LETTER

https://doi.org/10.1038/s41586-018-0449-8

Allergic inflammatory memory in human respiratory epithelial progenitor cells

Jose Ordovas-Montanes[1,2,3,4,5,6,14], Daniel F. Dwyer[7,8,14], Sarah K. Nyquist[1,2,3,4,5,9,10], Kathleen M. Buchheit[7,8], Marko Vukovic[1,2,3,4,5], Chaarushena Deb[1,2,3,4,5], Marc H. Wadsworth II[1,2,3,4,5], Travis K. Hughes[1,2,3,4,5], Samuel W. Kazer[1,2,3,4,5], Eri Yoshimoto[7,8], Katherine N. Cahill[7,8], Neil Bhattacharyya[8,11], Howard R. Katz[7,8], Bonnie Berger[10,12,13], Tanya M. Laidlaw[7,8], Joshua A. Boyce[7,8], Nora A. Barrett[7,8,15]* & Alex K. Shalek[1,2,3,4,5,13,15]*

Barrier tissue dysfunction is a fundamental feature of chronic human inflammatory diseases[1]. Specialized subsets of epithelial cells—including secretory and ciliated cells—differentiate from basal stem cells to collectively protect the upper airway[2–4]. Allergic inflammation can develop from persistent activation[5] of type 2 immunity[6] in the upper airway, resulting in chronic rhinosinusitis, which ranges in severity from rhinitis to severe nasal polyps[7]. Basal cell hyperplasia is a hallmark of severe disease[7–9], but it is not known how these progenitor cells[2,10,11] contribute to clinical presentation and barrier tissue dysfunction in humans. Here we profile primary human surgical chronic rhinosinusitis samples (18,036 cells, $n = 12$) that span the disease spectrum using Seq-Well for massively parallel single-cell RNA sequencing[12], report transcriptomes for human respiratory epithelial, immune and stromal cell types and subsets from a type 2 inflammatory disease, and map key mediators. By comparison with nasal scrapings (18,704 cells, $n = 9$), we define signatures of core, healthy, inflamed and polyp secretory cells. We reveal marked differences between the epithelial compartments of the non-polyp and polyp cellular ecosystems, identifying and validating a global reduction in cellular diversity of polyps characterized by basal cell hyperplasia, concomitant decreases in glandular cells, and phenotypic shifts in secretory cell antimicrobial expression. We detect an aberrant basal progenitor differentiation trajectory in polyps, and propose cell-intrinsic[13], epigenetic[14,15] and extrinsic factors[11,16,17] that lock polyp basal cells into this uncommitted state. Finally, we functionally demonstrate that ex vivo cultured basal cells retain intrinsic memory of IL-4/IL-13 exposure, and test the potential for clinical blockade of the IL-4 receptor α-subunit to modify basal and secretory cell states in vivo. Overall, we find that reduced epithelial diversity stemming from functional shifts in basal cells is a key characteristic of type 2 immune-mediated barrier tissue dysfunction. Our results demonstrate that epithelial stem cells may contribute to the persistence of human disease by serving as repositories for allergic memories.

The type 2 immunity (T2I) module[6] regulates homeostatic processes[18] (metabolism), host defence[19] (against parasites, venoms, allergens and toxins), and inflammatory tissue repair[11]. However, this module may become self-reinforcing in allergic inflammation, leading to substantial alterations in gross tissue architecture[20] as observed in polyps[7]. To investigate how the overall tissue cellular ecosystem shifts in composition and states during chronic respiratory T2I disease in humans, we used Seq-Well for single-cell RNA sequencing (scRNA-seq)[12] to profile the ethmoid sinus (EthSin) of patients spanning the chronic rhinosinusitis (CRS) spectrum (Fig. 1a, Supplementary Table 1, Methods; Supplementary Discussion I; $n = 12$ samples: 6 non-polyp, 6 polyp). Deconstructing these tissues into their component cells provides a unique lens into the cellular ecosystem of human T2I, helping us to: 1. characterize each major cell type without the biases that are typically introduced by pre-selection of markers; 2. evaluate cell types and/or states with disease-associated transcriptional differences; and, 3. reconstruct tissue-level dynamics.

We derived a unified cells-by-genes expression matrix (18,036 cells) and performed dimensionality reduction and graph-based clustering (Fig. 1a, Extended Data Fig. 1a, b, Supplementary Table 2; Methods). Using complete lists of cluster-specific genes to identify epithelial[2], stromal[7,20] and immune cells[4,6], we recovered a reproducible distribution of cell types within patient groups (Fig. 1b, c, Extended Data Figs. 1c–e, 2a–e, Supplementary Table 3; Methods; Supplementary Discussion II). We highlight the major cell types recovered (with hallmark expressed genes in parentheses): basal (*KRT5*) and apical (*KRT8*) cells—which orient the pseudostratified epithelial division, further specialized ciliated (*FOXJ1*) and glandular[3] (*LTF*) cells, and supportive endothelial cells (*DARC*, also known as *ACKR1*), fibroblasts (*COL1A2*), plasma cells (*CD79A*), myeloid cells (*HLA-DRA*), T cells (*TRBC2*) and mast cells (*TPSAB1*) (Extended Data Fig. 1e). For each cell type, sub-clustering revealed further, potentially meaningful heterogeneity, providing a useful reference atlas for studying human inflammatory diseases of barrier tissues (Extended Data Fig. 3a–c; Supplementary Discussion III).

Next, we charted the cell-of-origin for chemokines and lipid mediators, which aid in the recruitment and positioning of lymphoid and myeloid cells in tissues during T2I[21] (Extended Data Fig. 4a; Supplementary Discussion IV). For example, we found mast cells specifically enriched for *HPGDS* and *PTGS2*, suggesting that they may be a dominant source of prostaglandin D2, which is implicated in activation of T helper 2 (Th2) cells[4]. Alongside these mediators, the production of instructive first-order cytokines primes recruitment and activation of effector mechanisms. In particular, IL-25, IL-33 and thymic stromal lymphopoietin (TSLP) are broadly regarded as epithelial-derived cytokines[4,5,16,20,22], yet little is known about which cells express them in human disease. *TSLP* expression was uniquely restricted to basal cells, suggesting a link between increased basal cell numbers in disease and activation of effector cells (Fig. 1d, Extended Data Figs. 3a, 4b, c; Supplementary Discussion IV).

Expression of second-order effector cytokines was identified in a subset of $CD4^+$ T cells expressing *IL4*, *IL5*, *IL13* and *HPGDS*, fitting the profile of allergen-specific Th2A cells[23] (Fig. 1d, Extended Data Fig. 4c–e;

[1]Institute for Medical Engineering and Science (IMES), Massachusetts Institute of Technology, Cambridge, MA, USA. [2]Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA, USA. [3]Koch Institute for Integrative Cancer Research, MIT, Cambridge, MA, USA. [4]Broad Institute of MIT and Harvard, Cambridge, MA, USA. [5]Ragon Institute of MGH, MIT and Harvard, Cambridge, MA, USA. [6]Division of Infectious Diseases and Division of Gastroenterology, Boston Children's Hospital, Boston, MA, USA. [7]Jeff and Penny Vinik Center for Allergic Disease Research, Division of Rheumatology, Immunology, and Allergy, Brigham and Women's Hospital, Boston, MA, USA. [8]Department of Medicine, Harvard Medical School, Boston, MA, USA. [9]Program in Computational and Systems Biology, Massachusetts Institute of Technology, Cambridge, MA, USA. [10]Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA, USA. [11]Department of Surgery, Brigham and Women's Hospital, Boston, MA, USA. [12]Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA, USA. [13]Harvard–MIT Division of Health Sciences & Technology, Cambridge, MA, USA. [14]These authors contributed equally: Jose Ordovas-Montanes, Daniel F. Dwyer. [15]These authors jointly supervised this work: Nora A. Barrett, Alex K. Shalek. *e-mail: nbarrett@bwh.harvard.edu; shalek@mit.edu

© 2018 Springer Nature Limited. All rights reserved.

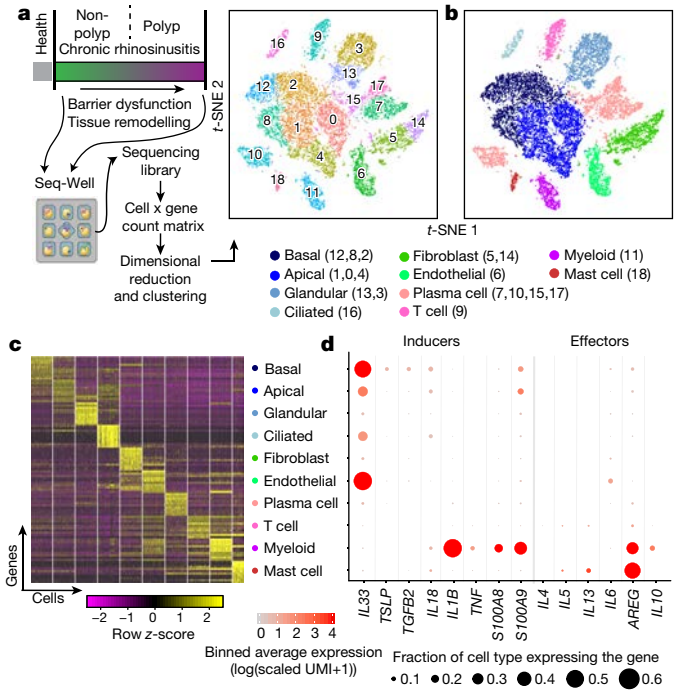


Fig. 1 | Mapping the T2I inflamed human sinus cellular ecosystem by scRNA-seq. a, **b**, Clinical disease spectrum ($n = 12$ samples) and experimental workflow leading to a *t*-distributed stochastic neighbour embedding (*t*-SNE) plot displaying 18,036 single cells, coloured by shared nearest neighbour (SNN) clusters (**a**) and cell types (**b**) (receiver operating characteristic (ROC) test; Supplementary Table 3; Methods) from respiratory tissue. **c**, **d**, Heat map of top 10 marker genes by ROC test (**c**) (area under curve (AUC) > 0.73) for indicated cell types; maximum 500 cells plotted per type (Extended Data Fig. 2a annotated; Supplementary Table 3 full gene list) and dot plot of T2I mediators mapped onto cell types (**d**) across all samples (Extended Data Fig. 4 shows this by disease state).

Supplementary Discussion V). Additionally, substantial numbers of mast cells expressed *IL5* and *IL13*, and, along with myeloid cells, were the main expressers of the tissue-reparative cytokine *AREG*[22]. Notably, patients with or without polyps showed consistent cells-of-origin for T2I-related chemokines, lipids and cytokines, with the exception of select mediators (Extended Data Fig. 4a, b; Supplementary Discussion IV). Expression of several genes implicated in allergic diseases[24] by genome-wide association studies (GWAS) was restricted to specific cell types. We therefore mapped the expression of candidate risk genes, including *GATA2*, *IL1RL1* (which encodes the IL-33 receptor ST2 subunit), *CDHR3*, *KIF3A*, *TMEM232* and *MYC* (Extended Data Fig. 4f; Supplementary Discussion VI). Cellular maps of tissues commonly affected by inflammatory disease should help to provide mechanistic insights into genotype–phenotype interactions.

We further analysed the epithelial clusters (Fig. 2a, Extended Data Fig. 5a–c), providing single-cell human transcriptomes[25] for basal, secretory, glandular and ciliated cell types from a T2I ecosystem (Fig. 2a, b, Extended Data Fig. 5, Supplementary Table 3). Analysis of epithelial marker genes identified conserved transcriptional programs in basal (three clusters), differentiating/secretory (three clusters), glandular (two clusters) and ciliated (one cluster) cell types[2,3] (Fig. 2a, b, Extended Data Fig. 5a–d, Supplementary Table 3; Supplementary Discussion VII).

On the basis of our observation of striking polyposis-related differences across clusters within cell types (Fig. 2c, Extended Data Fig. 5e; Supplementary Discussion VIII), we quantified the numerical over-representation of cells from the non-polyp and polyp ecosystems within each cluster and type. The clusters comprising basal, differentiating/secretory, and glandular cells showed the most significant links to the disease state (*P* values by Fisher's least-significant difference test; Fig. 2c, Supplementary Table 3). We compared transcriptomes of differentiating/secretory cells[3] (containing *KRT8*-expressing secretory and apical goblet cells), noting that secretory cells from polyps appear to supplant antimicrobial function with tissue repair (Fig. 2d, Supplementary Table 3; Supplementary Discussion VIII).

Of note, we observed expression of *MUC5B* within glandular mucus cells (cluster 13), whereas *MUC5AC* was expressed in a distinct subset of secretory goblet cells co-expressing *SCGB1A1* and *FOXA3* (Fig. 2b, Extended Data Fig. 5f, g; Supplementary Discussion IX). This suggests that the goblet cell program is overlaid on a secretory cell base[2]. We also assessed glandular heterogeneity, identifying five discrete subsets with variegated antimicrobial expression[3] (Fig. 2a, Extended Data Fig. 6a, b, Supplementary Table 3; Supplementary Discussion IX). This compartmentalization may represent a mechanism for regulated secretion, with imbalances in cell types or states affecting innate host defence.

To contextualize shifts associated with disease state, we turned to sinonasal scrapings as a method of sampling healthy apical cells through Seq-Well (Extended Data Fig. 6c, d, Supplementary Tables 3, 6; 18,704 additional cells: $n = 3$ healthy inferior turbinate (InfTurb), $n = 4$ polyp-patient InfTurb, $n = 2$ EthSin-polyp directly). We recovered immune cells and differentiating/secretory and ciliated epithelial cells from the InfTurb of patients with polyposis and healthy controls, but basal cells were found only in polyp scrapings (Extended Data Fig. 6d–f, Supplementary Table 3; Supplementary Discussion X). By combining all epithelial cells from the surgical resections with scrapings (Fig. 2a–c, e), we identified a conserved core secretory gene set that was present in all sites sampled, as well as healthy, CRS-InfTurb, CRS-EthSin-non-polyp and CRS-EthSin-polyp specific gene signatures. Overall, we note a shift from IFN-α/IFN-γ-induced genes to IL-4/IL-13-induced genes with increasing disease severity (Fig. 2e–g, Supplementary Table 3; Supplementary Discussion XI). Secretory cells from involved CRS-EthSin tissue differ markedly from those of the InfTurb, and secretory cells in non-polyp and polyp EthSin reach distinct states in which altered functionality may be linked to severity of disease.

As specialized epithelial cell types arise from basal progenitors[2,10], we formally examined their distribution in each sample (Fig. 3a, Extended Data Fig. 7a). Our data indicate a significant loss of epithelial ecological diversity in nasal polyps by Simpson's index (see Methods), largely driven by glandular and ciliated cell depletion, and an enrichment in basal cells (Fig. 3a, b, Extended Data Fig. 7a–d; Supplementary Discussion XII). This altered diversity tracked closely with rank-ordered pathology of patient tissue samples, which correlated positively with basal cell frequency ($r = 0.6252$) and negatively with epithelial diversity ($r = -0.6824$; Extended Data Fig. 7e). We hypothesize that alterations in the immune compartment in polyps may represent an overcorrection in attempting to balance epithelial shifts (Extended Data Fig. 7f; Supplementary Discussion XII).

To confirm our findings on epithelial cell types, we applied complementary approaches. Using flow cytometry[10], we demonstrated that the frequency of basal cells significantly increased in polyps at the expense of differentiated epithelial cells in 13 additional patients (Fig. 3c, Extended Data Fig. 7g, h). Using histology (which, unlike scRNA-seq or flow cytometry, is not subject to dissociation-induced artefacts), we confirmed[8] a significant increase of $p63^+$ cells per 1,000 μm^2 of epithelial area and a striking loss of glands in polyps (Fig. 3d, e, Extended Data Fig. 7i, j). We also used marker genes for specialized lineages to deconvolve bulk EthSin-tissue RNA sequencing (RNA-seq) from an additional cohort of 27 individuals. We identified four patient clusters and confirmed glandular enrichment in non-polyps, and shifts in secretory cell states and the progressive acquisition of basal-associated transcripts in polyps (Fig. 2d, f, 3f, g, h, Supplementary Tables 1, 3; Supplementary Discussion XIII). Finally, we validated these findings with publicly-available RNA-seq datasets containing normal human sinus tissue and polyps (Extended Data Fig. 7k, l; Supplementary Discussion XIII).

To investigate mechanisms that might account for the reduced epithelial diversity in polyps, we compared the transcriptomes of

© 2018 Springer Nature Limited. All rights reserved.

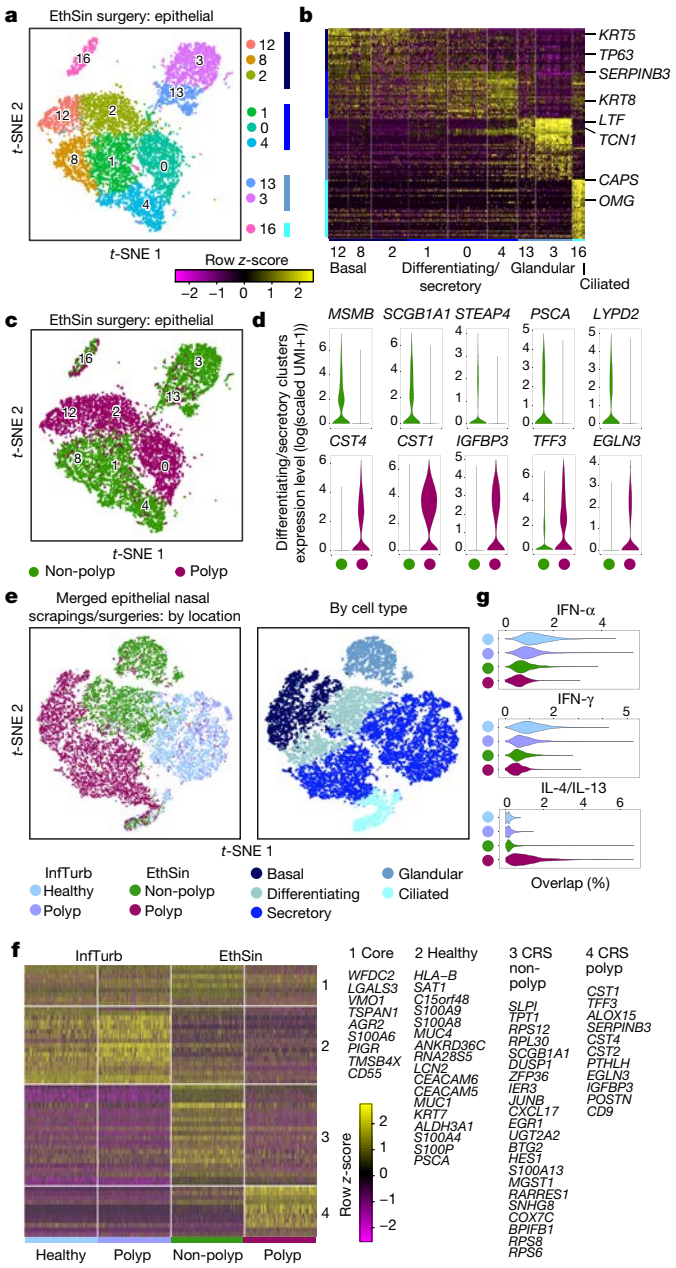


Fig. 2 | Single-cell transcriptomes of epithelial cells in T2I highlight shifts in secretory cell states across health and disease. a, **b**, *t*-SNE plot of 10,274 epithelial cells ($n = 12$ samples), coloured by SNN clusters (Fig. 1; re-clustered in Extended Data Fig. 6) with blue colour bars representing cell types determined according to Extended Data Fig. 5 (**a**) and heat map of marker genes by ROC (**b**) (AUC > 0.65; full list in Supplementary Table). **c**, **d**, *t*-SNE plot (**c**) coloured by disease ($n = 6$ non-polyp samples, $n = 6$ polyp samples) and violin plots (**d**) (all violins generated using standard Seurat implementation with default smoothing, density generated at >25% positive values, widest aspect centre of positive measures, minima and maxima within scale representing all points) for differentially expressed genes across disease state in differentiating or secretory cells. 2,566 cells, $n = 6$ non-polyp samples; 1,796 cells, $n = 6$ polyp samples. Bimodal test, all $P < 2.03 \times 10^{-55}$ or less with Bonferroni correction (exact values in Supplementary Table 3). **e**, *t*-SNE plot of 18,325 re-clustered single cells from merged nasal scrapings ($n = 9$) and surgical samples ($n = 12$) by location (left) (healthy InfTurb (3,681 cells, $n = 3$ samples), polyp-bearing patient InfTurb (1,370 cells, $n = 4$ samples), non-polyp EthSin surgical samples (5,928 cells, $n = 6$ samples), and polyp surgical and scraping samples directly from polyp in EthSin (7,346 cells, $n = 8$ samples)), and cell type (right; for immune cells, see Extended Data Fig. 6) (3,152 basal, 3,089 differentiating, 8,840 secretory, 1,105 ciliated, and 2,139 glandular cells). **f**, Heat map of secretory cells (1,000 cells per location) displaying select genes (AUC > 0.65; Supplementary Table 3). **g**, Violin plots of IFN-α, IFN-γ, and IL-4/IL-13 gene signatures for secretory cells. Healthy InfTurb (3,414 cells), polyp-bearing patient InfTurb (1,239 cells), non-polyp EthSin surgical samples (1,048 cells), polyp surgical and scraping samples directly from polyp in EthSin (3,139 cells); effect size −1.16 (IFN-α), −1.05 (IFN-γ) and 1.32 (IL-4/IL-13), polyp EthSin versus healthy. Mann–Whitney *U*-test, $P < 2.2 \times 10^{-16}$.

