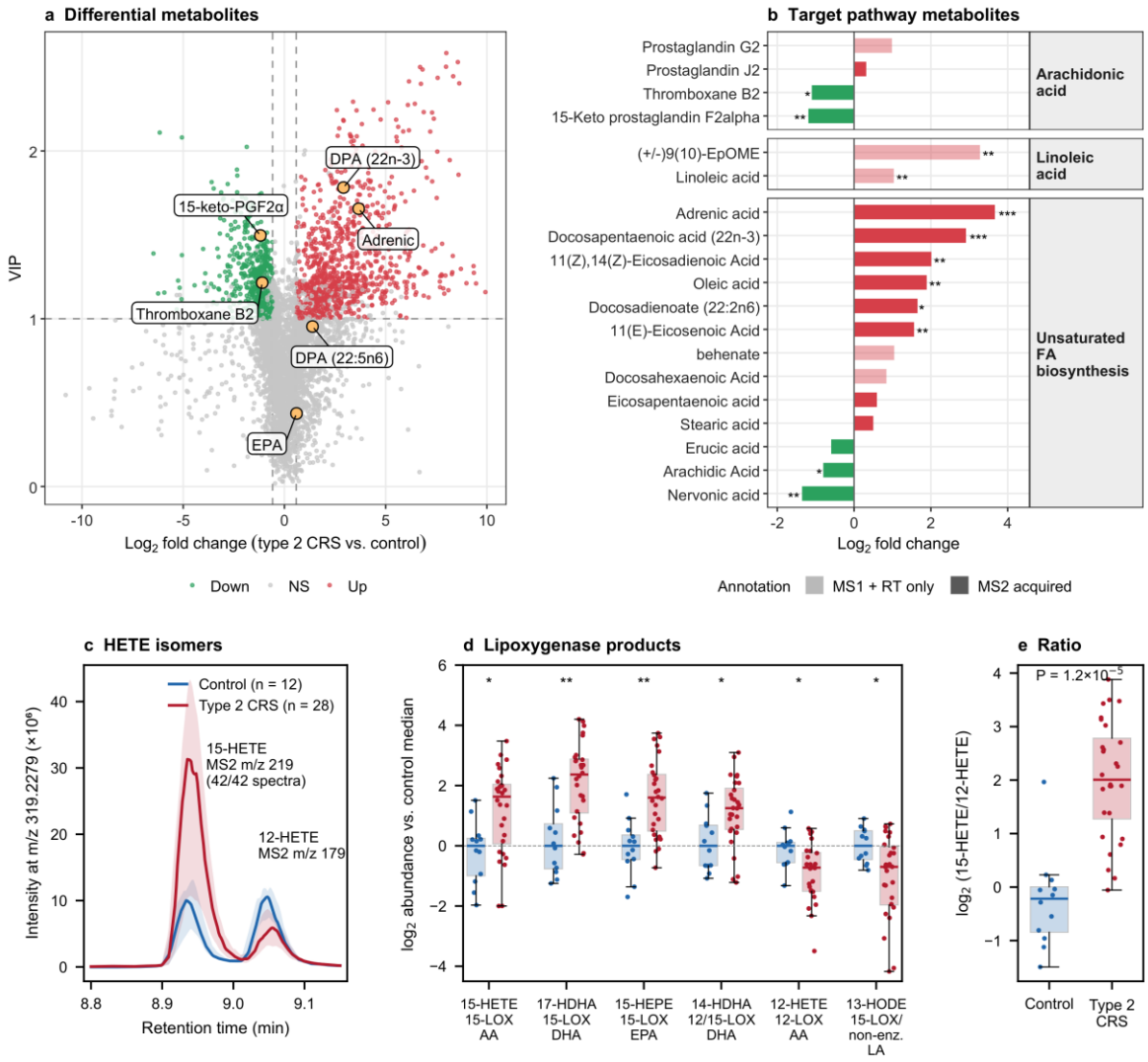
Supplementary Figures



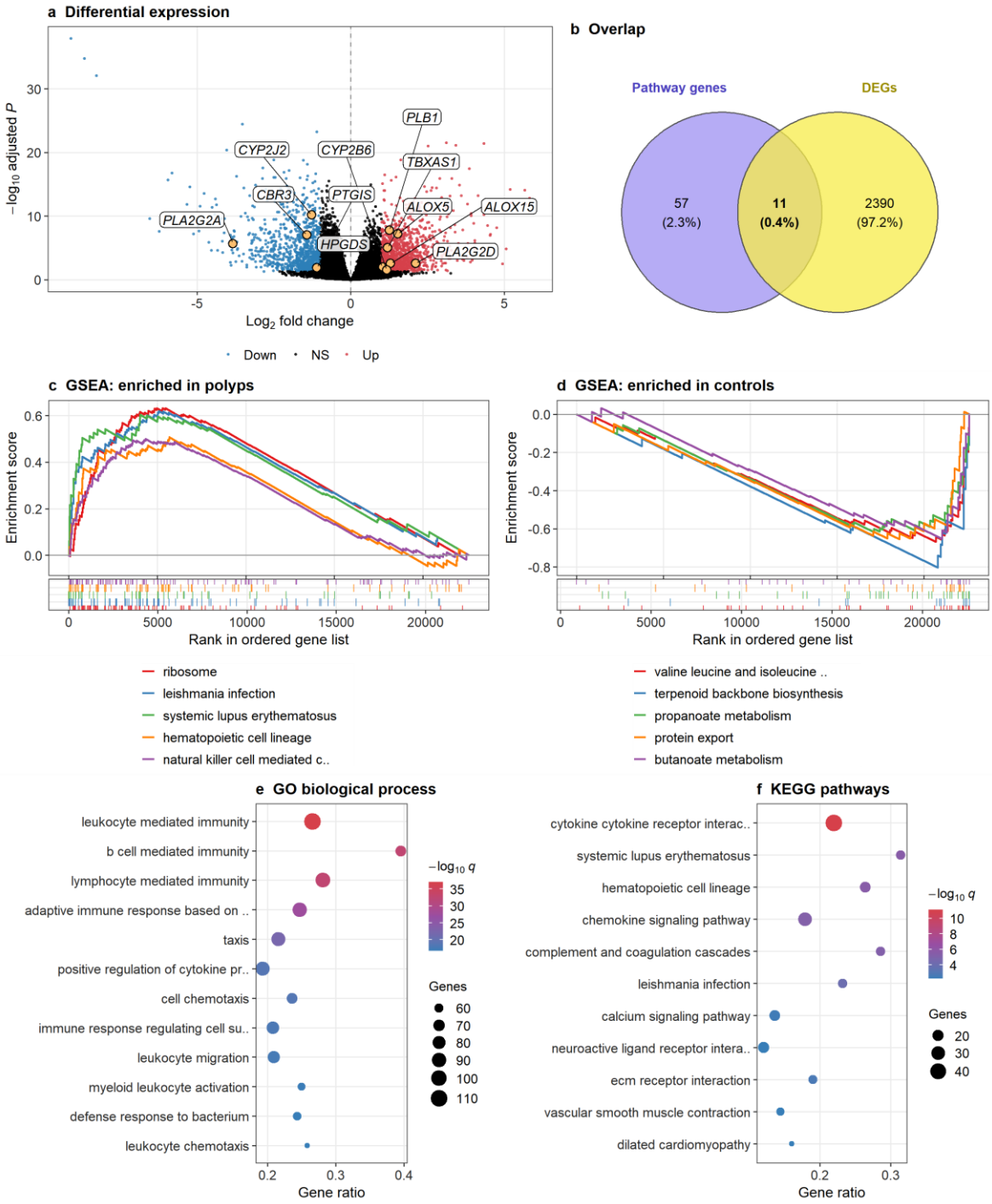
Supplementary Figure S1. Study design. The same pathway was examined at five levels: genetics, local metabolites in nasal secretions, enzyme expression in polyp tissue, single-cell localization and response to IL-4Rα blockade.



Supplementary Figure S2. Genetic analyses. (a) MR estimates for four fatty acids; blue marks estimates after removal of MR-PRESSO outliers. (b) Gene-level SMR for CRSwNP (red) and asthma (grey); filled circles pass both Bonferroni correction and the HEIDI test. (c) Overlap of the arachidonate and CRSwNP association signals on chromosome 11. (d) Linoleate as a negative control. (e) Colocalization posterior probabilities for each fatty acid.



