



RsiingSI · Independent Research

A peptidomics-informed computational assessment of porcine-spleen transfer factor against allergic rhinitis: the parent-protein layer is coordinately suppressed in diseased nasal epithelium, but no drug–target evidence emerges

Agent Team¹ · scalingX¹

¹RsiingSI — the open-source org scaling recursive self-improvement for bioscience.

Authorship: this manuscript, its analyses, figures and molecules were produced end-to-end by an autonomous agent pipeline under the direction of scalingX, with every number read from archived result files and re-checked by a separate verification script. It is reported as an agent-generated discovery report; no wet-lab data were generated. All code and derived tables are archived with the manuscript.

Computational hypothesis-generating study · 2026-10-20 · All analyses run in Python 3.11 (pydeseq2 0.5.4, no R/Bioconductor); every number in this manuscript is machine-generated from [results/paper_facts.json](#) and reproducible from the deposited GEO data.

Abstract. Transfer factor (TF) is a low-molecular-weight peptide mixture dialysed from leukocyte lysates; the marketed oral preparation is produced from porcine spleen. Because TF is a mixture without a single defined chemical structure, the classical “compound–target” network-pharmacology pipeline cannot be applied. We therefore inverted the design: the analysis starts from published peptidomics of porcine-spleen TF (Dua et al. 2021: 163 peptides mapping to 23 proteins, from which we curate a 19-gene core parent-protein set, UniProt-verified) and terminates in disease transcriptomics (three public cohorts: nasal epithelium microarrays GSE19187 n=38 and GSE43523 n=12, and nasopharyngeal RNA-seq GSE329964 n=25). The parent-protein layer behaves as a **coordinated programme**: in the discovery cohort its mean moderated t is -0.93 versus 0.06 for all other genes ($\Delta = -0.99$, gene-label permutation $p < 5 \times 10^{-5}$; weighted KS enrichment score -0.65, $p = 0.0014$; 13/17 genes negative, sign test $p = 0.025$), i.e. the programme is **down-regulated in allergic nasal epithelium**. The direction is reproduced in an independent nasopharyngeal allergic cohort ($\Delta = -0.54$, ns) and is absent in the non-allergic nasopharyngeal contrast ($\Delta = -0.02$, $p = 0.9486$) and in an independent adult seasonal-AR cohort ($\Delta = -0.13$, $p = 0.6042$). No individual parent gene survives FDR anywhere, and the overlap of the TF set with FDR-defined DEGs does not exceed the permutation null. In the exploratory in-silico peptidome, 696 tryptic peptides of the parent proteins fall inside the 5–24 aa window actually observed in porcine-spleen TF, and 13.1% of their 9-mers match the canonical HLA-A*02:01 anchor pattern. We conclude that the analysis generates one testable, directionally specific hypothesis — nasal epithelium of allergic rhinitis coordinately under-expresses the parent proteins of TF peptides — and explicitly does **not** constitute evidence of efficacy; no cohort contains TF exposure, and parent-protein transcripts are not peptide activity.

Keywords: transfer factor; allergic rhinitis; peptidomics; competitive gene-set test; moderated t; ssGSEA; diagnostic ranking; hypothesis generation.

What this study is not. This is a computational, hypothesis-generating analysis of public transcriptomes. It contains no TF exposure experiment, no animal model, no clinical endpoint, and no claim that transfer factor is effective against rhinitis. The only defensible reading of a significant programme-level signal is directional: if the parent-protein layer is suppressed in disease tissue, then replenishing it is *consistent with* — not evidenced by — the observed data.

1 Background

Transfer factor is historically defined by function — the dialysable leukocyte extract that transfers delayed-type hypersensitivity — and commercially produced from porcine spleen as a peptide mixture (<10 kDa). Two properties make it resistant to the standard “new use for an old drug” computational workflow: it has no single chemical structure, and no curated target spectrum exists in DrugBank/ChEMBL. The pragmatic way out, adopted here, is to anchor the analysis in what is published: the peptide composition of porcine-spleen TF.

Dua et al. characterized a commercial porcine-spleen TF (Imuno TF®) by LC-MS/MS and reported 163 peptides assigned to 23 proteins (haemoglobin chains, transferrin, talin-1, L-plastin, PRMT1, ubiquitin/ribosomal fusions, and others) [1]. A Chinese veterinary-pharmaceutical study of porcine-spleen TF reported 102 immune-related peptides mapping to 64 immune-active proteins, 5–24 aa, 599–2508 Da [2]; that table has no machine-readable public version, so we use it only as a qualitative anchor for the peptide-length window, never as a gene source. Functional peptide-level work (enzyme-digested porcine-spleen peptides rescuing cyclophosphamide-immunosuppressed mice) exists but carries no mappable parent genes [3]. The analysis is therefore a single forward chain: *published TF peptides* → *parent-protein gene set* → *behaviour of that gene set in allergic-rhinitis transcriptomes*.

2 Methods

2.1 Design

Three questions are kept strictly separate (Fig 1): (i) is the parent-protein layer *measured* in the target tissues; (ii) does it move *as a programme* between disease and control; (iii) is its overlap with differentially expressed genes larger than chance? Question (ii) uses competitive gene-set statistics calibrated by gene-label permutation because the discovery cohort has essentially no FDR-significant genes and single-gene thresholding would be meaningless.

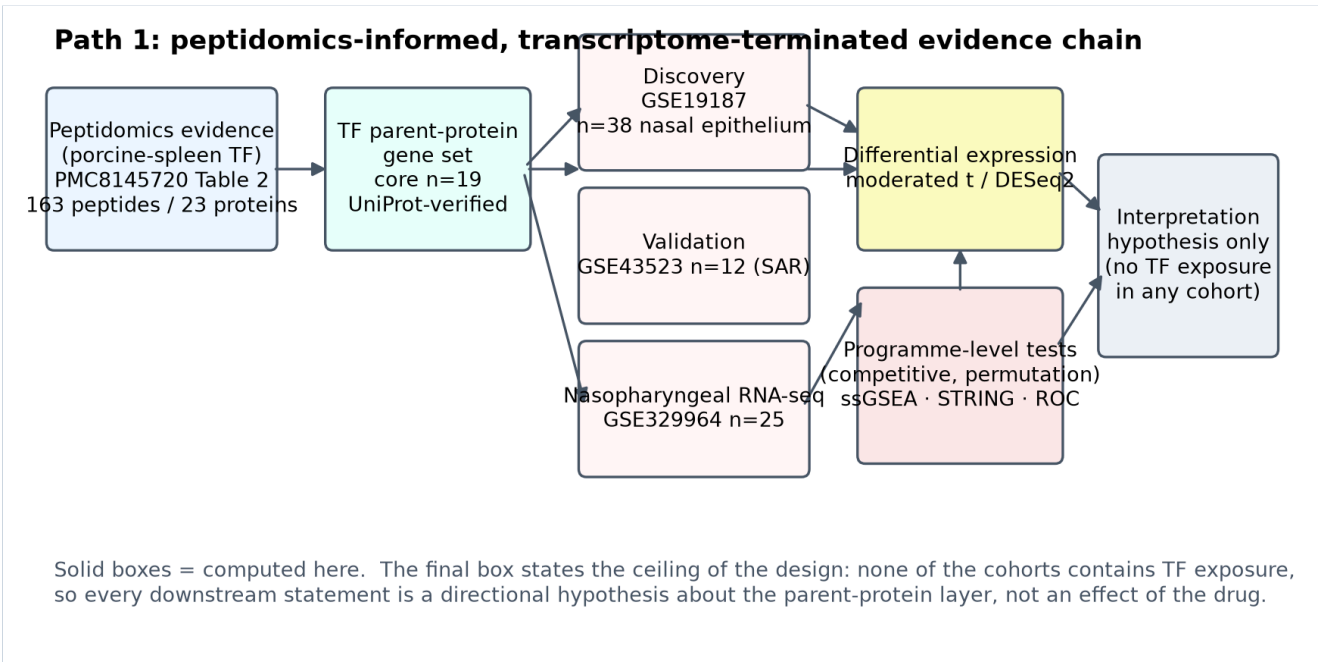


Figure 1. Evidence chain. Published TF peptidomics → audited parent-protein gene set → three public disease-transcriptome cohorts → programme-level statistics. The red block states the design ceiling: no cohort contains TF exposure.

