

CHRFAM7A-assigned transcript detection is enriched in a LAMP3+CCR7+ dendritic-cell state in esophageal squamous cell carcinoma

Zhao-wei Gao

zwgao1@gmail.com

Anyang Tumor Hospital

Xue-li Gao

Shanghai Jiao Tong University

Shu-min Lun

Anyang Tumor Hospital

Li-juan Duan

Anyang Tumor Hospital

Xiao-xiao Wang

Anyang Tumor Hospital

Fang Zhao

Anyang Tumor Hospital

Sha-sha Cao

Anyang Tumor Hospital

Yao-wen Zhang

Anyang Tumor Hospital

Research Article

Keywords: CHRFAM7A, esophageal squamous cell carcinoma, single-cell RNA sequencing, dendritic cells, LAMP3, CCR7

Posted Date: August 12th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10398840/v1>

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

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Background

CHRFAM7A is a human-restricted fusion gene related to CHRNA7 whose sequence similarity and sparse single-cell detection complicate interpretation. We examined the myeloid distribution of CHRFAM7A-assigned transcripts in esophageal squamous cell carcinoma (ESCC) and assessed matrix-to-raw-read concordance.

Methods

We reanalyzed GSE160269, comprising 19,273 myeloid cells from 64 samples and 60 patients. Tumor myeloid states were manually annotated and evaluated using mixed-effects and sample-stratified conditional logistic regression. Robustness analyses used higher UMI thresholds and abundance-matched sparse-gene controls. Raw reads from P31T and a rule-frozen replication sample, P126T, underwent transcript competition, barcode-UMI linkage, and structural comparison with annotated CHRFAM7A exon models.

Results

CHRFAM7A-assigned counts occurred in 1,578/19,273 cells (8.19%). Among 18,261 tumor core-myeloid cells, detection was highest in LAMP3 + CCR7+ dendritic cells (289/946, 30.55%). Within-sample enrichment remained strong versus other myeloid cells (odds ratio 5.04, 95% CI 4.39–5.78) and versus cDC/pDC cells (odds ratio 4.35, 95% CI 3.65–5.17). The direction persisted at ≥ 2 and ≥ 3 UMI thresholds and ranked at the 98th percentile among 100 abundance-matched genes. P31T and P126T contained 35 and 49 high-confidence sequences structurally compatible with annotated CHRFAM7A transcripts; raw-read support occurred in 17/20 versus 0/72 and 11/15 versus 0/15 matrix-positive and matrix-negative cells. Two identical 100-bp sequences were shared.

Conclusions

CHRFAM7A-assigned transcript detection shows a technically supported enrichment in a LAMP3 + CCR7+ dendritic-cell state in ESCC. These findings do not establish CHRNA7 expression, alpha7 receptor activity, vagal signaling, or causality.

Introduction

Esophageal cancer remains a major global health burden, with esophageal squamous cell carcinoma (ESCC) accounting for most cases in high-incidence regions. The biological and clinical heterogeneity of ESCC is shaped not only by malignant epithelial cells but also by a complex stromal and immune

microenvironment [1–4]. Single-cell studies have resolved diverse myeloid, lymphoid, fibroblast, and epithelial states in ESCC and have emphasized that apparently similar tumors can contain markedly different immune ecosystems [2–4].

Dendritic cells are central coordinators of antitumor immunity because they integrate antigen sensing, migration, T-cell priming, and immune regulation. Recent single-cell studies across cancers have identified LAMP3- and CCR7-expressing mature or migratory dendritic-cell states that can contain both stimulatory and regulatory programs [5–10]. These states are often described as mature regulatory dendritic cells or migratory dendritic cells; however, transcriptomic state labels do not by themselves establish lineage or function. Their molecular heterogeneity and potential relevance to immunotherapy make careful characterization in individual tumor types important.

The alpha7 nicotinic acetylcholine receptor, encoded by *CHRNA7*, has been implicated in inflammatory regulation in several experimental systems. *CHRFAM7A* is a human-restricted fusion gene that incorporates a partial duplication of *CHRNA7* and can modify alpha7 receptor biology in human cells [11, 12]. Previous experimental work linked vagal-alpha7 signaling to macrophage anti-inflammatory responses during influenza infection, and separate studies implicated *CHRFAM7A* in macrophage polarization [13, 14]. Nevertheless, these observations cannot be directly extrapolated to human tumors. *CHRFAM7A* and *CHRNA7* share extensive sequence similarity, *CHRFAM7A* is absent from conventional nonhuman models, and sparse short-read measurements can be vulnerable to annotation and mapping ambiguity [11, 12].

We therefore asked a narrower and testable question: in a large ESCC myeloid single-cell data set, is *CHRFAM7A*-assigned transcript detection associated with a specific myeloid state after accounting for sample structure and technical depth, and can the matrix-level signal be supported by raw reads? We combined manual cell-state annotation, mixed and conditional logistic models, UMI-threshold and abundance-matched controls, a two-sample raw-read sequence audit, and paired bulk-tissue context. The analysis was deliberately framed around *CHRFAM7A*-assigned transcript detection rather than *CHRNA7* expression or alpha7 receptor activity.

Materials and methods

Study design and public data sets

This was a secondary analysis of deidentified, publicly available transcriptomic data. The primary single-cell data set was GSE160269, which contains myeloid cells from ESCC tumors and adjacent normal tissues. The processed expression matrix included 19,273 cells from 64 samples representing 60 patients: 18,514 cells from 60 tumor samples and 759 cells from four normal samples. The primary inferential analysis was restricted to 18,261 tumor core-myeloid cells from 60 patients. Raw sequencing files corresponding to P31T and P126T were used for sequence-level technical validation. GSE23400, an

Affymetrix GPL96/GPL97 study containing 53 paired ESCC and adjacent-normal specimens, was used as external bulk-expression context [20].

Table 1
Data sets and analytical roles.

Data source	Material	Units analyzed	Primary role	Key boundary
GSE160269 processed single-cell matrix	ESCC myeloid cells	19,273 cells; 64 samples; 60 patients. Primary tumor set: 18,261 core-myeloid cells from 60 tumors	Cell-state annotation and sample-aware CHRFAM7A-assigned transcript-detection analysis	CHRNA7 and ARG1 were absent from the processed matrix; cell-level counts are not independent biological replicates
GSE160269 P31T raw reads	LAMP3 + CCR7+ cells from one tumor sample	92 cells in the matrix-to-raw-read comparison	Exploratory discovery and rule development for sequence-level technical validation	Discovery thresholds were exploratory/post hoc
GSE160269 P126T raw reads	LAMP3 + CCR7+ cells from a second tumor sample	30 cells in the matrix-to-raw-read comparison	Technical replication using frozen Q1-Q3 rules without sample-specific retuning	Internal technical replication within the same public cohort, not independent biological validation
GSE23400 GPL96	Paired ESCC and adjacent-normal bulk tissue	53 complete pairs; 47-pair QC sensitivity analysis	External paired-tissue expression context	Probe 210123_s_at maps to both CHRNA7 and CHRFAM7A; neither locus is separately measurable

ESCC, esophageal squamous cell carcinoma; QC, quality control.