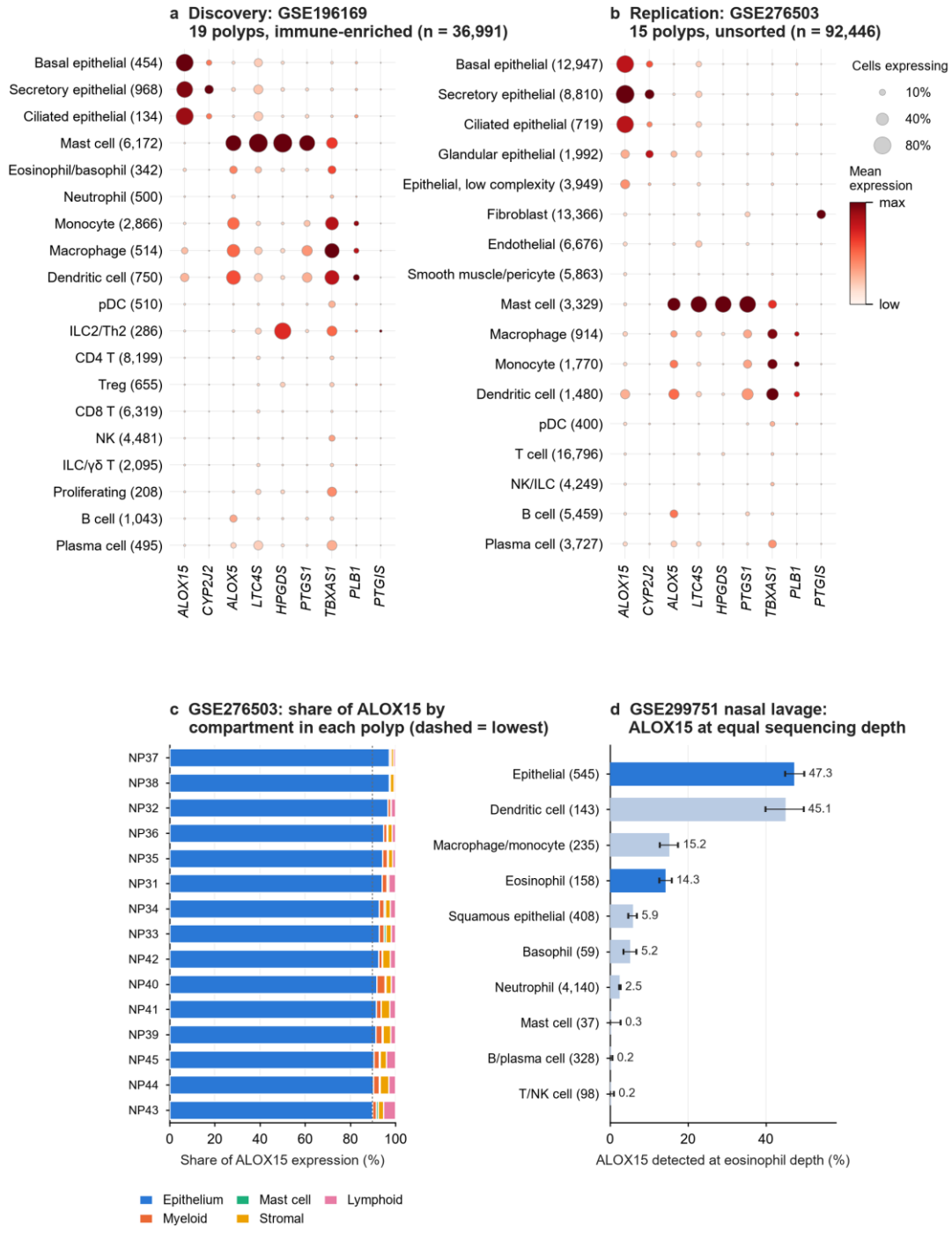
Supplementary Figure S3. Nasal secretion metabolome. (a) Differential metabolites; circled points are target metabolites with MS2 support. (b) Metabolites detected in the three target pathways; shading indicates whether the annotation has MS2 support. (c) Extracted ion chromatogram at m/z 319.2279 (group median and interquartile range); the two peaks were assigned to 15-HETE and 12-HETE from diagnostic MS2 fragments. (d) Abundance of lipoxygenase products relative to the control median. (e) Within-sample 15-HETE/12-HETE ratio.



Supplementary Figure S4. Polyp tissue transcriptome (GSE136825). (a) Volcano plot of differentially expressed genes. (b) Overlap between pathway genes and differentially expressed genes. (c,d) GSEA running-score curves. (e,f) GO and KEGG enrichment.

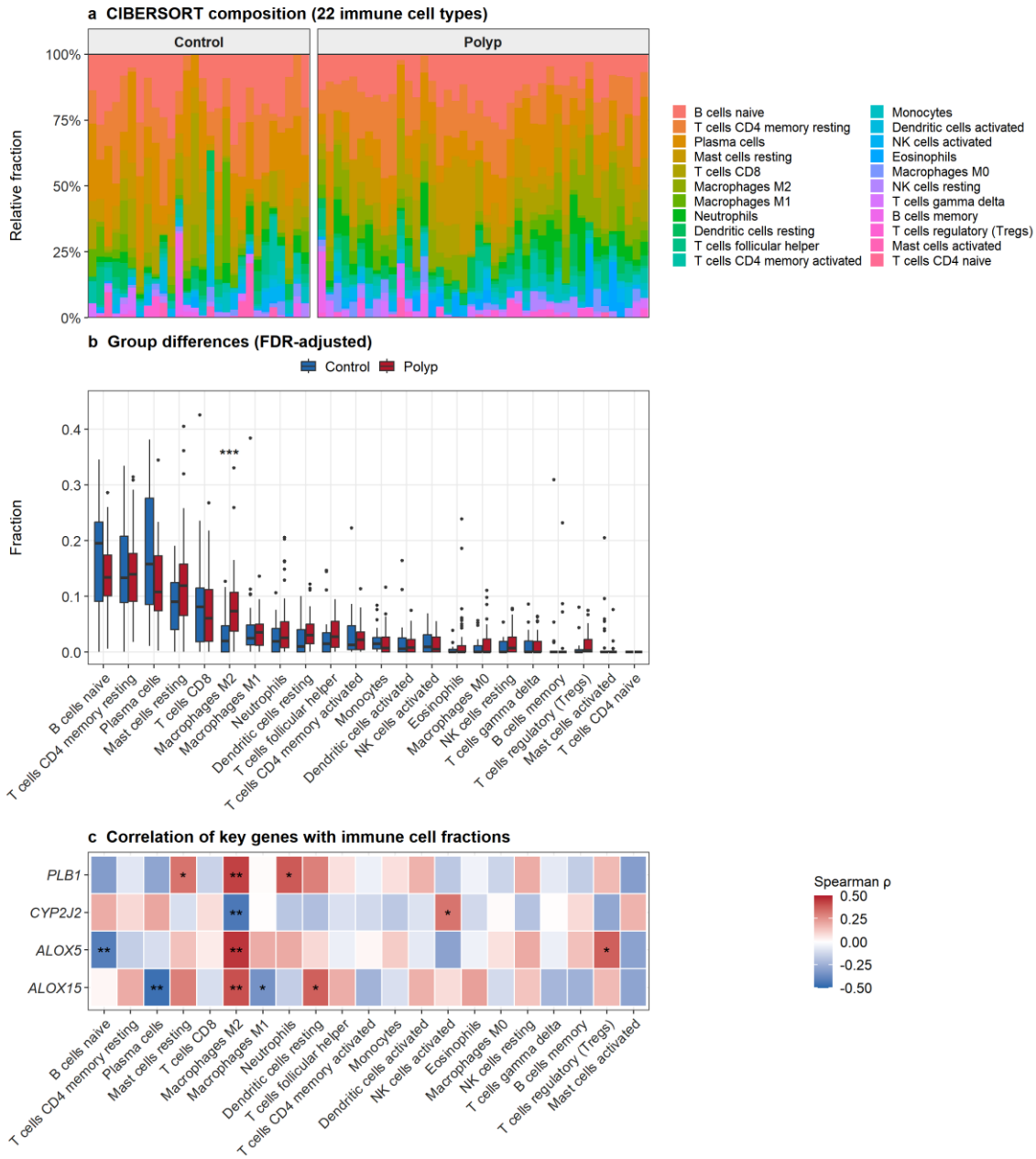


Supplementary Figure S5. Single-cell localization and lipid-mediator circuits (GSE196169). (a) 36,991 cells from 19 polyps; immune cells were enriched by flow sorting, so proportions do not reflect tissue composition. (b) Expression of synthases and receptors across cell types (dot size, fraction of expressing cells; color, relative mean expression). (c–f) Distribution of ALOX15, CYP2J2, ALOX5 and TBXAS1. (g) Intercellular circuits inferred from synthase and receptor expression.