2.2 Parent-protein gene set and its provenance audit

From the 23 proteins of [1], we map porcine accessions to human orthologues via UniProt. Accessions that could not be resolved to an HGNC symbol (5 entries: two Ensembl-porcine scaffolds, one LOC symbol, one reported human symbol absent from UniProtKB, one MAGE-like without HGNC symbol) were dropped. The audit also flagged 5 accessions whose UniProt entries are *human* despite the source paper stating a porcine search database — recorded as a provenance observation, not silently corrected. The core set contains 19 symbols (ACYP2, AGPAT5, EMID1, FAM184A, FAM19A2, GIMAP7, HBA1, HBA2, HBB, HPCA, LCP1, PRMT1, RBM25, RPS27A, SREK1, TFAA2, TF, TLN1, UBA52) — 18 distinct HGNC genes, because FAM19A2 is the former symbol of TFAA2 (one deposited protein, A0A287ALO8, carries both aliases; the two symbols are mutually exclusive in current platform annotations, so no gene is counted twice within any platform). Four additional genes (TMSB4X, TIAL1, UBB, UBC) that the literature associates with TF but that we could not verify against [1] are kept in a separate extended set (23 genes) and are excluded from every headline result.

Deposited accession	Reported gene	UniProt organism	Human symbol	Human UniProt	Decision	Reason
P01965	HBA	Sus scrofa	HBA1	P69905	KEEP	pig HBA -> human alpha-globin paralog family (HBA1/HBA2); counted once
P02067	HBB	Sus scrofa	HBB	P68871	KEEP	human ortholog resolved (reviewed)
F1SFZ8	TLN1	Sus scrofa	TLN1	Q9Y490	KEEP	human ortholog resolved (reviewed)
A0A287AZA7	RPS27A	Sus scrofa	RPS27A	P62979	KEEP	human ortholog resolved (reviewed)
A0A5K1UHC4	AGPAT5	UPID not found in UniProtKB	AGPAT5	Q9NUQ2	KEEP	human ortholog resolved (reviewed)
Q06AT1	HPCA	Sus scrofa	HPCA	P84074	KEEP	human ortholog resolved (reviewed)
A0A480JNE4	FAM184A	Sus scrofa	FAM184A	Q8NB25	KEEP	human ortholog resolved (reviewed)
P49756	RBM25	Homo sapiens	RBM25	P49756	KEEP	human ortholog resolved (reviewed)
P09571	TF	Sus scrofa	TF	P02787	KEEP	human ortholog resolved (reviewed)
Q8WXA9	SREK1	Homo sapiens	SREK1	Q8WXA9	KEEP	human ortholog resolved (reviewed)
Q96A84	EMD1	Homo sapiens	EMID1	Q96A84	KEEP	human ortholog resolved (reviewed)
A0A287AAR0	GIMAP7	Sus scrofa	GIMAP7	Q8NHV1	KEEP	human ortholog resolved (reviewed)
LOC/100622504	LOC100622504	UPID not found in UniProtKB	-	-	DROP	no HGNC gene symbol / uncharacterized protein -> no transcriptome-level
P00819	ACYP2	Sus scrofa	ACYP2	P14621	KEEP	human ortholog resolved (reviewed)
P63053	UBA52	Sus scrofa	UBA52	P62987	KEEP	human ortholog resolved (reviewed)
P13796	LCP1	Homo sapiens	LCP1	P13796	KEEP	human ortholog resolved (reviewed)
A0A5G2Q920	ENSSSCG00000049439	UPID not found in UniProtKB	-	-	DROP	no HGNC gene symbol / uncharacterized protein -> no transcriptome-level
Q99873	PRMT1	Homo sapiens	PRMT1	Q99873	KEEP	human ortholog resolved (reviewed)

Deposited accession	Reported gene	UniProt organism	Human symbol	Human UniProt	Decision	Reason
A0A287AL08	TAF2	Sus scrofa	TAF2	Q7TPG7	KEEP	human ortholog resolved (reviewed)
A0A5G2QDH3	ENSSSCG00000051012	UPID not found in UniProtKB	-	-	DROP	no HGNC gene symbol / uncharacterized protein -> no transcriptome-level
F1SVA9	SH2D6	Sus scrofa	SH2D6	-	DROP	no human UniProtKB entry for reported symbol
A0A287AHD6	MAGE-like	Sus scrofa	-	-	DROP	no HGNC gene symbol / uncharacterized protein -> no transcriptome-level

Table 1. Full provenance record of the published parent proteins: deposited accession, reported name, the organism of the resolved UniProt entry (five entries are human despite the source paper's stated porcine search), the human symbol used here, and the curation decision. The four extension genes are excluded from this table and from all headline tests.

2.3 Cohorts

Cohort	Accession	Platform	Tissue / site	Groups	Genes analysed	Reference
GSE19187	GSE19187	GPL6244 (Affymetrix HuGene-1.0-ST)	nasal airway epithelium brush (children)	C n=11, CA n=7, R n=14, UA n=6	20003	Guillon et al., PLoS One 2011, PMID 22005912
GSE43523	GSE43523	GPL6883 (Illumina HumanRef-8 v3)	nasal epithelial cells (adults, seasonal AR)	CTRL n=5, SAR n=7	17544	GEO series matrix
GSE329964	GSE329964	GPL24676 (Illumina RNA-seq, raw counts)	nasopharyngeal swab (children)	AH n=10, AHAR n=9, HC n=6	23870	Pu et al. 2025 (nasopharyngeal transcriptome), PMID 42591655

GSE19187 (dust-mite-allergic children, nasal epithelial brush) serves as the discovery cohort; GSE43523 (adult seasonal AR, nasal epithelium) as an independent array validation; GSE329964 (nasopharyngeal swabs) separates allergic (AHAR) from non-allergic (AH) inflammation. Primary contrast: AR = isolated rhinitis + rhinitis-with-controlled-asthma (n=21) versus healthy (n=11), covariates age and sex; uncontrolled-asthma samples are analysed separately as a severity stratum. GSE329964 raw counts are filtered (sum ≥ 10, >1 count in ≥3 samples).

2.4 Differential expression

Array data (log2 values as deposited; per-sample quantile spread 0.09–0.37 — no further normalisation) are analysed with a linear model `expression ~ group + age + sex` and a from-scratch Smyth (2004) empirical-Bayes moderated t [6]: prior degrees of freedom $d_0 = 4.39$, prior variance $s_0^2 = 0.0626$ (moment-matched with the corrected trigamma/digamma equations; validated to $<1 \times 10^{-9}$ against `scipy.special`). RNA-seq counts are analysed with pydeseq2 0.5.4 (negative-binomial GLM, Wald test) [7]. Decision rule: $FDR < 0.05$ & $|\log_2FC| > 0.5$ (BH [11]), with $|\log_2FC| > 1$ as sensitivity. Moderated and ordinary t correlate at $\rho = 0.997$, so moderation does not manufacture significance.

