Single-cell molecular profiling provides a high-resolution map of basophil and mast cell differentiation

Fiona K. Hamey[1,2], Winnie W.Y. Lau[1], Iwo Kucinski[1], Xiaonan Wang[1], Evangelia Diamanti[1], Nicola K. Wilson[1], Berthold Göttgens[1,*], Joakim S. Dahlin[1,3,*]

1 Department of Haematology, University of Cambridge, Wellcome–Medical Research Council Cambridge Stem Cell Institute, Cambridge, United Kingdom.

2 Present address: JDRF/Wellcome Diabetes and Inflammation Laboratory, Wellcome Centre for Human Genetics, University of Oxford, Oxford, UK.

3 Department of Medicine, Karolinska Institutet and Karolinska University Hospital, Stockholm, Sweden

*: Corresponding authors

B. Göttgens; Address: University of Cambridge, Department of Haematology, Wellcome–Medical Research Council Cambridge Stem Cell Institute, Jeffrey Cheah Biomedical Centre, Cambridge Biomedical Campus, Cambridge CB2 0AW, UK; E-mail: bg200@cam.ac.uk; Tel. +44-1223-336829; FAX +44-1223-762670

J. Dahlin; Address: NKS BioClinicum J7:30, Visionsgatan 4, 171 76 Solna, Sweden; E-mail: joakim.dahlin@ki.se; Tel. +46-8-517 761 69

Abstract word count: 222

Text word count: 4478

27 No of figures: 4

28 No of tables: 0

29 No of references: 36

30

Abstract

Differentiation of hematopoietic stem and progenitor cells ensure a continuous supply of mature blood cells. Recent models of differentiation are represented as a landscape, in which individual progenitors traverse a continuum of multipotent cell states before reaching an entry point that marks lineage commitment. Basophils and mast cells have received little attention in these models and their differentiation trajectories are yet to be explored. Here, we have performed multicolor flow cytometry and high-coverage single-cell RNA sequencing analyses to chart the differentiation of hematopoietic progenitors into basophils and mast cells in mouse. Analysis of flow cytometry data reconstructed a detailed map of the differentiation, including a bifurcation of progenitors into two specific trajectories. Molecular profiling and pseudotime ordering of the single cells revealed gene expression changes during differentiation, with temporally separated regulation of mast cell protease genes. We validate that basophil and mast cell signature genes increased along the trajectories into their respective lineage, and we demonstrate how genes critical for each respective lineage are upregulated during the formation of the mature cells. Cell fate assays showed that multicolor flow cytometry and transcriptional profiling successfully predict the bipotent phenotype of a previously uncharacterized population of basophil-mast cell progenitor-like cells in mouse peritoneum. Taken together, we provide a detailed roadmap of basophil and mast cell development through a combination of molecular and functional profiling.

Introduction

Hematopoietic stem and progenitor cells (HSPCs) constitutively generate blood cells, including erythrocytes, platelets, granulocytes, macrophages, and lymphocytes. The hierarchical model of hematopoiesis, with distinct megakaryocyte-erythroid, granulocyte-monocyte, and lymphoid branches, was the dominating representation of the differentiation process before the introduction of single-cell RNA sequencing.[1] Recent groundbreaking studies that couple single-cell transcriptomics and cell fate assays reveal that blood cell differentiation more likely represents a landscape of cell states with continuous progression from progenitors into the mature cell lineages.[2-6]

Mast cells and basophils constitute two cell types of the hematopoietic system, whose differentiation trajectories are yet to be deciphered. Mast cells are sentinel cells that are strategically positioned throughout the body and allow rapid triggering of the immune system upon infection.[7] Along with basophils, their activation results in prompt release of proteases and histamine from the cytoplasmic granules as well as synthesis of cytokines and chemokines. These mediators in turn cause inflammation, vasodilation, and leukocyte recruitment to the site of triggering.[7] A similar cascade is initiated following IgE-allergen-mediated cell activation that causes allergic symptoms in patients.

Bone marrow HSPCs give rise to basophil and mast cells,[8] and single-cell transcriptomics of Lin^- $c\text{-}Kit^+$ mouse bone marrow progenitors recently uncovered the gene expression changes during the transition from hematopoietic stem cells to common bipotent basophil-mast cell progenitor (BMCP).[3] However, the further progression of BMCPs to basophils and mast cells is yet to be delineated.

Here, we combine multicolor FACS index sorting with high-coverage single-cell RNA sequencing to investigate the basophil-mast cell bifurcation and the differentiation into each respective lineage. We demonstrate that molecular profiling and pseudotime ordering of single cells highlights genes that are critical for cell differentiation and maturation. The analysis is accompanied with the generation of a user-friendly web resource that allows gene expression to be explored across the single-cell landscape. Finally, we use cell-fate assays to show that single-cell transcriptomics and protein epitope data analysis successfully predict the fate potential of the previously uncharacterized BMCP-like cell population in the peritoneal cavity. Taken together, the current resource provides a detailed roadmap of two rare and developmentally related hematopoietic cells, whose activation contributes to a broad range of human diseases.

Methods

Cell isolation and flow cytometry

Experiments involving mice were performed according to the United Kingdom Home Office regulations. PBS with 2 % fetal calf serum (Sigma-Aldrich, St Louis, MO) and 1 mM EDTA was injected into the peritoneal cavity of euthanized C57BL/6 mice. The fluid was aspirated following vigorous massage, and the cells were prepared for fluorescence-activated cell sorting (FACS). Peritoneal lavage samples with excessive blood contamination were discarded before data acquisition. Bone marrow cells were extracted by flushing or crushing the femurs, tibias, and/or ilia. Red blood cells were lysed and the remaining cells were prepared for FACS. The cells were sorted with a

105 BD Influx cell sorter (BD Biosciences, San Jose, CA). Cell doublets were excluded
106 with the width parameters. BMCP-like cells and mast cells were sorted two
107 consecutive times for cell culture experiments. The cells were sorted into Terasaki
108 plates (Greiner Bio-One, Kremsmünter, Austria) or 96-well plate wells. Visual
109 inspection determined colony sizes following culture, and the size was set to 1 if no
110 live cells were observed in a particular well. Flow cytometry was typically performed
111 on colonies constituting at least 20 cells as described previously.[3] Cultured cells were
112 analyzed with the BD Fortessa flow cytometers (BD Biosciences).
113
114 *Antibodies and cell staining*
115 Primary cells were incubated with the antibodies mouse hematopoietic progenitor cell
116 isolation cocktail, integrin β7 (clone FIB504), CD34 (RAM34), Sca-1 (D7), CD16/32
117 (93), c-Kit (2B8), FcεRI (MAR-1), IL-33Rα/ST2 (DIH9), and/or CD49b (DX5).
118 Cultured cells were stained with c-Kit, FcεRI, CD49b, with or without TER119
119 (TER119). Fc-block (clone 93) was used where appropriate. The antibodies were
120 from STEMCELL Technologies (Vancouver, Canada), BD Biosciences, Biolegend
121 (San Diego, CA), and Thermo Fisher Scientific (Waltham, MA). DAPI (BD
122 Biosciences) or 7-AAD (Thermo Fisher Scientific) were used to exclude dead cells.
123
124 *Cell culture*
125 The cells were cultured for 6-7 days in IMDM (Sigma-Aldrich) with 20 % heat-
126 inactivated fetal calf serum (Sigma-Aldrich), 100 U/ml penicillin (Sigma-Aldrich),
127 0.1 mg/ml streptomycin (Sigma-Aldrich), 50-200 μM β-mercaptoethanol (Thermo
128 Fisher Scientific). The medium was supplemented with 20 ng/ml IL-3 and 100 ng/ml
129 stem cell factor, or 80 ng/ml stem cell factor, 20 ng/ml IL-3, 50 ng/ml IL-9, and 2

U/ml erythropoietin. All cytokines were recombinant mouse cytokines (Peprotech, Rocky Hill, NJ) except the erythropoietin (Eprex; Janssen-Cilag, High Wycombe, UK), which was human.

Flow cytometry analysis

FlowJo v10 (Treestar, Ashland, OR) produced the flow cytometry plots. Diffusion map and principal component analysis plots of flow cytometry data were generated using the R programming environment. Briefly, the flow cytometry events were down-sampled according to the population with the least number of events. Duplicate entries were removed, and the parameters representing fluorescent markers log-transformed. Variables were z-scored and diffusion map plots generated using the *destiny* and *ggplot2* packages. Principal component analysis (PCA) was calculated using the prcomp function. Data projection was performed using the predict function.

Single-cell RNA sequencing and analysis

Single-cells were FACS index sorted into lysis buffer, and single-cell RNA sequencing was performed based on the Smart-Seq2 protocol.[9] Protocols and single-cell RNA sequencing data generated for this article have been deposited in the Gene Expression Omnibus database (accession numbers GSE128003 and GSE128074). Single-cell RNA sequencing data of bone marrow BMCPs, analyzed in Dahlin et al (2018),[3] are available through GSE106973.