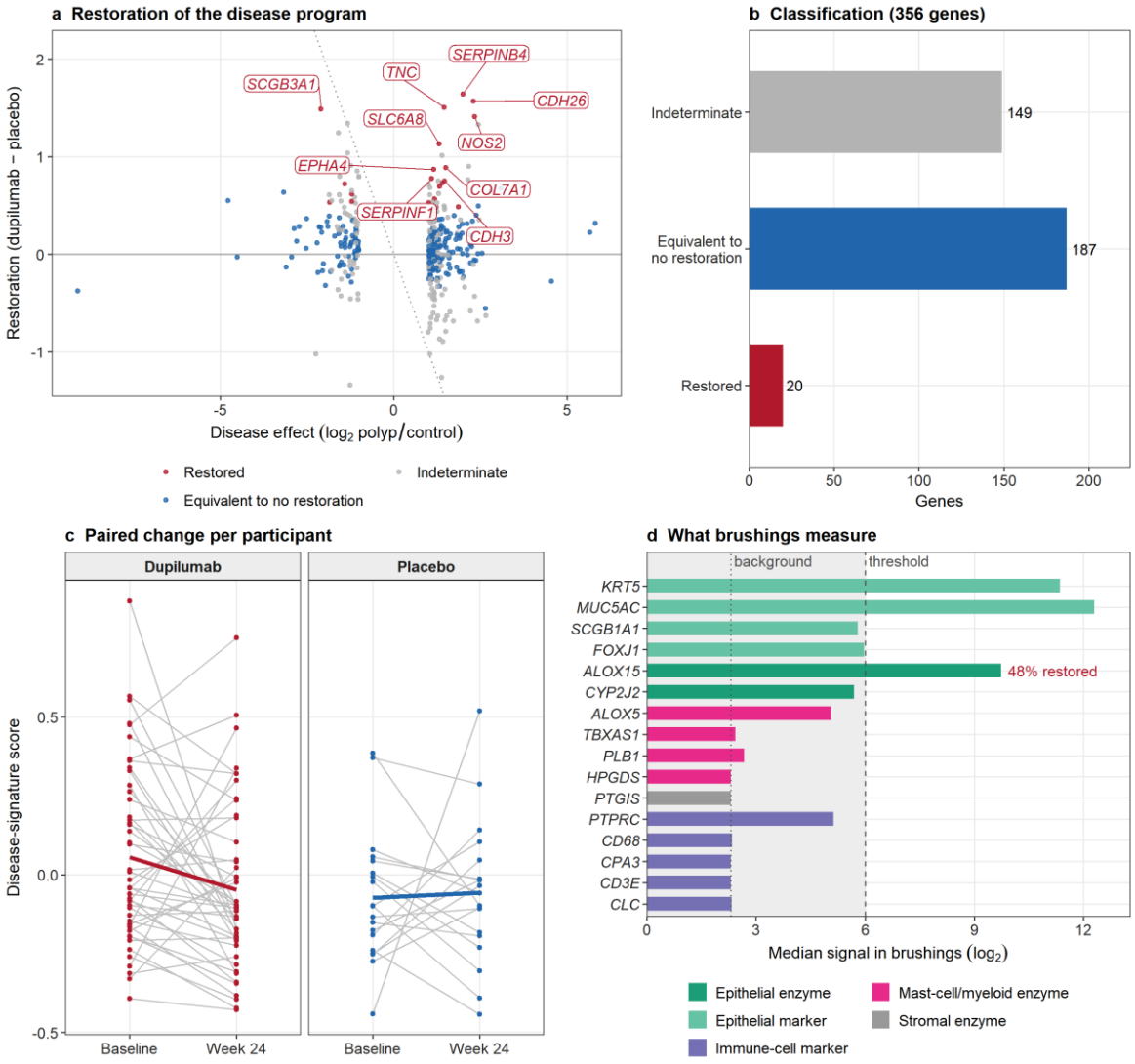


Supplementary Figure S6. Independent replication of single-cell localization. (a,b) Localization of the same enzymes in the discovery set (GSE196169, immune cells enriched by flow sorting) and the replication set (GSE276503, unsorted whole tissue); dot size is the fraction of expressing cells, color the mean expression scaled to the gene's maximum within each dataset, and cell numbers are given in parentheses. (c) Share of

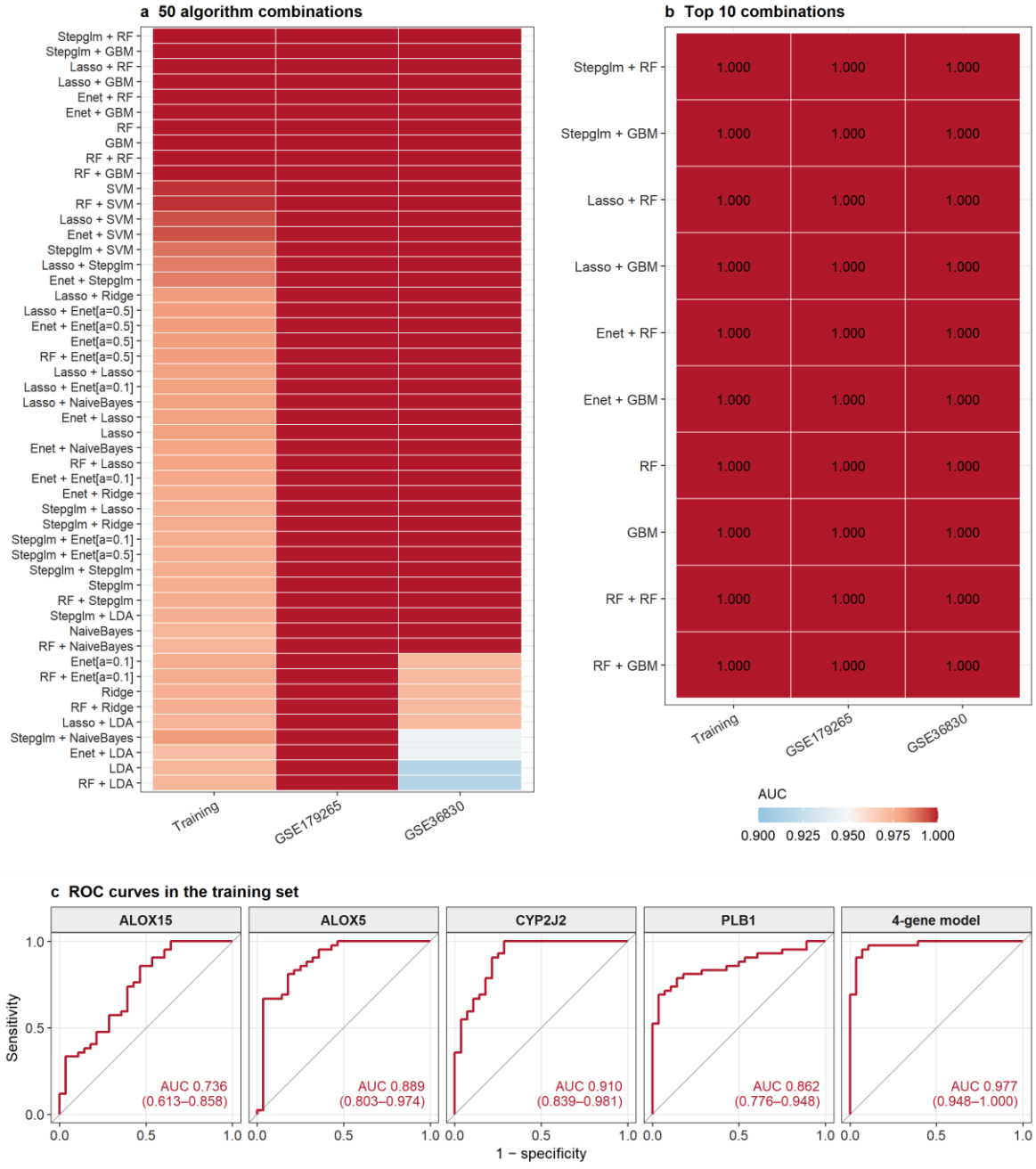
ALOX15 expression by compartment in each GSE276503 polyp; the dashed line marks the lowest epithelial share (89.8%). (d) ALOX15 detection in nasal lavage (GSE299751) after binomial down-sampling of each cell type to eosinophil sequencing depth (median 590 UMIs); bars show the mean of 200 draws and whiskers the 95% range.



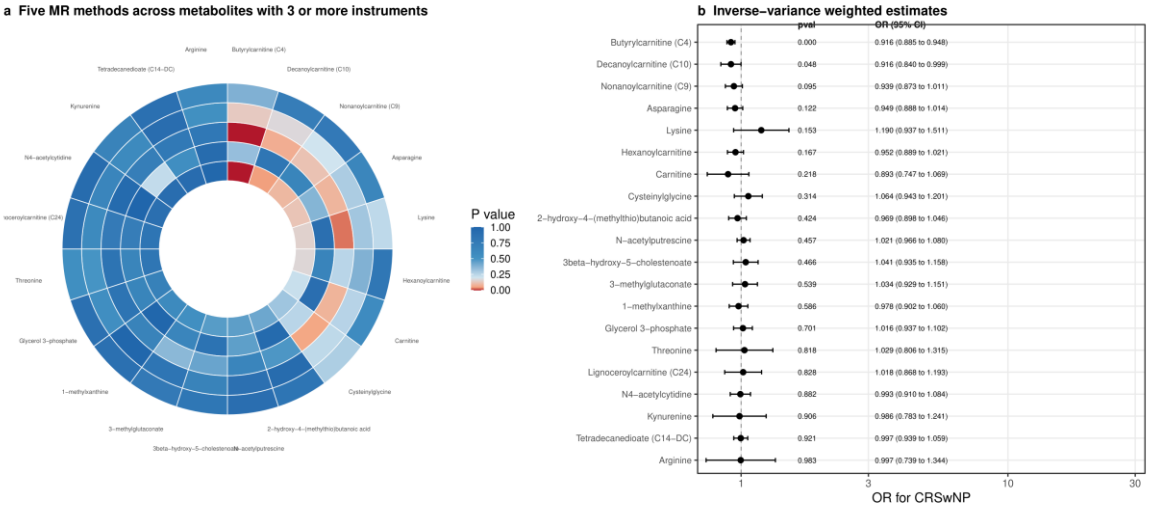
Supplementary Figure S7. Immune infiltration (GSE136825). (a) Composition of 22 immune cell types. (b) Differences between groups (asterisks, FDR-corrected). (c) Correlation of key genes with immune cell fractions.