The analysis was organized in three layers. First, the distribution of matrix-assigned CHRFAM7A counts was evaluated across manually reviewed myeloid states with sample-aware statistical models. Second, robustness was assessed using higher raw-UMI thresholds and abundance-matched sparse-gene controls. Third, raw reads from P31T and P126T were audited against CHRFAM7A and CHRNA7 transcript references and annotated genomic structures. P31T served as an exploratory discovery sample; the Q1-Q3 transcript-level rules were frozen before application to P126T, making P126T a prospective technical replication with respect to the analysis rule, not a prospectively collected clinical sample.

Single-cell annotation and analysis set

Cell clusters were manually reviewed using canonical lineage and state markers. The final core-myeloid annotation contained monocyte states (Mono01-Mono03), tumor-associated macrophage states

(TAM01-TAM04), mast cells, conventional dendritic cells (cDC), plasmacytoid dendritic cells (pDC), and a LAMP3 + CCR7+ dendritic-cell state labeled tDC in the source analysis. The latter state was supported by LAMP3, CCR7, CCL19, CCL22, and FSCN1. Two clusters dominated by T-cell or plasma/B-cell markers were treated as contaminants. A rare SIGLEC15 + CTSK+ osteoclast-like cluster was retained as provisional but excluded from the primary core-myeloid analysis. In this manuscript, tDC refers only to the transcriptomically defined LAMP3 + CCR7+ dendritic-cell state and is not intended as an ontogenetic lineage assignment.

A cell was considered CHRFAM7A-positive when the processed raw-count matrix contained at least one UMI assigned to CHRFAM7A. Because most positive cells contained very few UMIs and because CHRFAM7A is closely related to CHRNA7, the primary outcome is termed CHRFAM7A-assigned transcript detection rather than quantitative expression. CHRNA7 and ARG1 were absent from the released processed matrix; this was treated as a measurement limitation rather than evidence of biological absence.

Sample-aware subtype models

The primary cell-level model was a binomial generalized linear mixed model with CHRFAM7A detection as the outcome, broad myeloid subtype as the main explanatory variable, standardized log₁₀ total UMI count, its quadratic term, and standardized mitochondrial transcript percentage as fixed covariates, and sample as a random intercept. A likelihood-ratio test compared models with and without subtype. Model-adjusted detection probabilities and pairwise subtype contrasts were estimated on the response scale; pairwise P values were adjusted by the Benjamini-Hochberg method. Sensitivity analyses excluded candidate high-count cells and replaced total UMI with detected-gene count.

To control all sample-level characteristics that were constant within a tumor, sample-stratified conditional logistic regression compared tDC cells with non-tDC myeloid cells while adjusting for standardized sequencing depth, its quadratic term, and mitochondrial percentage. A DC-restricted model compared tDC with cDC and pDC cells. Patient/sample-level detection fractions were also compared in samples containing sufficient cells in both compartments. These approaches were selected to avoid treating individual cells as independent biological replicates [16, 17].

UMI-threshold and abundance-matched robustness analyses

The primary detection threshold was at least one raw UMI. Sensitivity analyses required at least two or at least three raw UMIs assigned to CHRFAM7A. At each threshold, conditional logistic models were refitted and sample-level tDC versus non-tDC detection fractions were summarized. The higher thresholds reduce sensitivity to isolated single-UMI events but also reduce the number of informative samples and events.

To determine whether tDC enrichment was unusual for a transcript with comparable sparsity and count abundance, 100 control genes were selected separately at the ≥ 1 -UMI and ≥ 2 -UMI thresholds by matching overall event frequency and mean raw count among event-positive cells. Every control gene was tested with the same sample-stratified conditional logistic model and technical covariates used for CHRFAM7A. The empirical one-sided P value was calculated as $(1 + \text{number of control odds ratios at least as large as the CHRFAM7A odds ratio}) / (1 + \text{number of valid control models})$.

Exploratory within-tDC transcriptional program

An exploratory discovery analysis had identified four candidate genes (ALOX15B, BIRC3, GAPDH, and LPP) associated with CHRFAM7A detection within tDC cells. Their replication was tested in 21 informative samples using 101 matched positive-negative cell pairs and 42 sample-level pseudobulk libraries. The analysis was considered supportive only if gene-level effects remained significant in the targeted replication model and showed consistent direction across patients. CHRFAM7A itself was excluded from biological interpretation because it defined the comparison groups.

Raw-read extraction and sequence-level classification

Raw paired-end reads for P31T and P126T were retrieved from the corresponding Sequence Read Archive runs. Reads associated with exact cell barcodes from the LAMP3 + CCR7+ state were extracted, and raw barcode-UMI combinations were retained. Candidate sequences were compared with CHRFAM7A and CHRNA7 RefSeq transcript libraries using BLAST+ [21]. The frozen Q1-Q3 scheme classified sequences as Q1 (full-length exact 100% identity in both transcript-reference comparisons), Q2 (full-length $\geq 99\%$ identity in both), or Q3 (full-length $\geq 95\%$ identity in both). Sequence-level read-pair, cell, and raw-UMI support were summarized after deduplication.

Structural interpretation used genome alignments and annotated exon models. A sequence was considered directly supported when its genomic alignment structure was compatible with annotated CHRFAM7A exons on the primary assembly; alternate/patch-only support was reported separately. Directly unresolved sequences could be rescued by projection of their top CHRFAM7A RefSeq transcript alignment onto multi-assembly GFF exon models [22]. These procedures establish structural compatibility with annotated CHRFAM7A transcripts but do not establish whole-genome absolute uniqueness.

Matrix-to-raw-read concordance was assessed within each sample by comparing the presence of validated raw-read/UMI evidence between cells with and without CHRFAM7A-assigned matrix counts. Fisher's exact test and Spearman correlations between matrix and raw-read counts were used as internal technical concordance measures. Exact full-length sequence overlap between P31T and P126T was descriptive because molecule recovery is sensitive to sequencing depth, UMI sampling, and cell composition.