non-polyp and polyp basal progenitors[2,10], identifying increased expression of transcripts involved in extracellular matrix remodelling and chemo-attraction of effector cells, and a decrease in protease-inhibitor expression and metabolic genes in polyps (Fig. 4a). As some of these upregulated genes are canonical IL-4/IL-13 responsive transcripts[7], we assessed cytokine-induced gene sets. A combined IL-4/IL-13 signature is strongly induced not only in differentiated polyp epithelium, but also in basal cells, with a large effect size between disease states (Fig. 4b, c, Extended Data Fig. 8a, Supplementary Table 4). IFN-α and IFN-γ signatures—indicative of a type 1 immune module[6]—had small effect sizes (Extended Data Fig. 8b, Supplementary Table 4). Furthermore, from specific hallmark genes, we observed altered balance between Wnt (*CD44*) and Notch (*HEY1*) signalling in polyp epithelium favouring Wnt[26,27] (Fig. 4a, c, Supplementary Table 4). We further contextualized our basal cell findings by defining alterations in the fibroblast niche that correlate with basal hyperplasia, and identifying significant changes in myeloid and endothelial cell gene expression (Extended Data Figs. 7b, 8c–f; Supplementary Discussion XIV, XV).

Next, we used diffusion pseudotime mapping (see Methods), aligning and reconstructing how basal cells differentiate to mature secretory cells to identify where basal cells become 'stuck' in polyps. In the non-polyp ecosystem, we observed that basal cells traverse a wider swath of common pseudotime, with the majority of secretory cells distributed towards the trajectory's terminus (Fig. 4d, e, Extended Data Fig. 9a). Conversely, in polyps, basal cells accumulate shy of the trajectory's midpoint, losing the true progenitor position occupied by cluster 8, yet failing to contribute towards later differentiation states (Fig. 4d, e, Extended Data Fig. 9a). Ordering cells along this common axis, we identified several genes that were dysregulated in polyps during epithelial cell differentiation (Extended Data Fig. 9b, Supplementary Table 3; Supplementary Discussion XVI).

As these data highlighted an impairment in differentiation of basal cells in polyp tissue, we sorted basal cells (Extended Data Fig. 7h) from three non-polyp and seven polyp tissues and performed Omni assay for transposase-accessible chromatin (ATAC) sequencing (Omni-ATAC-seq) to identify intrinsic epigenetic changes from the integration of extrinsic cellular signalling events[28], and subsequent bulk RNA-seq to confirm and extend our findings (Methods). Polyp basal cells were enriched in peaks for bZIP transcription factor motifs, including various AP-1 family members[11] such as JUN, along with FOXA1, ATF3, KLF5 and p63, which have been associated with the maintenance of an undifferentiated state, chromatin opening and oncogenesis (Fig. 4f, Extended Data Fig. 9b–f, Supplementary Table 5; Supplementary Discussion XVII). Clustering of enriched motifs revealed changes in correlation according to disease state (Extended Data Fig. 9c–f; Supplementary Discussion XVII). We further identified expressed candidate transcription factors that may bind to these accessible sites (Fig. 4f, g, Extended Data Fig. 9e, f; Supplementary Discussion XVIII). Collectively, our transcriptomic, pseudotemporal and epigenetic studies led us to hypothesize that during chronic T2I, basal cell differentiation is intrinsically impaired through the influence of extrinsic cues (notably, activation of IL-4/IL-13 and Wnt pathways).

To functionally test for intrinsically altered differentiation potential in vitro, we first seeded basal cells from non-polyp or polyp tissue into air–liquid interface (ALI) cultures (Fig. 5a, Extended Data Fig. 9g, Supplementary Table 3; Supplementary Discussion XIX). Our results suggest that basal cells from polyps can be released from their 'stuck' state and differentiate towards a mixed-tissue secretory cell phenotype

© 2018 Springer Nature Limited. All rights reserved.

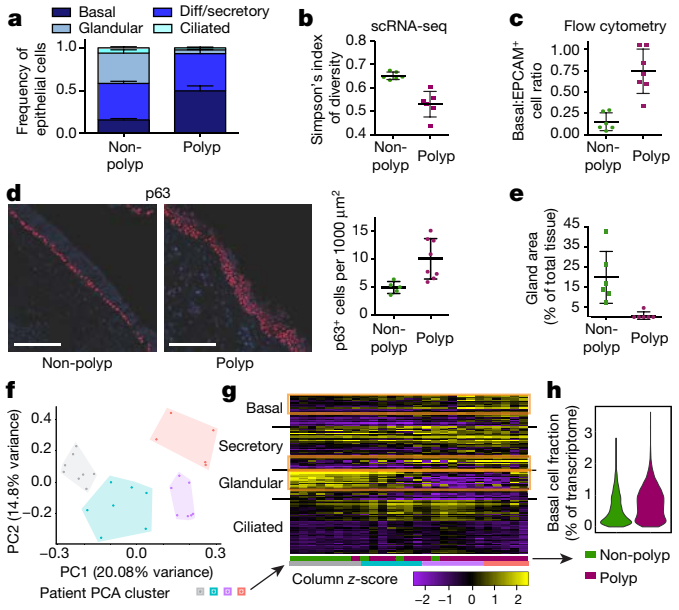


Fig. 3 | Reduced epithelial ecological diversity and basal cell hyperplasia in nasal polyps. a, scRNA-seq cell frequency (data from Fig. 2a; Extended Data Fig. 7a shows individual points) calculated for each sample. Basal cell, $P = 0.00023$; glandular, $P < 0.0001$; ciliated, $P = 0.0387$. non-polyp versus polyp. **b**, Simpson's index (Methods, $P = 0.0021$). $n = 6$ non-polyp samples, $n = 6$ polyp samples. Two-sided *t*-test; data are mean ± s.e.m. **c**, Quantification of cells by flow cytometry (Extended Data Fig. 7, full gating). $n = 6$ non-polyp samples, $n = 7$ polyp samples. Two-sided *t*-test; $P = 0.0005$; data are mean ± s.e.m. **d**, Immunofluorescence (left) and quantification (right) of p63$^+$ basal cells normalized to 1,000 µm^2 of epithelium. $n = 5$ non-polyp patients, 10 sections; $n = 8$ polyp patients, 41 sections. Mann–Whitney *U*-test, $P = 0.0282$. Data are mean ± s.d. Scale bar, 100 µm. **e**, Quantification of gland area. $n = 6$ non-polyp, $n = 6$ polyp patients. Mann–Whitney *U*-test, $P = 0.0022$ (Extended Data Fig. 7, isotype and representative). **f**, **g**, Bulk-tissue RNA-seq deconvolution by PCA (**f**) and heat map (**g**) over epithelial subset-specific genes (rows) with *k*-nearest neighbour (*k*NN) clusters ($n = 4$; Methods), from $n = 10$ non-polyp samples, $n = 17$ polyp samples (columns). **h**, Violin plot of basal cell gene fraction in scRNA-seq epithelium (Methods, Supplementary Table 4). 5,928 cells, $n = 6$ non-polyp; 4,346 cells, $n = 6$ polyp. Effect size 0.457, Mann–Whitney *U*-test, $P < 2.2 \times 10^{-16}$.

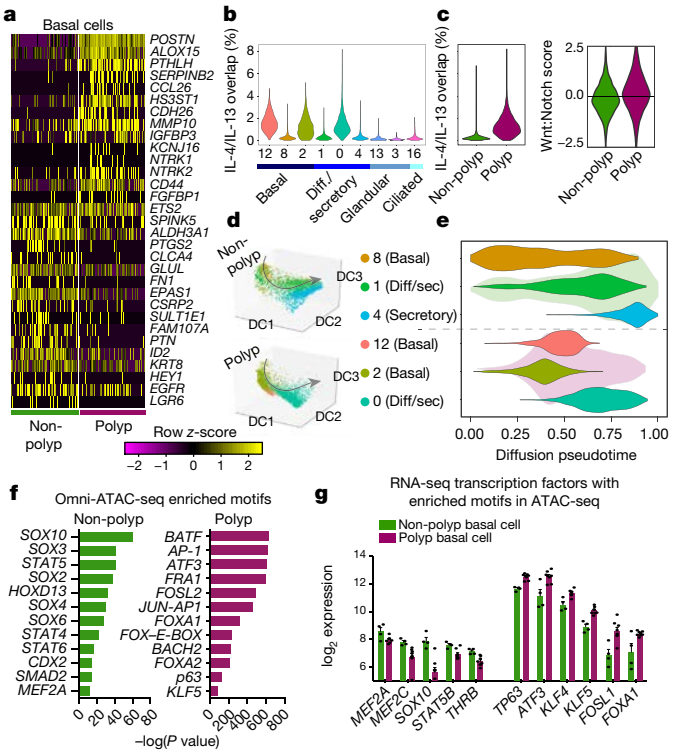


Fig. 4 | T2I cytokines and developmental pathways converge at the epigenetic level in basal cells to intrinsically impair differentiation in vivo. a, Heat map of select genes over basal cell clusters 8 and 12. $n = 6$ non-polyp samples, 860 cells; $n = 6$ polyp samples, 858 cells. Bimodal test, all displayed genes $P < 1.97 \times 10^{-39}$ or less with Bonferroni correction (Supplementary Table 3). **b**, Violin plot (Extended Data Fig. 8; cell numbers in Supplementary Table 3, genes listed in Supplementary Table 4) for commonly-induced IL-4/IL-13 gene signature. Mann–Whitney *U*-test, $P < 1.76 \times 10^{-15}$, relative to mean score, with Bonferroni correction. **c**, Violin plots of shared IL-4/IL-13 signature (Mann–Whitney *U*-test, $P < 2.2\text{x}10^{-16}$, effect size 1.305) and Wnt:Notch target gene ratio. Two-sided *t*-test, $P < 2.2 \times 10^{-16}$, effect size 0.334. Note the truncated axis; zero indicates equal scores. $n = 6$ non-polyp samples, 5,928 cells; $n = 6$ polyp samples, 4,346 cells. **d**, **e**, Diffusion pseudotime over epithelial cells with unified gene list (**d**) (Supplementary Table 3 cell numbers, gene list) and violin plot of pseudotime component with green ($n = 6$ non-polyp samples) and purple ($n = 6$ polyp samples) underlying distribution (**e**). DC, diffusion component. **f**, **g**, Omni-ATAC-seq and HOMER motif enrichment (**f**) over background peaks. All *Q* values < 0.0002, Benjamini correction. Transcription factors from low-input RNA-seq (**g**) on sorted basal cell populations ($\log_2$ expression rlog normalized using DESeq2). Two-sided *t*-test, $P < 0.05$ or less; Holm–Sidak correction; data are mean ± s.e.m. $n = 3$ non-polyp samples (4 RNA-seq analyses); $n = 7$ polyp samples.

if provided with strong and sustained extrinsic cues, even in the presence of IL-13 (Fig. 5b; Extended Data Figs. 7h, 9f, h, i; Supplementary Discussion XIX).

Second, since ALI cultures enforced strong terminal differentiation, we directly tested how IL-4/IL-13 act to induce rapid expression of genes in basal cells cultured for 5 weeks ex vivo, hypothesizing that polyp basal cells would respond more vigorously to exogenous cytokines than non-polyp basal cells[14]. Surprisingly, we identified 482 genes that were induced in non-polyp basal cells, but only 42 in polyps (Fig. 5c, Supplementary Table 3). Principal component analysis (PCA) highlighted that whereas unstimulated non-polyp basal cells grouped together, polyp basal cells were distributed along PC1, which captured cytokine stimulation (Fig. 5c). Identifying overlaps in genes that are significantly induced by cytokine treatment in non-polyp basal cells with genes that are upregulated at baseline in polyp versus non-polyp basal cells resolved 132 genes (Fig. 5c, Supplementary Table 3).

We focused on the central overlap of these three differential expression tests, which included *CTNNB1* (β-catenin), the key effector of Wnt pathway activation[26,27] (Fig. 5c). We highlight the fundamental finding that *CTNNB1* was robustly induced in non-polyp and polyp basal cells in a dose-sensitive fashion to IL-4 and IL-13. Moreover, baseline *CTNNB1* expression in polyp basal cells was equivalent to the levels induced by cytokine treatment of non-polyp cells (Fig. 5d). Wnt-pathway target genes were significantly upregulated across the tested doses, confirming overall activation of the pathway, and of specific factors (such as *CTGF*) (Fig. 5d, Extended Data Fig. 9j, Supplementary Table 4; Supplementary Discussion XX). On the basis of polyp epithelial gene signatures (Fig. 4c) and our functional testing for IL-4/IL-13 induced genes 'remembered' by polyp basal cells (Fig. 5c, d), we propose that chronic IL-4/IL-13 exposure in vivo can lead to persistent expression of Wnt/β-catenin target genes in a cell-intrinsic fashion, even in the absence of exogenous cytokines.

One polyp patient sampled through scraping commenced treatment with a monoclonal antibody targeting the shared IL-4Rα subunit of the IL-4 and IL-13 receptors to treat atopic dermatitis, which provided an opportunity to examine the in vivo relevance of our observational, mechanistic and functional data on how T2I cytokines influence basal cell states (Fig. 5e, Extended Data Fig. 10a, b). We compared cells recovered from pre- and 6-week-post-antibody treatment scrapings, and through surgical intervention at 7 weeks after antibody treatment (Fig. 5e, Extended Data Fig. 10a, b, Supplementary Table 7; Supplementary Discussion XXI). We identified basal cells and

© 2018 Springer Nature Limited. All rights reserved.

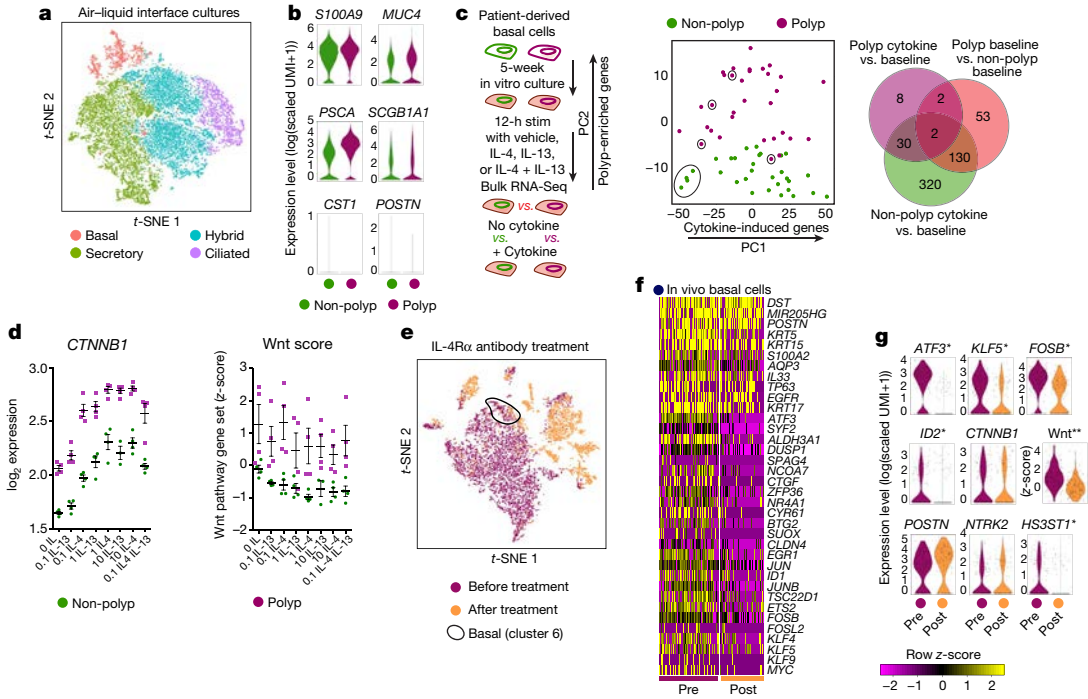


Fig. 5 | Transcriptional memory of IL-4/IL-13 exposure retained by basal cells ex vivo; in vivo IL-4Rα blockade partially resets state in polyps. a, *t*-SNE plot of cell types from air–liquid interface (ALI) cultures (Extended Data Fig. 9g) over 16,173 single cells ($n = 2$ non-polyp donors, 8,483 cells; $n = 2$ polyp donors, 7,690 cells). **b**, Violin plots for secretory genes (from Fig. 2f) in secretory cells from ALI culture (3,277 non-polyp cells, 3,143 polyp cells). None were statistically significant, except *MUC4*, *PSCA* and *SCGB1A1*, which were more highly expressed in polyp secretory cells. Bimodal $P < 4.04 \times 10^{-16}$, Bonferroni correction. **c**, Basal cell inflammatory memory of IL-4/IL-13 stimulation (left, Methods), displayed as PCA over variable genes (middle; circle denotes no exogenous cytokine) and Venn diagram showing overlaps of differential expression (right). Two-tailed *t*-test, Bonferroni-corrected $P < 0.05$; full gene lists in Supplementary Table 3. **d**, *CTNNB1* expression and Wnt pathway *z*-score (Methods) in basal cells from **c**. $n = 4$ samples per dose. Dose expressed in ng ml^{-1}. Two-way ANOVA; $P < 0.0001$ for *CTNNB1*; $P < 0.0282$ for Wnt pathway. $n = 2$ basal cell donors each for non-polyp and polyp samples (**a**–**d**). **e**–**g**, scRNA-seq for an individual treated with IL-4Rα antibody. *t*-SNE plot of 8,764 single cells from nasal polyps coloured by pre-treatment (5,731 cells) and post-treatment (3,033 cells) samples (**e**) (see also Extended Data Fig. 10), heat map of select genes (AUC > 0.68 core; $P < 2.46 \times 10^{-5}$ or less with Bonferroni correction) over basal cells (**f**) (200 pre, 151 post), and violin plots for select genes (**g**). *$P < 0.00087$, bimodal with Bonferroni correction; Wnt score pre-treatment versus post-treatment, **$P < 2.2 \times 10^{-16}$, two-tailed *t*-test, effect size = 0.942; otherwise not significant. Full gene list in Supplementary Table 3.

generated a heat map containing their top marker genes, agnostic to treatment, followed by genes that were differentially expressed pre- and post-treatment, leveraging myeloid cells to identify basal-specific changes (Fig. 5f, Extended Data Fig. 10a–c, Supplementary Table 3; Supplementary Discussion XXI).