2.5 Programme-level tests

For each contrast and each gene set (core, extended): (a) Δ mean moderated t (set – rest); (b) Δ mean rank; (c) weighted Kolmogorov–Smirnov enrichment score on the moderated-t ranking. Each statistic is calibrated by 20,000 gene-label permutations drawing sets of equal size from the tested universe; permutation p-values carry the +1/(n+1) correction and are never reported as zero.

2.6 Enrichment

Local copies of Enrichr libraries (KEGG 2021, GO 2023 ×3, Reactome 2022, MSigDB Hallmark 2020) are tested with a self-implemented hypergeometric test ($5 \leq$ set size ≤ 500 , BH-corrected). Because the array cohorts yield almost no FDR-defined DEGs, enrichment of disease expression uses the nominal top-300 genes by |t| and is labelled as such; the RNA-seq contrast uses its FDR list (n=2,504).

2.7 Immune deconvolution

Single-sample GSEA [8] with 22 hand-curated marker sets (Bindea 2013 / Charoentong 2017 / MCP-counter style [9,10]; the marker lists are ours, not downloaded — a stated limitation, since several sets contain only 3–6 genes). Signature differences use Mann–Whitney U (BH-corrected across signatures); the association of the TF parent-protein programme score (mean gene-wise z) with each signature uses Spearman correlation. The T-cell exhaustion/checkpoint panel (CD96, CD200R1, TRAF1 plus classical checkpoint genes) is reported group-wise; the CD96/CD200R1/TRAF1 axis is taken from the project brief (diagnostic AR markers) with the citation flagged as to-be-verified.

2.8 Network, diagnostic ranking and overlap null

A functional-association network is downloaded live from STRING v12 (combined score ≥ 0.400; nodes = TF layer + exhaustion panel + a bounded disease layer) [12]; hubs are ranked by degree, and TF-to-hub proximity by shortest paths. Diagnostic value is ranked by ROC AUC with bootstrap 95% CI and a 2,000-permutation label test per gene. The $TF \cap DEG$ overlap is additionally tested against a permutation null (5,000 random sets of the same size from the tested universe).

2.9 Exploratory in-silico peptidome

Reviewed human sequences for 21/23 parent genes (10,162 aa) are digested in silico (trypsin: cleave after K/R not before P) and restricted to the 5–24 aa window observed in real porcine-spleen TF [2]. The 9-mers inside these peptides are screened against the canonical HLA-A*02:01 anchor pattern (P2 hydrophobic AND C-terminal hydrophobic/aromatic) [13]. This is a rule-based feasibility screen — **not** a validated binding predictor; NetMHCpan could not be installed in the analysis environment and the actual peptide sequences of the commercial product are not public, so no per-peptide binding claim is made.

2.10 Reproducibility

All analyses run in Python 3.11.17 (pandas 2.3.3, numpy 2.4.6); no R/Bioconductor is installed, so every statistical method is either from-scratch (with unit checks) or a published faithful re-implementation (pydeseq2). Random seeds are fixed (20251018 for gene-set permutations, 777 for ROC); scripts are numbered 01–10 and re-run end-to-end to produce every table and figure.

3 Results

3.1 The discovery cohort carries almost no FDR-level signal

Of 20,003 genes tested in GSE19187 (AR vs healthy), 1 reaches $FDR < 0.05$ — CST1 ($\log_2FC$ 5.79, $p = 3.7e-09$, $FDR = 7.4e-05$) — while 1,017 genes have nominal $p < 0.05$ and the empirical null fraction is ≈ 0.96 (Fig 2). The p-value histogram is close to uniform. This is the single most important number in the study: any analysis that dichotomises this cohort at $FDR < 0.05$ is analysing one gene, so all programme-level conclusions are drawn from the continuous statistics instead. GSE43523 ($n=12$) yields zero FDR DEGs — expected at this sample size and reported as such. GSE329964 yields 2,504 DEGs for AHAR vs HC (2170 up / 334 down) and 264 for the non-allergic AH contrast.

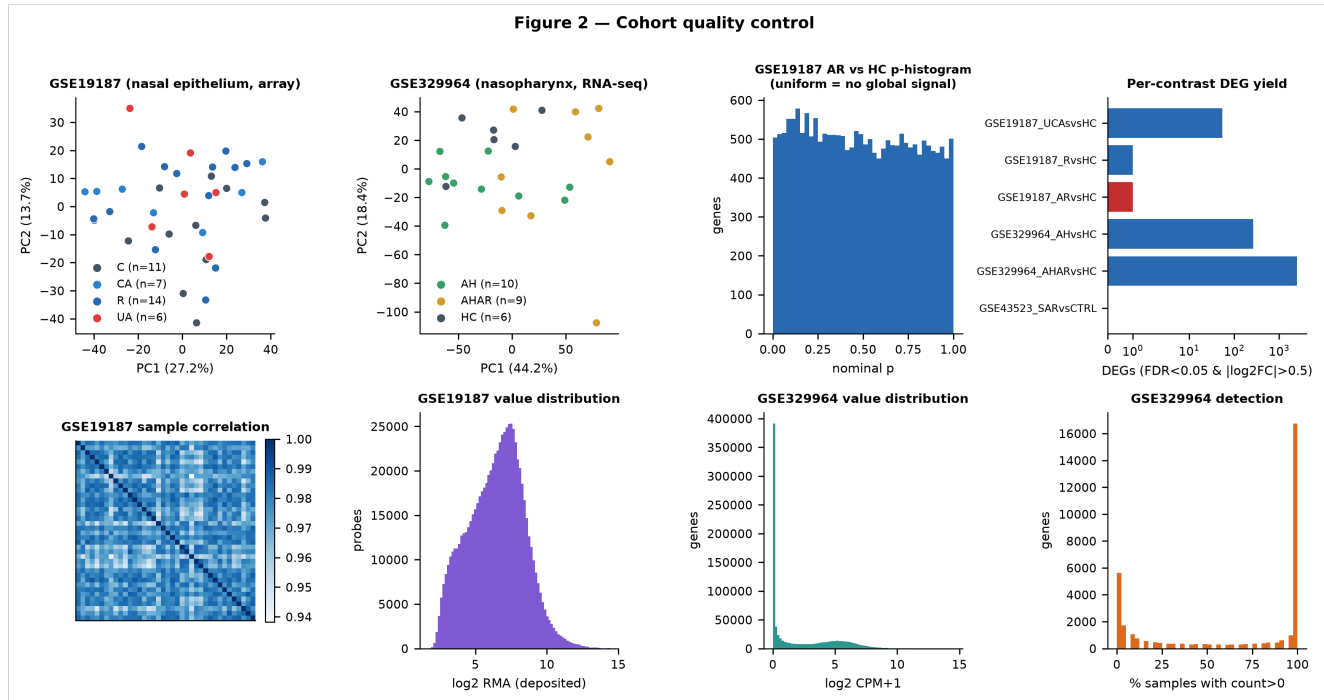


Figure 2. Cohort QC. Left: sample PCA of both cohorts coloured by diagnostic group. Middle: p-value histogram of the primary contrast (near-uniform = no global signal at FDR level) and per-contrast DEG yields on a symlog scale. Right: sample-correlation heatmap and value/detection distributions.

3.2 The TF parent-protein layer moves as a programme — downwards

In the discovery cohort the mean moderated t of the 17-gene core set is -0.93 versus 0.06 for all other tested genes ($\Delta = -0.99$; permutation $p < 5 \times 10^{-5}$; KS-ES -0.65 , $p = 0.0014$). Both statistics agree, and the effect is not driven by outliers: 13 of 17 measured genes move downwards (sign test $p = 0.025$), with the largest single effects on the haemoglobin genes (HBA2 $\log_2FC -0.98$, HBB -1.11 , both nominal $p < 0.03$). The same programme-level downward shift appears for isolated rhinitis alone and for the uncontrolled-asthma stratum (Table), i.e. it is not an asthma artefact.