Paired bulk-expression context

For GSE23400, processed GPL96 expression values were analyzed in a paired tumor-versus-adjacent-normal design using limma [19, 20]. All 53 complete pairs were included in the primary analysis; a sensitivity analysis excluded six samples flagged by conservative quality-control criteria, leaving 47 pairs. Probes with ambiguous multi-gene annotation were excluded before gene-level collapse.

The GPL96 probe 210123_s_at maps to both CHRNA7 and CHRFAM7A. Consequently, neither gene was separately measurable in the unique-gene matrix, and no direct tumor-normal differential-expression claim was made for either locus. The bulk analysis was used only to describe the wider ESCC transcriptional context and to document this measurement boundary.

Statistical analysis and reproducibility

Tests were two-sided unless explicitly identified as empirical one-sided tests. Effect estimates are reported with 95% confidence intervals. Multiple pairwise comparisons used Benjamini-Hochberg false-discovery-rate adjustment. Analyses were conducted in R using Seurat-compatible objects, lme4, emmeans, survival/conditional-logistic modeling utilities, edgeR, and limma, and in Python for raw-read sequence integration; BLAST + was used for sequence alignments [15, 18, 19, 21]. Final P31T, P126T, and cross-sample outputs were frozen with file-size and SHA256 manifests. No survival model was used to support the primary conclusions.

Detailed subtype, UMI-threshold, empirical-null, negative-replication, and raw-read concordance results are provided in Supplementary Tables S1-S7 (Online Resource 1).

Results

A manually reviewed ESCC myeloid atlas identifies a LAMP3 + CCR7+ dendritic-cell state

The processed GSE160269 matrix contained 19,273 myeloid cells from 64 samples and 60 patients. Manual review assigned 19,014 cells to the core-myeloid set, 86 cells to a rare provisional SIGLEC15 + CTSK+ population, and 173 cells to non-myeloid contaminant clusters. The primary tumor analysis included 18,261 core-myeloid cells from 60 tumors. The LAMP3 + CCR7+ state contained 975 cells overall, including 946 tumor cells and 29 cells from three normal samples (Fig. 1; Table 1).

Across all myeloid cells, CHRFAM7A-assigned counts were detected in 1,578 cells (8.19%) with 1,849 total UMIs. The mean among detected cells was 1.17 UMIs, confirming that this was a sparse transcript-detection phenotype rather than a high-abundance quantitative expression signal. CHRNA7 and ARG1 were not represented in the processed matrix.

CHRFAM7A-assigned transcript detection is concentrated in the LAMP3 + CCR7+ state

Within the 18,261 tumor core-myeloid cells, 1,488 cells (8.15%) were CHRFAM7A-positive. The raw detection fraction was highest in the LAMP3 + CCR7+ state (289/946, 30.55%), followed by TAM04 (107/673, 15.90%), cDC (257/2,567, 10.01%), TAM02 (393/4,589, 8.56%), and lower fractions in the remaining populations (Fig. 2A; Table 2).

The primary mixed model was not singular and showed a strong overall subtype effect (likelihood-ratio chi-square = 520.92, $P = 1.49 \times 10^{-105}$). At average technical covariate values, the adjusted detection probability for the LAMP3 + CCR7+ state was 27.34% (95% CI 24.21%-30.72%), compared with 2.37%-8.72% across the other broad myeloid subtypes. Pairwise odds ratios for the LAMP3 + CCR7+ state versus individual comparator subtypes ranged from 3.94 (versus mast cells) to 15.48 (versus Mono01), and all Benjamini-Hochberg-adjusted P values were below 1.2×10^{-23} . The direction and ranking were preserved after excluding candidate high-count cells and after adjustment for detected-gene count instead of total UMI (Fig. 2B).

Within-sample models and higher UMI thresholds support the enrichment pattern

A sample-stratified conditional logistic model confirmed enrichment within individual tumors. Relative to all other core-myeloid cells, the LAMP3 + CCR7+ state had an odds ratio of 5.04 (95% CI 4.39–5.78; $P = 5.58 \times 10^{-117}$) for CHRFAM7A detection after cell-level technical adjustment. The estimate was similar after excluding high-count cells (odds ratio 5.05) and increased in the detected-gene-adjusted sensitivity model (odds ratio 6.02). Restricting the comparator to cDC and pDC cells retained a strong association (odds ratio 4.35, 95% CI 3.65–5.17; $P = 7.19 \times 10^{-62}$).

Among 44 tumors eligible for sample-level paired comparison, the LAMP3 + CCR7+ detection fraction exceeded the non-tDC fraction in 43, was equal in none, and was lower in one (paired Wilcoxon $P = 9.70 \times 10^{-9}$). At the ≥ 2 -UMI threshold, 95/946 LAMP3 + CCR7+ cells (10.04%) from 41 samples remained positive; 38 of 44 eligible tumors had a higher tDC fraction ($P = 3.31 \times 10^{-8}$). At the ≥ 3 -UMI threshold, 31/946 cells (3.28%) from 21 samples were positive; despite the reduced event count, the sample-level direction remained nonrandom ($P = 7.61 \times 10^{-5}$) (Fig. 3A).

Table 2
Principal single-cell association and robustness results.

Analysis	Comparison/threshold	Result	Interpretation
Raw subtype summary	LAMP3 + CCR7+ state	289/946 positive cells (30.55%); highest among 11 tumor core-myeloid subtypes	Descriptive enrichment; not adjusted
Mixed-effects logistic model	Subtype omnibus test	Chi-square = 520.92; P = 1.49×10^{-105}	Subtype contributes beyond technical covariates and sample random intercept
Mixed-effects logistic model	Adjusted LAMP3 + CCR7+ probability	27.34% (95% CI 24.21%-30.72%)	Model-based probability at average covariate values
Conditional logistic model	LAMP3 + CCR7+ vs all other core myeloid	OR 5.04 (95% CI 4.39–5.78); P = 5.58×10^{-117}	Within-tumor enrichment after cell-level technical adjustment
Conditional logistic model	LAMP3 + CCR7+ vs cDC/pDC	OR 4.35 (95% CI 3.65–5.17); P = 7.19×10^{-62}	Enrichment persists within dendritic-cell compartment
Patient/sample-level paired summary	≥ 1 UMI	43/44 tumors higher in tDC; Wilcoxon P = 9.70×10^{-9}	Direction is consistent across tumors
UMI-threshold sensitivity	≥ 2 UMI	95/946 tDC positive; 38/44 tumors higher; P = 3.31×10^{-8}	Not explained solely by single-UMI events
UMI-threshold sensitivity	≥ 3 UMI	31/946 tDC positive; 20 higher, 19 equal, 5 lower; P = 7.61×10^{-5}	Direction persists with reduced event count
Abundance-matched null	≥ 1 UMI	OR 6.64; 98th percentile; empirical P = 0.0297	Unusual among 100 similarly sparse genes, but not unique
Abundance-matched null	≥ 2 UMI	OR 25.07; 98th percentile; empirical P = 0.0297	Robust at higher count threshold
Targeted four-gene program	ALOX15B, BIRC3, GAPDH, LPP	No gene met targeted replication criteria	No validated CHRFAM7A-associated transcriptional signature

CI, confidence interval; OR, odds ratio; tDC, transcriptomically defined LAMP3 + CCR7+ dendritic-cell state; UMI, unique molecular identifier.