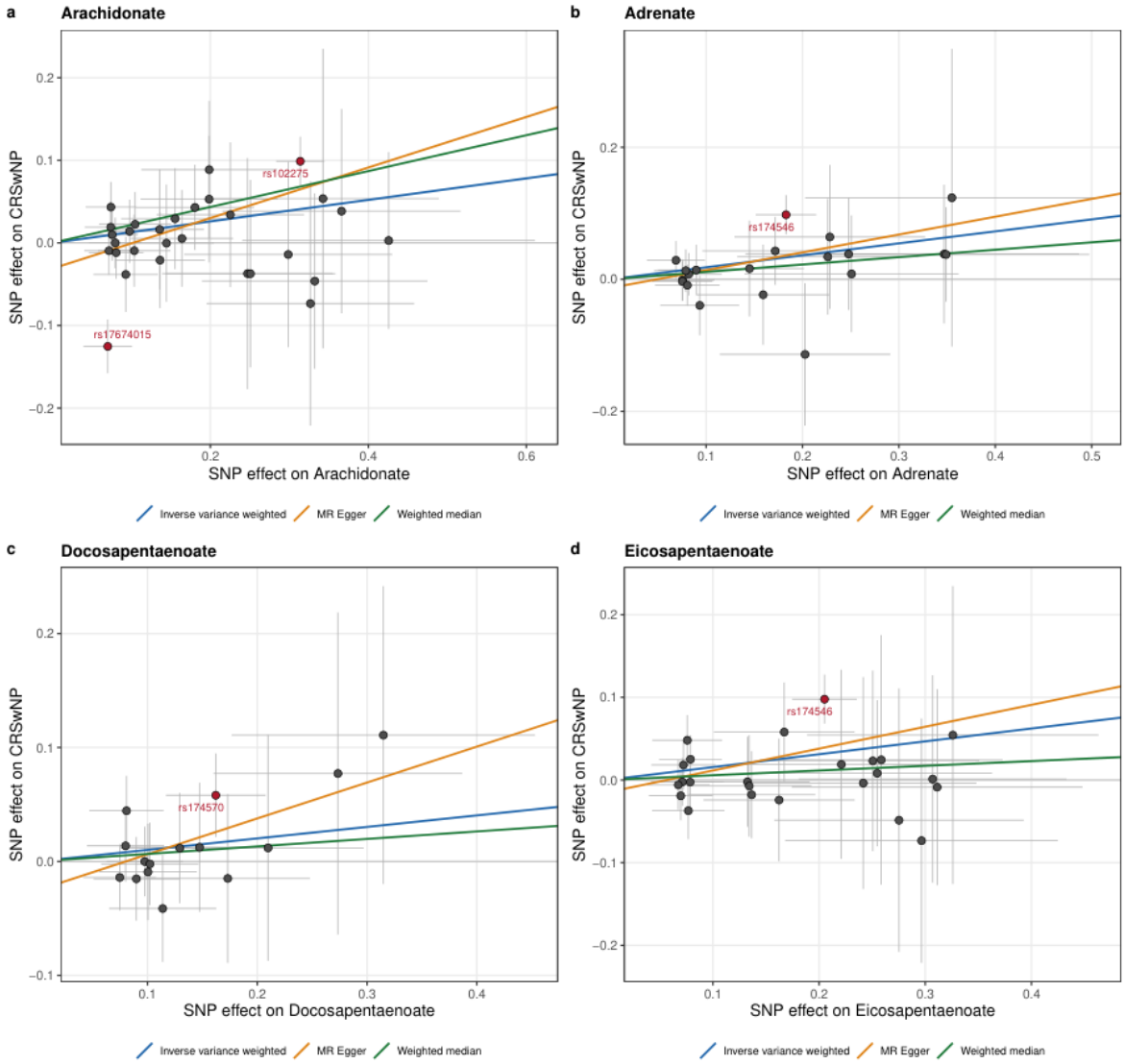
Supplementary Figure S8. Restoration after treatment (GSE267331). (a) Disease effect and restoration for each gene (red, restored; blue, equivalent to no restoration; grey, indeterminate). (b) Number of genes in each category. (c) Paired change in each participant. (d) Median expression of marker genes and key enzymes in nasal brushings; the grey zone lies below the expression threshold, and ALOX15 was the only enzyme above it.



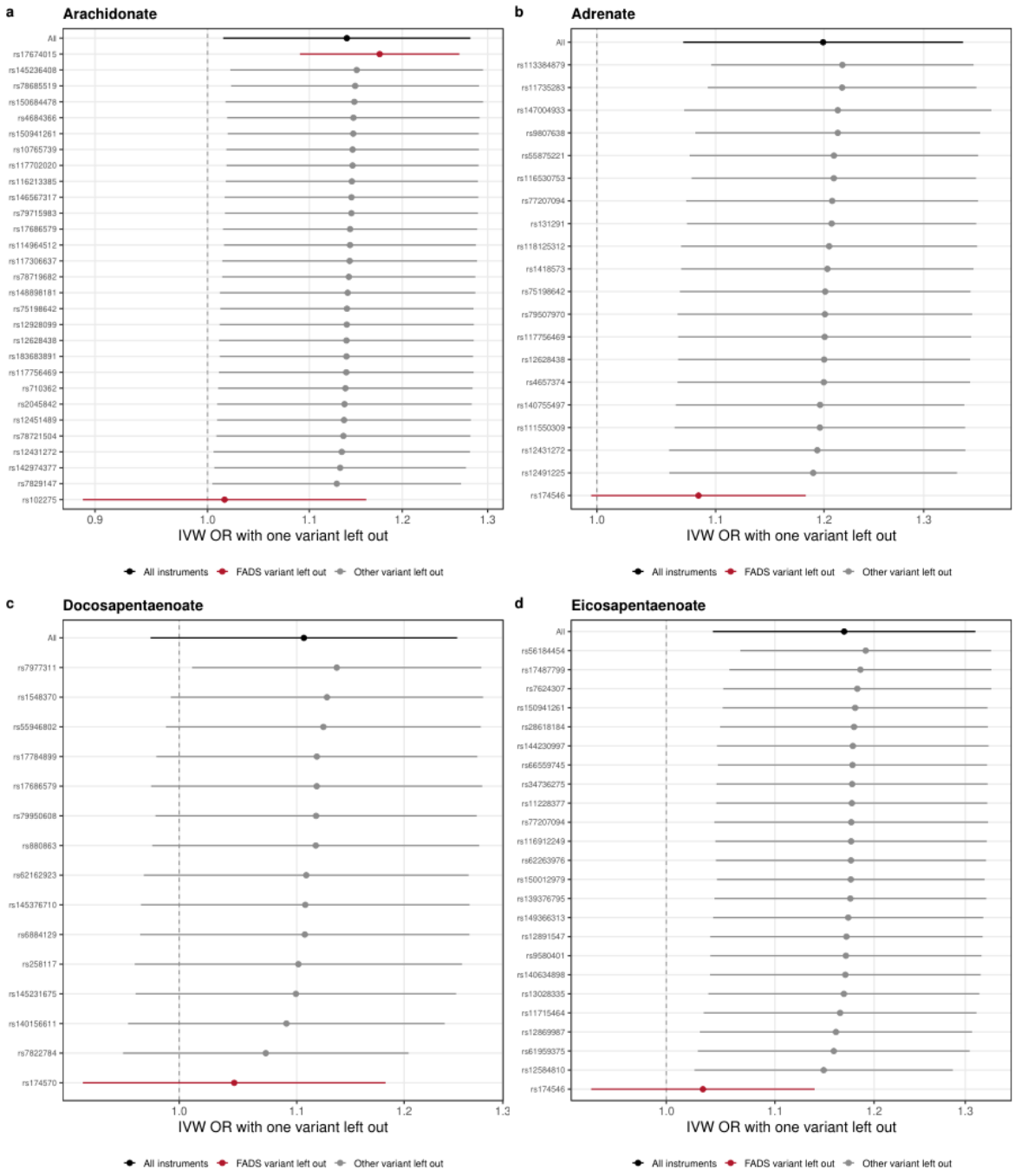
Supplementary Figure S9. Diagnostic model (secondary analysis). (a) AUC of 50 algorithm combinations in three cohorts. (b) Top 10 combinations. (c) ROC curve in the training set.