In the context of our earlier findings, these results enabled us to identify several key gene sets, including a conserved core set of basal cell genes (Extended Data Fig. 10d). Several transcription factors that are upregulated in polyp basal cells identified through Omni-ATAC-seq and RNA-seq (*ATF3, KLF5* and *FOSB*) were significantly down-regulated after treatment (Fig. 5g). Whereas Wnt pathway target gene expression was globally reduced, *CTNNB1* expression was retained, as was expression of genes that are upregulated both in vitro and in vivo in polyp basal cells, suggesting that some disease-associated genes in this patient, and at this time point, persist (Fig. 5g, Extended Data Fig. 10d; Supplementary Discussion XXI).

Lastly, we sought to understand how changes in the basal epithelium propagated through to secretory cells. Within secretory cells recovered from scrapings of both InfTurb and accessible polyp tissue pre- and post-treatment, our data suggest that even though CRS-EthSin samples have unique secretory cell signatures (Fig. 2f), cytokine blockade leads to expression of genes associated with healthy InfTurb secretory cells, even in polyp tissue (Extended Data Fig. 10e–h, Supplementary Table 8; Supplementary Discussion XXI).

One goal of understanding the cellular and molecular pathways activated in T2I is to provide mechanisms that explain persistent chronic allergic inflammatory disease[29]. Using scRNA-seq applied to patients across the CRS spectrum, our study provides descriptive, mechanistic and functional insights into an enigmatic basal cell state and productive differentiation of a barrier tissue. We reveal differences in expression of antimicrobial genes by secretory cells relative to healthy tissue, a loss of glandular cell heterogeneity, and strong induction of a transcriptional program by IL-4/IL-13 at the level of basal progenitor cells[15]. Our data may help to explain why nasal polyposis is associated with infections by specific microorganisms[7], and how a monoclonal antibody that targets the shared IL-4/IL-13 receptor can reduce nasal polyp burden.

Together with recent work in the murine intestinal tract and skin[11,13,16,17,30], our results provide evidence in humans for the emerging paradigm of stem cell dysfunction altering the set point of barrier tissues, highlighting substantial overlap among putative driving transcription factors (ATF3, AP-1, p63 and KLF5)[13]. This demonstrates that the principle of inflammatory memory[28] underlying barrier tissue adaptation is a generalizable phenomenon that is observed in distinct anatomical locations, inflammatory modules, and species. We build on these findings by culturing basal cells ex vivo and identifying the indelible mark of IL-4/IL-13 as a baseline induction of the Wnt pathway. We propose that basal cells form 'memories' of chronic exposure to an inflammatory T2I environment, shifting the entire cellular ecosystem away from productive differentiation and propagating disease. Future work will seek to determine the relative contributions of memory stored in distinct cellular compartments to develop the most effective mechanisms by which to erase them.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at https://doi.org/10.1038/s41586-018-0449-8.

© 2018 Springer Nature Limited. All rights reserved.

Received: 5 November 2017; Accepted: 4 July 2018;
Published online: 22 August 2018

1. Schleimer, R. P. & Berdnikovs, S. Etiology of epithelial barrier dysfunction in patients with type 2 inflammatory diseases. *J. Allergy Clin. Immunol.* **139**, 1752–1761 (2017).
2. Hogan, B. L. et al. Repair and regeneration of the respiratory system: complexity, plasticity, and mechanisms of lung stem cell function. *Cell Stem Cell* **15**, 123–138 (2014).
3. Whitsett, J. A. & Alenghat, T. Respiratory epithelisal cells orchestrate pulmonary innate immunity. *Nat. Immunol.* **16**, 27–35 (2015).
4. Iwasaki, A., Foxman, E. F. & Molony, R. D. Early local immune defences in the respiratory tract. *Nat. Rev. Immunol.* **17**, 7–20 (2017).
5. Holtzman, M. J., Byers, D. E., Alexander-Brett, J. & Wang, X. The role of airway epithelial cells and innate immune cells in chronic respiratory disease. *Nat. Rev. Immunol.* **14**, 686–698 (2014).
6. Iwasaki, A. & Medzhitov, R. Control of adaptive immunity by the innate immune system. *Nat. Immunol.* **16**, 343–353 (2015).
7. Schleimer, R. P. Immunopathogenesis of chronic rhinosinusitis and nasal polyposis. *Annu. Rev. Pathol.* **12**, 331–357 (2017).
8. Zhao, L. et al. Increase of poorly proliferated $p63^{+}$/$Ki67^{+}$ basal cells forming multiple layers in the aberrant remodeled epithelium in nasal polyps. *Allergy* **72**, 975–984 (2017).
9. Hansel, F. K. Clinical and histopathological studies of the nose and sinuses in allergy. *J. Allergy* **1**, 43–70 (1929).
10. Rock, J. R. et al. Basal cells as stem cells of the mouse trachea and human airway epithelium. *Proc. Natl Acad. Sci. USA* **106**, 12771–12775 (2009).
11. Karin, M. & Clevers, H. Reparative inflammation takes charge of tissue regeneration. *Nature* **529**, 307–315 (2016).
12. Gierahn, T. M. et al. Seq-Well: portable, low-cost RNA sequencing of single cells at high throughput. *Nat. Methods* **14**, 395–398 (2017).
13. Naik, S. et al. Inflammatory memory sensitizes skin epithelial stem cells to tissue damage. *Nature* **550**, 475–480 (2017).
14. Netea, M. G. et al. Trained immunity: a program of innate immune memory in health and disease. *Science* **352**, aaf1098 (2016).
15. Rochman, M. et al. Neurotrophic tyrosine kinase receptor 1 is a direct transcriptional and epigenetic target of IL-13 involved in allergic inflammation. *Mucosal Immunol.* **8**, 785–798 (2015).
16. von Moltke, J., Ji, M., Liang, H. E. & Locksley, R. M. Tuft-cell-derived IL-25 regulates an intestinal ILC2-epithelial response circuit. *Nature* **529**, 221–225 (2016).
17. Lindemans, C. A. et al. Interleukin-22 promotes intestinal-stem-cell-mediated epithelial regeneration. *Nature* **528**, 560–564 (2015).
18. Cheng, L. E. & Locksley, R. M. Allergic inflammation–innately homeostatic. *Cold Spring Harb. Perspect. Biol.* **7**, a016352 (2014).
19. Palm, N. W., Rosenstein, R. K. & Medzhitov, R. Allergic host defences. *Nature* **484**, 465–472 (2012).
20. Gieseck, R. L. III, Wilson, M. S. & Wynn, T. A. Type 2 immunity in tissue repair and fibrosis. *Nat. Rev. Immunol.* **18**, 62–76 (2017).
21. von Andrian, U. H. & Mackay, C. R. T-cell function and migration. *Two sides of the same coin. N. Engl. J. Med.* **343**, 1020–1034 (2000).
22. Allakhverdi, Z. et al. Thymic stromal lymphopoietin is released by human epithelial cells in response to microbes, trauma, or inflammation and potently activates mast cells. *J. Exp. Med.* **204**, 253–258 (2007).
23. Wambre, E. et al. A phenotypically and functionally distinct human T_H2 cell subpopulation is associated with allergic disorders. *Sci. Transl. Med.* **9** (2017).
24. Portelli, M. A., Hodge, E. & Sayers, I. Genetic risk factors for the development of allergic disease identified by genome-wide association. *Clin. Exp. Allergy* **45**, 21–31 (2015).
25. Zuo, W. L. et al. Ontogeny and biology of human small airway epithelial club cells. *Am. J. Respir. Crit. Care Med.* (2018).
26. Boscke, R. et al. Wnt signaling in chronic rhinosinusitis with nasal polyps. *Am. J. Respir. Cell Mol. Biol.* **56**, 575–584 (2017).
27. Nusse, R. & Clevers, H. Wnt/β-catenin signaling, disease, and emerging therapeutic modalities. *Cell* **169**, 985–999 (2017).
28. Ostuni, R. et al. Latent enhancers activated by stimulation in differentiated cells. *Cell* **152**, 157–171 (2013).
29. Lambrecht, B. N. & Hammad, H. The immunology of the allergy epidemic and the hygiene hypothesis. *Nat. Immunol.* **18**, 1076–1083 (2017).
30. Beyaz, S. et al. High-fat diet enhances stemness and tumorigenicity of intestinal progenitors. *Nature* **531**, 53–58 (2016).

Acknowledgements We thank S.L. Carroll for technical support with Seq-Well experiments; H. Raff for RNA extraction; J. Lai for histology; A. Chicoine of the Brigham and Women's Human Immunology Flow Core for assistance with isolating cells; L. Ludwig, J. Hammelman and J. Buenrostro for advice on reagents and analysis for ATAC-seq; D. Lingwood, U.H. von Andrian, B. Walker, S. Pillai, N. Yosef, S. Rakoff-Nahoum, S. Beyaz, C. Borges, M.B. Cole, N. Yosef, R. Satija and C. Bingle for discussions and comments on the manuscript; Shalek Laboratory members for experimental and computational advice; and M. Morrison for administrative support. A.K.S. was supported by the Searle Scholars Program, the Beckman Young Investigator Program, the Pew-Stewart Scholars, a Sloan Fellowship in Chemistry, NIH grants 1DP2OD020839, 2U19AI089992, 1U54CA217377, P01AI039671, 5U24AI118672, 2RM1HG006193, 1R33CA202820, 2R01HL095791, 1R01AI138546, 1R01HL126554, 1R01DA046277, 2R01HL095791, and Bill and Melinda Gates Foundation grants OPP1139972 and BMGF OPP1116944; N.A.B. by NIH R01HL120952 and Steven and Judy Kaye Young Innovators Award; T.M.L. by NIH R01HL128241; J.A.B. by NIH U19AI095219, R01AI078908, R01AI136041, R01HLI136209; D.F.D. by T32AI007306 (to J.A.B.); K.M.B. by NIH AADCRC Opportunity Fund Award U19AI070535; and K.N.C by NIH K23AI118804. S.K.N. was supported by NIH 2R01GM081871-09 to B.B. Support was also provided from the Koch Institute Support (core) Grant P30-CA14051 from the NCI, and Ragon Institute NIH-funded Centers for AIDS Research (P30 AI060354, Harvard University Center for AIDS Research), supported by NIH co-funding and participating Institutes and Centers: NIAID, NCI, NICHD, NHLBI, NIDA, NIMH, NIA, FIC, and OAR. J.O.M. is a HHMI Damon Runyon Cancer Research Foundation Fellow (DRG-2274-16), and thanks S. Montanes-Ordovas for encouraging him to work on human allergic disease.

Reviewer information *Nature* thanks R. Schleimer and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions J.O.-M., D.F.D., T.M.L., J.A.B., N.A.B. and A.K.S. designed the study. N.B. performed surgeries. J.O.-M., D.F.D., C.D., M.V., K.M.B., K.N.C. and E.Y. collected patient samples and performed single-cell experiments. J.O.-M., M.V., D.F.D. and E.Y. performed in vitro experiments. M.H.W. and T.K.H provided the Seq-Well platform and expertise. H.R.K. and E.Y. performed histologic analyses. B.B. provided supervision and analysed epigenetic experiments. J.O.-M., D.F.D., S.K.N. and S.W.K. analysed data. J.O.-M., D.F.D., S.K.N., N.A.B. and A.K.S. interpreted data. J.O.-M., D.F.D., N.A.B. and A.K.S. wrote the manuscript with input from all authors.

Competing interests The authors declare no competing interests.

Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41586-018-0449-8.
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-018-0449-8.
Reprints and permissions information is available at http://www.nature.com/reprints.
Correspondence and requests for materials should be addressed to N.A.B. or A.K.S.
Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© 2018 Springer Nature Limited. All rights reserved.

METHODS

Study participants and design for single-cell study from ethmoid sinus tissue. 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. Ethmoid sinus (EthSin) tissue was collected at the time of elective endoscopic sinus surgery from patients with physician-diagnosed CRS with and without nasal polyps on the basis of established guidelines[31]. Patients with polyps included patients with aspirin-tolerant chronic rhinosinusitis with nasal polyps (CRS polyp) and individuals with aspirin-exacerbated respiratory disease (AERD), both of which are referred to as CRS-EthSin-polyp for the purposes of this study. Patients were suspected of having AERD if they had asthma, nasal polyposis, and a history of respiratory reaction on ingestion of a COX 1 inhibitor, with confirmation via a graded oral challenge to aspirin. Subjects with cystic fibrosis and unilateral polyps were excluded from the study. No distinctions were made between these two disease endotypes in our study, as both present with polyposis, but we present the clinical diagnoses in Supplementary Table 1.

A tissue segment (one per patient) for bulk tissue RNA-seq was immediately placed in RNAlater (Qiagen) for RNA extraction, For patient samples loaded on Seq-Well and for flow-sorting to Omni-ATAC-seq/RNA-seq, tissue was received in-hand, placed in RPMI (Corning) with 10% FBS (ThermoFisher 10082-147) and immediately put on ice for transport. Details of the subjects' characteristics included in scRNA-seq cohort, tissue RNA-seq cohort, and basal cell flow cytometry/ATAC-seq/RNA-seq cohort (including age, gender, medication use, and disease severity) are included in Supplementary Table 1.

Originally, we enrolled a healthy control subject with no known history of CRS or nasal polyposis who was undergoing sinus surgery for concha bullosa. However, this subject upon pathology evaluation was found to have mild eosinophilia. A chart review revealed a history of allergic rhinitis and asthma, and their diagnosis was updated to CRS non-polyp clinically by the surgeon upon follow-up visits so we updated their status accordingly in our study. Additionally, non-polyp patient 6 was sampled twice (denoted as 6A and 6B), representing distinct cells that were captured on two different Seq-Well arrays. As such, they should not be viewed as a technical replicate and are referred to as distinct samples.

Collection of inferior turbinate and nasal polyp samples through nasal scraping. Nasal samples were collected from the inferior turbinate (InfTurb) of healthy control subjects and from the inferior turbinate and accessible polyp tissue in subjects with CRS-EthSin-polyps using the Rhino-Pro Curette, a sterile, disposable, mucosal collection device, as described[32,33]. One sample was taken from the right and left mid-inferior portion of the inferior turbinate using a gentle scraping motion. In two subjects with CRS polyp, with accessible nasal polyp tissue, the polyp tissue was sampled using the Rhino-Pro Curette under direct visualization. The nasal scrapings were placed directly in RPMI with 10% FBS and immediately put on ice for transport before loading on Seq-Well arrays. Details of the subjects' characteristics (including age, gender, medication use and disease severity) are included in Supplementary Table 1.

Nasal scraping allows for access to the superficial epithelial cell layer of the inferior turbinate[34]; by contrast, the surgical resections from EthSin that we utilize as the central dataset of this paper contain both epithelial cells and underlying tissue, including sub-mucosal glands[34] (Extended Data Fig. 6c). Since scraping samples a proximal but distinct anatomical location with a distinct technique, in addition to collecting InfTurb scrapings from healthy controls ($n=3$), we also collected InfTurb scrapings from individuals with polyps ($n=4$), and, from two of these individuals, from accessible polyps protruding beyond the middle meatus ($n=2$).

One subject with CRS polyps and co-morbid severe atopic dermatitis was started on dupilumab[35], a human monoclonal antibody that binds to the IL-4Rα subunit, which is approved for severe atopic dermatitis[36], and in a randomized, double-blind, placebo-controlled parallel-group study was shown to significantly reduce endoscopic nasal polyp burden after 16 weeks[37]. The inferior turbinate and nasal polyp tissue was sampled with the Rhino-Pro Curette pre- and post-treatment with 3 doses of dupilumab, and through endoscopic sinus surgery as noted above.

Tissue digestion. Single-cell suspensions from collected surgical specimens were obtained using a modified version of a previously published protocol[38], described below in detail. Each specimen was received directly in hand and processed directly with an average time from patient to loading onto the Seq-Well platform of 3 total hours, and never exceeding 4 h. Surgical specimens were collected into 30 ml of ice cold RPMI (Corning). Specimens were finely minced between two scalpel blades and incubated for 15 min at 37 °C in a rotisserie rack with end-over-end rotation in 25 ml digestion buffer supplemented with 600 U/ml collagenase IV (Worthington) and 20 μg/ml DNase 1 (Roche) in RPMI with 10% fetal bovine serum. After 15 min, samples were triturated five times using a syringe with a 16G needle and returned to the rotisserie rack for another 15 min. At the conclusion of the second digest period, samples were triturated an additional five times using a syringe with a 16G needle, at which point the digest process was stopped via the addition of EDTA to 20mM. Nasal scrapings were dissociated with one 15 min dissociation using collagenase and the 16G needle trituration was omitted and instead replaced with P1000 pipette trituration, as cell yields were typically <20,000 total cells. Processing downstream remained identical. Samples were typically fully dissociated at this step and were filtered through a 70-μm cell strainer and spun down at 500*g* for 10 min followed by a rinse with ice-cold PBS (ThermoFisher 10010023, Ca/Mg-free) to 30 ml total volume. Red blood cells (RBCs) were lysed using ACK buffer (ThermoFisher A1049201) for 3 min on ice to remove RBCs, even if no RBC contamination was visibly seen in order to maintain consistency across patient groups. Cells were then washed with sterile PBS and spun down at 500*g* for 5 min, resuspended in complete RPMI medium with 2% FCS (RPMI1640 (ThermoFisher 61870-127), 100 U/ml penicillin (ThermoFisher 15140-122), 100 μg/ml streptomycin (ThermoFisher 15140-122), 10 mM HEPES (ThermoFisher 15630-080), 2% FCS (ThermoFisher 10082-147), 50 μg/ml gentamicin (ThermoFisher 15750-060)), and counted to adjust concentration to 100,000 cells per ml for loading onto Seq-Well arrays.

Flow cytometry, cell sorting, and analysis. Single-cell suspensions in FACS Buffer (HBSS (ThermoFisher 14170161, Ca/Mg-free) supplemented with 2% FCS) were pre-incubated with Fc-Block (BD 564220) before staining for surface antigens. 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). Cells were stained for 30 min on ice in FACS buffer and then washed for immediate sorting. Cells were sorted on a BD FACSAria Fusion cell sorter using BD FACSDiva software. Up to 10,000 basal cells were sorted into 100 μl BAM banker (Wako chemicals) for Omni-ATAC-seq and cooled to −80 °C using a Mr. Frosty freezing container (Thermo scientific). Samples were stored at −80 °C until thawed for Omni-ATAC-seq. For bulk RNA-seq, 1,000 basal cells were sorted directly into 5 μl TCL buffer (Qiagen). FlowJo v10 by TreeStar was used to generate plots.

Histologic analyses. Biopsies were fixed in 4% paraformaldehyde, embedded in paraffin, and 6-μm sections were prepared and stained with haematoxylin and eosin for quantification of glandular areas. Photomicrographs encompassing the entire area of each biopsy were taken. Total and glandular areas were measured with ImageJ software and expressed as glandular area as a percentage of total area. For p63 immunofluoresence, sections were quenched for 10 min in 1 mg/ml sodium borohydride in PBS. For antigen retrieval, slides were placed in a Coplin jar with preheated citrate target retrieval buffer (DAKO) at 95 °C and transferred to a steamer for 60 min. Slides were cooled for 20 min at room temperature and then transferred to distilled water followed by PBS. Samples were blocked with serum free protein block (DAKO) containing 5% normal donkey serum for 60 min. Samples were incubated overnight at 4 °C with purified anti-TP63 antibody (Biolegend, clone W15093A). After three washes in PBS-T, samples were incubated with 1:500 Alexa Fluor 647-conjugated donkey anti-mouse IgG (Jackson immunoresearch, catalogue# 715-605-150) and 1:10,000 Hoescht nuclear dye. Quantification of $p63^+$ cells was performed in a blinded fashion and involved counting of $p63^+$ nuclei relative to background staining with an isotype control primary antibody. As the epithelium can vary in length across sections, we normalized our quantification of total positive nuclei per 1,000 μm^2 area of epithelium as measured in ImageJ and report the final value as $p63^+$ cells per 1,000 μm^2 of epithelium.