The 29 normal LAMP3 + CCR7+ cells showed a descriptively similar raw detection fraction (9/29, 31.03%), but they came from only three normal samples. The available data therefore do not support a claim that the state-associated pattern is tumor-specific.

The enrichment is unusual, but not unique, among abundance-matched sparse genes

In abundance-matched null analyses, CHRFAM7A had an odds ratio of 6.64 (95% CI 5.73–7.68) at the ≥ 1 -UMI threshold and 25.07 (95% CI 17.82–35.26) at the ≥ 2 -UMI threshold. At each threshold, only two of 100 matched control genes had odds ratios at least as large. CHRFAM7A therefore ranked at the 98th percentile, with an empirical one-sided P value of 0.0297 at both thresholds (Fig. 3B). This result supports the view that the tDC enrichment is not merely a generic property of low-abundance genes; however, it also shows that comparably strong enrichment can occur for a small number of other sparse genes.

The exploratory four-gene transcriptional program did not meet prespecified replication criteria. Directional agreement across 21 informative samples was 12/21 for ALOX15B, 14/21 for BIRC3, 9/21 for GAPDH, and 11/21 for LPP, and none remained significant in the targeted replication model after adjustment. These genes were therefore not used to define a CHRFAM7A-associated mechanism or signature.

Raw reads provide two-sample technical concordance for CHRFAM7A-assigned matrix counts

The P31T discovery set contained 35 Q1-Q3 sequences (15 Q1, 11 Q2, and nine Q3). Thirty-two had direct primary-assembly genomic structures compatible with annotated CHRFAM7A exons, and three were resolved by RefSeq-to-GFF projection. These sequences represented 37 validated read pairs and 28 reported barcode-raw-UMI molecules across 17 of 92 tDC cells. Validated support was found in 17 of 20 matrix-positive cells and in none of 72 matrix-negative cells (two-sided Fisher $P = 8.09 \times 10^{-16}$). Matrix and raw-read counts were correlated across all cells (Spearman $\rho = 0.914$) and within matrix-positive cells ($\rho = 0.686$).

Application of the frozen Q1-Q3 criteria to P126T without sample-specific retuning yielded 49 sequences (28 Q1, 18 Q2, and three Q3), 68 validated read pairs, and 16 globally deduplicated barcode-raw-UMI molecules across 11 of 30 tDC cells. Forty sequences had direct primary-assembly support, one had direct alternate/patch support, and eight were resolved by transcript-to-GFF projection. Raw-read support was found in 11 of 15 matrix-positive cells and in none of 15 matrix-negative cells (two-sided Fisher $P = 5.00 \times 10^{-5}$). Spearman correlations were 0.856 across all cells and 0.820 among matrix-positive cells (Fig. 4; Table 3).

P31T and P126T shared two identical full-length 100-bp sequences. Both retained the same query identifier, Q1 quality class, direct-primary structural class, preferred CHRFAM7A transcript

(XM_011522155.4), and projection class. Exact overlap represented 5.71% of the P31T set and 4.08% of the P126T set (Jaccard index 0.0244). Because exact molecule recovery is sampling-sensitive, this limited overlap was treated as additional molecule-level concordance rather than as the main replication criterion.

Table 3
Frozen raw-read technical-concordance results.

Measure	P31T discovery	P126T rule-frozen replication
Cells evaluated	92 tDC cells	30 tDC cells
Frozen Q1-Q3 sequences	35 (Q1 15; Q2 11; Q3 9)	49 (Q1 28; Q2 18; Q3 3)
Validated read pairs	37	68
Reported barcode-raw-UMI molecules	28	16 globally deduplicated
Cells with raw-read support	17/92	11/30
Matrix-positive cells with support	17/20 (85.0%)	11/15 (73.3%)
Matrix-negative cells with support	0/72	0/15
Fisher exact P	8.09x10 ⁻¹⁶	5.00x10 ⁻⁵
Spearman rho, all cells	0.914	0.856
Direct primary-assembly structural support	32	40
Direct alternate/patch structural support	0	1
RefSeq-to-GFF projection-resolved	3	8
Valid annotated transcript projections	35/35	49/49
Exact cross-sample sequence overlap	2 shared 100-bp sequences; 5.71% of P31T set	2 shared 100-bp sequences; 4.08% of P126T set

Q1, full-length exact 100% transcript-level identity under the frozen comparison scheme; Q2, full-length > = 99%; Q3, full-length > = 95%; UMI, unique molecular identifier. Structural compatibility does not establish whole-genome absolute uniqueness.

Paired bulk data provide ESCC context but cannot quantify CHRNA7 and CHRFAM7A separately

The GSE23400 primary analysis included 53 paired ESCC and adjacent-normal specimens. After collapse of uniquely annotated probes, 12,501 genes were represented; 243 were upregulated and 258 were downregulated under the primary thresholds. A 47-pair sensitivity analysis showed strong agreement with the primary results (Pearson $r = 0.989$, Spearman $\rho = 0.982$, 95.75% directional concordance), with 474 stable significant genes.

However, the relevant GPL96 probe 210123_s_at maps to both CHRNA7 and CHRFAM7A and was excluded from the unique-gene matrix. Thus, the bulk data do not independently validate differential expression of either locus. They serve only as paired-tissue transcriptional context and as a demonstration of why target-specific interpretation requires sequence-aware analysis.

Discussion

This study identifies a consistent concentration of CHRFAM7A-assigned transcript detection in a LAMP3 + CCR7+ dendritic-cell state within ESCC myeloid data. The association was evident descriptively, persisted after adjustment for sequencing depth and mitochondrial percentage, remained strong in within-sample comparisons and within the dendritic-cell compartment, and survived increasingly stringent UMI thresholds. An abundance-matched null analysis placed the effect at the 98th percentile among equally sparse genes. Raw-read analyses in P31T and a rule-frozen P126T technical replication sample further showed that matrix-positive cells were preferentially supported by high-confidence sequences structurally compatible with annotated CHRFAM7A transcripts.