Sequencing data were aligned using GSNAP[10] to Ensembl genome build 81[11] and gene counts were obtained using HT-Seq[12]. Quality control filtering and normalization was performed in the R programming environment. Quality control was

155 performed to exclude cells with fewer than 500,000 reads mapping to nuclear genes
156 or with over 25% of mapped reads mapping to ERCC spike-ins. For the peritoneal
157 cells, PCA of the quality control-filtered samples showed that two cells separated
158 from the rest of the sample in principal component (PC) 1 (Figure S2A). Genes with
159 high PC1 loadings were highly significantly enriched for B cell related genes, so these
160 two outlier cells were suspected to be contaminating B cells and excluded from
161 further analysis. Cells were then normalized using Scran[13] and highly variable genes
162 were identified using the ERCC spike-ins to estimate technical variance[14]. This
163 identified 3330 highly variable genes for the basophil dataset and 1832 highly
164 variable genes for the mast cell dataset.
165
166 Downstream analysis was performed using the scanpy v1.4 python module[15]. Inbuilt
167 scanpy functions were used for PCA and diffusion map dimensionality reduction.
168 Differential expression between cell types was performed using the
169 *rank_genes_groups* function with the t-test_overestim_var option for testing. P-values
170 were adjusted using the benjamini-hochberg method for correcting for multiple
171 testing, and genes with adjusted p-value < 0.01 were considered significant. Gene list
172 enrichment analysis was performed using the *enrichr* function from the *gseapy* python
173 module[16,17]. Cell cycle scoring was performed on scaled data using the scanpy
174 *score_genes_cell_cycle* function with S phase and G2/M phase gene lists downloaded
175 from Macosko et al[18]. Bone marrow BMCP cells from Dahlin et al (2018)[3] were
176 projected into the PCA space of the peritoneal cells and the k=10 closest peritoneal
177 neighbors of each bone marrow cell were identified in these co-ordinates. Figure 3B
178 was then generated by scoring how frequently each peritoneal cell was the nearest
179 neighbor of a bone marrow BMCP. Peritoneal mast cells were ordered in pseudotime

using the diffusion pseudotime (DPT) scanpy implementation[19]. Due to cell cycle effects confounding the diffusion map and DPT analysis, basophil progenitor cells were ordered in pseudotime by ordering cells along PC1. Genes with dynamic expression in pseudotime were identified following the method of Tusi et al.[2] Briefly, gene expression was first smoothed along pseudotime using a sliding window of size 20. For each ordering, the windows with minimum and maximum gene expression were identified, and a t-test performed between the values in each of these windows, giving a p-value for each gene. To generate a background distribution, this analysis was repeated for a random shuffling of cells along pseudotime. The adjusted p-value for each gene was then calculated as the fraction of shuffled p-values across all genes that were less than the p-value of the gene in question for non-permuted data. Genes with adjusted p-value < 0.01 were then treated as dynamic across pseudotime and plotted in the heatmaps. Gene expression in the heatmap was smoothed using a sliding window of size 20 and z-score transformed for each gene. To identify groups of genes with different pseudotime dynamics, genes were clustered using Louvain clustering[20] with the scanpy implantation, with the nearest neighbor matrix calculated on the full pseudotime expression matrix. The resolution was chosen to obtain 2 clusters for each dataset: downregulated and upregulated genes. Mast cell and basophil signature gene sets were obtained from Dwyer et al,[21] who used microarray analysis on bulk samples to characterize gene expression specific to these mature cell types. Statistical overlap between gene lists was calculated using a hypergeometric test. Panther v14.1 was used to identify the dynamic mast cell genes annotated as proteases (Panther category PC00190)[22]. To plot gene expression trends along pseudotime the genSmooth Curves function from the monocle R package[23] was used to fit smooth spline curves for the expression of each gene against pseudotime. When

all genes were plotted together expression values of each gene were scaled by dividing values by the maximum of that gene along pseudotime to account for the very different dynamic ranges across genes. Interactive websites for plotting gene expression and flow cytometry data are hosted at http://128.232.224.252/bas/ and http://128.232.224.252/per/ for the basophil and mast cell dataset, respectively.

Results

Multicolor flow cytometry analysis reveals the basophil and mast cell differentiation trajectories

Basophil and mast cell differentiation are closely linked, and the cells share a common bipotent progenitor (Figure 1A). Here, we used multicolor flow cytometry to map these branching trajectories at the single-cell level. Flow cytometry analysis of mouse bone marrow cells captured BMCPs and cells of the basophil differentiation trajectory (Figure 1Bi,ii).[3,24] As the late mast cell differentiation takes place at peripheral sites, parallel analysis of peritoneal cells identified BMCP-like cells and mast cells (Figure 1Biii). Dimensionality reduction with a diffusion map algorithm enabled 2-dimensional visualization of the flow cytometry single-cell datasets, which covered 5 cell populations recorded with 9 fluorescent and 2 light scatter parameters. The diffusion map visualization revealed a bifurcation at the BMCP stage, establishing the putative entry points to the basophil and mast cell trajectories (Figure 1C). The diffusion map embedding further visualized the progression from BMCP, through basophil progenitors, to basophils. The mast cell trajectory exhibited a similar pattern, with differentiation of BMCP-like cells to mature mast cells.

Plotting individual surface markers in the diffusion map allowed us to investigate how the proteins are expressed during differentiation. For example, loss of CD34 in combination with downregulation of c-Kit marked the progression from BMCPs to basophils (Figure 1D), and loss of integrin β7 in c-Kit$^+$ cells was associated with differentiation along the trajectory from BMCPs to mast cells (Figure 1D). Taken together, the flow cytometry dataset provides a template of basophil-mast cell

differentiation at single-cell level and highlights the bifurcation towards the two lineages.

Single-cell profiling captures progression of basophil differentiation in the bone marrow

Analysis by flow cytometry suggested that the flow cytometry gating strategies we used could be capturing a continuum of differentiation towards basophils and mast cells. To first identify changes in gene expression programs during basophil differentiation, we performed single-cell RNA-sequencing of basophil progenitor (BaP) cells and basophil (Ba) cells from mouse bone marrow. Both principal component analysis (PCA) and diffusion maps showed separation between the majority of cells from the two sorting gates (Figure 2A, Figure S1A). To investigate which genes were driving this separation, we performed differential gene expression analysis, identifying 212 genes upregulated and 833 genes downregulated in Ba cells compared to BaPs (Table S1). Enrichment analysis of these gene lists revealed that upregulated genes were enriched for granulocyte immune response terms (Figure S1B). Downregulated genes were enriched for cell cycle related terms (Figure 2B), suggesting a difference in cell cycle behavior throughout the differentiation process. This observation is in line with other hematopoietic differentiation pathways, where progenitors commonly loose proliferative capacity as they mature into the fully differentiated cell types.

To further explore this, we then performed analysis to computationally assign cell cycle state to the single-cell profiles.[18] Consistent with the gene list enrichment analysis, the majority of cells in the BaP gate were assigned to S and G2M states

(69%), whereas 87% of cells in the Ba gate were assigned to G1 state (Figure 2C, D). The effect of cell cycle status was clear in the diffusion map dimensionality reduction (Figure S1C), confounding attempts to order cells using pseudotime algorithms such as diffusion pseudotime (DPT). Instead, downregulation of progenitor marker genes such as *Cd34* and *Kit* indicated that ordering cells along the first principal component (PC) could be used to arrange cells in pseudotime (Figure S1D). Visualization of cell surface markers measured by index sorting also showed clear dynamics of the different surface markers along PC1 (Figure 2E). As expected, CD34 and c-Kit protein expression showed a negative correlation with pseudotime (compare Figure 1D and 2E), which indicates their downregulation during basophil differentiation. In addition, the basophil marker CD49b (DX5) showed a positive correlation with pseudotime ordering (Figure 2E).