Single-cell RNA-seq with Seq-Well. Once a single-cell suspension was obtained from freshly resected EthSin tissue, or scrapings from InfTurb, we used the Seq-Well platform for massively parallel scRNA-seq to capture transcriptomes of single cells on barcoded mRNA capture beads. Full methods on implementation of this platform are available[12]. In brief, 20,000 cells were loaded onto one array preloaded with barcoded mRNA capture beads (ChemGenes). The loaded arrays containing cells and beads were then sealed using a polycarbonate membrane with a pore size of 0.01 μm, which allows for exchange of buffers but retains biological molecules confined within each nanowell. Subsequent exchange of buffers allows for cell lysis, transcript hybridization, and bead recovery before performing reverse transcription en masse. Following reverse transcription using Maxima H Minus Reverse Transcriptase (ThermoFisher EP0753) and an Exonuclease I treatment (NewEngland Biolabs M0293L) to remove excess primers, PCR amplification was carried out using KAPA HiFi PCR Mastermix (Kapa Biosystems KK2602) with 2,000 beads per 50 μl reaction volume. Libraries were then pooled in sets of six (totaling 12,000 beads) and purified using Agencourt AMPure XP beads (Beckman Coulter, A63881) by a 0.6X SPRI followed by a 0.7X SPRI and quantified using Qubit hsDNA Assay (Thermo Fisher Q32854). Quality of WTA product was assessed using the Agilent hsD5000 Screen Tape System (Agilent Genomics) with an expected peak >1,000 bp tailing off to beyond 5000 bp, and a small or

© 2018 Springer Nature Limited. All rights reserved.

non-existent primer peak, indicating a successful preparation. Libraries were constructed using the Nextera XT DNA tagmentation method (Illumina FC-131-1096) on a total of 600 pg of pooled cDNA library from 12,000 recovered beads using index primers with format as in Gierahn et al.[12]. Tagmented and amplified sequences were purified at a 0.6X SPRI ratio yielding library sizes with an average distribution of 650–750 base pairs in length as determined using the Agilent hsD1000 Screen Tape System (Agilent Genomics). Two arrays were sequenced per sequencing run with an Illumina 75 Cycle NextSeq500/550v2 kit (Illumina FC-404-2005) at a final concentration of 2.2–2.8 pM. The read structure was paired end with read 1 starting from a custom read 1 primer[12] containing 20 bases with a 12-bp cell barcode and 8-bp unique molecular identifier (UMI) and read 2 containing 50 bases of transcript information.

Single-cell RNA-seq computational pipelines and analysis. Read alignment was performed as in Macosko et al.[39]. In brief, for each NextSeq sequencing run, raw sequencing data was converted to demultiplexed FASTQ files using bcl2fastq2 based on Nextera N700 indices corresponding to individual samples/arrays. Reads were then aligned to Hg19 genome using the Galaxy portal maintained by the Broad Institute for Drop-Seq alignment using standard settings. Individual reads were tagged according to the 12-bp barcode sequenced and the 8-bp UMI contained in Read 1 of each fragment. Following alignment, reads were binned onto 12-bp cell barcodes and collapsed by their 8-bp UMI. Digital gene expression matrices (for example, cells-by-genes tables) for each sample were obtained from quality filtered and mapped reads, with an automatically determined threshold for cell count. UMI-collapsed data was used as input into Seurat[40] (https://github.com/satijalab/seurat) for further analysis. Before incorporating a sample into our merged dataset, we individually inspected the cells-by-genes matrix of each as a Seurat object.

For analysis of all sequenced surgical ethmoid sinus resection samples, we merged UMI matrices across all genes detected in any condition and generated a matrix retaining all cells with at least 500 UMI detected (19,196 cells and 31,032 genes). This table was then used to set up the Seurat object in which any cell with at least 300 unique genes was retained and any gene expressed in at least 5 cells was retained (Supplementary Information; an R Script is included from this point to set up the Seurat object and walk reader through dimensionality reduction and basic data visualization). The object was initiated with log-normalization, from a UMI + 1 count matrix, scaling, and centering set to True. The total number of cells passing these filters captured across all patients was 18,624 cells with 22,575 genes, averaging 1,503 cells per sample with a range between 789 cells and 3,109 cells (Extended Data Fig. 1a, b, Supplementary Table 2). Before performing dimensionality reduction, data was subset to include cells with less than 12,000 UMI, and a list of 1,627 most variable genes was generated by including genes with an average normalized and scaled expression value greater than 0.13 and with a dispersion (variance/mean) greater than 0.28. We then performed principal component analysis (PCA) over the list of variable genes. For both clustering and *t*-stochastic neighbour embedding (*t*-SNE), we used the first 12 principal components, as upon visual inspection of genes contained within, each contributed to a non-redundant cell type and this reflected the inflection point of the elbow plot. We used FindClusters within Seurat (which utilizes a shared nearest neighbour (SNN) modularity optimization based clustering algorithm) with a resolution of 1.2 and *t*-SNE set to fast with the Barnes–Hut implementation to identify 21 clusters across the 12 input samples.

For analysis of all sequenced InfTurb scraping samples, the object was initiated with log-normalization, from a UMI + 1 count matrix, scaling, and centering set to true. The total number of cells passing these filters captured across all patients was 18,704 cells with 24,842 genes, averaging 2,078 cells per sample with a range between 65 cells and 5,625 cells (note: The 65-cell sample was a very mucus-laden polyp inferior turbinate sample, perhaps explaining the low cell yield, but clustered well within the three other samples each containing 253, 599, and 1,381 cells). Before performing dimensionality reduction, data was subset to include cells with less than 10,000 UMI, and a list of 1,499 most variable genes was generated by including genes with an average normalized and scaled expression value greater than 0.22 and with a dispersion (variance/mean) greater than 0.26. We then performed PCA over the list of variable genes. For both clustering and *t*-SNE, we used the first 16 principal components, as upon visual inspection of genes contained within, each contributed to a non-redundant cell type and this reflected the inflection point of the elbow plot. We used FindClusters (which utilizes an SNN modularity optimization based clustering algorithm) with a resolution of 1 and *t*-SNE set to fast with the Barnes–Hut implementation to identify 18 clusters across the 9 input samples.

For analysis of all sequenced ALI cultures, the object was initiated with log-normalization, from a UMI + 1 count matrix, scaling, and centering set to true. The total number of cells passing these filters captured across all patients was 16,173 cells with 27,396 genes, averaging 2,448 cells per sample with a range between 1,980 cells and 3,009 cells. Before performing dimensionality reduction, data was subset to include cells with less than 25,000 UMI, and a list of 1,670 most variable genes was generated by including genes with an average normalized and scaled expression value greater than 0.35 and with a dispersion (variance/mean) greater than 0.35. We then performed PCA over the list of variable genes. For both clustering and *t*-SNE, we used the first 16 principal components, as upon visual inspection of genes contained within, each contributed to a non-redundant cell state and this reflected the inflection point of the elbow plot. We used FindClusters (which utilizes an SNN modularity optimization based clustering algorithm) with a resolution of 0.6 and *t*-SNE set to fast with the Barnes–Hut implementation to identify 11 clusters across the 4 input samples.

Cell type identification and within cell type analysis. To identify genes which defined each cluster, we performed a ROC test implemented in Seurat with a threshold set to an area under the curve of 0.65. Top marker genes with high specificity were used to classify cell clusters into cell types (Fig. 1a–c; Extended Data Fig. 1e) based on existing biological knowledge. Three clusters were considered doublets (588 cells) based on co-expression of markers indicative of distinct cell types at ~1/2 the expression level detected in the parent cell cluster (for example, T cell and myeloid cell) and removed from further analyses yielding a matrix with 18,036 cells used in all subsequent steps. Closely related clusters were merged to cell types based on biological curation and analysis of hierarchical cluster trees yielding ten total cell types (Fig. 1a–c; Extended Data Fig. 1e). We identified a much smaller number of eosinophils than expected in our single-cell data. Specifically, if we do not place bulk tissue immediately into RNA-later within 10 min, we cannot reliably detect eosinophil associated transcripts. However, flow cytometrically we recover from 0.5% to 5% of total cells fitting eosinophil profiles from polyps, and focused single-cell studies on granulocytes at the expense of the full ecosystem are possible and the topic of future work (data not shown). With the gentler tissue dissociation required for scrapings, we recovered a greater frequency of eosinophils from polyps in line with flow data (0.31% to 4.6% of cells; Extended Data Fig. 6d). We also did not find a distinct cluster of ILCs, as they are around 0.01 to 0.1% of CD45 cells across the CRS spectrum, per existing literature[41], and extrapolating to the number of CD45 cells we captured, we would have detected between 0.8 and 8 ILCs. To investigate further granularity present within cell types, such as T cells, myeloid cells, fibroblasts, endothelial cells, and epithelial cells, we subset these cells from the Seurat object and re-ran dimensionality reduction and clustering (Extended Data Figs. 3, 4 and 6). The process used for clustering and subset identification was adapted for each cell type to optimize the parameters of variable genes, principal components, and resolution of clusters desired. Canonical correlation analysis[42] (CCA) was also performed to validate epithelial cell type classification across disease states (Extended Data Fig. 5; Supplementary Information).

Differential expression and fractional contribution of gene set to transcriptome. To identify differentially expressed genes within cell types across non-polyp and polyp disease states, we used the 'bimod' setting in FindMarkers implemented in Seurat based on a likelihood ratio test designed for single-cell differential expression incorporating both a discrete and continuous component[43]. To determine the expression contribution to a cell's transcriptome of a particular gene list, we summed the total log-normalized expression values for genes within a 'list of interest' and divided by the total amount of log-normalized transcripts detected in that cell, giving the proportion of a cell's transcriptome dedicated to producing those genes. For comparison of Wnt and Notch signalling, we *z*-scored the expression contribution metric and subtracted the value of Notch from Wnt yielding a metric centred on zero if both scores are equivalent, or weighted in the positive direction if enriched in Wnt. For reference gene lists used, including basal cell[44]; genes induced by IFN-α, IFN-γ, IL-4, IL-13, IL-4/IL-13[45]; Wnt and Notch please see Supplementary Table 4.

Simpson's index of diversity, and fibroblast gene correlation with basal cell frequency. To measure the 'richness' of the epithelial ecosystem[46], we employed Simpson's index of diversity (D), which we present as (1−D), and ranges between 0 and 1, with larger values indicating larger sample diversity[47]. We used Simpson's index to characterize the composition of epithelial cells across basal, differentiating/secretory, glandular, and ciliated groupings in the non-polyp and polyp ethmoid sinus tissue ecosystems, as this metric accounts for both the number of distinct cell types present (for example, species), and the evenness of the cellular composition across those cell types (for example, relative abundance of species to each other). This measure takes into account the total number of members of a cell type, the number of cell types, and the total number of cells present. We calculate (1−D) for each sample. To determine genes correlated in specific cell types (for example, fibroblasts) with the frequency of basal cells present in a cellular ecosystem, we correlated the average log-normalized single-cell count data for each gene to the rank of samples determined by increasing frequency of basal cells in each ecosystem (8.2% to 19.1% for non-polyp and 27.9% to 70.1% for polyp samples, Extended Data Fig. 7b).

Tissue and sorted basal cell RNA-seq. Population RNA-seq was performed using a derivative of the Smart-Seq2 protocol for single cells[48]. In brief, tissue was collected

© 2018 Springer Nature Limited. All rights reserved.

directly into RNAlater (Qiagen) in the surgical suite and stored at −80 °C until RNA isolation. RNA was isolated from 30 patients using phenol/chloroform extraction and normalized to 5 ng as the input amount for a 2.2X SPRI ratio cleanup using Agencourt RNAClean XP beads (Beckman Coulter, A63987). RNA-seq was performed on a bulk population of sorted basal cells using Smart-Seq2 chemistry, starting with a 2.2X SPRI ratio cleanup. After oligo-dT priming, Maxima H Minus Reverse Transcriptase (ThermoFisher EP0753) was used to synthesize cDNA with an elongation step at 52 °C before PCR amplification (15 cycles for tissue, 18 cycles for sorted basal cells) using KAPA HiFi PCR Mastermix (Kapa Biosystems KK2602). Sequencing libraries were prepared using the Nextera XT DNA tagmentation kit (Illumina FC-131-1096) with 250 pg input for each sample. Libraries were pooled post-Nextera and cleaned using Agencourt AMPure SPRI beads with successive 0.7X and 0.8X ratio SPRIs and sequenced with an Illumina 75 Cycle NextSeq500/550v2 kit (Illumina FC-404-2005) with loading density at 2.2 pM, with paired end 35 cycle read structure. Tissue samples were sequenced at an average read depth of 7.98 million reads per sample and 3 samples not meeting quality thresholds were excluded from further analyses yielding 27 total useable samples. Sorted basal cell samples were sequenced at an average read depth of 21.15 million reads per sample and all samples met quality thresholds regarding genomic and transcriptomic alignment.

Tissue and sorted basal cell RNA-seq data analysis. Tissue and sorted basal cell samples were aligned to the Hg19 genome and transcriptome using STAR[49] and RSEM[50]. Three samples were excluded for low transcriptome alignment (<25%), so we retained 27 samples for further analyses. Differential expression analysis was conducted using DESeq2 package for R[51]. Genes regarded as significantly differentially expressed were determined based on an adjusted *P* value using the Benjamini–Hochberg procedure to correct for multiple comparisons with a false discovery rate <0.05. We performed ingenuity pathway analysis (IPA, Qiagen) through an instance available through the Broad Institute on the top 1,000 differentially expressed genes (all adjusted $P < 0.05$) from our DESeq2 analysis, taking into account corresponding log-fold change for each gene. We also subset the tissue RNA-seq matrix based on genes found in Supplementary Table 3, which, from our single-cell marker discovery, were specific for basal, differentiating/secretory, glandular, or ciliated cells. We then ran PCA and *k*NN clustering implemented in R over these genes in order to identify the greatest vectors of variance across samples within the epithelial cell compartment (Fig. 3f, g).

For re-analysis of published data, we used two publicly-available RNA-seq datasets: one profiling normal human olfactory mucosa and the other assessing differences in gene expression between healthy, non-eosinophilic nasal polyps and eosinophilic nasal polyps[7,52,53]. Note that analysis is done on a per sample basis and as such no comparisons are made across the datasets or samples.

Diffusion pseudotime mapping for differentiation analysis. Using diffusion pseudotime[54] mapping, which seeks to provide the most likely reconstruction for the developmental progression of a set of cells we built a trajectory for cells within basal and differentiating/secretory epithelial clusters (non-polyp clusters: 8-basal, 1-differentiating/secretory, 4-secretory; and polyp clusters: 12-basal, 2-basal, 0-differentiating/secretory; running several iterations starting from a random seed cell in cluster 8), over the combined basal and apical marker gene list (Fig. 4d; Extended Data Fig. 9a, Supplementary Table 3). By calculating a pseudotime trajectory for cells from both non-polyps and polyps together, we were then able to ask where cells from each disease state fall along a shared inferred temporal axis (Fig. 4d, e; Extended Data Fig. 9a). Diffusion pseudotime[54] was calculated using the scanpy Python package 'dpt' function on log-normalized data for clusters 8, 1 and 4 (predominantly non-polyp, Supplementary Table 3) and 12, 2, and 0 (predominantly polyp, Supplementary Table 3) together. A random root cell was chosen from cluster 8, as this was the basal cell cluster representative of the non-polyp (for example, less aberrant) state, and we also ran iterations with random root cells chosen from the entire set of clusters and it assigned cluster 8 as the cluster most enriched at the beginning of the diffusion map, regardless. Plots were created with the seaborn, matplotlib and pandas packages. Pearson correlations were then calculated for all genes in all cells tested, or for all genes in non-polyp cells and all genes in polyp cells, relative to pseudotime (Extended Data Fig. 9b). Differential correlation testing was performed using the cocor package to identify significance for the difference between correlation coefficients using Fisher's 1925 *z*-statistic, accounting for number of cells.

Epigenetic profiling of basal cells using Omni-ATAC-seq. Accessible chromatin profiling[55] using the Omni-ATAC-seq protocol as described in Corces et al.[56] was performed on basal cells stored in 100 μl BAMBanker freezing media from 12 patients ($n = 4$ non-polyp (3 retained after data quality filtering) and $n = 8$ polyp). Cells (ranging from 1,000 to 10,000) were thawed quickly in a 37 °C rock bath and 900 μl of ice-cold PBS supplemented with Roche Complete Mini Protease inhibitor was added immediately. Cells were split into two 1.5-ml Eppendorf DNA lo-bind tubes to serve as technical replicates. Cells were centrifuged at 500*g* for 5 min at 4 °C, washed once in PBS with protease inhibitor, centrifuged at 500*g* for 5 min at 4 °C and supernatant was removed completely using two separate pipetting steps with extreme caution taken to avoid resuspension (for example, smooth and consistent aspiration). The transposition reaction consisted of 20-μ*L* total volume of the following mixture (10 μl 2× TD Buffer, 1 or 0.5 μl TDEnzyme, 0.1 μl of 2% digitonin, 0.2 μl of 10% Tween 20, 0.2 μl of 10% NP40, 6.6 μl of 1× PBS and 2.3 μl of nuclease-free water). We performed replicates with two distinct concentrations of TDE since, when dealing with minute clinical samples, flow sorting can sometimes give variable cell numbers, and the ratio of TDE to cells is critical in determining the frequency with which cuts are made in the genome. We optimized in pilot experiments that for basal cell inputs in the range of 500 to 10,000 cells, the aforementioned two ratios gave expected patterns of nucleosome banding in gels (data not shown). We performed two reactions and then later, during in silico analysis, pooled peaks together for downstream analysis. The cells were resuspended into the transposition mixture and incubated at 37 °C for 30 min in an Eppendorf Thermomixer with agitation at 300 r.p.m. Transposed DNA was purified using a Qiagen MinElute Reaction Cleanup Kit with elution in 15 μl. Libraries were constructed from 10 μl of DNA using a 50 μl total reaction volume of NEB HF 2X PCR Master Mix with custom Nextera N700 and N500 index primers to barcode samples (also used in Smart-Seq2 protocol). We performed 14 cycles of PCR amplification and SPRI purified at 1.8× ratio. Based on the molarity of each library, we adjusted the number of subsequent PCR cycles to either 3, 4 or 5 more for each sample. We then performed a 0.25× reverse SPRI to remove larger fragments followed by a 1.7× SPRI to purify libraries for sequencing. Libraries were sequenced on an Illumina NextSeq with paired-end 38-cycle read structure at a loading density of 1.95 pM.