The cellular state highlighted here is consistent with the broad family of LAMP3 + CCR7+ mature or migratory dendritic-cell states described in multiple cancers [5–10]. These cells can express migration, antigen-presentation, inflammatory, and regulatory programs simultaneously, and their function may vary with tissue context. We therefore avoided equating the tDC label with a single lineage or with a uniformly immunosuppressive phenotype. The present result is a mapping of a transcript-detection phenotype onto a transcriptomic state; it does not establish how CHRFAM7A affects dendritic-cell differentiation, migration, antigen presentation, or T-cell regulation.

The distinction between CHRFAM7A and CHRNA7 is central. CHRFAM7A is human-restricted and can alter alpha7 receptor biology, but its genomic architecture and sequence similarity to CHRNA7 complicate short-read quantification [11, 12]. The processed GSE160269 matrix contained a CHRFAM7A row but no CHRNA7 row, and the GPL96 bulk probe in GSE23400 could not separate the two genes. Accordingly, our results cannot be interpreted as evidence that the LAMP3 + CCR7+ state expresses functional CHRNA7, that alpha7 receptor signaling is active, or that vagal input shapes the observed cells. Experimental evidence for vagal-alpha7 signaling in infected lung macrophages and for CHRFAM7A effects on macrophage polarization provides biological context [13, 14], but not mechanistic validation in ESCC dendritic cells.

Several analytic features strengthen the inference that the observed pattern is not solely a technical artifact. First, the primary mixed model accounted for sample clustering, library depth, nonlinear depth

effects, and mitochondrial percentage. Second, conditional logistic regression compared cell states within the same tumor, controlling all sample-level covariates that were constant within a sample. Third, the direction persisted after excluding single-UMI-only events and after matching CHRFAM7A to genes with similar event frequency and event-positive abundance. These steps address common forms of pseudoreplication and detection-depth confounding in single-cell analysis [16, 17]. They do not, however, eliminate every form of mapping error, ambient RNA, or state-dependent RNA-content bias.

The raw-read audit offers an orthogonal check on the processed matrix. In both P31T and P126T, validated evidence was concentrated in cells with CHRFAM7A-assigned matrix counts and absent from matrix-negative cells in the evaluated tDC sets. The P126T rules were frozen before application, reducing the risk of sample-specific threshold tuning. Structural analysis separated direct primary-assembly support, alternate/patch support, and transcript-projection rescue, and all retained sequences had a valid annotated CHRFAM7A projection. Nonetheless, these are still short-read data. Structural compatibility with CHRFAM7A transcripts is not equivalent to proof of whole-genome absolute uniqueness, and the two samples constitute internal technical concordance rather than independent biological replication.

Several limitations require explicit consideration.

Limitations

- This is a secondary analysis of public transcriptomic data, and no new clinical cohort or experimental validation was generated.
- CHRFAM7A and CHRNA7 share extensive sequence similarity; therefore, matrix-level results are described as CHRFAM7A-assigned transcript detection rather than definitive locus-specific expression.
- Raw-read technical concordance was evaluated in only two tumor samples from the same public cohort and does not constitute independent biological or population-level replication.
- GSE160269 contained only four normal samples, which was insufficient for a reliable tumor-versus-normal inference, and GSE23400 could not quantify CHRNA7 and CHRFAM7A separately because of probe ambiguity.
- The study did not measure protein abundance, receptor activity, spatial localization, vagal input, or causal function; the exploratory four-gene within-tDC program did not replicate.

Future work should use assays designed for the CHRNA7-CHRFAM7A locus. Targeted long-read RNA sequencing, junction-specific digital PCR, capture-based sequencing, and allele-aware reference construction could establish isoform and locus specificity. Spatial or multiplexed in situ assays could determine whether CHRFAM7A-assigned transcripts localize to LAMP3 + CCR7+ cells in intact ESCC tissue. Protein-level and functional experiments would then be required to test receptor assembly, cholinergic responsiveness, antigen-presentation capacity, migratory behavior, and effects on T-cell activation. Such studies should include human-relevant models because CHRFAM7A is not represented in standard mouse genomes.

In summary, the data support a technically robust tDC-enriched CHRFAM7A-assigned transcript-detection phenotype in ESCC. The result provides a focused human-specific hypothesis for future validation while remaining separate from claims about CHRNA7 expression, alpha7 receptor function, vagal signaling, or causality.

Conclusions

CHRFAM7A-assigned transcript detection is enriched in a LAMP3 + CCR7+ dendritic-cell state in ESCC myeloid single-cell data. The association is supported by sample-aware models, higher-UMI and abundance-matched robustness analyses, and two-sample raw-read technical concordance under frozen sequence-classification rules. These findings define a focused human-specific hypothesis for future validation but do not establish CHRNA7 expression, alpha7 receptor function, vagal signaling, or causality.

Declarations

Ethics approval and consent to participate

This study used only deidentified data available in public repositories and involved no new participant recruitment, intervention, or access to identifiable information. Therefore, no new institutional ethics approval or informed consent was required for this secondary analysis.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Funding:

This work was supported by the 2022 Henan Provincial Medical Science and Technology Research Program Joint Co-construction Project (grant no. LHGJ20220980).

Author Contribution

Z.-W.G. and X.-L.G. contributed equally to this work. Z.-W.G. and X.-L.G. conceived the study, curated the data, performed the analyses, interpreted the results, and drafted the manuscript. S.-M.L., L.-J.D., X.-x.W., F.Z., and S.-s.C. contributed to data interpretation and critically revised the manuscript. Y.-W.Z. supervised the study, contributed to study design and interpretation, and critically revised the

manuscript. All authors read and approved the final manuscript and agree to be accountable for the work.

Acknowledgements

The authors thank the investigators who generated and publicly shared GSE160269 and GSE23400. An AI-assisted language tool was used during manuscript preparation to improve English expression and document organization; all scientific content, analyses, interpretations, and references were critically reviewed and verified by the authors.

Data Availability

The publicly available datasets analysed in this study are available from the National Center for Biotechnology Information Gene Expression Omnibus under accession numbers GSE160269 and GSE23400. Raw sequencing data used for targeted read-level validation are available from the National Center for Biotechnology Information Sequence Read Archive under BioProject PRJNA672851. Derived results and analysis code supporting the findings of this study are available from the corresponding authors upon reasonable request.