Figure 4 — The TF parent-protein layer behaves as a programme, not as isolated genes

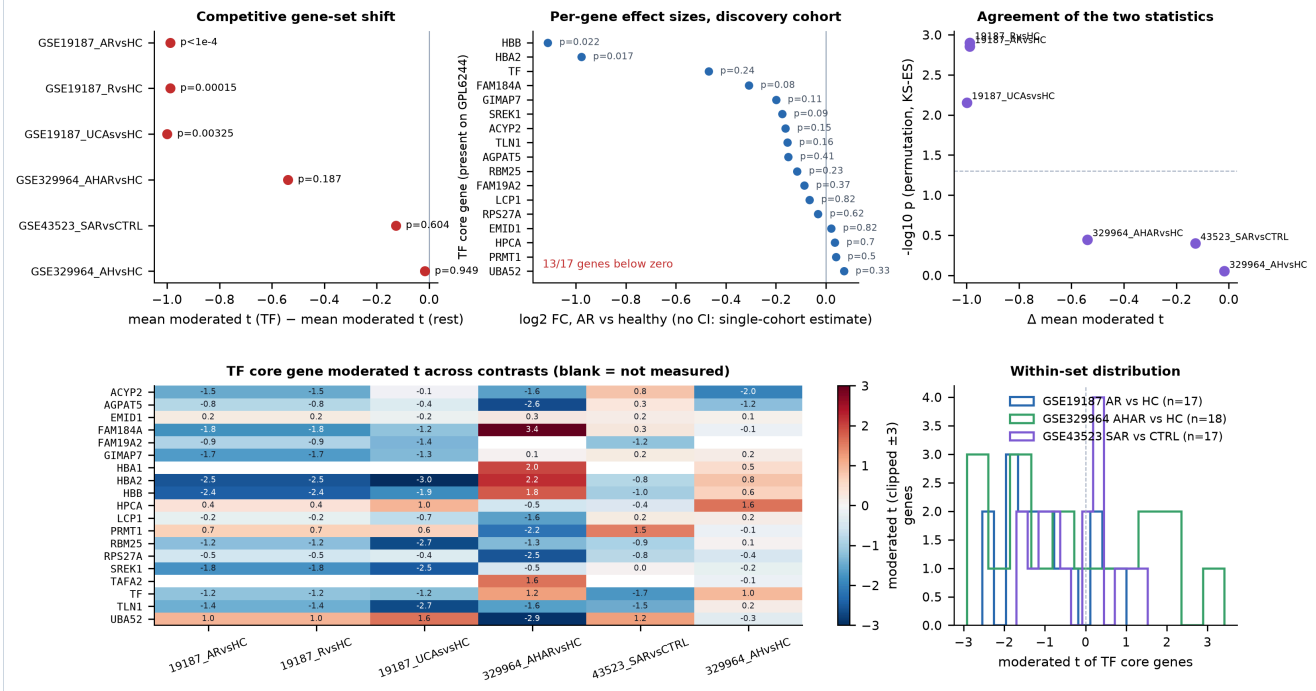


Figure 4. The parent-protein programme. Left: Δ mean moderated t with permutation p -values. Centre: per-gene $\log_2 FC$ in the discovery cohort (13/17 below zero). Right: agreement of the two competitive statistics; bottom: moderated t of every measured core gene across contrasts, and the within-set distribution per cohort.

Cross-cohort consistency is partial — and informative. The direction reproduces in the independent nasopharyngeal allergic contrast ($\Delta = -0.54$, $p = 0.1871$), but vanishes in the same cohort's non-allergic contrast ($\Delta = -0.02$, $p = 0.9486$) and in the adult seasonal-AR array ($\Delta = -0.13$, $p = 0.6042$). The honest summary: a coherent, permutation-significant downward shift in one tissue and age group; direction-consistent but non-significant in a second allergic tissue; no signal in non-allergic inflammation or in the small adult validation.

Contrast	Set size	mean t (set)	mean t (rest)	Δ mean t	p (perm, Δt)	KS-ES	p (perm, ES)
GSE19187_ARvsHC	17	-0.927	0.061	-0.988	$<5 \times 10^{-6}$	-0.650	0.0014
GSE19187_RvsHC	17	-0.927	0.061	-0.988	0.00015	-0.650	0.00125
GSE19187_UCAsvsHC	17	-0.956	0.043	-1.000	0.00325	-0.614	0.007
GSE329964_AHARvsHC	18	-0.266	0.273	-0.539	0.1871	-0.351	0.3583
GSE43523_SARvsCTRL	17	-0.205	-0.077	-0.127	0.6042	-0.410	0.3999
GSE329964_AHvsHC	18	0.040	0.057	-0.017	0.9486	-0.241	0.8891

Table. Competitive gene-set tests, core set. Permutation p -values carry the $+1/(n+1)$ correction; " $<5 \times 10^{-6}$ " denotes zero exceedances in 20,000 permutations.

3.3 The overlap with FDR-defined DEGs is not larger than chance

Across contrasts the intersection of the TF set with FDR-defined DEG lists is 0–2 genes and never exceeds the permutation expectation (AHAR vs HC: core-set overlap 2 against 1.89 expected by Fisher; the 23-gene extended set gives 3 observed against 2.42 expected under permutation, $p = 0.4474$; uncontrolled asthma: 0 observed, 0.055 expected). At the nominal level the only recurring names are the haemoglobin genes (HBA2/HBB in the discovery cohort) and, in the RNA-seq contrast, FAM184A and UBA52. This is a negative result and is reported as a negative result: the TF set does **not** concentrate among the strongest disease genes, and the "TF genes $\cap$ DEGs $\rightarrow$ pathway" narrative that the standard workflow would have produced is not supported here.

Figure 3 — Differential expression landscape (TF parent proteins circled)

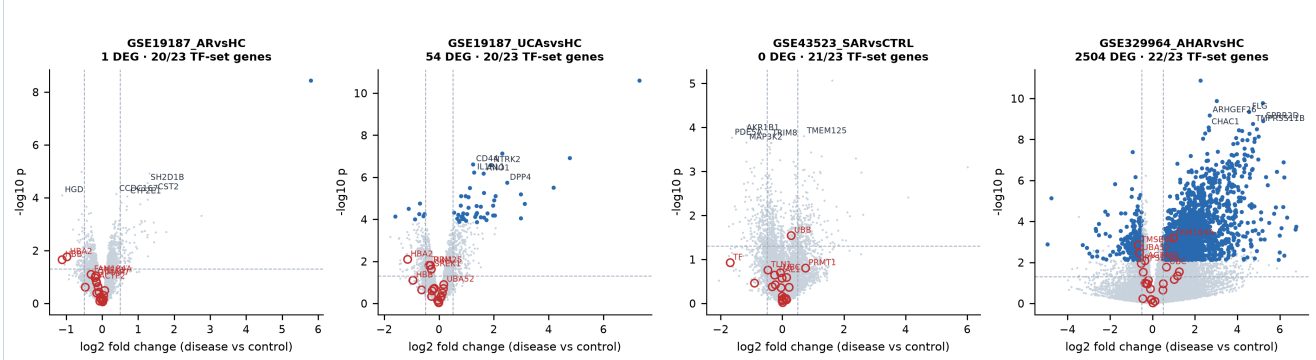


Figure 3. Volcano plots with TF parent proteins circled. No parent gene reaches FDR significance in any contrast; the discovery cohort contributes a single FDR DEG (CST1).