Using the PC1 pseudotime ordering, we then identified genes that dynamically changed during differentiation (Figure 2F). Clustering sorted these dynamic genes into two groups: one increasing and one decreasing with differentiation (Table S2). Basophil differentiation was associated with upregulation of *Hdc*, which is associated with histamine synthesis, and increased expression of the basophil gene E-cadherin (*Cdh1*). We further observed downregulation of the proteases *Mcpt8*, *Prss34* and *Ctsg* and upregulation of the transcription factors *Cebpa*, *Stat5b*, and *Spi1* (Figure 2G). To validate the full lists of dynamically regulated genes, we compared these to mast cell and basophil signature genes identified using bulk microarray analysis.[21] Genes upregulated during basophil differentiation exhibited a significant overlap with the previously described basophil signature genes ($p = 4.0 \times 10^{-29}$, hypergeometric test, Figure S1Ei), whereas genes that were downregulated during differentiation had

286 significant overlap with the previously described mast cell signature genes ($p = 1.3 \times$
287 10^{-4}, hypergeometric test, Figure S1Eii).
288
289 Together, this analysis offers a description of the dynamics of gene expression during
290 basophil differentiation and highlights changes in cell cycle activity as one of the
291 major occurrences during this maturation process.
292
293 *Single-cell gene expression analysis suggests a continuum of mast cell differentiation*
294 *in the peritoneal cavity*
295 After exploring the basophil progenitors, we next decided to focus on mast cell
296 differentiation in the peritoneal cavity. The flow cytometry data suggested the
297 existence of both BMCP-like peritoneal cells and peritoneal mast cells (Figure 1), so
298 we performed single-cell RNA-sequencing on these populations to characterize them
299 based on gene expression. A subset of the BMCP-like cells clustered separately from
300 the mast cells in the diffusion map plot, demonstrating a difference between the
301 transcriptome of these cells and the peritoneal mast cells (Figure 3A). In previous
302 work we characterized bone marrow BMCPs at the single-cell gene expression level.[3]
303 To examine the similarity of these bone marrow progenitors to the peritoneal mast
304 cell differentiation, single-cell bone marrow BMCP profiles from Dahlin et al[3] were
305 projected onto the peritoneal dataset (Figure 3B). This demonstrated that the BMCP-
306 like peritoneal cells furthest from the peritoneal MCs were most similar to the bone
307 marrow BMCPs, supporting that these were the most immature cells in the dataset.
308
309 To understand expression changes during mast cell maturation, we then performed
310 pseudotime ordering of the peritoneal cells using DPT (Figure 3C). As expected,

311 interrogation of cell surface markers along the pseudotime ordering showed a strong
312 downregulation of integrin β7 and strong upregulation of markers such as Sca1 and
313 ST2 (compare Figure 1D and 3D). Genes exhibiting dynamic expression patterns
314 were identified and clustered as for the basophil trajectory (Table S3, Figure 3E).
315 Annotation from the Panther database[22] was used to interrogate the two gene clusters
316 for overlap with specific annotated gene sets such as proteases. Protease genes
317 downregulated during mast cell differentiation included *Mcpt8* and *Gzmb*, whereas
318 *Cpa3*, *Cma1*, *Mcpt1*, *Mcpt4*, *Tpsb2,* and *Tpsab1* increased with differentiation (Figure
319 3F). To investigate the temporal induction and loss of protease genes, we changed
320 visualization method and scaled the gene expression according to the cell with
321 maximum expression (instead of z-scoring genes across the dataset). Plotting the
322 maximum value-scaled gene expression revealed the gene dynamics across the
323 pseudotime trajectory. Early onset proteases included *Cpa3*, followed by *Tpsb2*, and
324 finally *Tpsab1*, indicating that the protease induction occurs in stages (Figure 3G, raw
325 values for individual genes shown in Figure S2C).
326
327 To validate the full lists of dynamically regulated genes in the peritoneal mast cell
328 dataset, we compared these to mast cell and basophil signature identified in Dwyer et
329 al.[21] The upregulated genes significantly overlapped with the mast cell signature
330 genes ($p = 3.7 \times 10^{-65}$, hypergeometric test, Figure S2Di). Upregulated genes included
331 *Ndst2*, *Meis2* and *Hdc* (Figure 3H). Some genes showed expression enrichment
332 mainly in the mast cells (*Meis2*), whereas others were expressed more evenly across
333 the trajectory except for lower expression at the beginning of pseudotime (*Ndst2*).
334 There was also a small overlap between the downregulated genes and basophil
335 signature genes ($p = 2.5 \times 10^{-5}$, hypergeometric test, Figure S2Dii). To investigate the

link between gene and protein expression we also interrogated the expression of *Itga4* and *Itgb7*, which encode subunits of Integrin β7. *Itga4* was significantly downregulated with a similar expression pattern to integrin β7 in the flow cytometry data whereas *Itgb7* was not significantly dynamically changing in pseudotime (Figure 3D, H).

BMCP-like cells in the peritoneal cavity exhibit basophil and mast cell-forming potential

The flow cytometry-based and transcriptional analyses revealed an immature cell population with BMCP-like characteristics in the peritoneal cavity. However, a population of bipotent peritoneal BMCPs has not previously been described at this site. We therefore explored whether the protein and transcriptional analyses successfully predicted the developmental state of the peritoneal BMCP-like cells and mast cells. FACS-isolated BMCP-like cells and mast cells were cytochemically stained with May-Grünwald Giemsa. Primary BMCP-like cells displayed little cytoplasm that contained no or few granules, consistent with the morphology of blasts (Figure 4A). In contrast, primary mast cells were filled with numerous metachromatic granules, in agreement with a mature morphology (Figure 4A).

We cultured the peritoneal cells to investigate whether the BMCP-like cell population exhibited capacity to generate basophils and mast cells. BMCP-like cells cultured with IL-3 and stem cell factor generated c-Kit$^-$ FcεRI$^+$ CD49b$^+$ basophils and c-Kit$^+$ FcεRI$^+$ mast cells, whereas primary mast cells only displayed mast cell-forming capacity (Figure 4B-D).

By contrast to bulk cultured cells, only cell-fate assays performed at the single-cell level have the potential to reveal whether the BMCP-like population consists of bipotent progenitors. Therefore, single BMCP-like cells and mast cells were index sorted into individual wells, the resulting colony sizes were measured, and the colonies were subjected to flow cytometry analysis and cytochemical staining. To visualize the cell culture data, we first performed principal component analysis of the flow cytometry data presented in Figure 1C, henceforth referred to as the reference dataset (Figure 4E). We then projected the FACS index sort data onto the principal component space of the reference dataset, and plotted colony size and colony type data in the same embedding (Figure 4F). Analysis of colony sizes showed that colonies derived from BMCP-like cells were large, whereas cells along the mast cell trajectory exhibited reduced proliferation rate (Figure 4F). Notably, the cell-fate assays revealed that primary BMCP-like cells formed pure basophil colonies, pure mast cell colonies or mixed basophil-mast cell colonies (Figure 4F, S3A). Colonies derived from single mast cells were too small to analyze with flow cytometry. However, mast cells cultured in bulk remained mast cells as expected (Figure 4C-D, S3B-C).

We also cultured the BMCP-like peritoneal cells in erythroid-promoting conditions, as the early basophil-mast cell differentiation is closely linked to the erythrocyte trajectory.[2] However, no erythroid output was observed (Figure S4), indicating that the BMCP-like cells indeed consisted of bipotent basophil-mast cell progenitors. Taken together, the cell culture assays revealed that the protein and gene expression analyses successfully predicted the differentiation state of the BMCP-like cell population. Our study therefore not only identifies a previously unknown bipotent

386 peritoneal progenitor, but also provides comprehensive molecular profiles for this
387 progenitor as well as the mast cell and basophil differentiation trajectories.
388
389 **Discussion**
390 Here, we combine flow cytometry analysis, single-cell transcriptomics, and cell fate
391 assays to chart the basophil and mast cell differentiation trajectories. Multicolor flow
392 cytometry analysis reveals a developmental bifurcation with bipotent BMCPs and
393 their progression into each respective mature lineage. High-coverage single-cell RNA
394 sequencing allows us to generate a molecular map of the cell differentiation, and
395 pseudotime ordering reveals dynamically regulated genes during development. We
396 further demonstrate how flow cytometry and transcriptomics analysis can successfully
397 predict cell-forming potential.
398
399 Single-cell transcriptomics coupled with index sorting of thousands of bone marrow
400 HSPCs has previously been used to chart the erythrocyte and granulocyte-monocyte
401 differentiation.[25,26] BMCPs represent a minor fraction of the bone marrow HSPCs,
402 and capturing the early basophil-mast cell axis therefore requires analysis of tens of
403 thousands of HSPCs.[3] The early differentiation of progenitors with mast cell-forming
404 capacity occurs in the bone marrow.[8] However, full mast cell differentiation and
405 maturation takes place at peripheral sites,[8] and we therefore specifically sorted Lin^- c-
406 Kit^+ $Fc\varepsilon RI^+$ cells extracted from the peritoneal cavity of mice to capture this process.
407 Unlike cell extraction from tissues such as the intestine or skin, the isolation of cells
408 from peritoneum does not require enzymatic digestion, thus minimizing external
409 stimuli during cell processing. Basophil differentiation takes place in bone marrow,
410 and we therefore analyzed basophils and their progenitors from this site. The