References

1. Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2024;74:229–63. <https://doi.org/10.3322/caac.21834>.
2. Zhang X, Peng L, Luo Y, Zhang S, Pu Y, Chen Y, et al. Dissecting esophageal squamous-cell carcinoma ecosystem by single-cell transcriptomic analysis. *Nat Commun*. 2021;12:5291. <https://doi.org/10.1038/s41467-021-25539-x>.
3. Zheng Y, Chen Z, Han Y, Han L, Zou X, Zhou B, et al. Immune suppressive landscape in the human esophageal squamous cell carcinoma microenvironment. *Nat Commun*. 2020;11:6268. <https://doi.org/10.1038/s41467-020-20019-0>.
4. Zheng S, Liu Y, Guan X. The role of tumor microenvironment in invasion and metastasis of esophageal squamous cell carcinoma. *Front Oncol*. 2022;12:911285. <https://doi.org/10.3389/fonc.2022.911285>.
5. Moon CY, Belabed M, Park MD, Mattiuz R, Puleston D, Merad M. Dendritic cell maturation in cancer. *Nat Rev Cancer*. 2025;25:225–48. <https://doi.org/10.1038/s41568-024-00787-3>.
6. Heras-Murillo I, Adan-Barrientos I, Galan M, Wculek SK, Sancho D. Dendritic cells as orchestrators of anticancer immunity and immunotherapy. *Nat Rev Clin Oncol*. 2024;21:257–77. <https://doi.org/10.1038/s41571-024-00859-1>.

7. Del Prete A, Salvi V, Soriani A, Laffranchi M, Sozio F, Bosisio D, et al. Dendritic cell subsets in cancer immunity and tumor antigen sensing. *Cell Mol Immunol*. 2023;20:432–47. <https://doi.org/10.1038/s41423-023-00990-6>.
8. Wculek SK, Cueto FJ, Mujal AM, Melero I, Krummel MF, Sancho D. Dendritic cells in cancer immunology and immunotherapy. *Nat Rev Immunol*. 2020;20:7–24. <https://doi.org/10.1038/s41577-019-0210-z>.
9. Gerhard GM, Bill R, Messemaker M, Klein AM, Pittet MJ. Tumor-infiltrating dendritic cell states are conserved across solid human cancers. *J Exp Med*. 2021;218:e20200264. <https://doi.org/10.1084/jem.20200264>.
10. Maier B, Leader AM, Chen ST, Tung N, Chang C, LeBerichel J, et al. A conserved dendritic-cell regulatory program limits antitumour immunity. *Nature*. 2020;580:257–62. <https://doi.org/10.1038/s41586-020-2134-y>.
11. Ihnatovych I, Saddler RA, Sule N, Szigeti K. Translational implications of CHRFAM7A, an elusive human-restricted fusion gene. *Mol Psychiatry*. 2024;29:1020–32. <https://doi.org/10.1038/s41380-023-02389-1>.
12. Szigeti K, Ihnatovych I, Birkaya B, Chen Z, Ouf A, Indurthi DC, et al. CHRFAM7A: a human specific fusion gene, accounts for the translational gap for cholinergic strategies in Alzheimer's disease. *EBioMedicine*. 2020;59:102892. <https://doi.org/10.1016/j.ebiom.2020.102892>.
13. Gao ZW, Li L, Huang YY, Zhao CQ, Xue SJ, Chen J, et al. Vagal- $\alpha 7$ nAChR signaling is required for lung anti-inflammatory responses and arginase 1 expression during an influenza infection. *Acta Pharmacol Sin*. 2021;42:1642–52. <https://doi.org/10.1038/s41401-020-00579-z>.
14. Li T, Chen W, Zhang Q, Deng C. Human-specific gene CHRFAM7A mediates M2 macrophage polarization via the Notch pathway to ameliorate hypertrophic scar formation. *Biomed Pharmacother*. 2020;131:110611. <https://doi.org/10.1016/j.biopha.2020.110611>.
15. Hao Y, Hao S, Andersen-Nissen E, Mauck WM 3rd, Zheng S, Butler A, et al. Integrated analysis of multimodal single-cell data. *Cell*. 2021;184:3573–e358729. <https://doi.org/10.1016/j.cell.2021.04.048>.
16. Zimmerman KD, Espeland MA, Langefeld CD. A practical solution to pseudoreplication bias in single-cell studies. *Nat Commun*. 2021;12:738. <https://doi.org/10.1038/s41467-021-21038-1>.
17. Squair JW, Gautier M, Kathe C, Anderson MA, James ND, Hutson TH, et al. Confronting false discoveries in single-cell differential expression. *Nat Commun*. 2021;12:5692. <https://doi.org/10.1038/s41467-021-25960-2>.
18. Bates D, Machler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. *J Stat Softw*. 2015;67:1–48. <https://doi.org/10.18637/jss.v067.i01>.
19. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015;43:e47. <https://doi.org/10.1093/nar/gkv007>.

20. Su H, Hu N, Yang HH, Wang C, Takikita M, Wang QH, et al. Global gene expression profiling and validation in esophageal squamous cell carcinoma and its association with clinical phenotypes. Clin Cancer Res. 2011;17:2955–66. <https://doi.org/10.1158/1078-0432.CCR-10-2724>.

21. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. BMC Bioinformatics. 2009;10:421. <https://doi.org/10.1186/1471-2105-10-421>.

