Supplementary Materials for

Human airway mast cells proliferate and acquire distinct inflammation-driven phenotypes during type 2 inflammation


*Corresponding author. Email: shalek@mit.edu (A.K.S.); jboyce@rics.bwh.harvard.edu (J.A.B.); nbarrett@bwh.harvard.edu (N.A.B.);

Published 26 February 2021, Sci. Immunol. 6, eabb7221 (2021)
DOI: 10.1126/sciimmunol.abb7221

The PDF file includes:

Supplementary Methods
- Fig. S1. In-depth characterization of nasal polyp MCs.
- Fig. S2. Characterization of nasal polyp MC protease phenotype and localization.
- Fig. S3. Characterization of sorted cells through scRNA-seq.
- Fig. S4. Immunohistologic localization of EBI2-expressing MCs.
- Fig. S5. Identification of MCs in three IPF scRNA-seq datasets.
- Fig. S6. Identification of an asthma-enriched CD38<sup>high</sup> MC cluster.
- Fig. S7. Influence of dupilumab on nasal polyp MC transcript expression.
- Fig. S8. Schematic representation of strategy for evaluating the transcriptome of Itgβ7<sup>high</sup> MCs.
- Fig. S9. Proposed model for MC homeostasis and hyperplasia in human sinus mucosa.

Legends for tables S1 to S7

Other Supplementary Material for this manuscript includes the following:

(available at immunology.sciencemag.org/cgi/content/full/6/56/eabb7221/DC1)

- Table S1 (.zip format). Patient information.
- Table S2 (Microsoft Excel format). Gene lists derived from literature and databases used for scoring.
- Table S3 (Microsoft Excel format). Cell-by-gene matrix (raw) for sorted MCs from six polyp donors.
- Table S4 (Microsoft Excel format). Cluster defining marker genes, MC subset defining genes, and differential expression full gene lists presented in manuscript.
Table S5 (Microsoft Excel format). Cluster defining marker genes for scRNA-seq datasets from the human IPF lung atlas and asthma versus healthy donors.
Table S6 (Microsoft Excel format). Differential gene expression analysis for human umbilical cord–derived MCs treated with IL-4 or vehicle for 96 hours.
Table S7 (Microsoft Excel format). Source data for figures.
Supplementary Methods

Sample preparation
Single-cell suspensions from collected ethmoid sinus surgical specimens were obtained using a previously published protocol (22), described below in detail. Each specimen was received directly in hand and processed immediately with an average time from patient to loading onto the Seq-Well platform of 3 total hours, and never exceeding 4 hours. Surgical specimens were collected into 30 mL of ice cold RPMI (Corning). Specimens were finely minced between two scalpel blades and incubated for 15 minutes 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 ug/mL DNase I (Roche) in RPMI with 10% fetal bovine serum. After 15 minutes, samples were triturated five times using a syringe with a 16G needle and returned to the rotisserie rack for another 15 minutes. After 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. Samples were typically fully dissociated at this step and were filtered through a 70um cell strainer and spun down at 500G for 10 minutes followed a rinse with ice-cold PBS (ThermoFisher, Ca/Mg free) to 30 mL total volume. RBCs were lysed using ACK buffer (ThermoFisher) for 3 minutes on ice to remove red blood cells, 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 500G for 5 minutes, resuspended in complete RPMI medium with 2% FCS (RPMI1640, 100 U/mL penicillin, 100 ug/mL streptomycin, 10 mM HEPES, 2% FCS, 50 ug/mL gentamycin (all reagents from ThermoFisher)), and counted to adjust concentration to 100,000 cells/mL for loading onto Seq-Well arrays. Surplus cells were frozen in CryoStor CS10 (Millipore Sigma), cooled to -80 C using a “Mr. Frosty” freezing container (Thermo scientific) and stored in liquid nitrogen to allow for batched flow cytometric evaluation. Single-cell suspensions from discarded skin samples were obtained using the above protocol with the addition of 2.5 mg/mL dispase (Gibco) to the digest buffer. Nasal polyp epithelium was isolated by incubating crudely chopped polyp segments with epithelial extraction media (RPMI1640, 2% FCS, 1 mM
EDTA, 1 mM dithiothreitol (DTT)) for 30 minutes at 37°C using a magnetic stirbar set to 500 RPM. After 30 minutes, the supernatant containing dispersed epithelial cells and intraepithelial leukocytes was collected and centrifuged at 500G for 5 minutes.

