nature research

Corresponding author(s): Nora A. Barrett and Alex K. Shalek

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated
☐	☒	Clearly defined error bars *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	STAR, RSEM, R, Seurat 1.4.0.1 (and associated functions therein as described in Methods), GraphPad Prism, DESeq2 1.10.1, Ingenuity Pathway Analysis, HOMER, macs2, FlowJo, scanpy, seaborn, matplotlib, pandas, bowtie2, samtools, ImageJ
Data analysis	STAR, RSEM, R, Seurat 1.4.0.1 (and associated functions therein as described in Methods), GraphPad Prism, DESeq2 1.10.1, Ingenuity Pathway Analysis, HOMER, macs2, FlowJo, scanpy, seaborn, matplotlib, pandas, bowtie2, samtools, ImageJ. Example walk-through script included as R Script in Supplementary Information for generating Analysis from core data set in Figure 1. Additional walk-through scripts available on http://shaleklab.com/resources.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The cells-by-genes matrix generated from ethmoid sinus surgical resections and analyzed during the current study is available along with the manuscript as Supplementary Table 2 along with R code for standard implementation of Seurat. A cells-by-genes matrix from inferior turbinate and polyp scraping data is also available as Supplementary Table 6. Dupilumab treatment cells-by-genes matrices as Supplementary Tables 7 and 8. A metadata table encompassing all scRNA-seq samples is provided as Supplementary Table 9. The count and TPM matrices and associated metadata from bulk tissue RNA-seq are available as Supplementary Tables 10, 11, and 12. FASTQ file format data will be available through dbGaP under an accession number to be assigned. Marker gene lists for cell types identified in Fig. 1a,b, and from resultant analyses in Fig. 2b, for frequencies of cell clusters and types in Fig. 2c, for cell types identified in Fig. 2e, Fig. 2f, Fig. 3g, Fig. 5a, Fig. 5e, Extended Data Fig. 3a,b,c, Extended Data Fig. 4c, Extended Data Fig. 5e, Extended Data Fig. 6b,d, Extended Data Fig. 10a, selected comparisons of differential expression in Fig. 2d, Fig. 4a, Fig. 5c, Fig. 5f, Extended Data Fig. 2c, Extended Data Fig. 10h, and pseudotime correlation Extended Data Fig. 9b, are available as tabs in Supplementary Table 3. Source Data for plots is available for Fig. 1d, Fig. 3a, Fig. 3b, Fig. 3c, Fig. 3d, Fig. 3e, Fig. 4f, Fig. 4g, Fig. 5d, Extended Data Fig. 1d, Extended Data Fig. 2d, Extended Data Fig. 2e, Extended Data Fig. 4a, Extended Data Fig. 4b, Extended Data Fig. 4f, Extended Data Fig. 7a, Extended Data Fig. 7b, Extended Data Fig. 7d, Extended Data Fig. 7e, Extended Data Fig. 7f, Extended Data Fig. 7k, Extended Data Fig. 7l, Extended Data Fig. 8c, Extended Data Fig. 9i, and Extended Data Fig. 9j. Differential peak calling from epigenetic profiling available in Supplementary Table 5. Additional R code for analyses available on http://shaleklab.com/resources/.