Omni-ATAC-seq data analysis. Reads were aligned using bowtie2 using the following flags: '-S -p 1 -X 2000 --chunkmbs 1000' then bams were created using samtools view with the following flags: 'samtools view -bS -F 4'. Duplicates were removed with picard. Forward reads were shifted 4 bp and negative reads were shifted 5 bp using a custom Python script and the pysam package as is recommended for ATAC-seq data. Samples for each patient were merged using samtools merge and all patients were downsampled to 3 million reads using custom python scripts and 'samtools view' with the '-b' and '-s' flags. MACS2 'callpeak' command was used to call peaks on each sample with flags '-f BAMPE -q 0.001 --nomodel --shift -100 --extsize 200 -B --broad'. Peaks from all samples were merged into one peakfile with bedtools and counts of reads per peak for each sample was generated with bedtools multicov. DESeq2 was run with the design ~polyp, testing for significant differences between polyp and non-polyp samples on this peak matrix and differential peaks with Benjamini–Hochberg adjusted *P* value less than 0.01 with 'greater' or 'less' null hypotheses were used in downstream analysis. Homer2 was run for known motif finding on differential peaks with the set of all peaks as background[57]. To determine a false discovery rate, Homer2 was run on sets of random peaks chosen with replacement from the set of all peaks.

Epithelial cell culture. Tissues were digested as described above from either non-polyp or polyp surgical resections from the ethmoid sinus. 1,000,000 digested cells were added to a 25-cm^2 tissue culture flask (Corning) pre-coated with 0.03 mg/ml Type I bovine collagen solution (StemCell Technologies) and cultured in PneumaCult-Ex media (StemCell Technologies, 05008). Media was changed every second day until cells reached confluence. Cells were subsequently frozen in 70% basal media with 20% FBS and 10% DMSO.

Air–liquid interface cultures. For air–liquid interface (ALI) cultures[58], 100,000 cultured epithelial cells per well were added to 0.4-μm pore 24-well polyester membrane inserts (Corning) pre-coated with 0.03 mg/ml Type I bovine collagen solution (StemCell Technologies) with Pneumacult-Ex media (StemCell Technologies, 05008) on both sides of the membrane. After 24 h, apical media was changed to remove dead cells. After 72 h, apical media was removed completely and basal media was changed to Pneumacult-ALI (StemCell Technologies, 05001) supplemented with 5 ml 100× penicillin-streptomycin (Fisher), 1 ml 500× gentamicin/amphotericin B (ThermoFisher), 1 ml 0.2% heparin sodium salt in PBS (StemCell Technologies) and 2.5 ml 200× hydrocortisone stock solution (StemCell Technologies) and 0, 0.1, 1 or 10 ng/ml IL-13 (Biolegend). Basal media was changed every 2–3 days for 21 days, after which membranes were removed and cells dissociated with Stempro Accutase Cell Dissociation Reagent (Gibco) for Seq-Well or flow cytometry. After following scRNA-seq data analysis pipelines described above, cell states recovered in ALI cultures (Fig. 5a; Extended Data Fig. 9g) were related to in vivo cell types[59].

Basal cell stimulation. Basal cells from non-polyp or polyp surgical resections from ethmoid sinus were placed into epithelial cell culture (for example, 'lateral expansion' in the absence of differentiation, see above, inspired by experiments in microglia[60]), passaged, and 10,000 cells seeded at passage 5 (for example, 5 weeks ex vivo) and cultured at confluence in 96-well flat-bottom collagen-coated tissue culture plates (Corning, 3799) for 48 h in Pneumacult-Ex serum-free media (StemCell Technologies, 05008). Cytokines were added for 12 h overnight at increasing doses (0, 0.1, 1, 10 ng/ml) of IL-4 (Biolegend 766205), IL-13 (Biolegend 571104), and

© 2018 Springer Nature Limited. All rights reserved.

(0.1, ng/ml) IL-4 + IL-13 in combination ($n = 32$ samples non-polyp and $n = 32$ samples polyp basal cells over all conditions, each condition run as a biological duplicate, and a technical duplicate therein), before lysis using RLT + 1% BME (Qiagen and Sigma, respectively). Bulk RNA-seq was performed as described for sorted basal cells starting from lysates. Basal cell stimulation samples were sequenced at an average read depth of 3 million reads per sample and all samples met quality thresholds regarding genomic and transcriptomic alignment.

Statistical analyses. No statistical methods were used to predetermine sample size. Number of samples included in analyses are listed throughout figure legends and all represent distinct biological samples. The same surgeon performed surgeries on all individuals and was blinded to study design. The same allergist/immunologist performed nasal scrapings on all samples and was blinded to study design. Quantification of histological sections was performed in a blinded fashion. No samples or cells meeting quality thresholds were excluded from analyses. Where single-cell data was analysed on a gene level, the statistics were performed over the number of cells. Statistical analyses were performed using GraphPad Prism v7.0a, Seurat 1.4.0.1 implemented in RStudio, DESeq2 1.10.1 package implemented in RStudio, and Ingenuity Pathway Analysis run through the Broad Institute, and macs2, DESeq2 and Homer2 for Omni-ATAC-seq. All violin plots were generated using standard Seurat code without modification to smoothing or density. Violin density only generated when >25% of cells in indicated sample have non-zero measurement for gene, widest aspect represents centre of positive measures, minima and maxima are represented within the scale with minima at 0 and maxima encompassing all points for the count-based expression level (log(scaled UMI + 1)) of each gene. Exact values for all genes displayed and tested available in Supplementary Table 3 organized by panel. All violin plots contain at minimum 100 individual cells in any one cluster (Supplementary Table 3 for precise numbers of cells per cluster and type, most are included in figure legends where space allows), and have points suppressed for ease of legibility. Some violin plots with less than 100 cells have individual data points displayed and corresponding statistical metrics are available in accompanying figure legend and Supplementary Table 3. As some scores followed non-normal distributions as tested for using a Lilliefors normality test, we used a Mann–Whitney *U*-test where indicated for determining statistical significance. For scores in single-cell data, we report effect sizes in addition to statistical significance as an additional metric for the magnitude of the effect observed. The calculation was performed as Cohen's *d* where: effect size $d = (\text{Mean}_1\text{-Mean}_2)/(\text{s.d. pooled})$. Unpaired two-tailed *t*-tests for direct comparisons and *t*-test with Holm–Sidak correction, Bonferroni correction, or Benjamini–Hochberg for multiple comparisons, depending on software package used, where appropriate. Mann–Whitney *U*-test for quantification of histological data due to non-normally distributed data. Pearson correlation thresholds were determined as significant through determination of asymptotic *P* values through use of rcorr function in Hmisc, but exact corrected *P* values by Holm–Sidak method for multiple comparisons are calculated for those highlighted in text using RcmdrMisc package. Comparison of Pearson correlation coefficients in pseudotime analyses was done using Fisher's 1925 *z*-statistic accounting for the number of cells.

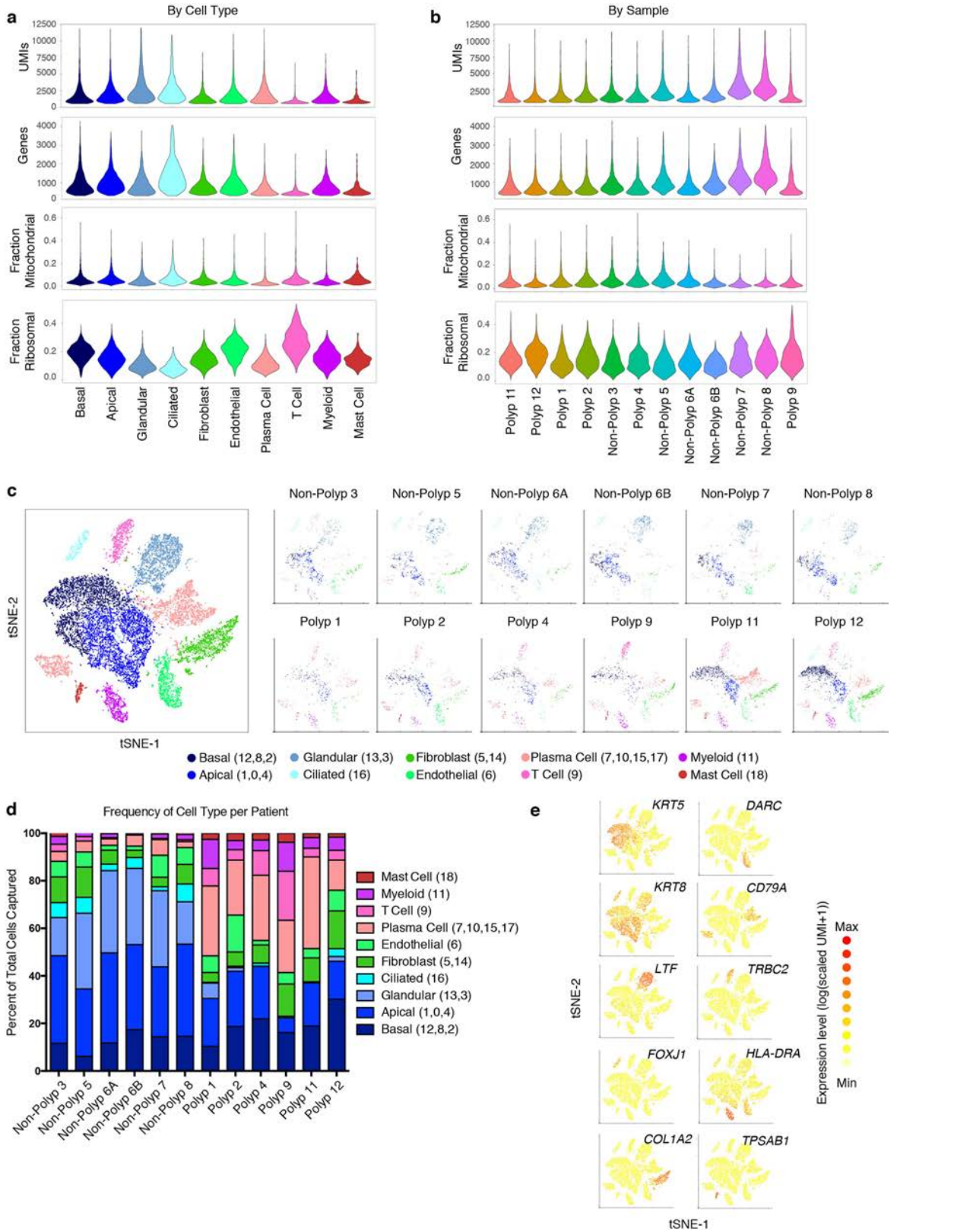
Reporting summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this paper.

Data availability. The cells-by-genes matrix generated from EthSin surgical resections and analysed during the current study is available along with the manuscript as Supplementary Table 2 alongside R code for standard implementation of Seurat. A cells-by-genes matrix from InfTurb and polyp scraping data are also available as Supplementary Table 6. Dupilumab treatment cells-by-genes matrices are shown in 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 have been deposited in and are available from the dbGaP database under dbGaP accession 30434 (https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs030434.v1.p1). 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, f, 3g, 5a, e, Extended Data Fig. 3a–c, 4c, 5e, 6b, d, 10a, selected comparisons of differential expression in Figs. 2d, 4a, 5c, f, Extended Data Fig. 2c, 10h, and pseudotime correlation in Extended Data Fig. 9b are available as tabs in Supplementary Table 3. Differential peak calling from epigenetic profiling is available in Supplementary Table 5. Additional R code for analyses is available at http://shaleklab.com/resources/. Aligned and quality-filtered data and complete statistical outputs for the figures are included as Supplementary Tables, with further information at http://shaleklab.com/resources/

.

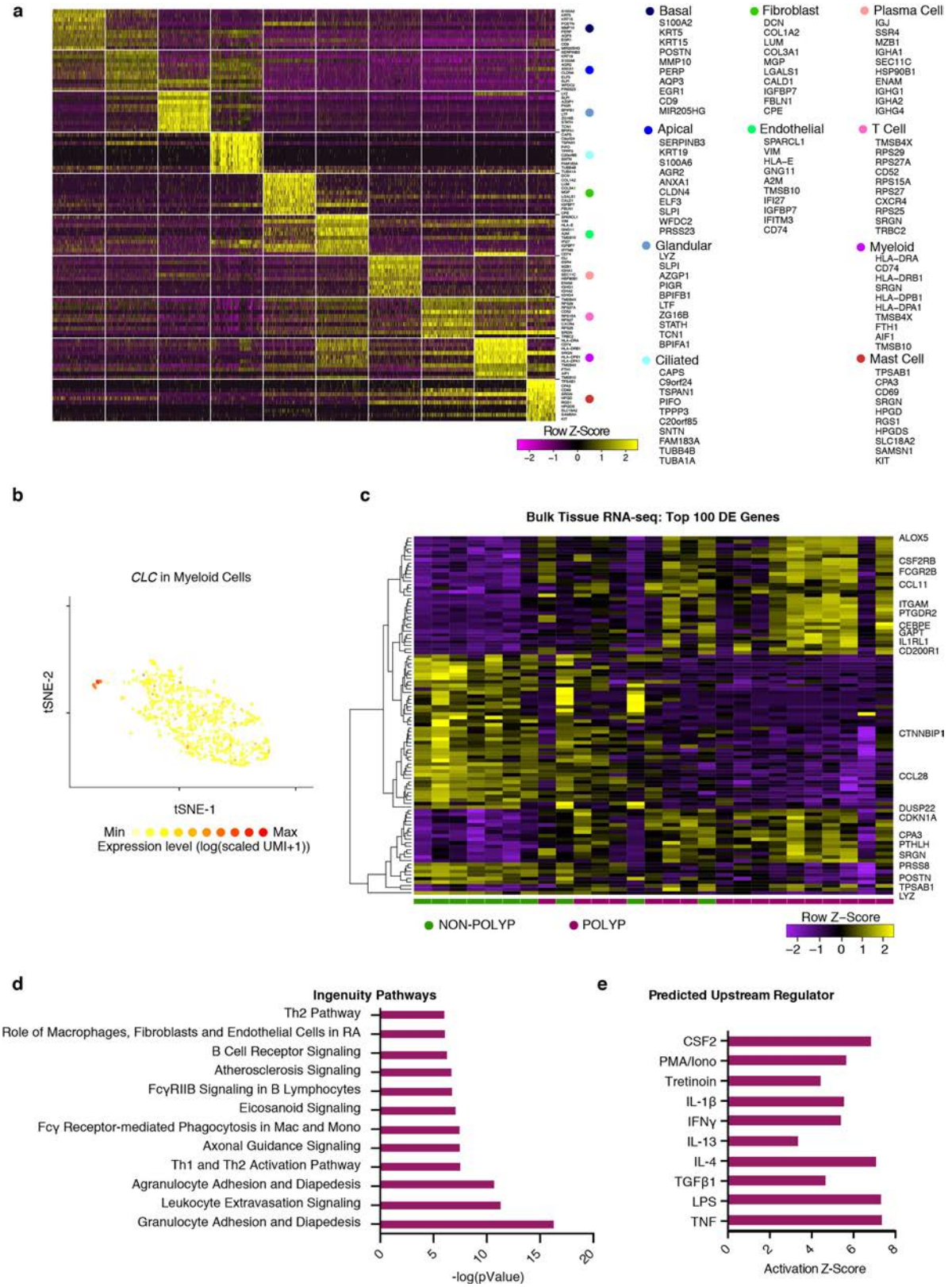
31. Meltzer, E. O. et al. Rhinosinusitis: establishing definitions for clinical research and patient care. *J. Allergy Clin. Immunol.* **114**, 155–212 (2004).
32. Dhariwal, J. et al. Mucosal type 2 innate lymphoid cells are a key component of the allergic response to aeroallergens. *Am. J. Respir. Crit. Care Med.* **195**, 1586–1596 (2017).
33. Proud, D., Sanders, S. P. & Wiehler, S. Human rhinovirus infection induces airway epithelial cell production of human β-defensin 2 both in vitro and in vivo. *J. Immunol.* **172**, 4637–4645 (2004).
34. Pipkorn, U. & Karlsson, G. Methods for obtaining specimens from the nasal mucosa for morphological and biochemical analysis. *Eur. Respir. J.* **1**, 856–862 (1988).
35. Wenzel, S. et al. Dupilumab in persistent asthma with elevated eosinophil levels. *N. Engl. J. Med.* **368**, 2455–2466 (2013).
36. Beck, L. A. et al. Dupilumab treatment in adults with moderate-to-severe atopic dermatitis. *N. Engl. J. Med.* **371**, 130–139 (2014).
37. Bachert, C. et al. Effect of subcutaneous dupilumab on nasal polyp burden in patients with chronic sinusitis and nasal polyposis: a randomized clinical trial. *J. Am. Med. Assoc.* **315**, 469–479 (2016).
38. Dwyer, D. F., Barrett, N. A., Austen, K. F. & Immunological Genome Project Consortium. Expression profiling of constitutive mast cells reveals a unique identity within the immune system. *Nat. Immunol.* **17**, 878–887 (2016).
39. Macosko, E. Z. et al. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **161**, 1202–1214 (2015).
40. Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* **33**, 495–502 (2015).
41. Poposki, J. A. et al. Group 2 innate lymphoid cells are elevated and activated in chronic rhinosinusitis with nasal polyps. *Immun. Inflamm. Dis.* **5**, 233–243 (2017).
42. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).
43. McDavid, A. et al. Data exploration, quality control and testing in single-cell qPCR-based gene expression experiments. *Bioinformatics* **29**, 461–467 (2013).
44. Hackett, N. R. et al. The human airway epithelial basal cell transcriptome. *PLoS One* **6**, e18378 (2011).
45. Giovannini-Chami, L. et al. Distinct epithelial gene expression phenotypes in childhood respiratory allergy. *Eur. Respir. J.* **39**, 1197–1205 (2012).
46. Naeem, S., Thompson, L. J., Lawler, S. P., Lawton, J. H. & Woodfin, R. M. Declining biodiversity can alter the performance of ecosystems. *Nature* **368**, 734–737 (1994).
47. Simpson, E. H. Measurement of diversity. *Nature* **163**, 688 (1949).
48. Trombetta, J. J. et al. Preparation of single-cell RNA-seq libraries for next generation sequencing. *Curr. Protoc. Mol. Biol.* **107**, 4.22.1–4.22.17 (2014).
49. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
50. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).
51. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
52. Olender, T. et al. The human olfactory transcriptome. *BMC Genomics* **17**, 619 (2016).
53. Wang, W. et al. Transcriptome analysis reveals distinct gene expression profiles in eosinophilic and noneosinophilic chronic rhinosinusitis with nasal polyps. *Sci. Rep.* **6**, 26604 (2016).
54. Haghverdi, L., Buttner, M., Wolf, F. A., Buettner, F. & Theis, F. J. Diffusion pseudotime robustly reconstructs lineage branching. *Nat. Methods* **13**, 845–848 (2016).
55. Buenrostro, J. D. et al. Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* **523**, 486–490 (2015).
56. Corces, M. R. et al. An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat. Methods* **14**, 959–962 (2017).
57. Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime *cis*-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).
58. Fulcher, M. L., Gabriel, S., Burns, K. A., Yankaskas, J. R. & Randell, S. H. Well-differentiated human airway epithelial cell cultures. *Methods Mol. Med.* **107**, 183–206 (2005).
59. Mead, B. E. et al. Harnessing single-cell genomics to improve the physiological fidelity of organoid-derived cell types. *BMC Biol.* **16**, 62 (2018).
60. Gosselin, D. et al. An environment-dependent transcriptional network specifies human microglia identity. *Science* **356**, eaal3222 (2017).

© 2018 Springer Nature Limited. All rights reserved.