22. O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Res. 2016;44:D733–45. <https://doi.org/10.1093/nar/gkv1189>.

Figures

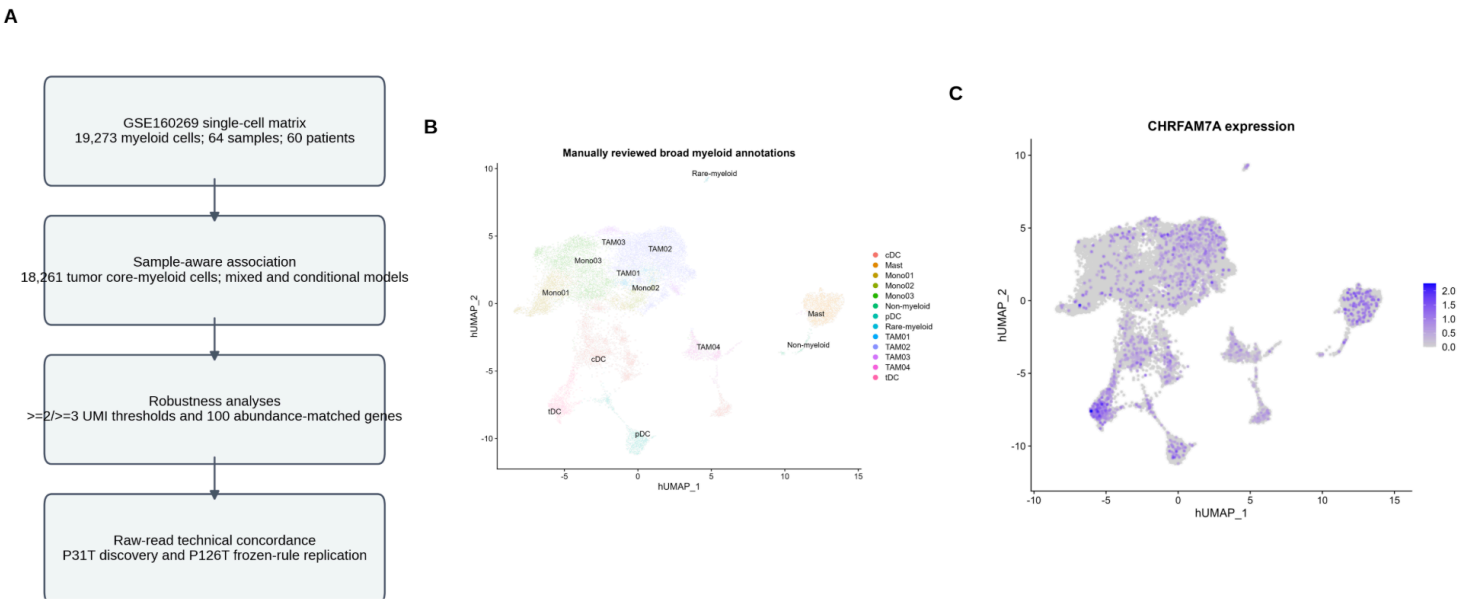


Figure 1

Study design and manually reviewed ESCC myeloid atlas. (A) Three-layer analytical workflow. (B) Manually reviewed broad myeloid annotations in GSE160269. (C) CHRFAM7A-assigned matrix counts across the myeloid embedding. The rare provisional and non-myeloid populations were excluded from the primary tumor core-myeloid analysis

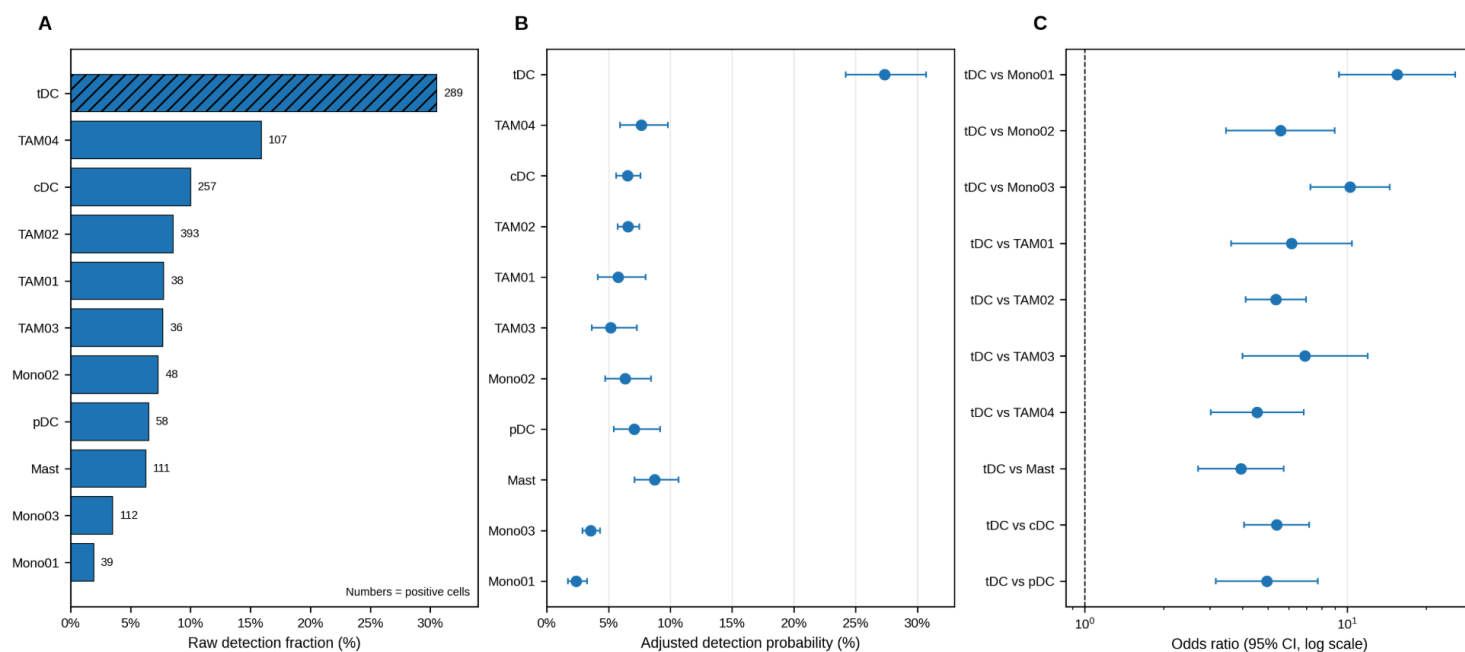


Figure 2

CHRFAM7A-assigned transcript detection across tumor myeloid states. (A) Raw detection fractions; numbers denote positive cells. (B) Model-adjusted probabilities with 95% confidence intervals at average technical covariate values. (C) Pairwise odds ratios comparing tDC with each other subtype in the primary mixed model