411 combined peritoneal and bone marrow datasets provide a high-resolution map
412 covering the BMCP bifurcation and the mast cell and basophil differentiation.
413
414 BMCPs have been described in the spleen and bone marrow,[3,24] and the presence of a
415 bipotent progenitor population indicates that there is a close association between the
416 basophil and mast cell differentiation trajectories. Recent data suggest that the
417 erythroid axis is coupled with the basophil and/or mast cell fates.[2,4,27] However, we
418 did not observe erythrocyte-forming potential among 163 sorted BMCP-like cells in
419 the peritoneum. In agreement with this, BMCPs in the spleen and bone marrow are
420 unable to generate erythrocytes,[3,24] altogether suggesting that loss of erythrocyte-
421 forming potential is a relatively early event along the differentiation trajectory from
422 hematopoietic stem cells to basophil and mast cells. A similar differentiation process
423 has been suggested in human hematopoiesis.[27]
424
425 Temporal ordering of the cells in the transcriptomic datasets allows exploration and
426 verification of molecular processes in differentiating basophils and mast cells. We
427 show that *Ndst2* (encoding *N*-deacetylase/*N*-sulphotransferase-2) is upregulated
428 during differentiation from BMCPs to mature mast cells, and this was also associated
429 with the appearance of numerous densely stained granules. In agreement with these
430 findings, dense May-Grünwald Giemsa staining of the peritoneal mast cell granules
431 requires sulphated heparin, which is dependent on *Ndst2* expression.[28] Histamine is
432 present in basophil and mast cells, and is quickly released upon cell activation. This
433 potent mediator causes allergic reactions, and we show that the expression of the
434 enzyme that catalyzes the histamine synthesis, histidine decarboxylase (*Hdc*),
435 increased upon differentiation of both basophils and mast cells. Analysis of the single-

436 cell transcriptomics data can also give insights into more complex regulatory
437 processes. For example, autologous expression of integrin β7 is important for mast
438 cell progenitor migration into tissues, such as lung and intestine.[29,30] Downregulation
439 of integrin β7 on the cell surface is a hallmark of terminal mast cell differentiation,[31]
440 which we confirmed with flow cytometry analysis. However, we did not observe
441 downregulation of the *Itgb7* gene expression during the transition from BMCPs to
442 mast cells. Notably, integrins constitute αβ heterodimers when localized to the cell
443 surface, and further investigation into the gene expression profile revealed decreased
444 expression of *Itga4*, the binding partner of the integrin β7 subunit, upon
445 differentiation. Thus, the loss of integrin α4 gene expression likely explains the loss
446 of integrin β7 protein expression on the cell surface during the BMCP to mast cell
447 transition.
448
449 During basophil differentiation, the transcription factors *Stat5b* and *Cebpa* are
450 upregulated along the progression of pseudotime. The expression of C/EBPα is
451 STAT5-dependent, and both genes are required for basophil formation.[24,32] Dynamic
452 expression of transcription factors with currently unknown functions in basophil and
453 mast cell differentiation was also recognized. For example, *Spi1*, which encodes
454 PU.1, is upregulated during late basophil differentiation. It is known to be involved in
455 neutrophil granulocyte maturation,[33,34] but the role of PU.1 in basophil differentiation
456 is yet to be delineated. During mast cell differentiation, we describe the increase of
457 the transcription factor *Meis2*. Primary mast cells from human skin express this
458 transcription factor,[35] but the potential function during mast cell differentiation is yet
459 to be described. Thus, the datasets can be explored to identify previously
460 unrecognized genes that may regulate basophil and mast cell differentiation.

Microarray analyses reported previously provide detailed gene expression patterns of mature hematopoietic cell populations, including bulk-sorted mature basophils and mast cells.[21] We observed that differentiation into basophils and mast cells involves activation of mutually exclusive lineage programs. However, a small subset of the previously reported signature genes is not unique to mature cells, but can also be observed in bipotent progenitors. For example, we show that *Mcpt8* expression is not restricted to basophils but is also expressed by BMCPs. This in fact provides an explanation to a major conundrum in the field. Basophils, identified as *Mcpt8*-expressing cells, have been reported to exhibit potential to transdifferentiate into mast cells.[36] Our results show that a more likely scenario is that a subset of the previously reported *Mcpt8*-expressing cells constitutes bipotent BMCPs that can give rise to mast cells.

Dimensionality reduction approaches are commonly applied to visualize single-cell transcriptomics data. Here, we show that the diffusion map embedding and PCA visualization successfully separate the basophil and mast cell differentiation trajectories based on the multidimensional flow cytometry data. We took advantage of the flow cytometry-based PCA visualization to interrogate single-cell fate assay data. The visualization showed that the in vitro proliferative capacity of the index-sorted peritoneal cells is quickly reduced as BMCP-like cells differentiate and enter the mast cell trajectory. In addition, only the most immature BMCP-like cells exhibit capacity to form basophil, mast cell, and mixed basophil-mast cell colonies. Taken together, dimensionality reduction techniques of flow cytometry data combined with cell fate assays provide new insights into basophil and mast cell differentiation.

In summary, here we have reported the generation of transcriptomic and flow cytometry data capturing the progression from bipotent progenitors toward basophil or mast cells. Our resource provides a detailed description of the expression changes occurring during this differentiation process at the single-cell level. A user-friendly interactive website has been created for the wider community to enable further exploration of the data.

Authorship contributions

J.S.D. and W.W.Y.L. performed experiments; E.D. mapped sequencing data; F.K.H., J.S.D., and X.W. analyzed single-cell RNA sequencing data; J.S.D. analyzed flow cytometry and cell culture experiments; N.K.W. contributed to important discussions; I.K. created the web resource; B.G. and J.S.D. supervised the study; B.G. secured funding; F.K.H. and J.S.D drafted the manuscript; and all authors contributed to final version of the manuscript.

Acknowledgements

We thank Chiara Cossetti, Gabriela Grondys-Kotarba, and Reiner Schulte at the Cambridge Institute for Medical Research Flow Cytometry Core for their assistance with cell sorting. J.S.D. is supported by funding from the Swedish Research Council, the Swedish Cancer Society, and Karolinska Institutet. Research in B.G.'s laboratory is supported by Bloodwise, Wellcome, CRUK, MRC and by core funding from Wellcome and MRC to the Wellcome-MRC Cambridge Stem Cell Institute. F.K.H. is funded by a MRC Physical Biology of Stem Cells PhD studentship and by part of a

510 Wellcome Strategic Award (105031/D/14/Z) awarded to W. Reik, S. Teichmann, J.

511 Nichols, B.D. Simons, T. Voet, S. Srinivas, L. Vallier, B.G. and J.C. Marioni.

512

513

514 **References**

515

516 1. Akashi K, Traver D, Miyamoto T, Weissman IL. A clonogenic common
517 myeloid progenitor that gives rise to all myeloid lineages. *Nature*. 2000;404:193-197.
518 2. Tusi BK, Wolock SL, Weinreb C, Hwang Y, Hidalgo D, Zilionis R, Waisman
519 A, Huh JR, Klein AM, Socolovsky M. Population snapshots predict early
520 haematopoietic and erythroid hierarchies. *Nature*. 2018;555:54-60.
521 3. Dahlin JS, Hamey FK, Pijuan-Sala B, Shepherd M, Lau WWY, Nestorowa S,
522 Weinreb C, Wolock S, Hannah R, Diamanti E, Kent DG, Gottgens B, Wilson NK. A
523 single-cell hematopoietic landscape resolves 8 lineage trajectories and defects in Kit
524 mutant mice. *Blood*. 2018;131:e1-e11.
525 4. Drissen R, Thongjuea S, Theilgaard-Monch K, Nerlov C. Identification of two
526 distinct pathways of human myelopoiesis. *Sci Immunol*. 2019;4.
527 5. Pellin D, Loperfido M, Baricordi C, Wolock SL, Montepeloso A, Weinberg
528 OK, Biffi A, Klein AM, Biasco L. A comprehensive single cell transcriptional
529 landscape of human hematopoietic progenitors. *Nat Commun*. 2019;10:2395.
530 6. Velten L, Haas SF, Raffel S, Blaszkiewicz S, Islam S, Hennig BP, Hirche C,
531 Lutz C, Buss EC, Nowak D, Boch T, Hofmann WK, Ho AD, Huber W, Trumpp A,
532 Essers MA, Steinmetz LM. Human haematopoietic stem cell lineage commitment is a
533 continuous process. *Nat Cell Biol*. 2017;19:271-281.
534 7. Nathan C. Points of control in inflammation. *Nature*. 2002;420:846-852.
535 8. Grootens J, Ungerstedt JS, Nilsson G, Dahlin JS. Deciphering the
536 differentiation trajectory from hematopoietic stem cells to mast cells. *Blood Adv*.
537 2018;2:2273-2281.