Extended Data Fig. 1 | Consistency of cell capture and identification in surgical EthSin scRNA-seq patient cohort. a, Number of unique molecular identifiers (nUMI) and genes identified, and fraction of reads mapping to mitochondrial or ribosomal genes across recovered cell types: 3,222 basal cells, 4,362 apical cells, 2,192 glandular cells, 498 ciliated cells, 835 T cells, 2,976 plasma cells, 1,724 fibroblasts, 1,143 endothelial cells, 811 myeloid cells and 273 mast cells. **b**, nUMI and genes identified, and fraction of reads mapping to mitochondrial or ribosomal genes across patient samples: 789 polyp 1 cells, 1,309 polyp 2 cells, 1,153 polyp 3 cells, 913 polyp 4 cells, 1,219 polyp 5 cells, 1,141 polyp 6A cells, 1,334 polyp 6B cells, 1,314 polyp 7 cells, 1,286 polyp 8 cells, 1,481 polyp 9 cells, 2,988 polyp 11 cells, 3,109 polyp 12 cells. **c**, *t*-SNE plot as in Fig. 1b coloured by cell types across all patients and then separated by sample: 18,036 single cells ($n = 12$ samples). **d**, The percentage of each cell type recovered within each sample. **e**, Select marker gene overlays displaying binned count-based UMI-collapsed expression level (log(scaled UMI + 1)) on a *t*-SNE plot from Fig. 1b for key cell types identified (see Supplementary Table 3 for full gene lists); AUC 0.998 to 0.7 for all markers displayed.

© 2018 Springer Nature Limited. All rights reserved.

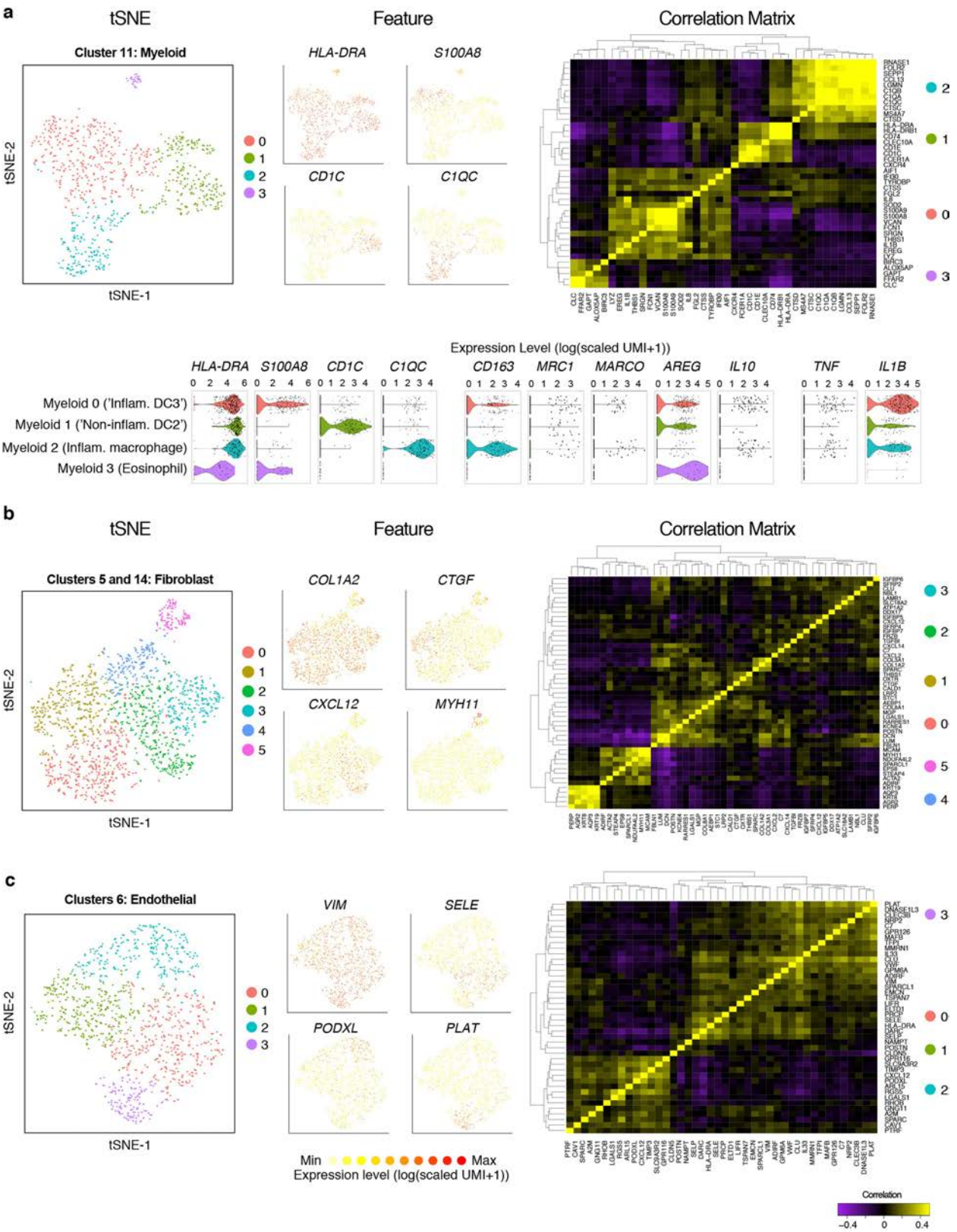


Extended Data Fig. 2 | See next page for caption.

© 2018 Springer Nature Limited. All rights reserved.

Extended Data Fig. 2 | Top marker genes for cell types by scRNA-seq and bulk tissue RNA-seq from EthSin recovers expected T2I and eosinophilic modules. a, Row-normalized heat map of the top 10 marker genes identified by ROC test (AUC > 0.73 for all) over all cell types (Fig. 1b, c) with select genes displayed on *y* axis and cells on *x* axis (see Supplementary Table 3 for full gene lists); maximum 500 cells per type. **b**, An overlay of *CLC* (a pathognomonic gene for eosinophils) displaying binned count-based expression level (log(scaled UMI + 1)) amongst myeloid cells. 811 myeloid cells from $n = 12$ samples. **c**, A row-normalized and row-clustered heat map over the top 100 positively and negatively differentially-expressed genes (50 in each direction) in bulk tissue RNA-seq of 27 samples from non-polyp ($n = 10$) and polyp ($n = 17$) tissue with select genes displayed. DESeq2 Wald test, all $P < 9.03 \times 10^{-5}$ for genes displayed, corrected for multiple comparisons by Benjamini procedure, samples ordered as in Fig. 3g (see Supplementary Table 4 for full gene list and associated statistics). **d**, The top differentially regulated pathways identified by ingenuity pathway analysis (see Methods) over the top 1,000 differentially expressed genes, as determined by $P < 0.05$ corrected for multiple comparisons by Benjamini procedure, across polyp and non-polyp tissue. **e**, Predicted upstream regulators based on differentially expressed gene modules in polyp tissue relative to non-polyp determined using ingenuity pathway analysis (see Methods).

© 2018 Springer Nature Limited. All rights reserved.

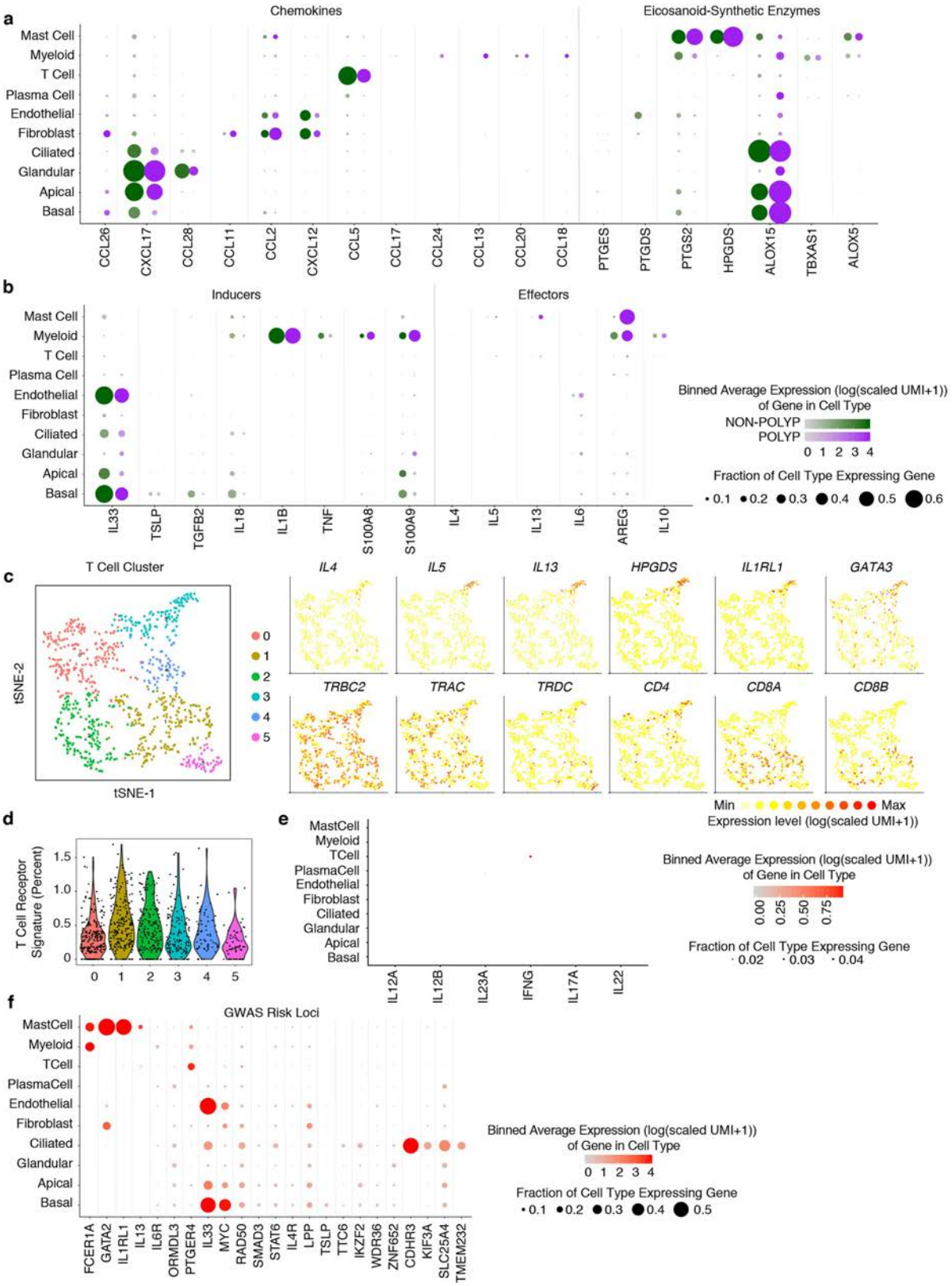


Extended Data Fig. 3 | See next page for caption.

© 2018 Springer Nature Limited. All rights reserved.

Extended Data Fig. 3 | Sub-clustering of myeloid, fibroblast and endothelial cell types from the EthSin T2I inflammatory ecosystem. **a**, *t*-SNE plot of 811 myeloid cells ($n = 6$ non-polyp, $n = 6$ polyp samples), coloured by clusters identified through shared nearest neighbour (SNN) analysis (Supplementary Table 3; Methods), from CRS-EthSin; select marker gene overlays displaying count-based (UMI-collapsed) expression level (log(scaled UMI + 1)) on a *t*-SNE plot (see Supplementary Table 3 for full gene lists; genes identified by ROC test with AUC 0.689 for *S100A8*, 0.763 for *CD1C*, 0.927 for *C1QC*); a clustered correlation matrix of marker genes identified in single-cell data from myeloid cells; and violin plots for the expression value (log(scaled UMI + 1)) of selected markers of myeloid activation state. **b**, *t*-SNE plot of 1,724 fibroblasts ($n = 6$ non-polyp, $n = 6$ polyp samples), coloured by clusters identified through shared nearest neighbour (SNN) analysis (Supplementary Table 3; Methods), from CRS-EthSin; select marker gene overlays displaying count-based (UMI-collapsed) expression level (log(scaled UMI + 1)) on a *t*-SNE plot (see Supplementary Table 3 for full gene lists; genes identified by ROC test with AUC 0.691 for *CTGF*, 0.683 for *CXCL12*, 0.726 for *MYH11*); and a clustered correlation matrix of marker genes identified in single-cell data from fibroblasts. Note, clusters 4 and 5 are likely to represent doublets with epithelial cells and endothelial cells, respectively. Although we exclude these clusters from further formal analyses, we note that there may be interesting biology within pairs of cells that are found to interact more frequently than by chance. **c**, *t*-SNE plot of 1,143 endothelial cells ($n = 6$ non-polyp, $n = 6$ polyp samples), coloured by clusters identified through shared nearest neighbour (SNN) analysis (Supplementary Table 3; Methods), from CRS-EthSin; select marker gene overlays displaying count-based (UMI-collapsed) expression level (log(scaled UMI + 1)) on a *t*-SNE plot (see Supplementary Table 3 for full gene lists; genes identified via ROC test with AUC 0.742 for *SELE*, 0.706 for *PODXL*, 0.822 for *PLAT*); and a clustered correlation matrix of marker genes identified in single-cell data from endothelial cells.

© 2018 Springer Nature Limited. All rights reserved.

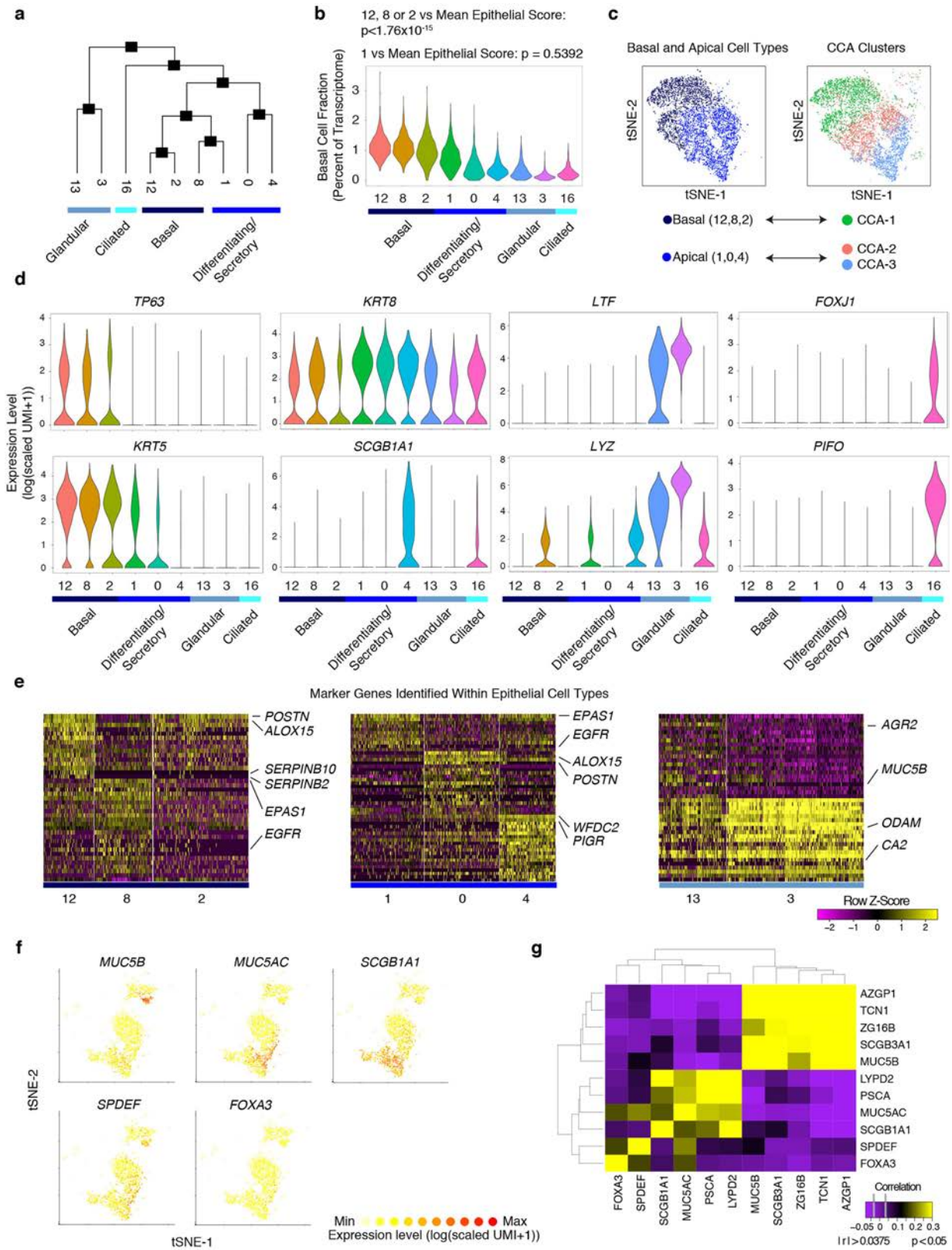


Extended Data Fig. 4 | See next page for caption.

© 2018 Springer Nature Limited. All rights reserved.

Extended Data Fig. 4 | Mapping T2I mediators within EthSin non-polyp or polyp ecosystems and the identities of T cells. a, Dot plots of chemokines and lipid mediators with known roles in T2I mapped onto cell types divided by non-polyp or polyp disease state. Dot size represents fraction of cells within that type expressing the gene, and colour intensity represents binned (log(scaled UMI + 1)) gene expression amongst expressing cells (related to Fig. 1d). **b**, Dot plot of inducers and effectors of T2I mapped onto cell types divided by non-polyp or polyp disease state. Dot size represents fraction of cells within that type expressing the gene, and colour intensity represents binned (log(scaled UMI + 1)) gene expression amongst expressing cells (related to Fig. 1d). **c**, *t*-SNE plot of re-clustered T cells with select gene overlays displaying binned count-based expression level (log(scaled UMI + 1)) for Th2A-specific genes (top row) and canonical T cell markers (bottom row); 835 T cells from $n = 6$ non-polyp and $n = 6$ polyp samples. **d**, Violin plot of five identified T cell clusters scored for expression of T cell receptor complex genes (for example, *TRAC* and *CD3E*, see Methods, Supplementary Table 4). Dots represent individual cells; 835 total T cells. **e**, Dot plot of inducers and effectors of Type 1 immunity across all cell types (note that *IL17F* was not detected). **f**, Dot plot of select GWAS risk alleles[41] for allergic disease, mapped onto cell types. Dot size represents fraction of cells within that type expressing the gene, and colour intensity represents binned (log(scaled UMI + 1)) gene expression amongst expressing cells (related to Fig. 1d).

© 2018 Springer Nature Limited. All rights reserved.