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	As our study was not a prospective clinical trial, we did not predetermine sample size. We enrolled patients meeting tight clinical criteria for non-polyp or polyp diagnoses, ran 12 Seq-Well arrays, and assessed reproducibility of cell clusters obtained as shown in Extended Data Fig. 1c,d with each subsequent patient analyzed. We determined given the reliability with which we sampled each cell type, that n=6 samples within each group was sufficient for this study to identify cell types and the significant transcriptional differences across disease. Furthermore, no cluster of cells detected and analyzed contains less than 273 unique measurements, with the average number of cells in any one cell type >1,800 (Supplementary Table 3). In the revised manuscript, we now provide an additional 9 Seq-Well derived samples, 3 of which were healthy controls, 4 of which were of polypoid individuals but from the anatomical region sampled in healthy controls, and 2 were polyp scrapings to control for method of isolation relative to our original study.
Data exclusions	Described in detail in Methods. Cells were pre-filtered during alignment, and further filtered with no cell accepted with less than 300 genes, 500 UMI, or more than 12,000 UMI. Furthermore, three cell doublet clusters were excluded (Methods) based on the co-occurrence of cell-type restricted markers. 3 bulk tissue RNA-seq samples were excluded based on poor transcriptome alignment (< 25%) and low read counts indicative of RNA degradation. The same filtering metrics were applied to our revision data set, of which all cells meeting these criteria were included.
Replication	Experimental findings were reliably reproduced across our single-cell cohort of 12 samples (6 non-polyp and 6 polyp), and confirmed and extended using distinct techniques. E.g. single-cell RNA-sequencing and flow cytometry are both susceptible to dissociation-induced bias in cell type recovery, so we elected to use histology and bulk tissue RNA-seq to confirm findings. Seq-Well experiments performed one year later using scrapings provided data that clustered with pre-existing cells from our original submission. This suggests that our cell type identification was robust to patient, sampling technique, and experimental bias.
Randomization	Study participants were not randomized and were assigned to non-polyp or polyp groups based on established diagnostic criteria.
Blinding	The surgeon was blinded to study questions, the allergist who performed scrapings was blinded, and histology was performed in a blinded fashion. All other experimental techniques were not blinded.

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
☐	☒ Unique biological materials
☐	☒ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☐	☒ Human research participants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Unique biological materials

Policy information about availability of materials

Obtaining unique materials: All unique materials are available upon request and users are welcome to explore the dataset generated with the matrix and code provided. Code updates will be provided on http://shaleklab.com/resources

Antibodies

Antibodies used: The following antibodies were used to identify basal cells via flow cytometry: FITC anti-human THY1 (Biolegend, clone 5E10), Brilliant Violet 421 anti-human CD45 (Biolegend, clone HI30), Brilliant Violet 650 anti-human EPCAM (Biolegend, clone 9C4), APC/Cy7 anti-human ITGA6 (Biolegend, clone GoH3), PE/Cy7 anti-human NGFR (Biolegend, clone ME20.4), APC anti-human PDPN (Biolegend, clone NC-08) and for histology: anti-TP63 antibody (Biolegend, clone W15093A), and AlexaFluor 647-conjugated donkey anti-mouse IgG (Jackson immunoresearch, catalog# 715-605-150) secondary.

Validation: All antibodies have been described and validated previously in publications such as Rock et al., 2009 identifying human basal cells, and isotype controls were used for determining positive signals in flow cytometry and histology.

Human research participants

Policy information about studies involving human research participants

Population characteristics: Supplementary Table 1 provides the full clinical characteristics of single-cell patients and tissue RNAseq patients, histology, flow cytometry, and ATAC-seq/RNA-seq patients describing the gender, age, diagnosis, nasal polyp grade, h/o atopy, asthma and/or sinus infection, number of previous surgeries, ASA tolerance, oral steroid use, topical steroid use, immunosuppressant use, leukotriene modification, and antibiotic use. We refer the interested reader to Supplementary Table 1 for this information given the difficulty in displaying it within the space provided here.

Recruitment: Subjects between the ages of 18 and 75 years were recruited from the Brigham and Women's Hospital (Boston, Massachusetts) Allergy and Immunology clinic and Otolaryngology clinic between May 2014 and March 2018 (Supplementary Table 1). The Institutional Review Board approved the study, and all subjects provided written informed consent. There was a low chance for recruitment bias as individuals were receiving necessary surgical care for their respective diagnosis.

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation: Single-cell suspensions in FACS Buffer (HBSS -/- supplemented with 2% FCS) were pre-incubated with Fc-Block before staining for surface antigens. Cells were stained for 30 minutes on ice in FACS buffer and then washed for immediate sorting.

Instrument: BD FACSAria Fusion

Software: BD FACSDiva Software for collection and FlowJo v10 by TreeStar for analysis

Cell population abundance: Purity of samples was determined from parallel samples sorted for functional assays as sorting into lysis buffer for ATAC-seq and RNA-seq precludes re-sorting to check purity.

Gating strategy | Extended Data Figure 7h for full gating strategy

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.