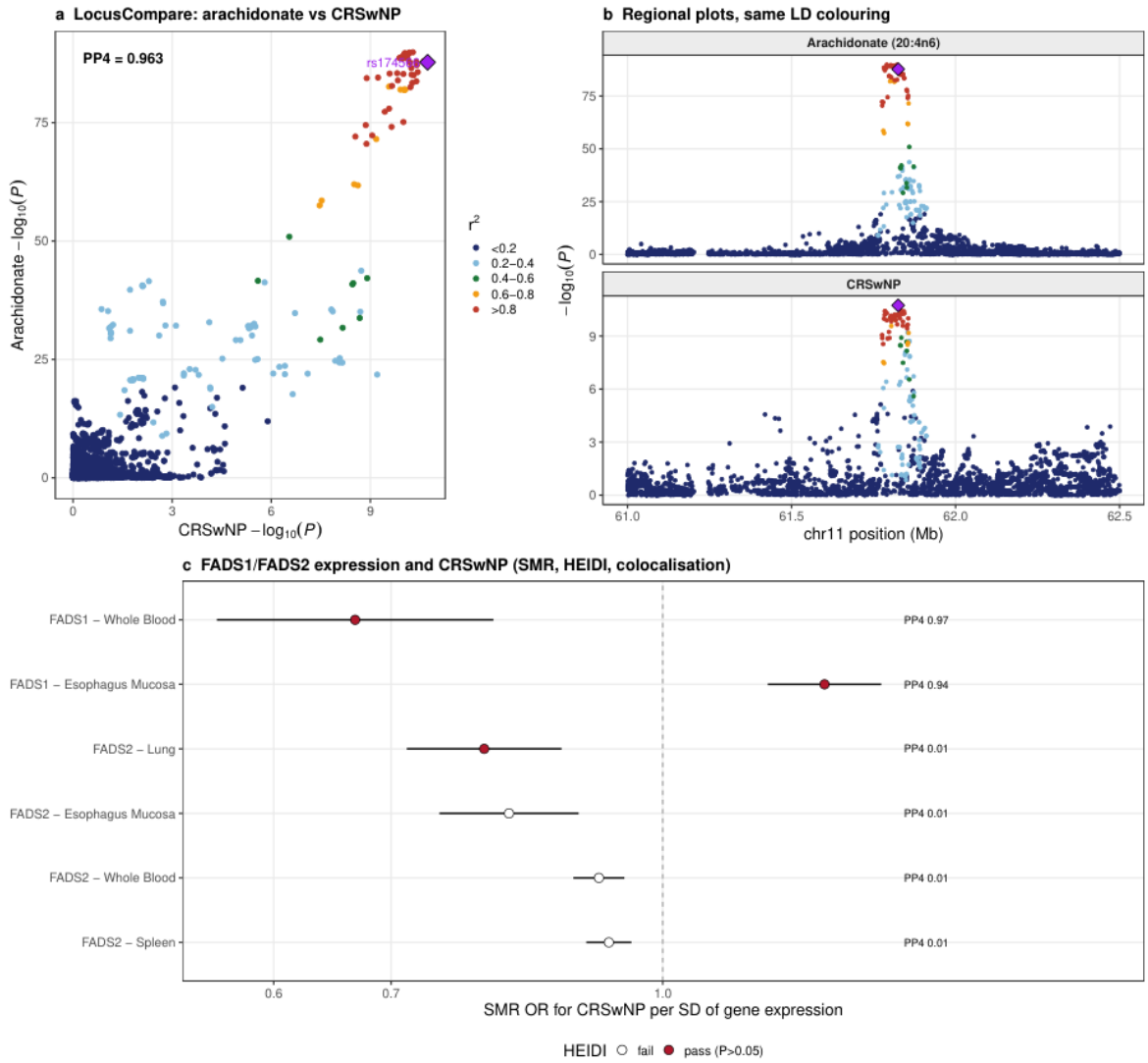
Supplementary Figure S10. Mendelian randomization screen of 95 nasal secretion metabolites against CRSwNP.



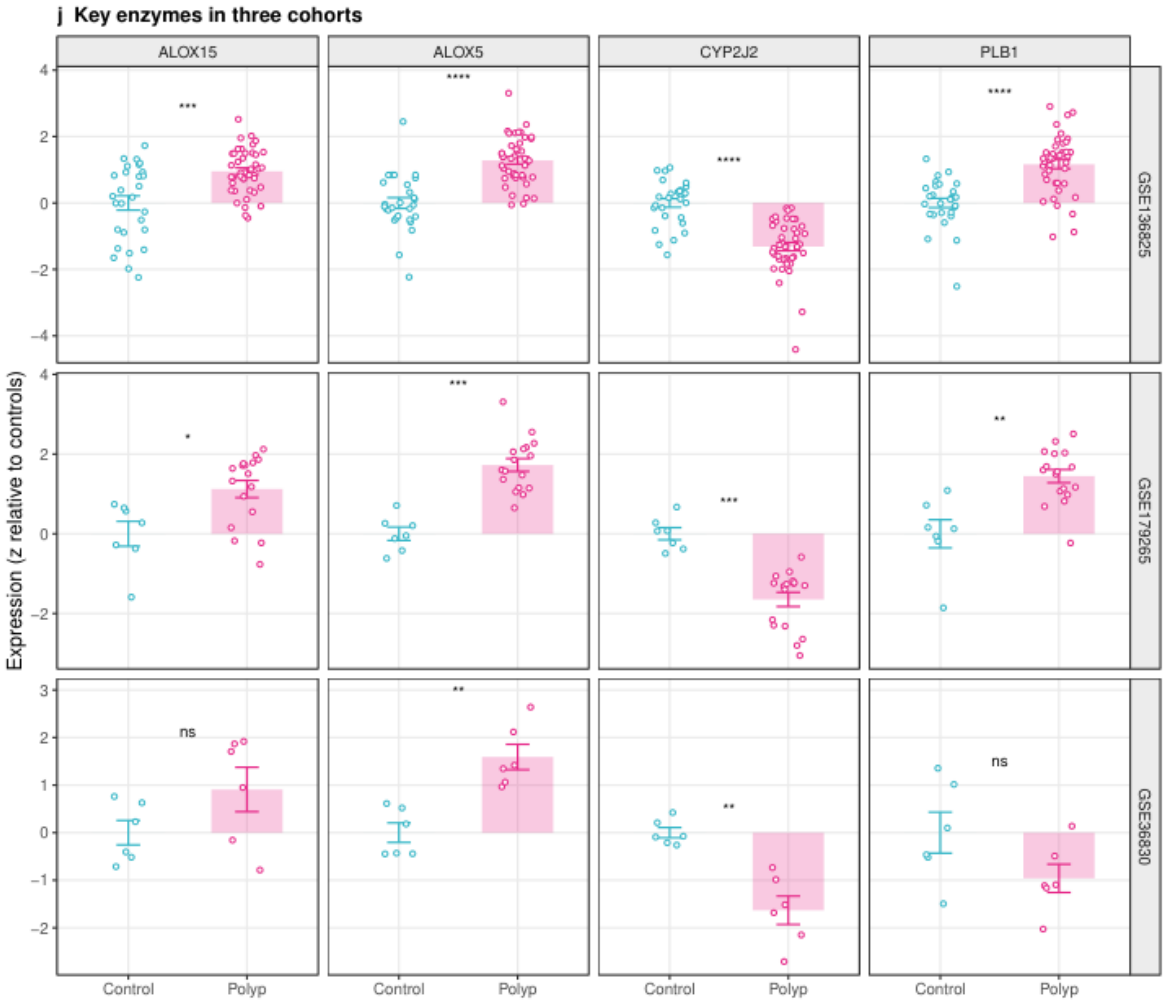
Supplementary Figure S11. MR scatter plots for the four long-chain PUFAs.



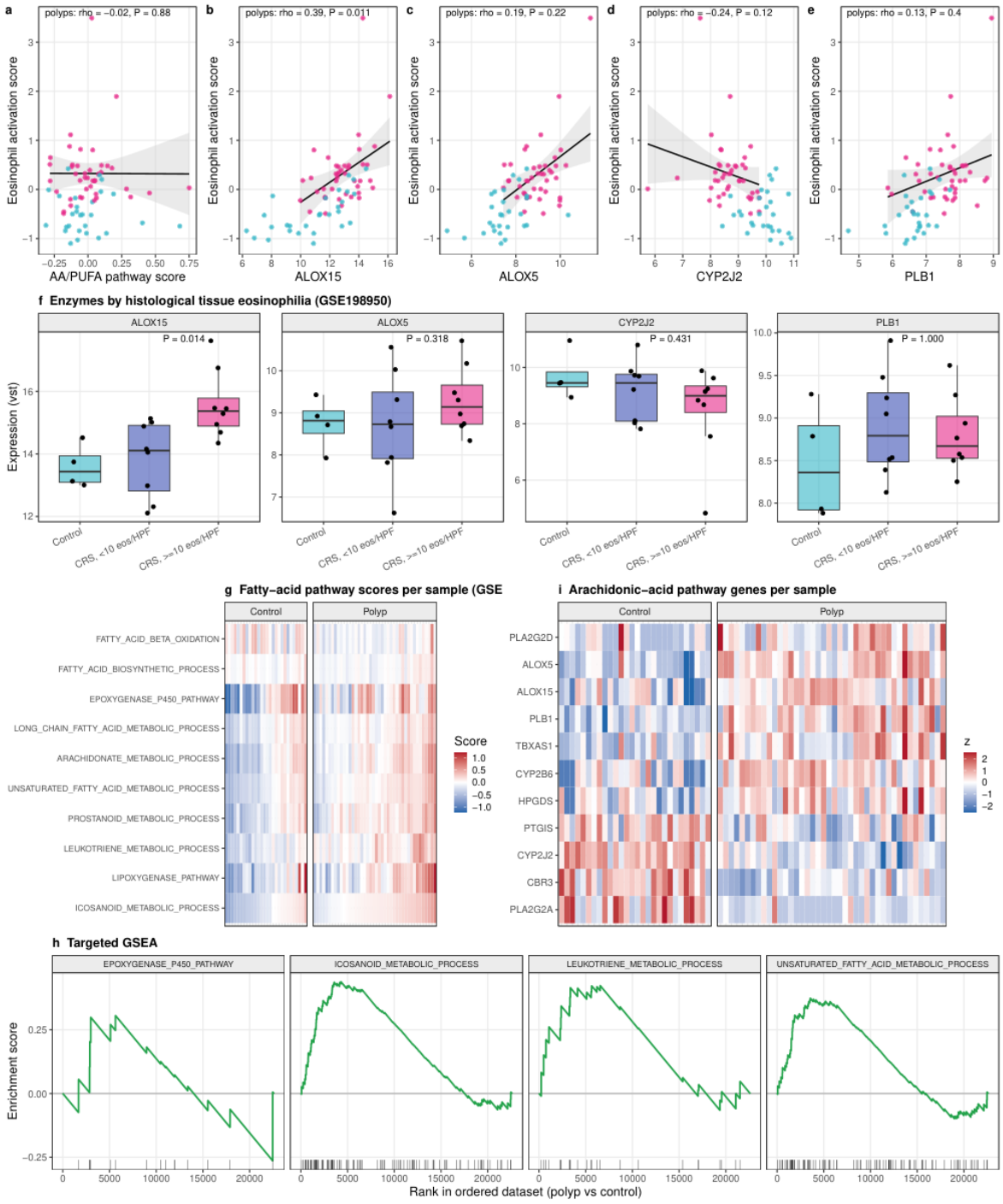
Supplementary Figure S12. Leave-one-out MR analyses.