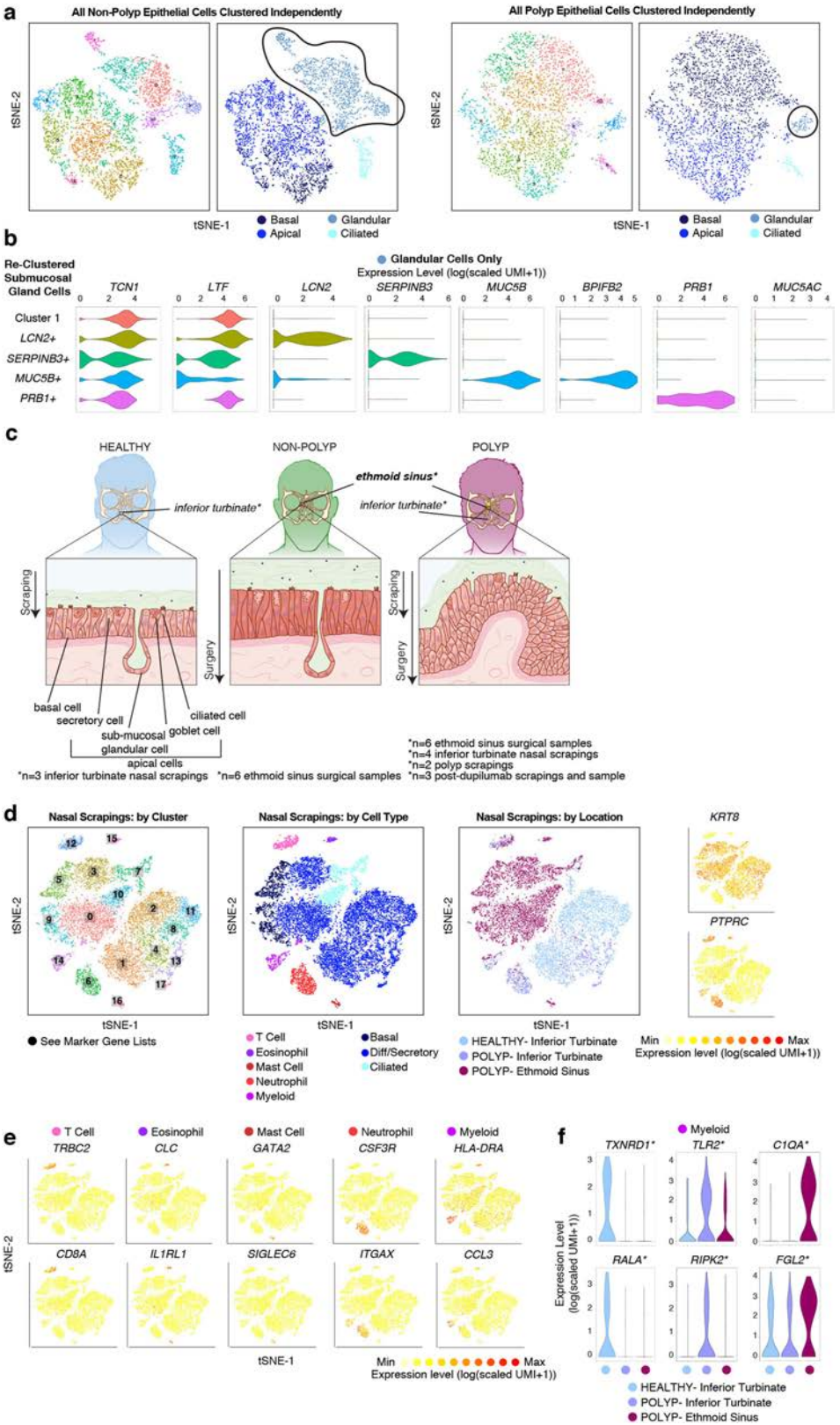


Extended Data Fig. 5 | See next page for caption.

© 2018 Springer Nature Limited. All rights reserved.

Extended Data Fig. 5 | Relationship of EthSin epithelial cell clusters and secretory/glandular distinctions. a, A phylogenetic tree based on the average cell from each cluster of epithelial cell clusters in gene-space. **b**, Violin plot of expression contribution to a cell's transcriptome of basal cell genes (see Methods and Supplementary Table 4) across all epithelial cells. Cluster 12, 794 cells; cluster 8, 924 cells; cluster 2, 1,504 cells; cluster 1, 1,561 cells; cluster 0, 1,600 cells; cluster 4, 1,201 cells; cluster 13, 725 cells; cluster 3, 1,467 cells; cluster 16, 498 cells; Mann–Whitney *U*-test, with Bonferroni correction, $P < 1.76 \times 10^{-15}$, cluster 12, cluster 8 or cluster 2 versus the mean score of basal/apical epithelial cells; $P = 0.5392$, cluster 1 versus the mean score. **c**, Canonical correlation analysis (CCA) displaying our cell type annotations for basal and apical cells derived through clustering and biological curation alongside CCA clusters in *t*-SNE space; 7,584 basal and apical cells. **d**, Violin plots for the count-based expression level (log(scaled UMI + 1)) of selected marker genes for each identified epithelial cell subset; cell numbers as in **b**. **e**, Row-normalized heat map of the top marker genes identified by ROC test (AUC > 0.6) within each cell type for each cell cluster with genes displayed on *y* axis and cluster annotations on *x* axis (see Supplementary Table 3 for full gene lists). **f**, Select overlays on clusters 0 and 4 (differentiating/secretory) and 13 (glandular) displaying binned count-based expression level (log(scaled UMI + 1)) in *t*-SNE space for canonical goblet (*MUC5B*, *MUC5AC*, *SPDEF*, *FOXA3*) and secretory (*SCGB1A1*) genes; 3,526 cells. **g**, A clustered correlation matrix of glandular, goblet, and secretory cell genes. Pearson's abs(r) > 0.038 is significant ($P < 0.05$) based on asymptotic *P* values.

© 2018 Springer Nature Limited. All rights reserved.

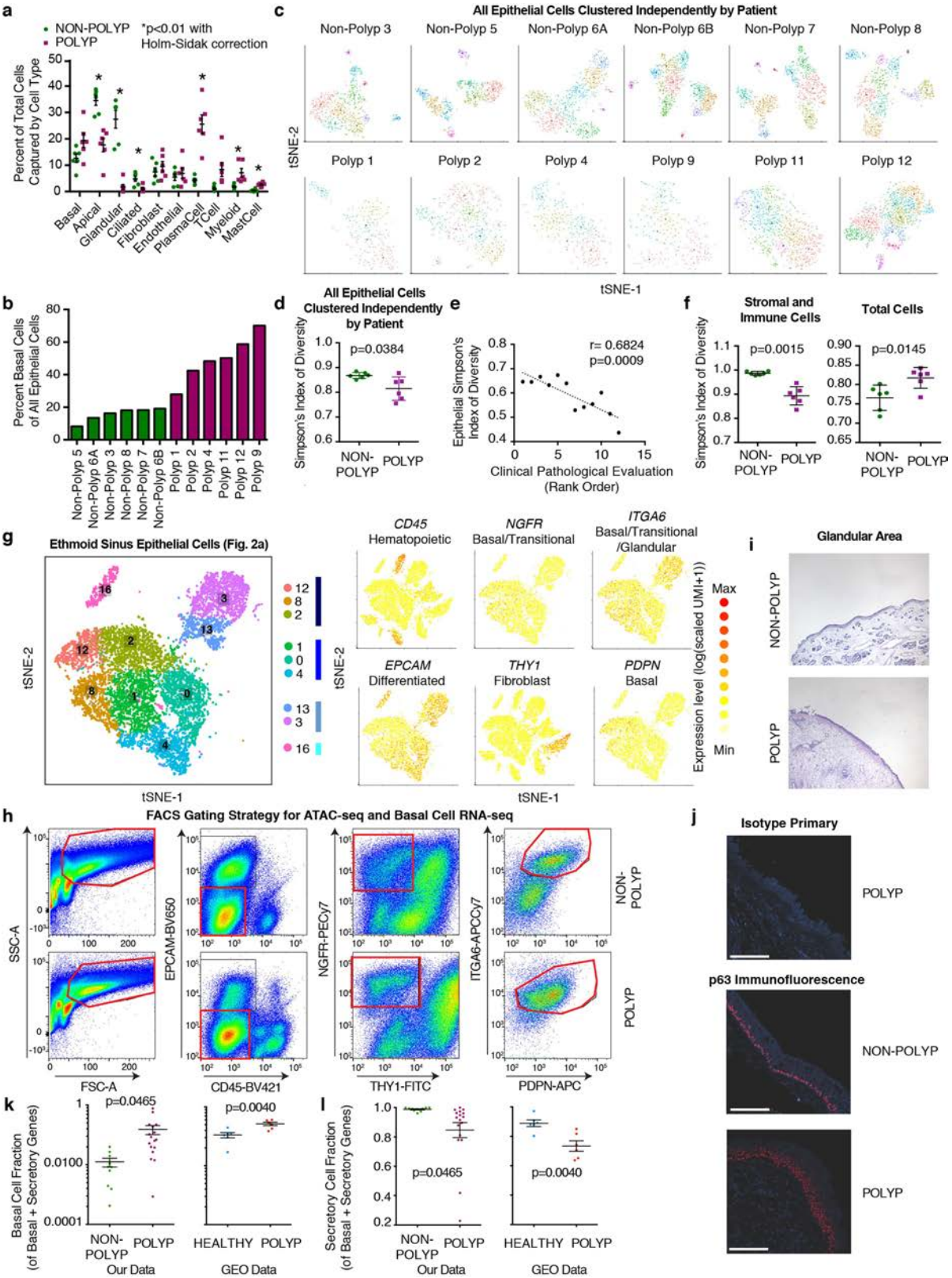


Extended Data Fig. 6 | See next page for caption.

© 2018 Springer Nature Limited. All rights reserved.

Extended Data Fig. 6 | Glandular cell subsets, their relationship to apical secretory cells, and immune cells recovered through nasal scrapings. a, *t*-SNE plots of 5,928 single epithelial cells ($n = 6$ non-polyp samples) and 4,346 single epithelial cells ($n = 6$ polyp samples) coloured by clusters identified through (left) shared nearest neighbour (SNN) analysis and (right) original biological curation of cell types (Supplementary Table 3; Methods) as illustrated in Fig. 2a. Note, cluster colours in left panels of each disease are not comparable but curated clusters in panels are, and glandular cells are highlighted for subsetting in next panel. **b**, Violin plots for the count-based expression level (log(scaled UMI + 1)) of selected marker genes identified through marker discovery (ROC test) for each subset of glandular cells; 2,114 total cells (cluster 1, 791 cells; *LCN2* cluster, 709 cells; *SERPINB3* cluster, 283 cells; *MUC5B* cluster, 209 cells; *PRB1* cluster, 183 cells) with representation of every non-polyp patient in each cluster of cells (for example, no cluster is unique to one patient) and AUC metric 0.800 for *LCN2*, 0.736 for *SERPINB3*, 0.985 for *MUC5B*, 0.973 for *BPIFB2*, and 0.908 for *PRB1*. **c**, Samples were acquired through the two distinct methods of nasal scraping and ethmoid sinus surgical intervention. This allowed for sampling of healthy tissue from InfTurb (scraping, top left), CRS-EthSin-non-polyp tissue (surgery, top middle), CRS-EthSin-polyp tissue (surgery, top right), InfTurb of polyp-bearing individuals (scraping, top right) and CRS-EthSin-polyp tissue accessible for scraping (scraping, top right). Bottom panels, anatomy of the nasal turbinates (healthy and CRS polyp) and ethmoid sinus (CRS non-polyp and CRS polyp) where samples were acquired, highlighting the depth of cells recovered from each site related to Fig. 2. Healthy tissue is annotated with basal and apical cell types, including sub-mucosal glands. **d**, Left, *t*-SNE plot of 18,704 single cells from nasal scrapings ($n = 9$ samples) coloured by clusters identified through shared nearest neighbour (SNN) analysis (Supplementary Table 3; Methods). Middle, *t*-SNE plot coloured by cell types identified through marker discovery (ROC test) and biological curation of identified clusters (Supplementary Table 3; Methods). Right, *t*-SNE plot coloured by disease and tissue of origin from healthy InfTurb (7,603 cells; $n = 3$ samples), polyp-bearing patient InfTurb (2,298 cells; $n = 4$ samples) and polyp scraping directly from EthSin-polyp (8,803 cells; $n = 2$ samples), with adjacent select marker gene overlays displaying count-based UMI-collapsed expression level (log(scaled UMI + 1)) for apical epithelial (*KRT8*) and haematopoietic (*PTPRC*) genes. **e**, Select marker gene overlays displaying count-based UMI-collapsed expression level (log(scaled UMI + 1)) on a *t*-SNE plot from **a** for key cell types identified (see Supplementary Table 3 for full gene lists); area under the curve (AUC) 0.946 to 0.705 for all markers displayed. **f**, Violin plots for the count-based expression level (log(scaled UMI + 1)) for key differentially expressed genes using ROC test within myeloid cells across disease states and tissues identified (Methods); 137 cells, $n = 3$ healthy inferior turbinate; 157 cells, $n = 4$ polyp inferior turbinate; 210 cells, $n = 2$ polyp ethmoid sinus samples; AUC 0.67 for *TXNRD1*, 0.615 for *RALA*, 0.647 for *TLR2*, 0.619 for *RIPK2*, 0.747 for *C1QA*, 0.674 for *FGL2*.

© 2018 Springer Nature Limited. All rights reserved.

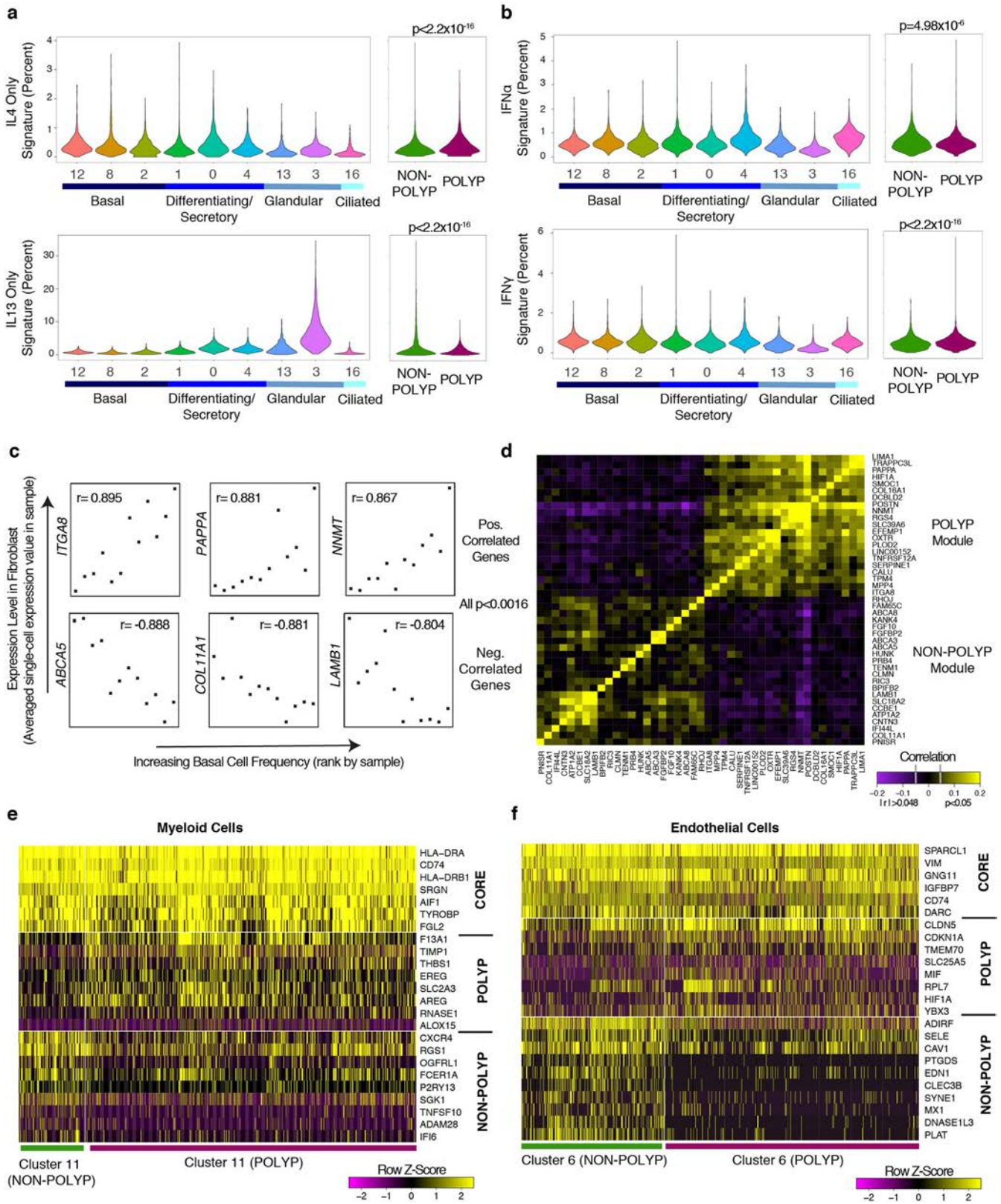


Extended Data Fig. 7 | See next page for caption.

© 2018 Springer Nature Limited. All rights reserved.

Extended Data Fig. 7 | Changes in cellular composition between EthSin-non-polyp and EthSin-polyp tissue by scRNA-seq and flow cytometric gating and histological strategy for quantification and isolation of basal cells. a, The frequency of each cell type recovered amongst all cells within each patient sample ($n = 6$ non-polyp, $n = 6$ polyp) grouped by disease state. Two-sided *t*-test; apical, $P = 0.0003$; glandular, $P < 0.0001$; ciliated, $P = 0.0047$; plasma cell, $P = 0.00014$; myeloid, $P = 0.0098$; mast cell, $P = 0.00018$; all non-polyp versus polyp with Holm–Sidak correction for multiple comparisons. Data are mean ± s.e.m. **b**, The frequency of basal cells amongst epithelial cells captured in scRNA-seq data displayed for each sample and coloured by non-polyp or polyp designation. **c**, *t*-SNE plots with each patient's cells clustered independently over a common list of most variable genes identified from all epithelial cells and with clustering parameters set constant to 12 principal components and resolution set to 1.4; minimum 789 cells in each plot; see Extended Data Fig. 1b and Supplementary Table 3 for specific cell numbers. **d**, Simpson's index of diversity, an indication of the total richness present within an ecosystem, over epithelial cell clusters identified in **c**, calculated for each patient; $n = 6$ non-polyp and $n = 6$ polyp samples. Two-tailed *t*-test, $P = 0.0384$. Data are mean ± s.e.m. **e**, Correlation of Simpson's index of diversity calculated over epithelial cells against the ranked order of samples based on clinical pathological evaluation; $n = 6$ non-polyp and $n = 6$ polyp samples; $r = 0.6824$, $P = 0.009$. **f**, Simpson's index of diversity over stromal and immune cell types and total cells, calculated for each sample ($n = 6$ non-polyp and $n = 6$ polyp). Points represent individual samples. Two-tailed *t*-test, $P = 0.0015$ (stromal and immune), $P = 0.0145$ (total cells), non-polp versus polyp. Data are mean ± s.e.m. **g**, Reproduced from Fig. 2a: *t*-SNE plot of 10,274 epithelial cells, coloured by clusters identified through SNN, with adjacent colour bars representing related cell clusters, and overlays displaying binned count-based expression level (log(scaled UMI + 1)) of selected genes used to negatively (*CD45*, *EPCAM*, *THY1*) and positively (*NGFR*, *ITGA6*, *PDPN*) identify basal cells. **h**, Full flow cytometric gating strategy for quantification and isolation of basal cells from non-polyp and polyp tissue (related to Fig. 3c). **i**, Representative histology (5× magnification) of the glandular area detected in haematoxylin and eosin stained tissue sections from non-polyp or polyp patients; quantification in Fig. 3e. **j**, Representative immunofluorescence of $p63^+$ cells (basal cell marker) relative to isotype control; quantification in Fig. 3d. Scale bar, 100 μm. **k**, Basal cell fraction of transcripts from bulk tissue RNA-seq data of our own dataset (related to Fig. 3g, h) and two GEO datasets containing healthy and healthy/polyp nasal mucosa biopsies. Our data: $n = 10$ non-polyp samples, $n = 17$ polyp samples. Reference data: $n = 6$ healthy, $n = 6$ polyp samples. Two-tailed *t*-test, $P = 0.0465$ (our data) and $P = 0.0040$ (GEO data). Data are mean ± s.e.m. **l**, Secretory cell fraction of transcripts from bulk tissue RNA-seq data of our own dataset (related to Fig. 3g, h) and two GEO datasets containing healthy and healthy/polyp nasal mucosa biopsies. Our data, $n = 10$ non-polyp samples, $n = 17$ polyp samples; reference data, $n = 6$ healthy, $n = 6$ polyp samples. Two-tailed *t*-test, $P = 0.0465$ (our data) and $P = 0.0040$ (GEO data). Data are mean ± s.e.m.

© 2018 Springer Nature Limited. All rights reserved.