Contrast	FDR DEGs	TF∩DEG	expected	perm. mean	perm. p	overlap genes
GSE19187_ARvsHC	1	0	0.00	n/a	n/a	—
GSE19187_RvsHC	1	0	0.00	n/a	n/a	—
GSE19187_UCAsvsHC	54	0	0.05	0.06	1	—
GSE43523_SARvsCTRL	0	0	n/a	n/a	n/a	—
GSE329964_AHARvsHC	2504	2	1.89	2.42	0.4474	FAM184A;UBA52
GSE329964_AHvsHC	264	0	0.20	0.27	0.2354	—

Table. TF∩DEG overlap against the permutation null (extended set, n=23; Fisher exact on the core set is reported in the results tables).

3.4 Pathway enrichment: what the disease signal, and the TF set, look like

The RNA-seq AHAR DEG list (n=2,504) enriches for epithelial and secretory programmes — keratinisation, cornification, antimicrobial/secretory function — rather than for classical type-2 inflammation modules, consistent with a swab-based epithelial sample (Table; Fig 5). The descriptive annotation of the TF parent-protein set itself shows, unsurprisingly, oxygen transport (HBA/HBB) and ubiquitin-dependent translation/cell-cycle terms (RPS27A/UBA52/UBB/UBC): the set is dominated by abundant housekeeping and blood proteins, which is exactly why its behaviour in epithelial transcriptomes must be interpreted with the blood/epithelium caveat below.

Library	Term	overlap	fold	p
KEGG_2021	Ribosome	38	2.8	2.0e-09
KEGG_2021	Coronavirus disease	47	2.2	1.3e-07
KEGG_2021	Tight junction	34	2.0	4.8e-05
KEGG_2021	IL-17 signaling pathway	20	2.2	0.0006619
KEGG_2021	Axon guidance	32	1.8	0.001046
MSigDB_Hallmark_2020	Estrogen Response Early	46	2.3	1.0e-07
MSigDB_Hallmark_2020	TNF-alpha Signaling via NF-kB	40	1.9	3.6e-05
MSigDB_Hallmark_2020	p53 Pathway	39	1.9	7.0e-05
MSigDB_Hallmark_2020	IL-2/STAT5 Signaling	37	1.8	0.0002888
MSigDB_Hallmark_2020	UV Response Dn	27	1.8	0.001372
Reactome_2022	Keratinization R-HSA-6805567	41	3.8	6.8e-15
Reactome_2022	Eukaryotic Translation Elongation R-HSA-156842	37	4.0	1.9e-14
Reactome_2022	Eukaryotic Translation Termination R-HSA-72764	35	3.8	6.9e-13
Reactome_2022	Peptide Chain Elongation R-HSA-156902	34	3.9	8.3e-13
Reactome_2022	Nonsense Mediated Decay (NMD) Independent Of Exon Junction Complex (EJ	35	3.7	1.5e-12

Table. Top hypergeometric hits for the AHAR vs HC FDR DEG list (local Enrichr libraries).

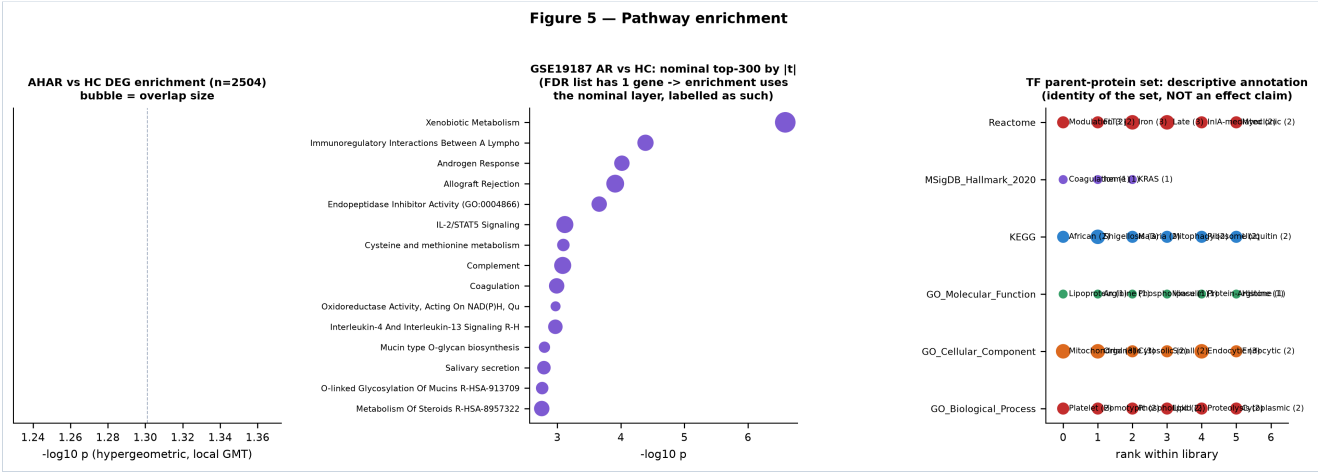


Figure 5. Enrichment landscapes: AHAR DEGs (left), the discovery cohort's nominal top-300 layer (centre), and the descriptive identity of the TF parent-protein set (right).

3.5 Immune microenvironment

Immune deconvolution of the discovery cohort shows goblet/secretory-cell expansion ($\Delta z = +1.20$, MW $p = 0.002564$) and dendritic-cell contraction in AR, with mast-cell, pDC and neutrophil shifts confined to the uncontrolled-asthma stratum (Table). In the nasopharyngeal cohort no signature survives BH correction. The exhaustion/checkpoint panel is expressed at low levels in nasal epithelium and shows no AR-specific pattern after correction; the literature-flagged diagnostic axis (CD96/CD200R1/TRAF1) does not reproduce as a tissue-level signal here, although CD200R1 is the strongest single-gene diagnostic marker of the allergic nasopharyngeal contrast (Section 3.7).

Figure 6 — Immune microenvironment and its relation to the TF parent-protein layer