538 9. Picelli S, Faridani OR, Bjorklund AK, Winberg G, Sagasser S, Sandberg R.
539 Full-length RNA-seq from single cells using Smart-seq2. *Nat Protoc*. 2014;9:171-
540 181.
541 10. Wu TD, Nacu S. Fast and SNP-tolerant detection of complex variants and
542 splicing in short reads. *Bioinformatics*. 2010;26:873-881.
543 11. Zerbino DR, Achuthan P, Akanni W, Amode MR, Barrell D, Bhai J, Billis K,
544 Cummins C, Gall A, Giron CG, Gil L, Gordon L, Haggerty L, Haskell E, Hourlier T,
545 Izuogu OG, Janacek SH, Juettemann T, To JK, Laird MR, Lavidas I, Liu Z, Loveland
546 JE, Maurel T, McLaren W, Moore B, Mudge J, Murphy DN, Newman V, Nuhn M,
547 Ogeh D, Ong CK, Parker A, Patricio M, Riat HS, Schuilenburg H, Sheppard D,
548 Sparrow H, Taylor K, Thormann A, Vullo A, Walts B, Zadissa A, Frankish A, Hunt
549 SE, Kostadima M, Langridge N, Martin FJ, Muffato M, Perry E, Ruffier M, Staines
550 DM, Trevanion SJ, Aken BL, Cunningham F, Yates A, Flicek P. Ensembl 2018.
551 *Nucleic Acids Res*. 2018;46:D754-D761.
552 12. Anders S, Pyl PT, Huber W. HTSeq--a Python framework to work with high-
553 throughput sequencing data. *Bioinformatics*. 2015;31:166-169.
554 13. Lun AT, Bach K, Marioni JC. Pooling across cells to normalize single-cell
555 RNA sequencing data with many zero counts. *Genome Biol*. 2016;17:75.
556 14. Brennecke P, Anders S, Kim JK, Kolodziejczyk AA, Zhang X, Proserpio V,
557 Baying B, Benes V, Teichmann SA, Marioni JC, Heisler MG. Accounting for
558 technical noise in single-cell RNA-seq experiments. *Nat Methods*. 2013;10:1093-
559 1095.
560 15. Wolf FA, Angerer P, Theis FJ. SCANPY: large-scale single-cell gene
561 expression data analysis. *Genome Biol*. 2018;19:15.

562 16. Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles GV, Clark NR,
563 Ma'ayan A. Enrichr: interactive and collaborative HTML5 gene list enrichment
564 analysis tool. *BMC Bioinformatics*. 2013;14:128.
565 17. Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z,
566 Koplev S, Jenkins SL, Jagodnik KM, Lachmann A, McDermott MG, Monteiro CD,
567 Gundersen GW, Ma'ayan A. Enrichr: a comprehensive gene set enrichment analysis
568 web server 2016 update. *Nucleic Acids Res*. 2016;44:W90-97.
569 18. Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, Tirosh I,
570 Bialas AR, Kamitaki N, Martersteck EM, Trombetta JJ, Weitz DA, Sanes JR, Shalek
571 AK, Regev A, McCarroll SA. Highly Parallel Genome-wide Expression Profiling of
572 Individual Cells Using Nanoliter Droplets. *Cell*. 2015;161:1202-1214.
573 19. Haghverdi L, Buttner M, Wolf FA, Buettner F, Theis FJ. Diffusion
574 pseudotime robustly reconstructs lineage branching. *Nat Methods*. 2016;13:845-848.
575 20. Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E. Fast unfolding of
576 communities in large networks. *Journal of Statistical Mechanics: Theory and*
577 *Experiment*. 2008;2008.
578 21. Dwyer DF, Barrett NA, Austen KF, Immunological Genome Project C.
579 Expression profiling of constitutive mast cells reveals a unique identity within the
580 immune system. *Nat Immunol*. 2016;17:878-887.
581 22. Mi H, Muruganujan A, Ebert D, Huang X, Thomas PD. PANTHER version
582 14: more genomes, a new PANTHER GO-slim and improvements in enrichment
583 analysis tools. *Nucleic Acids Res*. 2019;47:D419-D426.
584 23. Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, Lennon
585 NJ, Livak KJ, Mikkelsen TS, Rinn JL. The dynamics and regulators of cell fate

586 decisions are revealed by pseudotemporal ordering of single cells. *Nat Biotechnol*.
587 2014;32:381-386.
588 24. Arinobu Y, Iwasaki H, Gurish MF, Mizuno S, Shigematsu H, Ozawa H,
589 Tenen DG, Austen KF, Akashi K. Developmental checkpoints of the basophil/mast
590 cell lineages in adult murine hematopoiesis. *Proc Natl Acad Sci U S A*.
591 2005;102:18105-18110.
592 25. Nestorowa S, Hamey FK, Pijuan Sala B, Diamanti E, Shepherd M, Laurenti E,
593 Wilson NK, Kent DG, Gottgens B. A single-cell resolution map of mouse
594 hematopoietic stem and progenitor cell differentiation. *Blood*. 2016;128:e20-31.
595 26. Paul F, Arkin Y, Giladi A, Jaitin DA, Kenigsberg E, Keren-Shaul H, Winter
596 D, Lara-Astiaso D, Gury M, Weiner A, David E, Cohen N, Lauridsen FK, Haas S,
597 Schlitzer A, Mildner A, Ginhoux F, Jung S, Trumpp A, Porse BT, Tanay A, Amit I.
598 Transcriptional Heterogeneity and Lineage Commitment in Myeloid Progenitors.
599 *Cell*. 2015;163:1663-1677.
600 27. Grootens J, Ungerstedt JS, Wu C, Hamberg Levedahl K, Nilsson G, Dahlin
601 JS. CD203c distinguishes the erythroid and mast cell-basophil differentiation
602 trajectories among human FcepsilonRI(+) bone marrow progenitors. *Allergy*. 2019:E-
603 pub ahead of print.
604 28. Forsberg E, Pejler G, Ringvall M, Lunderius C, Tomasini-Johansson B,
605 Kusche-Gullberg M, Eriksson I, Ledin J, Hellman L, Kjellen L. Abnormal mast cells
606 in mice deficient in a heparin-synthesizing enzyme. *Nature*. 1999;400:773-776.
607 29. Gurish MF, Tao H, Abonia JP, Arya A, Friend DS, Parker CM, Austen KF.
608 Intestinal mast cell progenitors require CD49dbeta7 (alpha4beta7 integrin) for tissue-
609 specific homing. *J Exp Med*. 2001;194:1243-1252.

610 30. Abonia JP, Hallgren J, Jones T, Shi T, Xu Y, Koni P, Flavell RA, Boyce JA,
611 Austen KF, Gurish MF. Alpha-4 integrins and VCAM-1, but not MAdCAM-1, are
612 essential for recruitment of mast cell progenitors to the inflamed lung. *Blood*.
613 2006;108:1588-1594.
614 31. Dahlin JS, Ding Z, Hallgren J. Distinguishing Mast Cell Progenitors from
615 Mature Mast Cells in Mice. *Stem Cells Dev*. 2015;24:1703-1711.
616 32. Qi X, Hong J, Chaves L, Zhuang Y, Chen Y, Wang D, Chabon J, Graham B,
617 Ohmori K, Li Y, Huang H. Antagonistic regulation by the transcription factors
618 C/EBPalpha and MITF specifies basophil and mast cell fates. *Immunity*. 2013;39:97-
619 110.
620 33. Scott EW, Simon MC, Anastasi J, Singh H. Requirement of transcription
621 factor PU.1 in the development of multiple hematopoietic lineages. *Science*.
622 1994;265:1573-1577.
623 34. Anderson KL, Smith KA, Pio F, Torbett BE, Maki RA. Neutrophils deficient
624 in PU.1 do not terminally differentiate or become functionally competent. *Blood*.
625 1998;92:1576-1585.
626 35. Motakis E, Guhl S, Ishizu Y, Itoh M, Kawaji H, de Hoon M, Lassmann T,
627 Carninci P, Hayashizaki Y, Zuberbier T, Forrest AR, Babina M, consortium F.
628 Redefinition of the human mast cell transcriptome by deep-CAGE sequencing. *Blood*.
629 2014;123:e58-67.
630 36. Metcalf D, Ng AP, Baldwin TM, Di Rago L, Mifsud S. Concordant mast cell
631 and basophil production by individual hematopoietic blast colony-forming cells. *Proc*
632 *Natl Acad Sci U S A*. 2013;110:9031-9035.
633
634

635

636

637

638 **Figure legends**

639 **Figure 1. Flow cytometry analysis reveals differentiation trajectories from**
640 **bipotent basophil-mast cell progenitors to basophils and mast cells.**

641 (A) Illustration outlining the basophil and mast cell differentiation trajectories. (B)
642 Flow cytometry-based gating strategies of (Bi) bipotent basophil-mast cell progenitors
643 (BMCPs) from bone marrow, (Bii) basophil progenitors (BaP) and basophils (Ba)
644 from bone marrow, and (Biii) BMCP-like cells and mast cells from peritoneal cavity.
645 (C) Diffusion map visualization of the flow cytometry data colored by cell type. (D)
646 Diffusion map visualization of the flow cytometry data colored by protein expression
647 or light scatter parameters. The surface expression parameters and light scatter
648 parameters are visualized on log-transformed and linear scales, respectively.
649 Expression of lineage markers and viability staining are not shown. The data are
650 representative of 4 independent experiments.