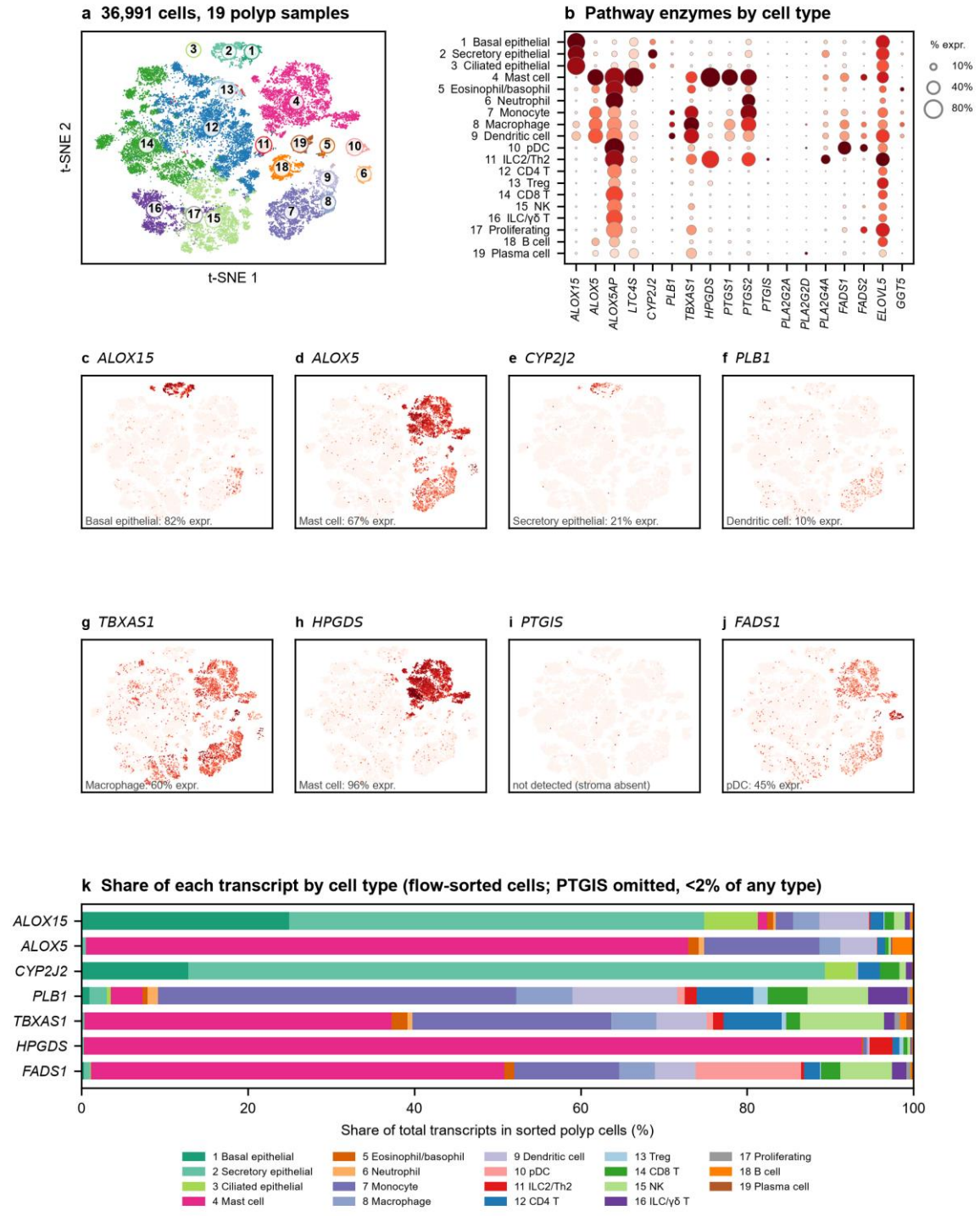
Supplementary Figure S13. LocusCompare, regional and eQTL colocalization plots at the FADS locus.



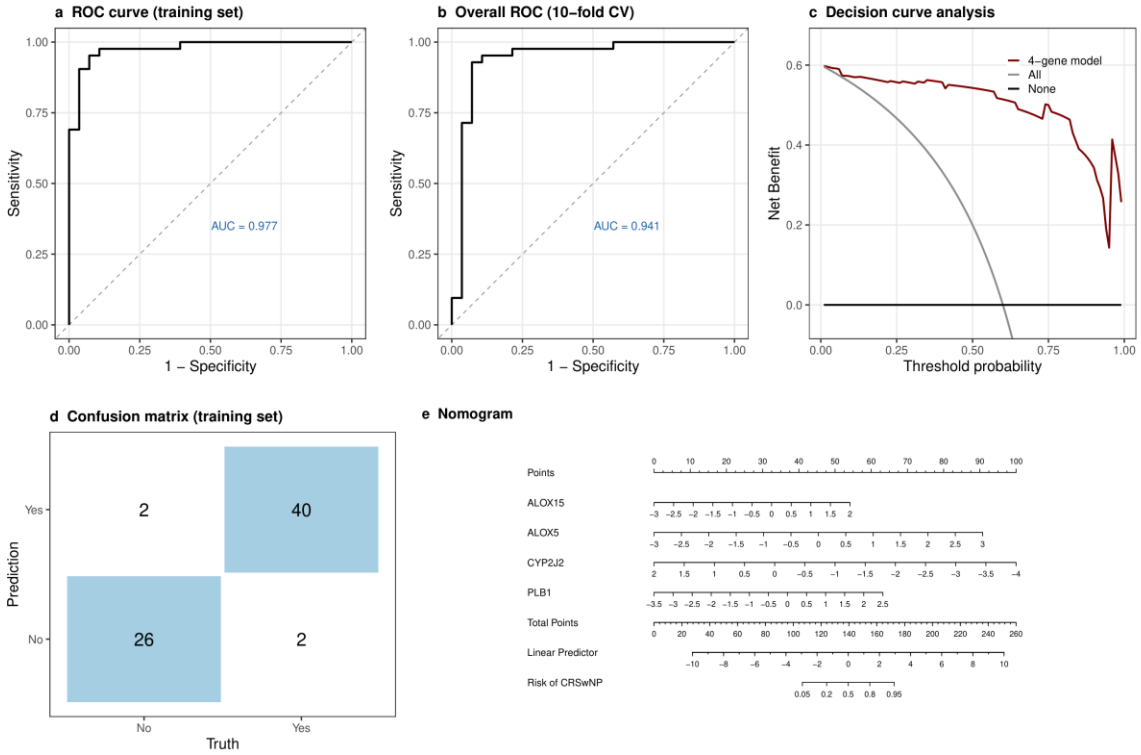
Supplementary Figure S14. Key enzymes in three polyp cohorts.