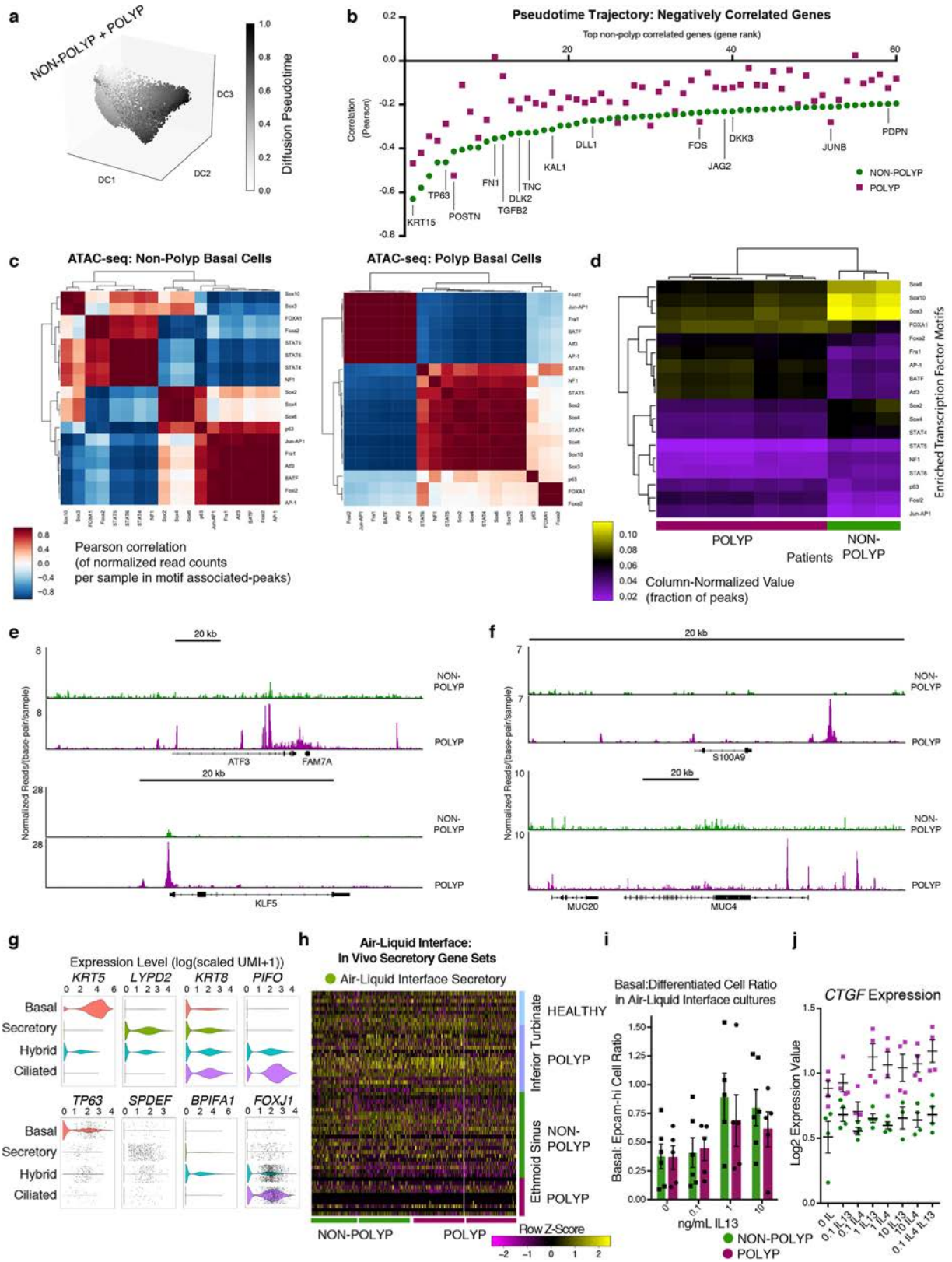


Extended Data Fig. 8 | See next page for caption.

© 2018 Springer Nature Limited. All rights reserved.

Extended Data Fig. 8 | Epithelial cytokine signatures from CRS-EthSin tissue demonstrate T2I pattern, discovery of gene modules in the fibroblast niche which correlate with basal cell hyperplasia, and differential expression within myeloid and endothelial cells by polyp status. a, Violin plots of IL-4- or IL-13-uniquely induced gene signatures in respiratory epithelial cell clusters or grouped by disease state presented as expression contribution to a cell's transcriptome (see Methods, Fig. 4b for shared genes, and Supplementary Table 4). Cluster 12, 794 cells; cluster 8, 924 cells; cluster 2, 1,504 cells; cluster 1, 1,561 cells; cluster 0, 1,600 cells; cluster 4, 1,201 cells; cluster 13, 725 cells; cluster 3, 1,467 cells; cluster 16, 498 cells. Mann–Whitney *U*-test, $P < 2.2 \times 10^{-16}$, 0.305 IL-4 effect size (polyp versus non-polyp) and −0.448 IL-13 effect size (polyp versus non-polyp). **b**, Violin plots of IFN-α- or IFN-γ-induced gene signatures in respiratory epithelial cell clusters or grouped by disease state presented as expression contribution to a cell's transcriptome (see Methods, and Supplementary Table 4); cell numbers as in **a**. Mann–Whitney *U*-test, $P = 4.98 \times 10^{-6}$, −0.156 IFN-α effect size (polyp versus non-polyp). Mann–Whitney *U*-test, $P < 2.2 \times 10^{-16}$, 0.161 IFN-γ effect size (polyp versus non-polyp). **c**, Selected genes detected in fibroblasts from single-cell data which correlate with the samples ranked by basal cell frequency detected within each ecosystem. Non-polyp, $n = 6$; polyp, $n = 6$. All genes used: Spearman correlation, abs(r) > 0.7651, $P < 0.0037$. To determine genes correlated in specific cell types (for example, fibroblasts) with the frequency of basal cells present in a cellular ecosystem, we correlated the average log-normalized single-cell count data for each gene to the rank of samples determined by increasing frequency of basal cells in each ecosystem (8.2% to 19.1% for non-polyp and 27.9% to 70.1% for polyp samples, Extended Data Fig. 7b). **d**, A clustered correlation matrix of genes identified as per **c** in single-cell data from fibroblasts; Pearson's abs(r) > 0.048 is significant ($P < 0.05$) based on asymptotic *P* values. **e**, Row-normalized heat map for myeloid cells from ethmoid sinus with select genes displayed on the *y* axis, including a core myeloid signature (ROC test myeloid cells versus rest of cells, AUC > 0.8), and genes differentially expressed (bimodal test) by disease state, with disease-state annotations on *x* axis. Bimodal test, all non-core genes $P < 0.0002$ or less with Bonferroni correction for multiple hypothesis testing based on number of genes tested. **f**, Row-normalized heat map for endothelial cells from ethmoid sinus with select genes displayed on *y* axis including a core basal signature (ROC test endothelial cells versus rest of cells, AUC > 0.75), and genes differentially expressed (bimodal test) by disease state, with disease-state annotations on *x* axis. Bimodal test, all non-core genes $P < 2.43 \times 10^{-6}$ or less with Bonferroni correction for multiple hypothesis testing based on number of genes tested.

© 2018 Springer Nature Limited. All rights reserved.

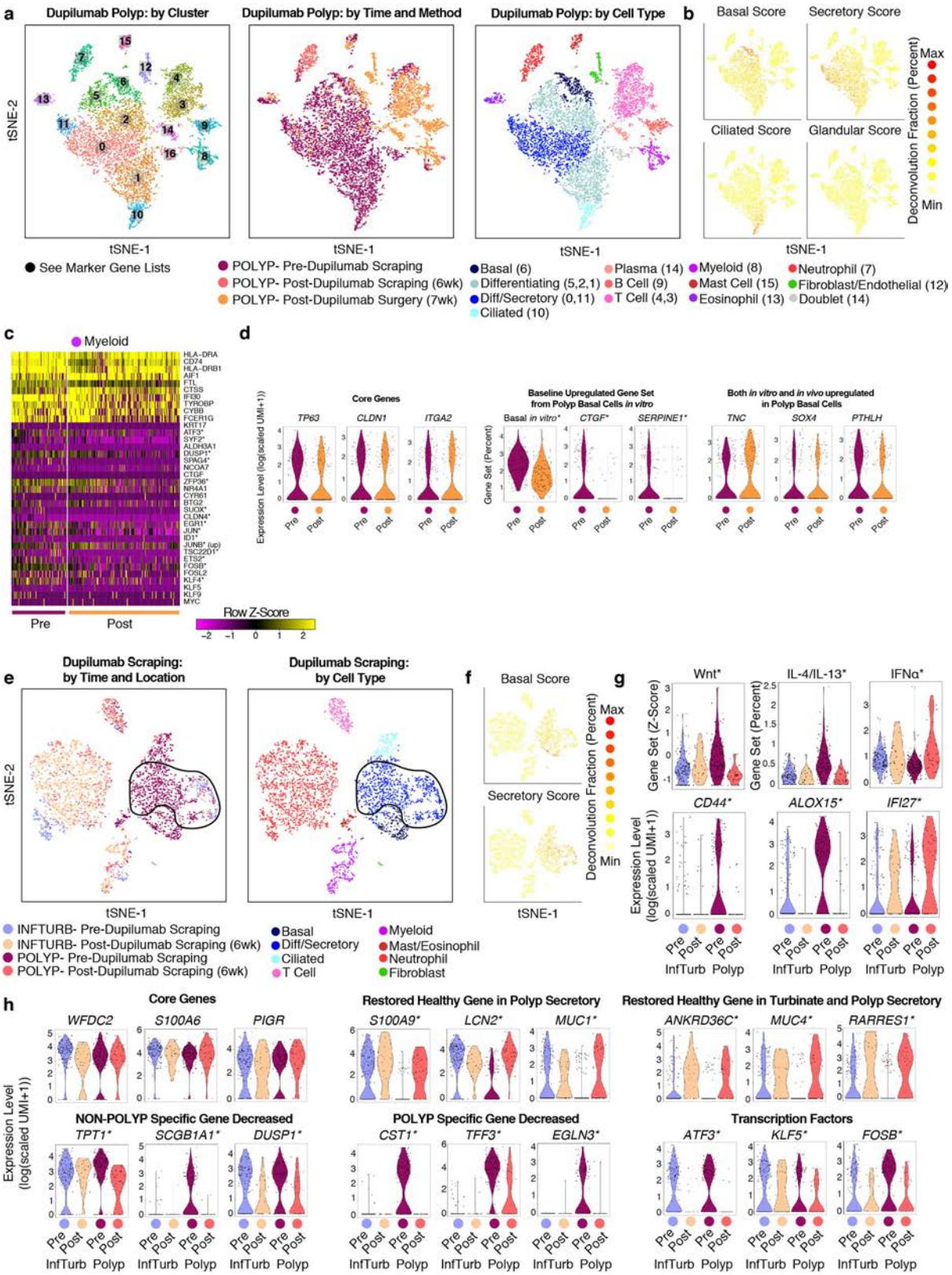


Extended Data Fig. 9 | See next page for caption.

© 2018 Springer Nature Limited. All rights reserved.

Extended Data Fig. 9 | Pseudotime analysis on basal and differentiating/secretory cell clusters from EthSin, transcriptional motif enrichments in non-polyp and polyp basal cells, and the identity of cell types in air–liquid interface cultures. **a**, Pseudotime analysis using diffusion mapping (see Methods) of selected clusters of epithelial cells, here displaying diffusion pseudotime (related to Fig. 4d). Clusters 8/1/4, 3,516 cells; clusters 12/2/0, 4,064 cells. $n = 6$ non-polyp, $n = 6$ polyp samples. Diffusion map and diffusion coefficients (DC) are calculated over the set of basal and apical marker genes identified in Fig. 1a (see Supplementary Table 3). **b**, The top 60 negatively correlated genes expressed in non-polyp cells with pseudotime trajectory and Pearson correlation values for genes in polyp cells also displayed; differential correlation coefficient analysis using Fisher's z-statistic, accounting for number of cells in each group (specific genes highlighted, all $>2z$; full results including Bonferroni corrected P values in Supplementary Table 3). **c**, Correlation matrices (row and column clustered) of the normalized read counts per sample in motif-associated peaks for non-polyp or polyp samples. Pearson correlation, $n = 3$ non-polyp, $n = 7$ polyp. **d**, A column-normalized heat map (row and column clustered) for the fraction of peaks with a motif corresponding to accessibility of the respective transcription factor displayed by patient. $n = 3$ non-polyp, $n = 7$ polyp. **e**, IGV tracks for *ATF3* and *KLF5* based on peaks detected and averaged by non-polyp and polyp samples from ATAC-seq profiling. **f**, IGV tracks for *S100A9* and *MUC4* based on peaks detected and averaged by non-polyp and polyp samples from ATAC-seq profiling. **g**, Violin plots for the count-based expression level (log(scaled UMI + 1)) for key marker genes using ROC test across cell types identified in Fig. 5a, Supplementary Table 3. 1,345 basal; 6,420 secretory; 6,381 hybrid; and 2,027 ciliated cells from $n = 2$ non-polyp and $n = 2$ polyp patients. AUC = 0.943 (*KRT5*), 0.667 (*TP63*), 0.644 (*LYPD2*), <0.55 (*SPDEF*), <0.55 (*KRT8*), 0.602 (*BPIFA1*), 0.813 (*PIFO*), 0.73 (*FOXJ1*). **h**, Row-normalized heat map for ALI secretory cells (subsampled to 300 cells per donor) as in Fig. 2f of the top in vivo secretory marker genes identified by ROC test (AUC > 0.662) with select genes displayed on y axis including a core secretory signature (ROC test, secretory cells versus rest of cells), and then within secretory cells, ROC test was used to identify marker genes within each disease/location category; and basal-cell derived annotations on x axis (see Supplementary Table 3 for full gene lists, all AUC > 0.65 for markers displayed in Fig. 2f). **i**, Quantification of flow cytometry for the ratio of basal to $Epcam^{hi}$ cells (gating as in Extended Data Fig. 7h) from ALI cultures at 21 days, stimulated with the indicated doses of IL-13. Points represent individual biological replicates; $n = 6$ non-polyp, $n = 5$ polyp samples for each dose. Two-way ANOVA; not significant between disease groups at any dose tested; Two-way ANOVA, $P = 0.0224$ for IL-13 dose. Data are mean ± s.e.m. **j**, Expression levels for *CTGF* ($\log_2$ expression value of log-normalized count data) in basal cells from non-polyp or polyp individuals across doses of cytokines displayed. $n = 4$ samples each dose. Two-way ANOVA $P < 0.0260$ for *CTGF*; all conditions non-polyp versus polyp except 0.1 ng ml^{-1} IL-4 dose for *CTGF*.

© 2018 Springer Nature Limited. All rights reserved.



Extended Data Fig. 10 | See next page for caption.

© 2018 Springer Nature Limited. All rights reserved.

Extended Data Fig. 10 | In vivo blockade with an anti-IL-4Rα monoclonal antibody shifts secretory cell state towards healthy-associated genes. a, Left, *t*-SNE plot of 8,764 single cells (related to Fig. 5e) from the nasal polyps of an individual treated with dupilumab (IL-4Rα monoclonal antibody) (1 patient, sampled at $n = 3$ time points), coloured by clusters identified through SNN analysis (Supplementary Table 3; Methods); middle, *t*-SNE plot coloured by time point and tissue of origin from polyp pre-dupilumab scraping (5,731 cells), from polyp post-dupilumab scraping (647 cells), and polyp post-dupilumab surgical sample (2,386 cells); right, *t*-SNE plot coloured by cell types identified through marker discovery (ROC test) and biological curation of identified clusters (Supplementary Table 3; Methods). **b**, Select cell-type specific score overlays for cell types indicated in original core dataset (see Supplementary Table 3 for full gene list). **c**, Row-normalized heat map for myeloid cells of the top marker genes identified by ROC test (AUC > 0.8) with select genes displayed on *y* axis including a core myeloid signature (ROC test myeloid cells versus rest of cells), and then genes found to be differentially expressed from Fig. 5f in basal cells, and treatment annotations on *x* axis. Bimodal test. *, differential genes in both basal cells and myeloid cells pre- versus post-treatment ($P < 0.003$ or less with Bonferroni correction for multiple hypothesis testing based on number of genes tested). **d**, Violin plots for basal cells (200 cells pre-dupilumab and 151 cells post-dupilumab, noted in **a**) for the count-based expression level (log(scaled UMI + 1)), except where indicated for gene scores, fraction of transcriptome and *z*-score (see Methods, Supplementary Table 4 for gene set used) for key basal cell genes for selected biological processes, or from the baseline upregulated gene set from polyp basal cells in vitro (Fig. 5c). Differential expression testing for decreased expression post-treatment using bimodal test not significant except where indicated ($*P < 0.00087$ or less with Bonferroni correction for multiple hypothesis testing based on number of genes tested; see Supplementary Table 3 for full list). Basal in vitro score pre versus post: two-tailed *t*-test, $P < 3.897 \times 10^{-15}$, effect size 0.822. **e**, *t*-SNE plot of 4,486 single cells (related to Figs. 2e, 5e) from the inferior turbinate or nasal polyps of an anti-IL-4Rα (dupilumab) treated individual ($n = 4$ samples) coloured by time point and tissue of origin (left) from inferior turbinate pre-dupilumab scraping (643 cells), from inferior turbinate post-dupilumab scraping (1,596 cells), polyp pre-dupilumab scraping (1,600 cells), and polyp post-dupilumab scraping (647 cells). *t*-SNE plot coloured by cell types (right) identified through marker discovery (ROC test) and biological curation of identified clusters (Supplementary Table 3; Methods); black outline indicates cells considered in **g**. **f**, Select deconvolution score overlays for cell types indicated in original core dataset (see Supplementary Table 3 for full gene list). **g**, Violin plot for the gene set score over Wnt pathway (*z*-score) and expression contribution to a cell's transcriptome over IFN-α- and IL-4/IL-13-commonly induced gene signature in secretory cells grouped as in **e** and sub-sampled to a maximum of 150 cells from each disease or location category from inferior turbinate pre-dupilumab scraping (150 cells), from inferior turbinate post-dupilumab scraping (23 cells), polyp pre-dupilumab scraping (150 cells) and polyp post-dupilumab scraping (38 cells) (see Methods, Supplementary Table 3, Supplementary Table 4 for gene lists used). Two-tailed *t*-test; Wnt score pre versus post polyp tissue, effect size 1.02, $P = 1.091 \times 10^{-14}$; Wnt score pre versus post inferior turbinate tissue, effect size −0.17, $P = 0.3706$; IL-4/IL-13 score pre versus post polyp tissue, effect size 1.17, $P < 2.2 \times 10^{-16}$; IL-4/IL-13 score pre versus post inferior turbinate tissue, effect size −0.17, $P = 0.163$; IFN-α score pre versus post polyp tissue, effect size −1.25, $P = 4.254 \times 10^{-5}$; IFN-α score pre versus post inferior turbinate tissue, effect size −0.304, $P = 0.2766$. $*P < 7.81 \times 10^{-6}$ or less between pre- and post-treated polyp, differential expression testing for decreased expression post-treatment using bimodal test. **h**, Violin plots of secretory cells grouped as in **e** and sub-sampled to a maximum of 150 cells from each disease or location category from inferior turbinate pre-dupilumab scraping (150 cells), inferior turbinate post-dupilumab scraping (23 cells), polyp pre-dupilumab scraping (150 cells) and polyp post-dupilumab scraping (38 cells) for the count-based expression level (log(scaled UMI + 1)) and for secretory cell genes from the gene set used in Fig. 2f affected by treatment within anatomical regions indicated by heading. $*P < 6.36 \times 10^{-5}$ or less except *KLF5* ($P = 0.0033$) and *FOSB* ($P = 0.0053$), differential expression testing for decreased expression post-treatment using bimodal test with Bonferroni correction for multiple hypothesis testing based on number of genes tested, see Supplementary Table 3 for all genes tested.

© 2018 Springer Nature Limited. All rights reserved.

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.