651

652 **Figure 2. Bone marrow basophil progenitors downregulate cell cycle genes**
653 **during differentiation.** (A) PCA of scRNA-seq profiles colored by cell surface
654 marker phenotype. PC, principal component. (B) Top 5 GO Biological Process terms
655 associated with the genes significantly upregulated in BaP cells compared to Ba cells,
656 ranked by adjusted p-value. Benjamini-Hochberg correction for multiple hypotheses
657 testing. Genes upregulated in Ba compared to BaP are presented in Figure S1B. (C)
658 Proportion of scRNA-seq profiles from each phenotype computationally assigned to
659 G1, S or G2M cell cycle states based on gene expression. (D) PCA colored by cell
660 cycle state. (E) Levels of cell surface markers for cells ordered by PC1 pseudotime.
661 Index data values were log-transformed, smoothed along pseudotime by using a
662 sliding window of size 20 and scaled between 0 and 1 for each marker. Correlation

values indicate the pearson correlation coefficient between pseudotime and the unsmoothed expression values for each surface marker. Colorbar at the top indicates the phenotypic cell type proportions within each window. Blue corresponds to entirely BaPs and orange to Ba cells. (F) Heatmap displaying the expression of genes dynamically expressed along the PC1 pseudotime ordering. The top colorbar indicates the cell type proportion in each window. Expression is smoothed along a sliding window and z-scored for each gene, and genes were clustered using Louvain clustering into groups showing different dynamics. (G) PCA colored by z-scored expression of specific genes. The data represents cells pooled from 4 individual mice.

Figure 3. Transcriptional profiling of peritoneal mast cell progenitors captures a differentiation continuum. (A) Diffusion map dimensionality reduction of scRNA-seq profiles colored by cell phenotype. DC, diffusion component. (B) Projection of bone marrow BMCP progenitor scRNA-seq profiles from Dahlin et al (2018)[3] to their most similar expression profiles from the peritoneal dataset. Each peritoneal cell is colored by its similarity to the projected bone marrow cells, see methods for details. (C) Diffusion map colored by pseudotime ordering of cells. DPT, diffusion pseudotime. (D) Levels of cell surface markers for pseudotime ordered cells. Index data values were log-transformed, smoothed along pseudotime by using a sliding window of size 20 and scaled between 0 and 1 for each marker. Correlation values indicate the pearson correlation coefficient between pseudotime and the unsmoothed expression values for each surface marker. Colorbar at the top indicates the phenotypic cell type proportions within each window. Green corresponds to entirely BMCP-like cells and purple to MCs. (E) Heatmap displaying the expression of genes dynamically expressed along the pseudotime ordering. The top colorbar indicates the

688 proportion of cell type in each window. Expression is smoothed along a sliding
689 window and z-scored for each gene, and genes were clustered using Louvain
690 clustering into groups showing different dynamics. (F) Heatmap of dynamically
691 regulated proteases showing z-scored gene expression along pseudotime. Genes were
692 ordered using the hierarchical clustering indicated by the dendrogram. Colorbar
693 indicates the Louvain cluster from (E) for each gene. (G) Expression trends of
694 specific genes along pseudotime. Genes are scaled by their maximum expression
695 value rather than z-scoring as in the heatmap. (H) Diffusion map colored by z-score
696 scaled expression of specific genes. The data represents cells pooled from 3 individual
697 mice.
698
699 **Figure 4. BMCP-like peritoneal cells exhibit potential to form basophils and**
700 **mast cells.** (A-B) May-Grünwald Giemsa staining of primary and in vitro cultured
701 BMCP-like cells and mast cells extracted from the peritoneal cavity. Ba, basophil;
702 MC, mast cell. Two or seven independent experiments revealed the morphology of
703 primary BMCP-like cells and mast cells, respectively. (C) Flow cytometry gating
704 strategy to identify basophils and mast cells cultured from primary BMCP-like cells
705 and mast cells. (D) Quantification of cell type output following bulk-culture and flow
706 cytometry analysis of BMCP-like cells and mast cells. Pooled data from 4
707 independent experiments per population are shown. The means and SEMs are shown.
708 Unpaired two-tailed Student *t*-tests; ****$P<0.0001$. (E) Principal component analysis
709 of the flow cytometry reference dataset, provided in Figure 1C, colored by cell type.
710 (F) Projection of index-sorted cells into the principal component space of the
711 reference dataset. The point size represents $\log_{10}$-transformed colony size and the

colors represent colony type following cell culture. Panel F shows data pooled from 2 independent experiments. The cells were cultured with IL-3 and stem cell factor.

Supplementary Figure legends

Figure S1

(A) Diffusion map of Ba and BaP cells colored by their phenotypic cell type. DC, diffusion component. (B) Top 5 GO Biological Process terms associated with the genes significantly upregulated in Ba cells compared to BaP cells, ranked by adjusted p-value. Benjamini-Hochberg correction for multiple hypotheses testing. (C) Diffusion map colored by computationally assigned cell cycle state. (D) Diffusion map colored by expression of specific genes. (E) Overlap of basophil (Ba) differentiation up- (i) or down- (ii) regulated genes with mast cell and basophil signature gene sets from Dwyer et al. Significance of overlap was tested using a hypergeometric test, with resulting p-values displayed in figure. The Venn diagrams show genes annotated in Ensembl genome build 81 and Dwyer et al.

Figure S2

(A) PCA of peritoneal single-cell RNA-seq profiles showing outlier cells in PC1. PC, principal component. (B) Diffusion map dimensionality reduction colored by computationally assigned cell cycle state. DC, diffusion component. (C) Expression trends of specific genes along pseudotime. Splines were fitted using the monocle R package function. (D) Overlap of mast cell (MC) differentiation up- (i) or down- (ii) regulated genes with mast cell and basophil signature gene sets from Dwyer et al.

737 Significance of overlap was tested using a hypergeometric test, with resulting p-
738 values displayed in figure. The Venn diagrams show genes annotated in Ensembl
739 genome build 81 and Dwyer et al.

740

741 Figure S3.

742 (A) Flow cytometry assessment of colonies derived from single BMCP-like peritoneal
743 cells. The same colonies were stained with May-Grünwald Giemsa. (B) Cultured
744 bulk-sorted peritoneal mast cells provided as reference to panel A. (C) Single-cells
745 were sorted into individual wells and the colony size was determined after 7 days in
746 culture. The numbers above each group represent the number of wells analyzed. Each
747 dot represents one well. Wells in which no viable cells were found were scored as 1.
748 The red lines represent geometric means. The data in panel C are pooled from 2
749 independent experiments. Two-tailed Mann-Whitney test; ****P<0.0001. The cells
750 were cultured with IL-3 and stem cell factor.

751

752 Figure S4.

753 (A) Flow cytometry gating strategy for determining colony type following culture for
754 6 days in erythrocyte-promoting conditions. E, erythroid; MC, mast cell; Ba, basophil.
755 Culture of bulk-sorted bone marrow (BM) Lin^- $c\text{-}Kit^+$ cells served as positive control
756 for erythroid-forming potential. (B) Colony type output of single index-sorted BMCP-
757 like cells projected into the principal component space of the reference dataset. The
758 point size represents $\log_{10}$-transformed colony size. (C) Single-cells were sorted into
759 individual wells and the colony size was determined after 6 days in culture. The
760 number represents the number of wells analyzed. Each dot represents one well. Wells
761 in which no viable cells were found were scored as 1. The red line represents

762 geometric mean. The data in panels B and C are pooled from 2 independent
763 experiments.
764