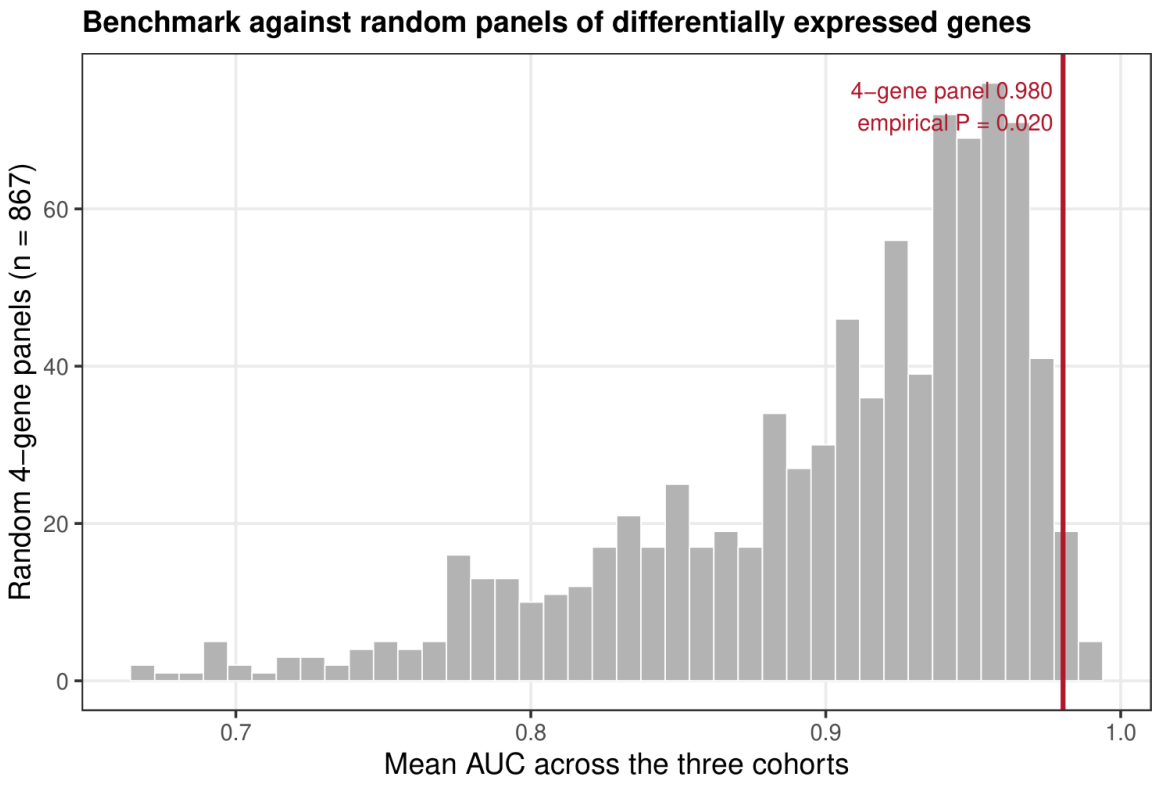
Supplementary Figure S15. Eosinophil activation, histological eosinophilia and fatty acid pathway scores.



Supplementary Figure S16. Single-cell annotation and enzyme localization in GSE196169.



Supplementary Figure S17. Diagnostic model details (ROC, cross-validation, decision curve, confusion matrix, nomogram).



Supplementary Figure S18. Random four-gene set benchmark.

Supplementary Tables

Supplementary Tables SVII–SXVIII are provided in Supplementary_Tables_SVII-SXVIII.xlsx.

Supplementary Table SI. Data sources

Layer	Data source	Samples	Platform	Use
Genetics (exposure)	Chen et al. 2023, GWAS of 1400 plasma metabolites (GWAS Catalog)	8299 individuals	Metabolon HD4	MR instruments
Genetics (outcome)	Saarentaus et al. 2025, CRSwNP (FinnGen and UK Biobank)	9626 cases/685,602 controls	GWAS meta-analysis	Outcome for MR, colocalization and SMR
eQTL reference	GTEx v8 significant cis-eQTLs, European ancestry	Whole blood, lung, esophageal mucosa, spleen, minor salivary gland, EBV-transformed lymphocytes	SMR/HEIDI (BESD format)	Gene-level SMR and colocalization
Nasal secretion metabolome	MTBLS13192 (MetaboLights); raw spectra re-quantified in this study	Type 2 CRS 28 vs. controls 12	UPLC-Q Exactive, DDA	Local metabolites; lipoxygenase products
Polyp transcriptome (discovery)	GSE136825	Polyps 42 vs. inferior turbinate controls 28	RNA-seq (GPL20301)	Differential expression, pathway scores, immune infiltration, model training
External validation 1	GSE179265	Polyps 17 vs. uncinate controls 7	RNA-seq (GPL24676)	Direction check; external validation
External validation 2	GSE36830	Polyps 6 vs. controls 6	Affymetrix HG-U133 Plus 2.0 (GPL570)	Direction check; external validation
Histological eosinophilia	GSE198950	CRS polyps with < 10 vs. ≥ 10 eosinophils/HPF	RNA-seq	ALOX15 and tissue eosinophilia
Single-cell transcriptome	GSE196169 (Bangert et al. 2022)	19 polyps (12 AERD, 7 CRSwNP); 36,991 cells after QC	10x Chromium v2/v3, flow-sorted	Cell localization; lipid-mediator circuits
Single-cell replication	GSE276503 (Xie et al. 2024)	15 polyps, unsorted whole tissue; 92,446 cells after QC	DNBelab C4	Independent replication of localization
Nasal lavage single-cell	GSE299751 (Doni Jayavelu et al. 2025)	6 nasal lavage samples from children with asthma; 6151 cells	10x Flex (fixed-RNA probes)	ALOX15 in eosinophils vs. epithelium
Treatment cohort	GSE267331 (SINUS-52, NCT02898454)	Dupilumab 47 vs. placebo 19; paired baseline and week 24	Affymetrix HG-U133 Plus 2.0 (GPL570)	Restoration of the disease program

All data analyzed in this study are public. Data sources: plasma metabolite GWAS [1]; CRSwNP and asthma GWAS [2]; GTEx v8 [7]; GSE136825 [11]; GSE179265 [14]; GSE36830 [15]; GSE196169 [17]; GSE276503 [18]; GSE299751 [19]; GSE267331 [21]. MTBLS13192 and GSE198950 have no associated publication and are cited by accession number.

Supplementary Table SII. Mendelian randomization and colocalization of four polyunsaturated fatty acids with CRSwNP.

Fatty acid	IVs (n)	Mean F	IVW OR (95% CI)	IVW P	WM OR (95% CI)	MR-Egger OR (95% CI)	Egger intercept P	Cochran's Q P	MR-PRESSO global P	Outliers removed	IVW OR after removal (95% CI)	P after removal	PP4
Arachidonate (20:4n6)	29	34.5	1.14 (1.01–1.28)	0.027	1.24 (1.13–1.37)	1.36 (1.11–1.66)	0.048	1.1×10^{-11}	<0.001	2	1.07 (0.99–1.15)	0.104	0.963
Adrenate (22:4n6)	20	26.7	1.20 (1.07–1.34)	0.00153	1.12 (0.99–1.27)	1.31 (1.04–1.65)	0.396	0.0021	0.001	1	1.08 (1.00–1.18)	0.064	0.959
Docosapentaenoate (n6 DPA; 22:5n6)	15	23.3	1.11 (0.98–1.25)	0.111	1.07 (0.92–1.24)	1.37 (0.95–1.99)	0.252	0.052	0.038	0	1.11 (0.98–1.25)	0.111	0.915
Eicosapentaenoate (EPA; 20:5n3)	24	27.7	1.17 (1.04–1.31)	0.00794	1.06 (0.93–1.21)	1.30 (1.03–1.65)	0.315	4.1×10^{-4}	<0.001	1	1.03 (0.94–1.14)	0.52	0.963

None of the associations remained significant after removal of outlying variants. CI, confidence interval; IVW, inverse-variance weighted; PP4, posterior probability of a shared causal variant.

Supplementary Table SIII. Metabolites reported from nasal secretions and their annotation confidence.

Metabolite	Ion mode	$\log_2$ fold change	P	q	VIP	Annotation confidence (MSI level)
Adrenic acid	NEG	+3.67	3.0×10^{-5}	6.1×10^{-4}	1.66	2 (MS2 acquired)
(+/-)9(10)-EpOME	NEG	+3.28	0.0011	0.00556	1.37	3/4 (MS1 + RT only)
Docosapentaenoic acid (22n-3)	NEG	+2.92	1.5×10^{-5}	4.8×10^{-4}	1.78	2 (MS2 acquired)
Oleic acid	NEG	+1.89	0.00166	0.0071	1.03	2 (MS2 acquired)
all-cis-4,7,10,13,16-Docosapentaenoic acid	POS	+1.38	0.00981	0.025	0.95	2 (MS2 acquired)
Linoleic acid	POS	+1.04	0.00248	0.00943	1.38	3/4 (MS1 + RT only)
Eicosapentaenoic acid	POS	+0.60	0.189	0.252	0.44	2 (MS2 acquired)
Thromboxane B2	POS	−1.10	0.00633	0.019	1.21	2 (MS2 acquired)
15-Keto prostaglandin F2alpha	POS	−1.19	4.7×10^{-4}	0.00321	1.50	2 (MS2 acquired)
Nervonic acid	POS	−1.35	1.0×10^{-3}	0.00515	0.92	2 (MS2 acquired)

MS2 acquisition means that a spectrum was recorded at the corresponding m/z and retention time; it does not imply a spectral library match. VIP, variable importance in projection.