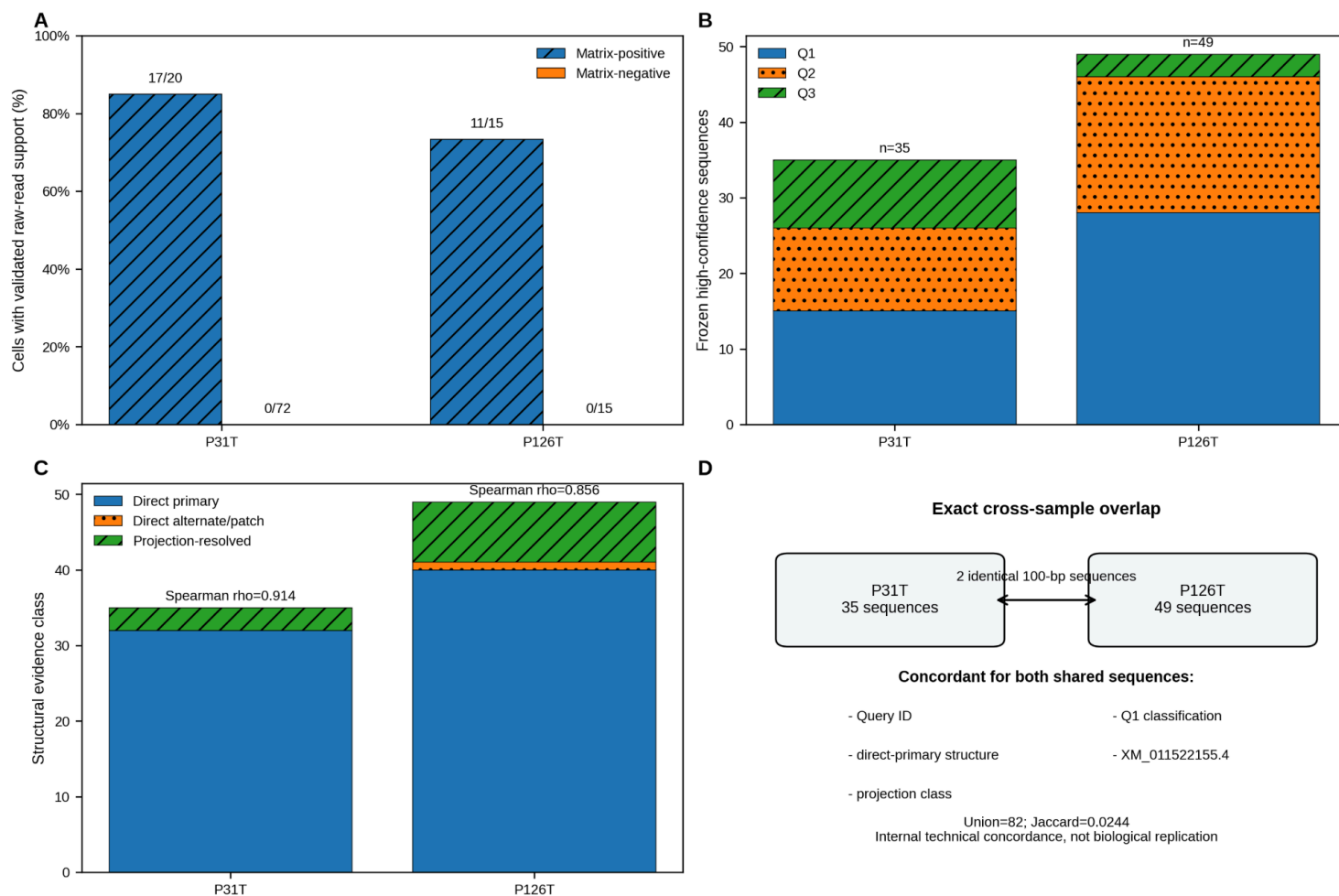


Figure 3

Raw-read and sequence-level technical concordance in P31T and P126T. (A) Validated raw-read support among matrix-positive and matrix-negative tDC cells. (B) Frozen Q1-Q3 sequence counts. (C) Final structural evidence classes; Spearman correlations compare matrix and raw-read molecule counts across cells. (D) Two identical 100-bp sequences were recovered in both samples with concordant classification. This is internal technical concordance and does not establish whole-genome absolute uniqueness or independent biological replication

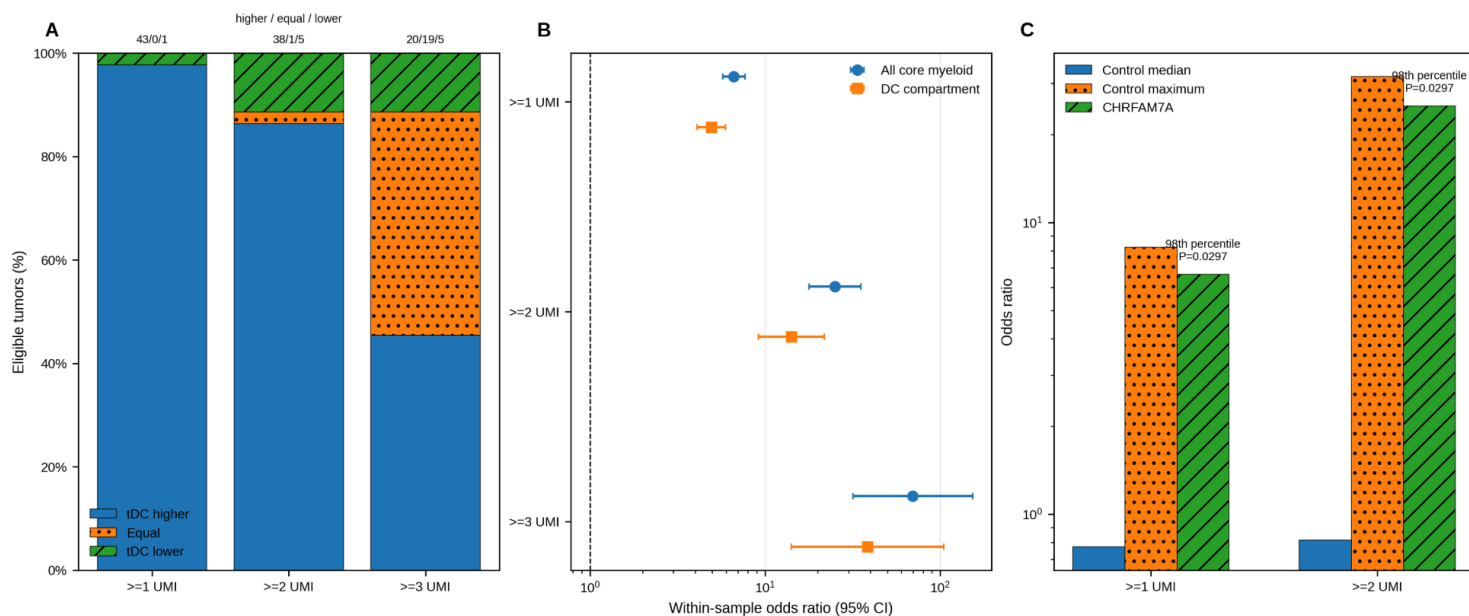


Figure 4

Within-sample and sparse-gene robustness analyses. (A) Proportions of eligible tumors with higher, equal, or lower tDC detection than non-tDC cells at three UMI thresholds; labels give counts. (B) Conditional-logistic odds ratios for all core-myeloid and dendritic-cell-restricted comparisons. (C) CHRFAM7A odds ratios compared with the median and maximum among 100 abundance-matched control genes; empirical P values are one-sided

Supplementary Files

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

- [DiscoverOncologySupplementaryInformation.docx](#)