For peripheral blood mononuclear cell (PBMC) isolation, peripheral blood was collected into 10mL Vacutainers (Beckton Dickenson) containing sodium heparin. Whole blood was diluted 1:1 with HBSS -/- containing 2 mM EDTA, layered on top of Ficoll Paque Plus (GE healthcare), and centrifuged for 30 minutes at 500 RCF without break. The interface was collected and washed twice with HBSS -/- containing 2 mM EDTA at 110 RCF to remove platelets, after which any remaining erythrocytes were lysed using ACK buffer.

**Cell Culture**

Human CBMCs were derived from umbilical cord mononuclear cells as previously described (66). Briefly, mononuclear cells were cultured in the presence of SCF (100 ng/mL), IL-6 (50 ng/mL), and IL-10 (10 ng/mL) (all cytokines purchased from R&D systems). After reaching greater than 95% purity based on staining with toluidine blue (6-9 weeks), CBMCs were transferred to new media containing SCF with IL-4 (10 ng/mL) or SCF alone for 72 hours (qPCR studies) or 96 hours (RNAseq), after which RNA was collected for analysis.

**Flow Cytometry Antibodies**

The following antibodies were used to identify mast cells via flow cytometry: anti-human FcεR1α (AER-37(CRA-1), Biolegend), anti-human CD117 (104D2, Biolegend), anti-human β7 integrin (FIB504, BD biosciences), anti-human αE integrin (BER-ACT8, Biolegend), anti-human α4 integrin (9F10, Biolegend), anti-human CD45 (2D1, biolegend), anti-human CD38 (HIT2, Biolegend), anti-human CD34 (561, Biolegend), anti-human MRGX2 (K125H4, Biolegend), anti-human CD4 (OKT4, Biolegend), anti-human CD8 (HIT8A, Biolegend), anti-human CD19 (HIB19, Biolegend), anti-human CD11b (ICRF44, Biolegend), anti-human CD11c (3.9, Biolegend). Non-specific antibody binding was blocked using Truestain FcX (Biolegend). Surface epitopes were
stained for 30 minutes on ice in FACS buffer and then washed. For intracellular avidin
staining, cells were next permeabilized using a BD Cytofix/Cytoperm
fixation/permeabilization solution kit (BD bioscience) and stained with FITC avidin
(Biolegend). Chymase immunostaining was conducted as previously described (67) with
modification. Anti-human chymase (clone B7, Chemicon International) was conjugated
to Alexa Fluor 647 in parallel with isotype control using an Alexa Fluor 647 antibody
labeling kit (Invitrogen) and used at a 1:5000 dilution. For nuclear staining, surface-
stained cells were permeabilized with a True-Nuclear Transcription Factor Buffer set
(Biolegend) and stained with anti-human Ki67 (Ki67, Biolegend). Single-cell
suspensions used for CITE-seq analysis were stained with anti-human CD38-A0389
(HIT2, Biolegend), anti-human CD117-A0061 (104D2, Biolegend), and anti-human b7-
integrin-A0214 (FIB504, Biolegend). For fluorescent staining of CD117 in CITE-seq
experiments, anti-human CD117 clone YB5.B8 (BD Bioscience) was used after
screening for clone compatibility.

**Immunofluorescence Antibodies**

For slide-based cytologic analysis, cells were stained with mouse anti-human tryptase
mAb AA1 (Agilent) and goat anti-human chymase polyclonal Ab (Abcam #111239).
Binding of Abs was detected with AF488-labeled F(ab’)2 fragments of donkey anti-
mouse IgG (Jackson Immunoresearch #715-546-150) and AF594-labeled F(ab’)2
fragments of donkey anti-goat IgG (Jackson Immunoresearch #715-586-147),
respectively. Cell nuclei were visualized using DAPI.

For single indirect immunofluorescence staining of paraffin embedded sections,,
sections were blocked with 10% donkey serum + 3% BSA in PBS-T followed by 4°C
overnight incubation with mouse anti-human Chymase mAb at 2 µg/ml (BIO-RAD, MAC
1930, clone CC1). Sections were then washed and stained with Alexa Fluor 594
AffiniPure F(ab’)2 fragment donkey anti-mouse IgG (H+L) secondary Abs (Jackson
ImmunoResearch Laboratories, code 715-586-150) at 1:500 with Hoechst nuclear dye
at 1:5000 in PBS at room temperature for 1 hour. Double indirect immunofluorescence
was performed with VectaFluor Duet Immunofluorescence Double Labeling Kit
(vectorlabs, DK-8818) according to manufacturer’s instruction. 1:67 unlabeled rabbit anti-human GPR183/EBI2 pAb(Invitrogen, PA5-33673) and 1:143 mouse anti-mast cell tryptase [AA1] (abcam, ab2378) were used as primary antibodies. DyLight 488 anti-Rabbit, and DyLight 594 anti-mouse from the kits were used as secondary antibodies. Nuclear staining was performed with 1:5000 Hoechst 33342 (invitrogen) diluted in PBS.