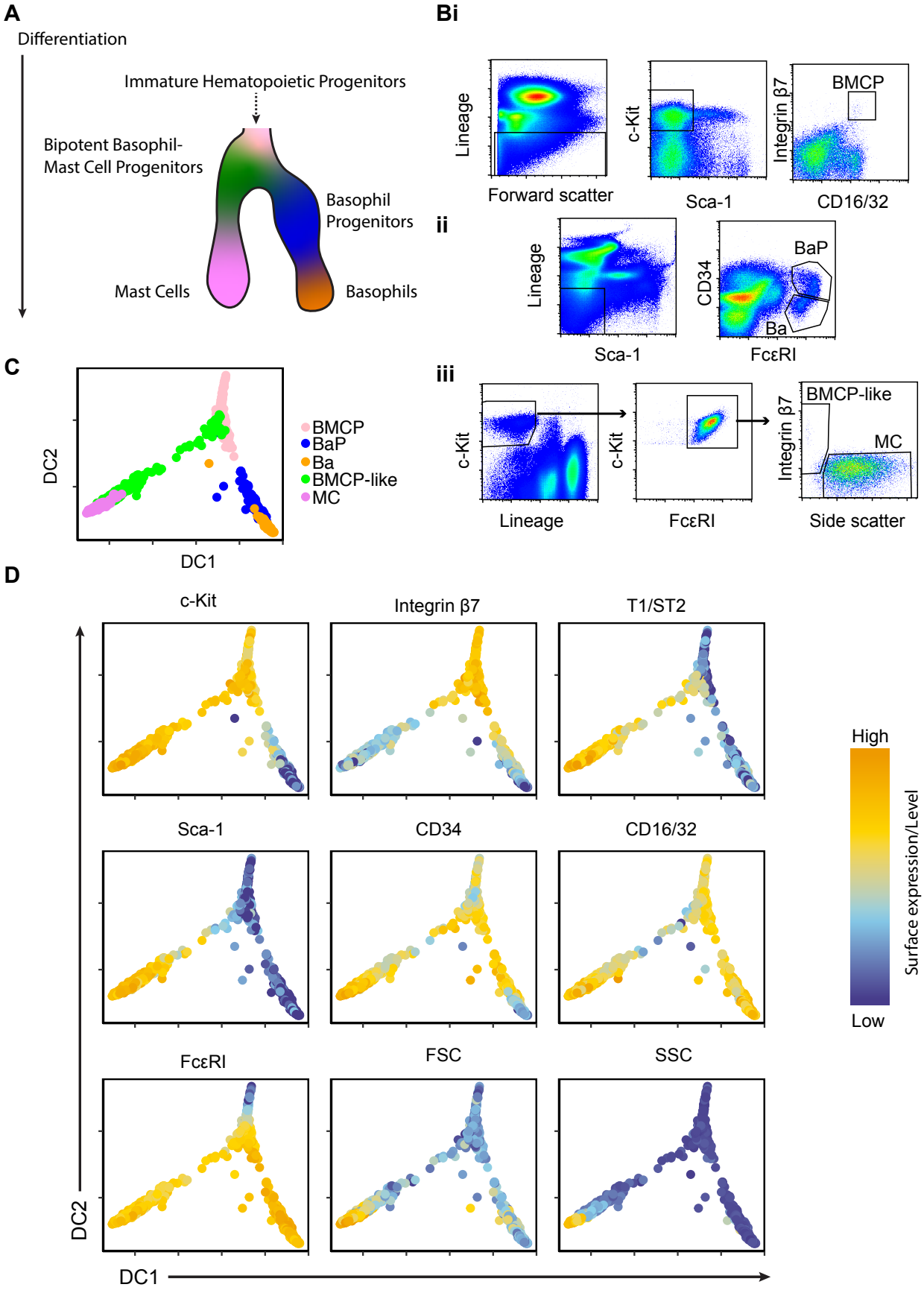


Figure 1

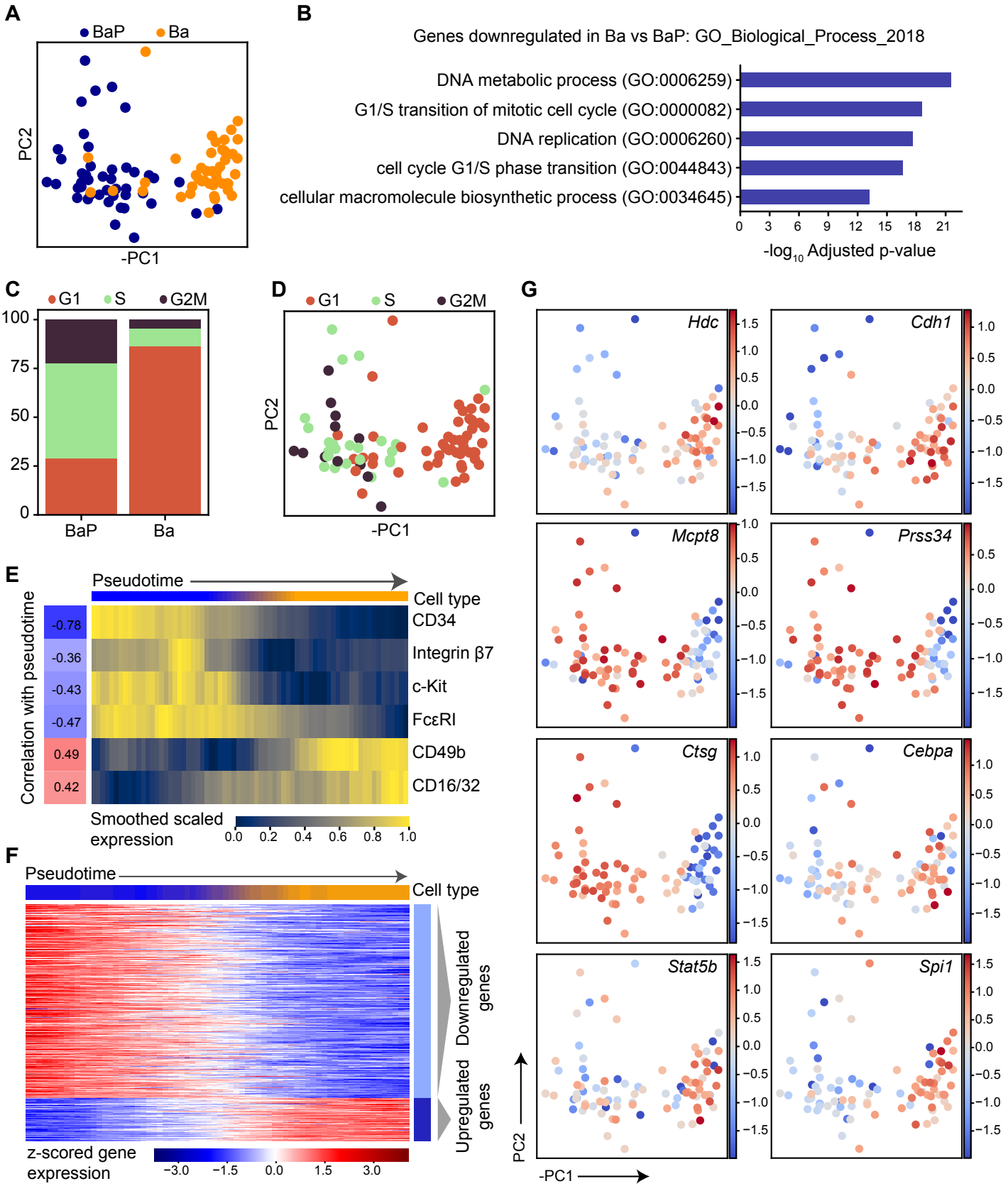


Figure 2