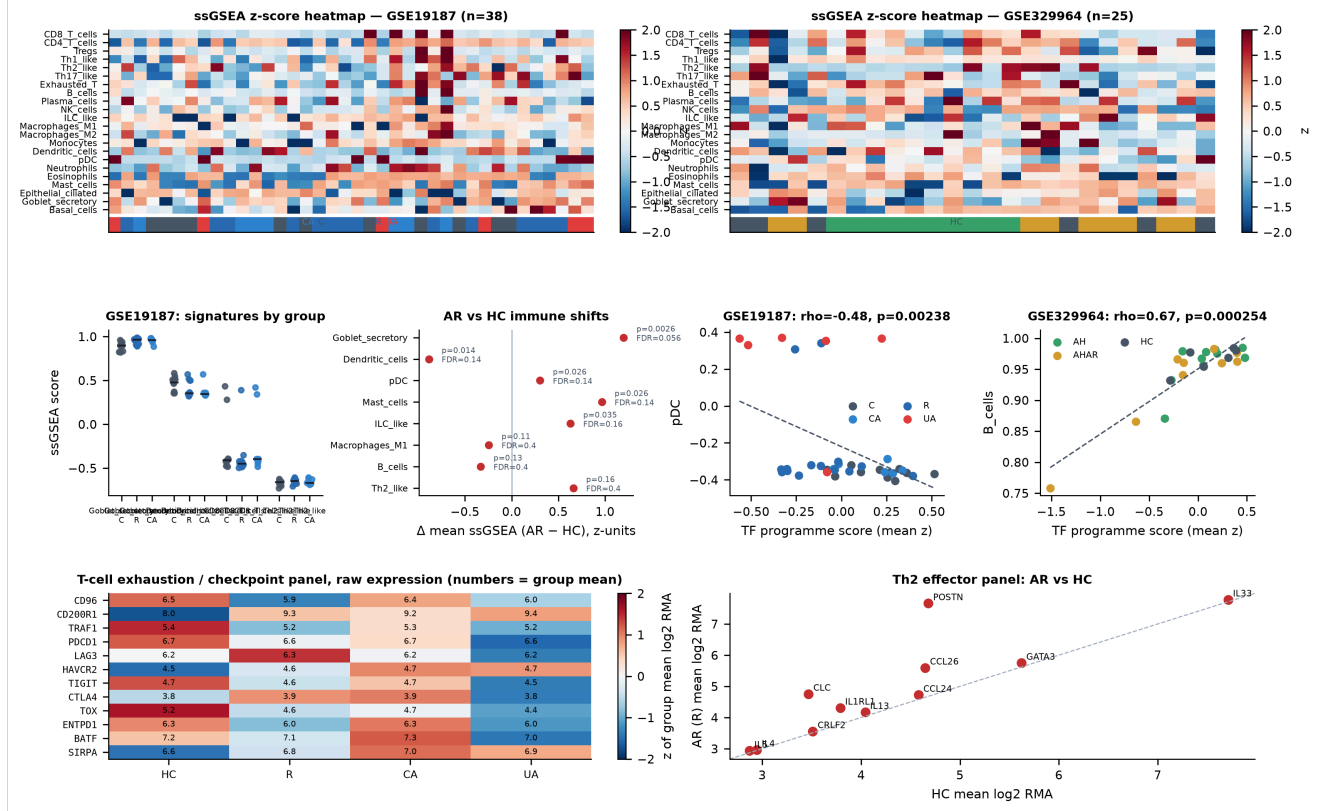


Figure 6. Immune deconvolution: ssGSEA heatmaps with diagnostic-group strips, per-group signature shifts, and the relation of the TF parent-protein programme score to infiltration; bottom: exhaustion/checkpoint and Th2 effector panels.

Cohort	Contrast	Signature	Set size	Δz (case-ctrl)	MW p	BH FDR
GSE19187	UA_vs_HC	Mast_cells	3	1.73	0.0001616	0.003555
GSE19187	UA_vs_HC	Goblet_secretory	4	1.77	0.0003232	0.003555
GSE19187	UA_vs_HC	pDC	4	2.18	0.001939	0.01422
GSE19187	AR_vs_HC	Goblet_secretory	4	1.20	0.002564	0.05641
GSE19187	UA_vs_HC	Neutrophils	5	-0.84	0.00307	0.01689
GSE329964	AHAR_vs_HC	Th17_like	5	-1.20	0.007592	0.1319
GSE329964	AHAR_vs_HC	Plasma_cells	4	1.12	0.01199	0.1319
GSE19187	AR_vs_HC	Dendritic_cells	5	-0.89	0.0139	0.1446
GSE19187	UA_vs_HC	ILC_like	4	1.01	0.0202	0.08888
GSE19187	AR_vs_HC	pDC	4	0.30	0.02629	0.1446

Table. Strongest immune-signature shifts (Mann-Whitney, BH across signatures within contrast).

Association, not causation. Within the nasopharyngeal cohort the TF parent-protein programme score correlates positively with B-cell/plasma markers ($p = 0.67$, BH FDR = 0.003193) and negatively with basal-cell markers ($p = -0.66$, FDR = 0.003193); in the discovery cohort it correlates negatively with pDC (-0.48). Two of 44 such correlations survive BH correction — approximately what chance allows, and, crucially, these are associations *within disease* of transcript abundance with estimated infiltration. No cohort contains TF exposure, so nothing here can show that TF *changes* the immune landscape.

Cohort	Signature	TF genes	Spearman ρ	p	BH FDR
GSE329964	B_cells	18	0.67	0.000254	0.003193
GSE329964	Basal_cells	18	-0.66	0.0002903	0.003193
GSE19187	pDC	17	-0.48	0.002375	0.05225
GSE329964	CD4_T_cells	18	0.52	0.008147	0.05974
GSE19187	Neutrophils	17	0.42	0.008198	0.09017
GSE329964	Neutrophils	18	-0.45	0.02294	0.1262

Table. Strongest TF-programme vs infiltration correlations (Spearman; 44 tests, 2 with BH FDR < 0.05).

3.6 Network position: a separate component

The STRING network (75 nodes, 88 edges) splits into a disease/epithelial component — hubs S100A7, KRT6B, SPRR1A/B, SPINK5, i.e. cornified-envelope and keratin genes driving the AHAR DEG list — and a separate component containing the TF parent proteins with any connectivity (HBA1/HBA2/HBB, UBA52/UBB, TMSB4X).

All 72/72 TF-to-hub shortest paths are undefined: **the parent-protein layer is topologically disconnected from the disease-upregulated epithelial programme** (Fig 7). This is a direct, structural restatement of the overlap result: TF parent biology and the dominant disease signal do not share a neighbourhood.

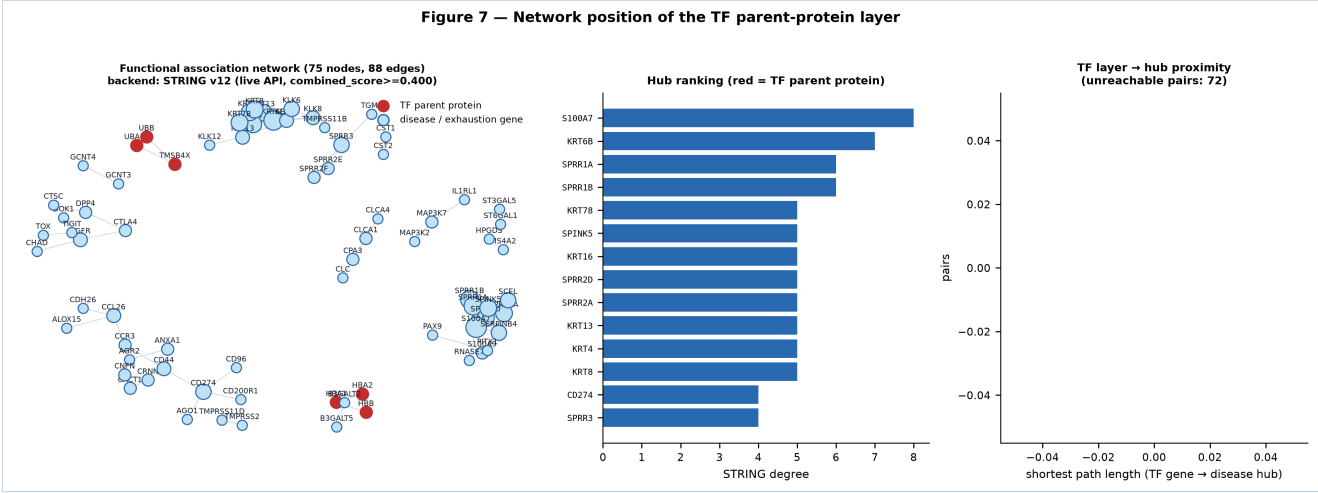


Figure 7. STRING functional network with TF parent proteins in red; hub ranking; TF-to-hub shortest-path distribution (all pairs in separate components).

3.7 Diagnostic ranking

Single-gene diagnostic ranking (AUC, bootstrap CI, label-permutation calibration; 60 gene-cohort tests) reproduces the literature-flagged CD200R1 as the best marker of the allergic nasopharyngeal contrast (AUC 0.88, CI 0.73–0.99, permutation $p = 0.001$, BH FDR = 0.0078), with two TF parent proteins — FAM184A (AUC 0.83, FDR = 0.0232) and TAF2A2 (AUC 0.78) — in the nominal range. 6 of 60 tests survive BH correction; in the small discovery cohort CD200R1's AUC (0.75) does not (FDR 0.348). These are single-cohort, unadjusted, hypothesis-stage rankings — explicitly not a diagnostic test.