**Single-cell RNA-sequencing with the Seq-Well platform**

For single-cell RNA sequencing, we utilized the Seq-Well platform for massively parallel scRNA-seq to capture transcriptomes of sorted MCs on barcoded mRNA capture beads (27). Briefly, 20,000 cells were loaded onto a single array preloaded with barcoded mRNA capture beads (ChemGenes). The loaded arrays containing cells and beads were 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 buffer allows for cell lysis, transcript hybridization, and bead recovery before performing pooling samples for reverse transcription en masse. Following reverse transcription using Maxima H Minus Reverse Transcriptase (ThermoFisher EP0753) and an Exonuclease I treatment (New England 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 >1000bp tailing off to beyond 5000bp, and a small/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 previously described(27). 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). For CITE-seq experiments, the antibody-derived tag (ADT) fraction was separated and libraries prepared as previously described (50), without modification. Two arrays were sequenced per sequencing run with an Illumina 75 Cycle NextSeq500/550v2 kit (Illumina FC-404-2005) at a final concentration of 2.8pM. The read structure was paired end with Read 1 starting from a custom read 1 primer containing 20 bases with a 12bp cell barcode.
and 8bp unique molecular identifier (UMI) and Read 2 containing 50 bases of transcript or ADT information.

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 the 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 (e.g. cells-by-genes tables) for each sample were obtained from quality filtered and mapped reads, with an automatically determined threshold for cell count.

**Differential expression and fractional contribution of gene sets to transcriptome**

To identify differentially expressed genes within cell types across clusters, we utilized Wilcoxon rank sum test implemented in Seurat through the FindAllMarkers command. To determine the contribution of a particular gene list to a cell’s overall transcriptome, we summed the total log-normalized expression values for genes within a “list of interest” and divided by the total amount of raw transcripts detected in that cell, giving the proportion of a cell’s transcriptome dedicated to producing those genes. Gene lists implemented in this manner included cell identity signatures generated from our prior study of nasal polyposis, a plasmacytoid dendritic cell (pDC) signature obtained from peripheral blood pDCs (68), MC activation signatures derived from a study of murine bone marrow-derived cultured MCs activated with IL-33 or IgE crosslinking (35). For reference gene lists used, please see Supplementary Table 2. Differentially expressed transcripts associated with the MC1 or MC3 clusters were derived from the MC1-4 cluster defining genes displayed in Supplemental Table 3.

**Cell type identification**

To identify contaminating cell populations within our sorted MC scRNA-seq dataset, all clusters were scored by identifying the percentage of their transcriptome consisting of
cellular identity signatures. Significant transcripts enriched within each cluster were secondarily visually inspected to confirm their identity. As cluster 5 did not show enrichment for any signature identified previously in nasal polyps, the most significantly enriched transcripts within the cluster were identified (PLAC8, IRF8, GZMB, IRF7). A literature search identified these transcripts as all highly enriched in human pDC (68), after which a pDC signature was constructed to confirm their identity. All non-MC clusters were excluded from subsequent analysis.

**Analysis of externally generated scRNA-seq datasets**

For analysis of MCs from fibrotic lung tissue, we used three datasets associated with the human IPF lung atlas (36). The Bankovich/Kropski dataset including IPF and ILD versus healthy control was accessed via GEO accession GSE135893 (37). The Kaminski/Rosas dataset of analysis of IPF and chronic obstructive pulmonary disease (COPD) versus healthy control was accessed via GEO accession GSE136831 (38). The Lafyatis dataset of IPF versus healthy control was accessed via GEO accession GSE128033(39). MCs from asthmatic lung tissue were derived from a previously published study of healthy and asthmatic lung(41). All datasets were normalized using the Harmony package for R (65) based on donor. The prominent MC cluster from each dataset, identified through TPSAB1 expression, was subsetted and re-clustered following a second round of Harmony normalization, and differential gene analysis was conducted using the Wilcox test implemented within Seurat.