Supplementary Table SIV. Differentially expressed pathway genes in polyp tissue (GSE136825).

Gene	Base mean	$\log_2$ fold change	Adjusted P	Branch
PLA2G2D	29	+2.11	0.00275	Pro-inflammatory / substrate release
ALOX5	421	+1.53	6.7×10^{-8}	Pro-inflammatory / substrate release
ALOX15	8907	+1.30	0.00282	Pro-inflammatory / substrate release
PLB1	134	+1.26	1.7×10^{-8}	Pro-inflammatory / substrate release
TBXAS1	231	+1.20	9.6×10^{-6}	Pro-inflammatory / substrate release
CYP2B6	486	+1.18	0.031	Pro-inflammatory / substrate release
HPGDS	43	+1.04	0.011	Pro-inflammatory / substrate release
PTGIS	164	−1.10	0.014	Resolving or vasodilatory
CYP2J2	647	−1.27	6.8×10^{-11}	Resolving or vasodilatory
CBR3	79	−1.43	9.5×10^{-8}	Resolving or vasodilatory
PLA2G2A	67	−3.83	2.2×10^{-6}	Resolving or vasodilatory

Adjusted P values are from DESeq2.

Supplementary Table SV. Cellular localization of key enzymes.

Gene	Enzyme	Basal (454)	Secretory (968)	Ciliated (134)	Mast (6172)	Eos/Baso (342)	Macro (514)	Mono (2866)	DC (750)	ILC2/Th2 (286)	Main site (GSE196169)	Highest fraction (GSE276503)	Top compartment, samples (GSE276503)
ALOX15	15-Lipoxygenase	82.2	72.7	82.8	0.6	2.9	11.7	1.7	20.0	1.0	Epithelium (> 70% in all three subsets)	Secretory epithelium 90.7%	Epithelium 15/15
CYP2J2	P450 epoxygenase	8.4	21.5	8.2	0.0	0.0	0.0	0.0	0.1	0.0	Secretory epithelium	Secretory epithelium 23.9%	Epithelium 15/15
ALOX5	5-Lipoxygenase	2.4	4.2	1.5	67.3	16.4	45.5	39.4	54.9	2.1	Mast cells	Mast cells 45.1%	Mast cells 15/15
LTC4S	LTC4 synthase	18.9	24.0	23.1	94.6	11.4	16.5	5.3	18.5	10.1	Mast cells	Mast cells 70.9%	Mast cells 15/15
HPGDS	PGD2 synthase	2.6	3.7	1.5	96.5	3.5	5.3	1.8	4.7	76.6	Mast cells; ILC2/Th2	Mast cells 72.4%	Mast cells 15/15
PTGS1	Cyclooxygenase-1	0.2	2.6	3.7	67.6	2.6	30.4	10.5	25.9	2.8	Mast cells	Mast cells 73.5%	Mast cells 15/15
TBXAS1	Thromboxane synthase	1.8	2.3	1.5	34.2	18.1	59.5	47.8	57.2	28.7	Myeloid cells	Dendritic cells 39.8%	Myeloid 15/15
PLB1	Phospholipase B1	1.5	1.6	3.0	0.3	0.3	7.6	6.9	10.4	1.7	Myeloid cells (low fraction)	Dendritic cells 8.0%	Myeloid 15/15
PTGIS	Prostacyclin synthase	0.0	0.3	0.0	0.1	0.0	0.2	0.0	0.0	1.7	Not detected (< 2%)	Fibroblasts 21.8%	Fibroblasts 15/15

Values are the percentage of cells in each cell type that express the gene; numbers in parentheses are cell counts. Basal, Secretory and Ciliated denote epithelial subsets; Eos/Baso, eosinophils/basophils; Macro, macrophages; Mono, monocytes; DC, dendritic cells. Cell-type columns come from GSE196169 (19 polyps, immune cells enriched by flow sorting; proportions do not reflect tissue composition). The last two columns give the replication in unsorted whole tissue (GSE276503, 15 polyps): the cell type with the highest fraction of expressing cells, and the number of samples in which that compartment had the highest expression.

Supplementary Table SVI. Restoration of key enzymes after dupilumab (GSE267331).

Gene	Median signal in brushings	Disease effect (log_2)	Restoration (log_2)	95% CI	Restored fraction	P (difference)	Equivalence margin	P (equivalence)	Interpretation
CYP2J2	5.7	−1.27	0.32	−0.19 to 0.84	25%	0.214	0.64	0.121	Not assessable (below expression threshold)
ALOX15	9.7	+1.30	0.62	0.11 to 1.14	48%	0.019	0.65	0.46	Partly restored
ALOX5	5.1	+1.53	0.18	−0.95 to 1.32	12%	0.744	0.77	0.162	Not assessable (below expression threshold)
PLB1	2.7	+1.26	−0.12	−0.30 to 0.05	−10%	0.168	0.63	2.1×10^{-7}	Not assessable (below expression threshold)
PTGIS	2.3	−1.10	−0.01	−0.03 to 0.00	−1%	0.171	0.55	2.7×10^{-34}	Not assessable (below expression threshold)
TBXAS1	2.4	+1.20	0.13	−0.32 to 0.58	11%	0.559	0.60	0.024	Not assessable (below expression threshold)
HPGDS	2.3	+1.04	0.10	−0.43 to 0.63	10%	0.702	0.52	0.065	Not assessable (below expression threshold)

The equivalence margin was 50% of the disease effect; genes with a median expression ≤ 6 in brushings (array background 2.31) were not interpreted.

Supplementary References

1. Chen Y, Lu T, Pettersson-Kymmer U, et al. Genomic atlas of the plasma metabolome prioritizes metabolites implicated in human diseases. Nat Genet 2023; 55: 44-53.

2. Saarentaus EC, Fischer-Rasmussen K, Sliz E, et al. 131 genetic loci highlight immunological pathways and tissues in nasal polyposis and asthma. Nat Commun 2025; 16: 9879.

3. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. Int J Epidemiol 2015; 44: 512-525.

4. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. Genet Epidemiol 2016; 40: 304-314.

5. Verbanck M, Chen CY, Neale B, Do R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. Nat Genet 2018; 50: 693-698.

6. Giambartolomei C, Vukcevic D, Schadt EE, et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. PLoS Genet 2014; 10: e1004383.

7. GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science 2020; 369: 1318-1330.

8. Zhu Z, Zhang F, Hu H, et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. Nat Genet 2016; 48: 481-487.

9. Dieterle F, Ross A, Schlotterbeck G, Senn H. Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in 1H NMR metabonomics. Anal Chem 2006; 78: 4281-4290.

10. Sumner LW, Amberg A, Barrett D, et al. Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). Metabolomics 2007; 3: 211-221.

11. Peng Y, Zi XX, Tian TF, et al. Whole-transcriptome sequencing reveals heightened inflammation and defective host defence responses in chronic rhinosinusitis with nasal polyps. Eur Respir J 2019; 54: 1900732.

12. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 2014; 15: 550.

13. Wu D, Smyth GK. Camera: a competitive gene set test accounting for inter-gene correlation. Nucleic Acids Res 2012; 40: e133.

14. Nakayama T, Lee IT, Le W, et al. Inflammatory molecular endotypes of nasal polyps derived from White and Japanese populations. J Allergy Clin Immunol 2022; 149: 1296-1308.e6.

15. Stevens WW, Ocampo CJ, Berdnikovs S, et al. Cytokines in Chronic Rhinosinusitis. Role in Eosinophilia and Aspirin-exacerbated Respiratory Disease. Am J Respir Crit Care Med 2015; 192: 682-694.

16. Newman AM, Liu CL, Green MR, et al. Robust enumeration of cell subsets from tissue expression profiles. Nat Methods 2015; 12: 453-457.

17. Bangert C, Villazala-Merino S, Fahrenberger M, et al. Comprehensive Analysis of Nasal Polyps Reveals a More Pronounced Type 2 Transcriptomic Profile of Epithelial Cells and Mast Cells in Aspirin-Exacerbated Respiratory Disease. Front Immunol 2022; 13: 850494.

18. Xie X, Wang P, Jin M, et al. IL-1β-induced epithelial cell and fibroblast transdifferentiation promotes neutrophil recruitment in chronic rhinosinusitis with nasal polyps. Nat Commun 2024; 15: 9101.

19. Doni Jayavelu N, Liu AH, Gaberino C, et al. Single-cell transcriptomic profiling of eosinophils and airway immune cells in childhood asthma. J Allergy Clin Immunol 2025; 156: 923-936.

20. Efremova M, Vento-Tormo M, Teichmann SA, Vento-Tormo R. CellPhoneDB: inferring cell-cell communication from combined expression of multi-subunit ligand-receptor complexes. Nat Protoc 2020; 15: 1484-1506.

21. Gayvert K, Desrosiers M, Laidlaw TM, et al. Nasal brushing molecular endotyping distinguishes patients with chronic rhinosinusitis with nasal polyps with better response to dupilumab. J Allergy Clin Immunol 2024; 154: 619-630.

22. Lakens D. Equivalence Tests: A Practical Primer for t Tests, Correlations, and Meta-Analyses. Soc Psychol Personal Sci 2017; 8: 355-362.