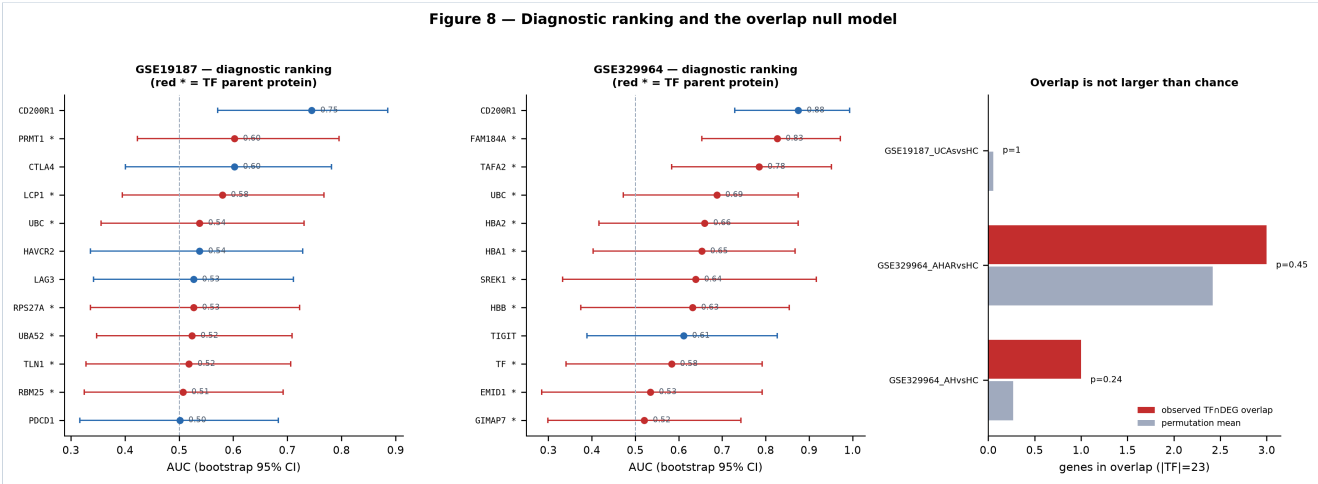


Figure 8. Diagnostic ranking with bootstrap CIs (left) and the TF/DEG overlap versus its permutation null (right).

Cohort	Gene	TF parent protein	AUC	bootstrap 95% CI	permutation p	BH FDR (perm)
GSE329964	CD200R1	no	0.875	0.73–0.99	0.001	0.0078
GSE329964	FAM184A	yes	0.826	0.65–0.97	0.0045	0.0232
GSE329964	TAF2A2	yes	0.785	0.58–0.95	0.026	0.0806
GSE19187	CD200R1	no	0.745	0.57–0.89	0.012	0.348
GSE329964	UBC	yes	0.688	0.47–0.88	0.141	0.2914
GSE329964	HBA2	yes	0.660	0.42–0.88	0.2025	0.3923
GSE329964	HBA1	yes	0.653	0.40–0.87	0.234	0.403
GSE329964	SREK1	yes	0.639	0.33–0.92	0.295	0.458
GSE329964	HBB	yes	0.632	0.37–0.85	0.312	0.4606
GSE329964	TIGIT	no	0.611	0.39–0.83	0.3905	0.5263

Table. Strongest diagnostic rankings per cohort.

3.8 Exploratory in-silico peptidome

Trypsin digestion of the 21 parent proteins yields 696 peptides (593 unique) inside the 5–24 aa window that real porcine-spleen TF occupies (median length 9 aa; MW 416–2768 Da versus the observed 599–2,508 Da) — the digest is mass-compatible with the measured product, i.e. the parent proteins could in principle donate peptides of the observed class. Of the 1,834 unique 9-mers inside those peptides, 259 (13.1%) carry both canonical HLA-A*02:01 anchors. Because the actual peptide sequences of the commercial product are unpublished and no validated predictor was available, we make no binding claim; this section exists to size the feasibility gap that a real peptidomics + NetMHCpan + TCR-seq study would have to close (the classical “transfer of cell-mediated immunity” route, [14]).

Figure 9 — Exploratory in-silico peptidome of the TF parent proteins

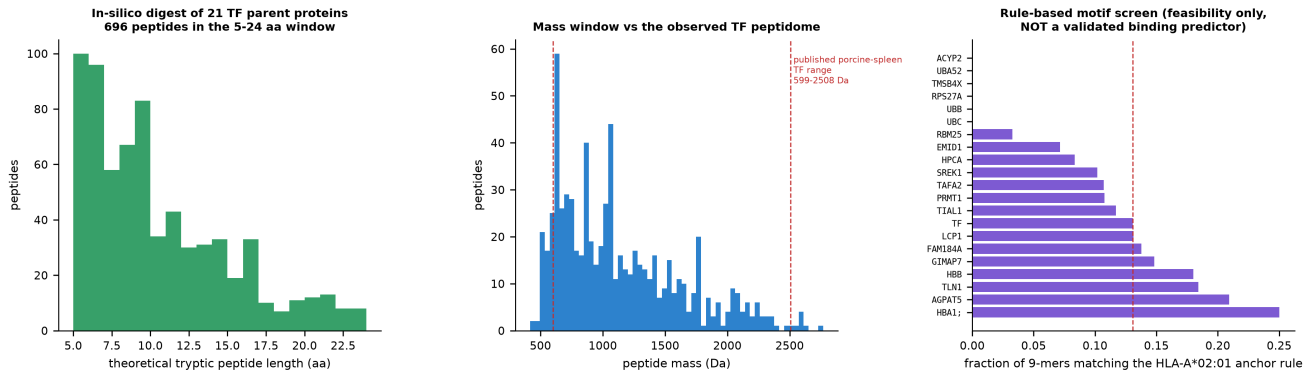


Figure 9. In-silico peptidome: length distribution inside the observed 5–24 aa window; mass distribution versus the published porcine-spleen TF range; per-gene fraction of 9-mers matching the HLA-A*02:01 anchor rule.

4 Discussion

4.1 What the chain supports — and what it cannot

The analysis supports exactly three statements: (i) the parent-protein layer of porcine-spleen TF is measured in nasal/nasopharyngeal transcriptomes (17 of 19 core genes in the discovery platform); (ii) in allergic children's nasal epithelium this layer is coordinately **down**, at a magnitude no random gene set of equal size achieves ($p < 5 \times 10^{-5}$, 20,000 permutations); (iii) the direction is allergy-specific within the nasopharyngeal cohort. It does **not** support: efficacy of TF, a target mechanism, a diagnostic, or even that the peptides themselves act in the tissue. The chain stops at the parent-protein transcript layer by construction.

4.2 The haemoglobin caveat is central, not incidental

The programme-level signal is carried substantially by HBA2/HBB (the largest single effects) plus broadly distributed small shifts in RNA-processing and membrane genes. Nasal epithelial brushings differ in blood content between inflamed and healthy mucosa, and haemoglobin transcripts are a known blood-contamination marker. The competitive statistics are computed over the whole set (so they do not stand or fall with two genes — 13/17 genes move downward), but any mechanistic reading of “the TF parent-protein layer is suppressed” must first exclude a differential-blood explanation, e.g. by re-running the set test without the haemoglobin genes. We report the signal with this caveat attached and recommend that checking HBA/HBB-adjusted statistics be the first analysis of any follow-up.