For healthy lung scRNA-seq analysis, we utilized a series of droplet-based human lung scRNA-seq datasets generated by the Krasnow Laboratory(47). Annotated Seurat objects generated by Travaglini et al. for three healthy donors were accessed through synapse (https://www.synapse.org/#!Synapse:syn21041850). Analysis was restricted to cells from the distal, medial and proximal lung. The resulting cells were merged using the Seurat package for R and normalized using Harmony based on donor. Upon independent identification of a MC cluster expressing tryptase tryptase (TPSAB1), this
population was subsetted further analysis of MC_{TC}-associated transcripts. Differential gene analysis was conducted using the Wilcox test implemented within Seurat.

**SmartSeq2 sequencing and analysis**

Sequencing of in vitro stimulated CBMCs was conducted by the Broad Institute Genomics Platform. SmartSeq2 libraries were prepared according to the SmartSeq2 protocol described by Trombetta et al (69). Briefly, total RNA was purified using RNA-SPRI beads. cDNA was generated from full-length mRNA transcripts using reverse transcriptase with terminal transferase activity. Combined with a second "template switch" primer, the cDNA was constructed to have two universal priming sequences. Following preamplification, the Nextera XT library construction kit was used to prepare 96 unique indexes specific to each sample. Barcoded cDNA fragments were then pooled prior to sequencing. Sequencing was carried out using an Illumina NextSeq500 as paired-end 2x38bp to a coverage of approximately ~4 million reads per well. Data for each lane was separated by barcode and concatenated. Transcripts were aligned to GRCh38 genome assembly with STAR (70) and quantified with HTseq(71). Data analysis for all samples was conducted using Bioconductor for R. Count normalization and differential was conducted using the DESeq2 software package (72). Genes were considered significantly differentially regulated based on a false discovery rate of <0.1 using a Benjamini-Hochberg adjusted \( p \) value to correct for multiple comparisons. Heatmaps were generated using the pheatmap software package.
Fig. S1. In-depth characterization of nasal polyp MCs.

(A) Full gating scheme to identify nasal polyp MCs from two donors, arrows indicate sequential gating. (B) Full gating scheme to identify CRSsNP MCs from two donors, arrows indicate sequential gating. (C) Characterization of CCR3 expression on nasal polyp MCs identified as in (S1A), arrows indicate sequential gating. (D) Plot showing CCR3 vs CD117 expression on polyp-derived CD45+ cells, from the same donor shown in (S1C). (E) Characterization of HLA-DRA expression on nasal polyp MCs identified as in (S1A), arrows indicate sequential gating. (F) Plot showing HLA-DRA vs CD117 expression on polyp-derived CD45+ cells, from the same donor shown in (S1E). (G) Characterization of CD34 expression on nasal polyp MCs identified as in (S1A), arrows indicate sequential gating.
Fig. S2. Characterization of nasal polyp MC protease phenotype and localization.

(A) Channel series showing protease staining on sorted polyp MCs from the epithelial enriched (CD117 low) gate. Blue:DNA; Green: tryptase; Red: chymase. (B) Channel series showing protease staining on sorted polyp MCs from the supepithelial (CD117 high) gate. Blue:DNA; Green: tryptase; Red: chymase. (C) Representative plots from two donors showing localization of chymase-expressing cells within the nasal polyp. White line shows border between epithelium and sub-epithelium, with epithelial side indicated by a white E. Green arrows indicate positive cells. (D) Quantification of localization of chymase-expressing cells within the nasal polyp for six individual donors. *** indicates P<0.001
Fig. S3. Characterization of sorted cells through scRNA-seq.

(A) Identification of 10 clusters derived from 7,355 cells sorted for MC surface marker expression (FcεRIα+, CD117+, CD45+, CD11b-, CD11c-, CD3-, CD19-) from 6 donors.

(B) Expression of signature genes associated with plasmacytoid dendritic cells, plasma cells, myeloid cells, epithelial cells, and fibroblasts in each cluster identified in Fig. 2A. Signature genes shown in Supplementary table 1. Clusters 3 and 7 showed the most significant enrichment for myeloid signature genes (p<1x10^{-10}), cluster 4 for plasma cell signature genes (p<1x10^{-10}), cluster 5 for plasmacytoid dendritic cell signature genes (p<1x10^{-10}), cluster 8 for epithelial cell signature genes (p<1x10^{-10}), and cluster 9 for fibroblast signature genes (p<1x10^{-10}) (ANOVA followed by Tukey’s test).

(C) Labeling of cellular identity of each cluster identified in Fig. 2A based on gene expression signatures shown in Fig. S1B.

(D) Heatmap of top transcripts associated with each population identified in (C).

(E) Number of unique molecular identifiers, total genes, percentage of mitochondrial RNA, and percentage of ribosomal RNA per each cluster.

(F) Stacked bar graphs denoting relative concentration of each observed cell population in 6 donors labeled by disease endotype.
Fig. S4. Immunohistologic localization of EBI2-expressing MCs.
Immunostaining for tryptase (red) and the polyp MC_T-associated marker EBI2 (green) in nasal polyp sections from four donors. Arrows indicate cells co-expressing EBI2 and tryptase.
Fig. S5. Identification of MCs in three IPF scRNA-seq datasets.
(A) Identification of MCs based on *TPSAB1* expression from the Bankovich/Kropski scRNAseq study of control lung, IPF, and ILD. (B) Identification of MCs based on *TPSAB1* expression from the Kaminski/Rosas scRNAseq study of healthy, IPF and COPD lung biopsy samples. (C) Identification of MCs based on *TPSAB1* expression from the Lafayatis scRNAseq study of healthy and IPF lung biopsy samples.
Fig. S6. Identification of an asthma-enriched CD38$^{\text{high}}$ MC cluster.

(A) Clustering of MCs from asthma and healthy dataset indicated a population (outlined in black, left panel) that was statistically enriched for the polyp MC3-associated transcript CD38 (center panel, padj < 0.03) and predominantly composed of MCs from asthma tissue relative control (right panel). (B) Expression of GPR183 and IL17RB in MCs from the lungs of asthma patients and control donors. Neither transcript was found to be significantly enriched.
Fig. S7. Influence of dupilumab on nasal polyp MC transcript expression.

(A) Previously defined clusters of nasal polyp cells sampled from a single donor sampled prior to undergoing therapy and following 6 weeks of dupilumab. (B) Violin plots showing expression of the MC3-associated transcripts GPR183 (ns) and IL17RB (padj < 1e-5) in nasal poly MCs prior to undergoing immunotherapy and following 6 weeks of dupilumab treatment.
Fig. S8. Schematic representation of strategy for evaluating the transcriptome of Itgβ7^high^ MCs.

MCs were subsetted into Itgβ7^high^ and Itgβ7^Low^ populations based on CITE-seq surface expression of Itgβ7, using an arbitrary normalized cutoff value of 1. Itgβ7^high^ and Itgβ7^Low^ Differential expression analysis was conducted on Itgβ7^high^ and Itgβ7^Low^ MCs falling within each cluster. Genes indicated on right-hand side indicate all differentially expressed transcripts between Itgβ7^high^ and Itgβ7^Low^ MCs within a cluster (FDR<0.05, Wilcox test)
Fig. S9. Proposed model for MC homeostasis and hyperplasia in human sinus mucosa.

(A) In CRSsNP (left), intermediate MCs and MC_{TC} are observed in the absence of MC_{T} and \( \beta_{7}^{\text{high}} \) MCs, suggesting that MC_{TC} arise from intermediate MCs. In polyposis (right), inflammation-associated MC_{T} are observed in the epithelium, subepithelial MCs transcriptionally intermediate between MC_{T} and MC_{TC} are expanded in part through local proliferation, and subepithelial MC_{TC} expand. \( \beta_{7}^{\text{high}} \) MCs expressing high CD38 and a range of CD117 are enriched in the epithelium and MC3 clusters, indicating a relationship with polyp MC_{T}, but are also observed in subepithelium. These \( \beta_{7}^{\text{high}} \) MCs are transcriptionally indistinct, suggesting they undergo rapid differentiation based on the compartment within which they reside, and their relationship with the intermediate and MC_{TC} populations is not clear. (B) Summary table linking flow cytometric characteristics, scRNA-seq defined clusters and select additional features across MC phenotypes defined within this study.
Table S1. Patient Information.

Table S2. Gene lists derived from literature and databases used for scoring.

Table S3. Cell-by-gene matrix (raw) for sorted MCs from six polyp donors.

Table S4. Cluster defining marker genes, MC subset defining genes, and differential expression full gene lists presented in manuscript.

Table S5. Cluster defining marker genes for scRNA-seq datasets from the human IPF lung atlas and asthma versus healthy donors.

Table S6. Differential gene expression analysis for human umbilical cord–derived MCs treated with IL-4 or vehicle for 96 hours.

Table S7. Source data for figures.