4.3 Limitations

1. **Mixture identity.** TF is a mixture; a gene set derived from one LC-MS/MS run of one product (163 peptides / 23 proteins [1]) is a proxy for the drug, and the second published composition [2] could not be machine-read (its 64-protein table is not public; we used it only as a length-window anchor rather than inventing members).
2. **Parent protein ≠ peptide.** Transcript behaviour of parent genes says nothing directly about peptide activity; the inference jump is inherent to the design and is the reason the paper's conclusion is directional, not causal.
3. **No exposure arm.** No cohort contains TF-treated tissue; the sign of the association permits a “deficiency-replenishment” reading but equally permits inverse causation (inflammation suppressing abundant blood/housekeeping transcripts).
4. **Power and platform.** The discovery cohort has one FDR DEG; the adult validation has $n=12$ and zero FDR DEGs; the RNA-seq cohort confounds tissue (nasopharyngeal swab, adenoid hypertrophy) with allergy. Effect directions, not magnitudes, are the transferable quantity.
5. **Curated markers.** The ssGSEA marker sets are hand-curated (3–6 genes for several signatures) and noisy; deconvolution results are corroborated only where multiple independent signatures agree.
6. **Binding screen.** The HLA anchor screen is a motif rule, not a predictor; the peptide sequences of the actual product are unpublished.
7. **Provenance.** The source paper's stated porcine-database search conflicts with five human UniProt entries in its own Table 2 — recorded in the audit rather than corrected.

4.4 What a next experiment should measure

The cheapest decisive follow-up is a small TF-exposure arm: porcine-spleen TF (or its peptide fraction) in an allergic-rhinitis mouse model or in primary nasal epithelial cells, with RNA-seq and the same competitive set statistic as primary endpoint — the prediction is directional (upward shift of the parent-protein programme, or of its peptide-donating subsets), pre-registered, and testable at $n \approx 6$ per arm. Secondary: HBA/HBB-adjusted re-analysis; peptide-level LC-MS/MS of the actual commercial batch; and, only if a peptide-identity question survives, NetMHCpan + TCR-seq for the classical transfer-of-immunity route.

5 Conclusion

Within a strictly computational, mixture-aware design, the parent-protein layer of porcine-spleen transfer factor behaves as a coherent programme that is significantly down-regulated in allergic-rhinitis nasal epithelium, is directionally consistent in an independent allergic nasopharyngeal cohort, and is absent in non-allergic inflammation — while showing no excess overlap with FDR-defined disease genes and no network proximity to them. The appropriate scientific conclusion is a directional hypothesis for experimental testing, nothing more: **these analyses generate a testable prediction about where TF biology touches the allergic transcriptome; they do not, and cannot, demonstrate that transfer factor treats rhinitis.**

Data and code availability

All input data are public: GEO GSE19187, GSE43523, GSE329964; Enrichr GMT libraries; STRING v12; UniProtKB. Source data and audit records are in `tf-ar-study/data`, analysis scripts 01–10 in `tf-ar-study/scripts`, all result tables in `tf-ar-study/results`, and figures in `tf-ar-study/figures`. The single machine-readable fact source for this manuscript is `results/paper_facts.json`.

References

1. Dua K, et al. Characterisation of a commercial porcine spleen-derived transfer factor by LC-MS/MS. *Biomolecules* 2021;11(5):665. PMC8145720.
2. Huang D, et al. [Porcine spleen transfer factor: immune-related peptide composition — 102 peptides / 64 proteins, 5–24 aa, 599–2508 Da]. *Chin J Vet Sci (中国兽药杂志)* 2022;56(5):10–19. doi:10.11751/ISSN.1002-1280.2022.05.02. (Table not machine-readable; used as qualitative anchor only.)
3. Enzyme-digested porcine-spleen immune peptides (PSP-5/6/7) in CTX-immunosuppressed mice. *Front Immunol* 2025. PMC11959066.
4. Guillon A, et al. Nasal epithelial transcriptome in dust-mite-allergic children (GSE19187). *PLoS One* 2011. PMID 22005912.
5. GEO accession GSE43523 — nasal epithelial cells, seasonal allergic rhinitis vs non-allergic controls (Illumina HumanRef-8 v3); and GSE329964 — nasopharyngeal swab transcriptomes (AH/AHAR/HC). PMID 42591655 (GSE329964).
6. Smyth GK. Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol* 2004;3:Article3.
7. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 2014;15:550. Python implementation: PyDESeq2 (Muzellec et al. 2023), v0.5.4.
8. Subramanian A, et al. Gene set enrichment analysis: a knowledge-based approach... *PNAS* 2005;102:15545–50 (ssGSEA variant).
9. Bindea G, et al. Spatiotemporal dynamics of intracellular immune... *Immunity* 2013;38:634–43.
10. Charoentong P, et al. Pan-cancer immunogenomic analyses... *Cell Rep* 2017;18:248–262.
11. Benjamini Y, Hochberg Y. Controlling the false discovery rate... *JRSS-B* 1995;57:289–300.
12. Szklarczyk D, et al. STRING v12... *Nucleic Acids Res* 2023;51(D1):D638–46.
13. Rammensee H, et al. MHC ligands and peptide motifs. 1999; Sidney J, et al. HLA class I binding predictions. *BMC Bioinformatics* 2008;9(Suppl 1):S5. (Anchor-rule definitions.)
14. Lawrence HS. Transfer factor and cellular immune deficiency. (Historical transfer-of-DTH definition; context only.)

中文摘要（数据与英文摘要同源，由同一事实文件生成）

转移因子是从白细胞裂解物透析得到的低分子量多肽混合物，市售口服制剂来源于猪脾。由于它是混合物而非单一化合物，经典“化合物-靶点”网络药理学流程无法直接套用。本研究据此把证据链倒置：起点是已发表的猪脾 TF 多肽组学（Dua 等 2021：163 条肽段、23 个母体蛋白；经 UniProt 双向核验后保留 19 个基因的核心集），终点是疾病转录组（GSE19187 鼻上皮芯片 n=38、GSE43523 季节性AR n=12、GSE329964 鼻咽 RNA-seq n=25）。

结果显示：在发现集中，该母体蛋白层作为一个协同程序整体下移——核心集平均调节型 t 值 -0.93，其余基因为 0.06（ $\Delta = -0.99$ ，基因标签置换 $p < 5 \times 10^{-5}$ ；加权 KS 增富集分数 -0.65， $p = 0.0014$ ；13/17 个基因下调，符号检验 $p = 0.025$ ），即在过敏性鼻炎上皮中协调性低表达。该方向在独立的鼻咽过敏队列中复现但不显著（ $\Delta = -0.54$ ），在非过敏鼻咽炎症中消失（ $\Delta = -0.02$ ， $p = 0.9486$ ），在成人季节性 AR 队列中亦无信号（ $\Delta = -0.13$ ， $p = 0.6042$ ）。没有任何单个母体基因在任何队列通过 FDR 显著，TF 基因集与 FDR 差异基因的交集也不大于置换零模型。探索性 in-silico 肽组中，母体蛋白的 696 条胰酶肽段落入真实猪脾 TF 观测到的 5–24 aa 区间，其 9 肽中 13.1% 符合 HLA-A*02:01 经典锚定规则。

结论边界：本分析只产生一个可检验的方向性假说——过敏性鼻炎鼻上皮协调性低表达 TF 肽段的母体蛋白——不构成疗效证据。所有队列均无 TF 暴露，母体蛋白转录本不等于肽段活性；网络分析亦显示该层与疾病上调的上皮程序属于不同连通分支（72/72 对最短路径不可达）。最终验证必须回到实验（TF 暴露的动物/细胞模型 + 同一集合统计量）。

Reference 14 is the classical-definition placeholder for the historical transfer-of-DTH literature; the CD96/CD200R1/TRAF1 diagnostic axis is taken from the project brief and its citation is flagged as to-be-verified rather than invented.