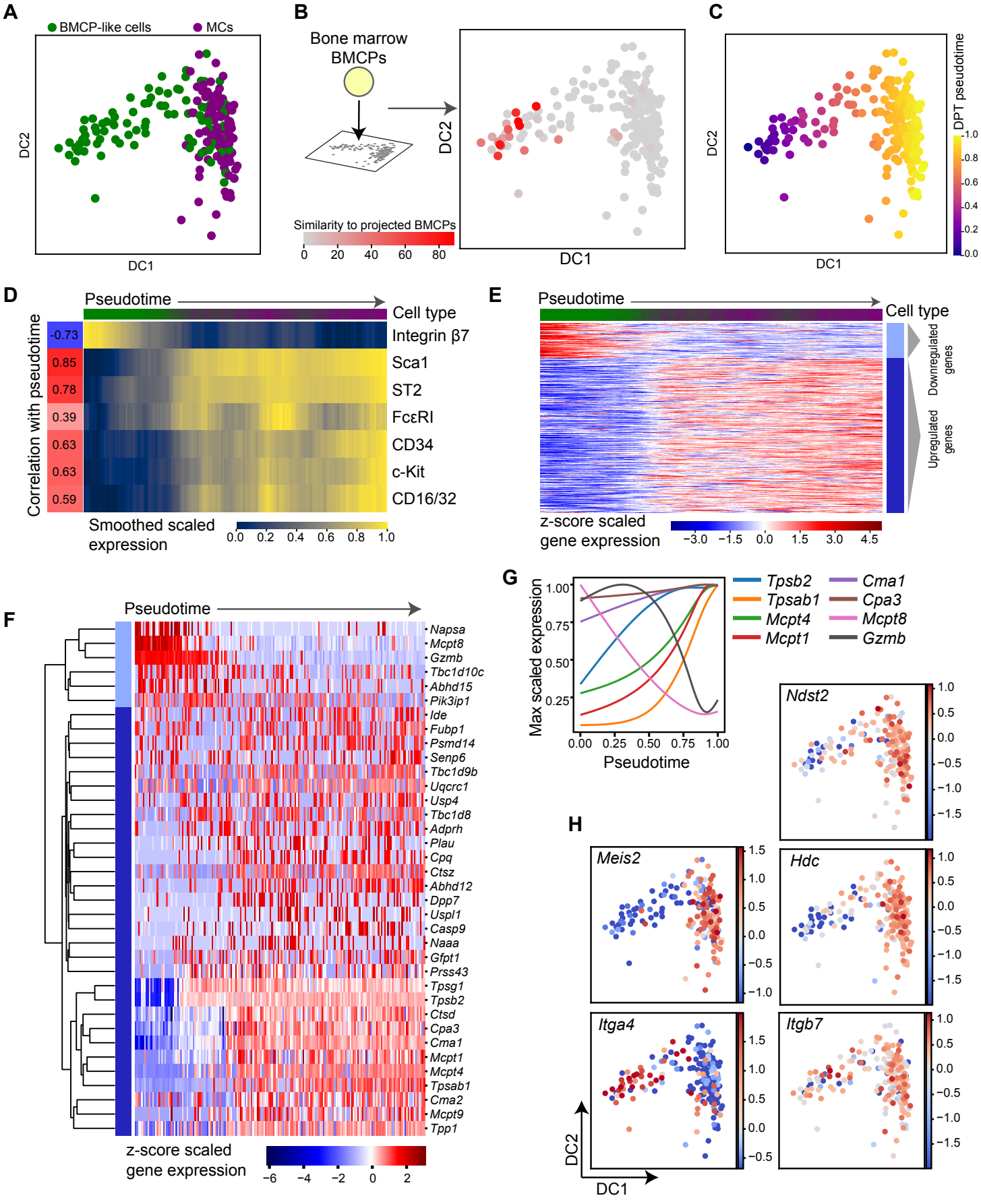


Figure 3

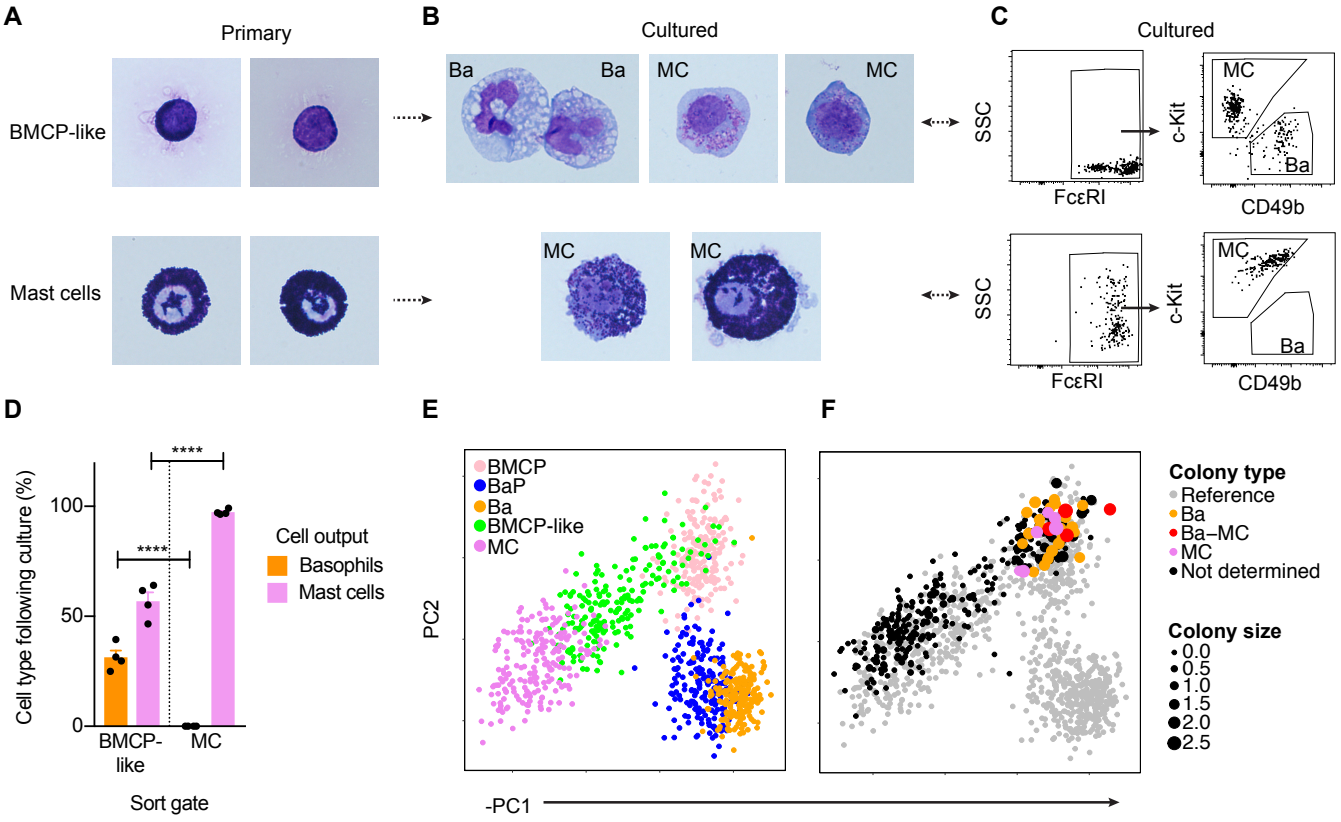


Figure 4

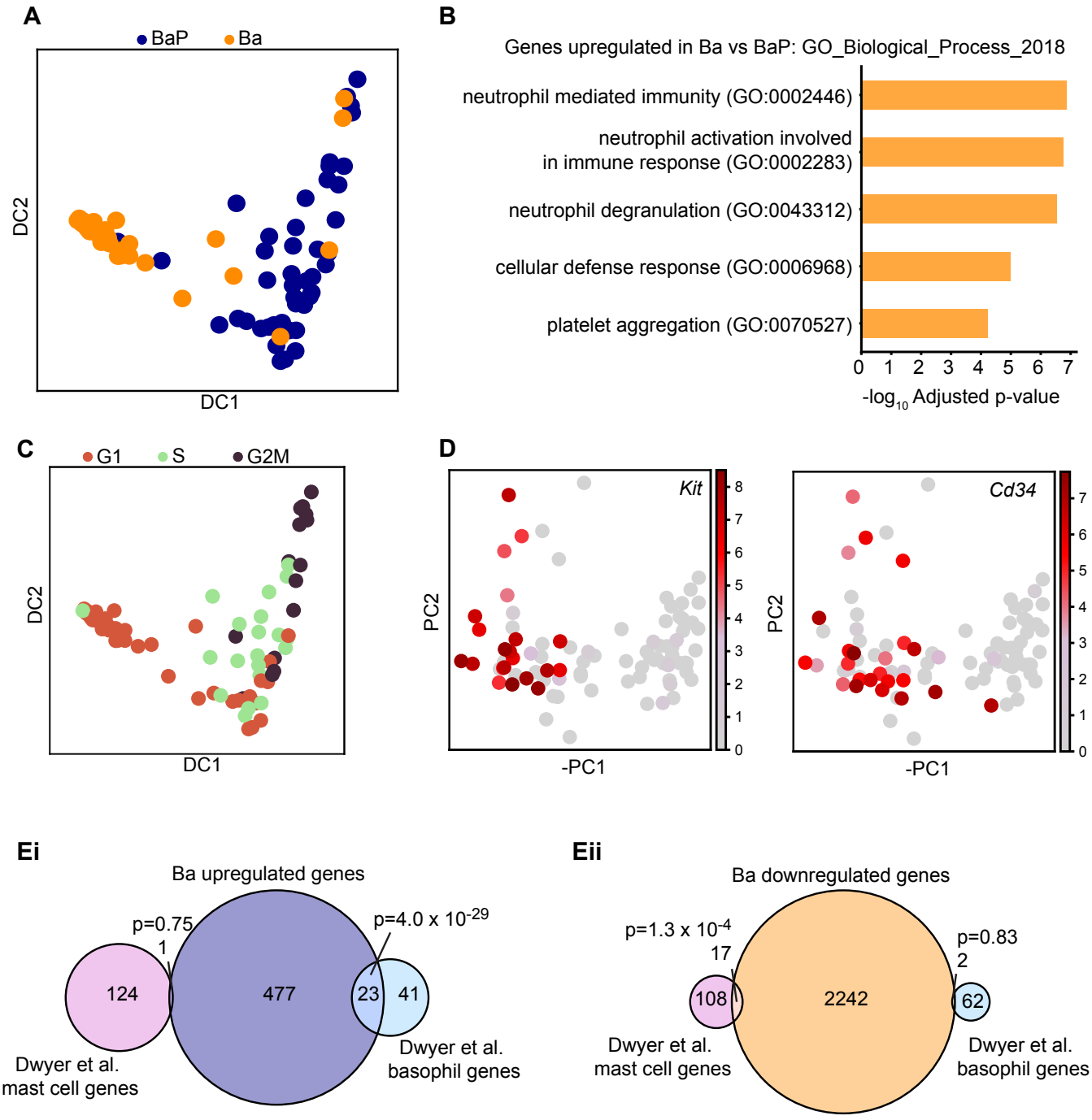
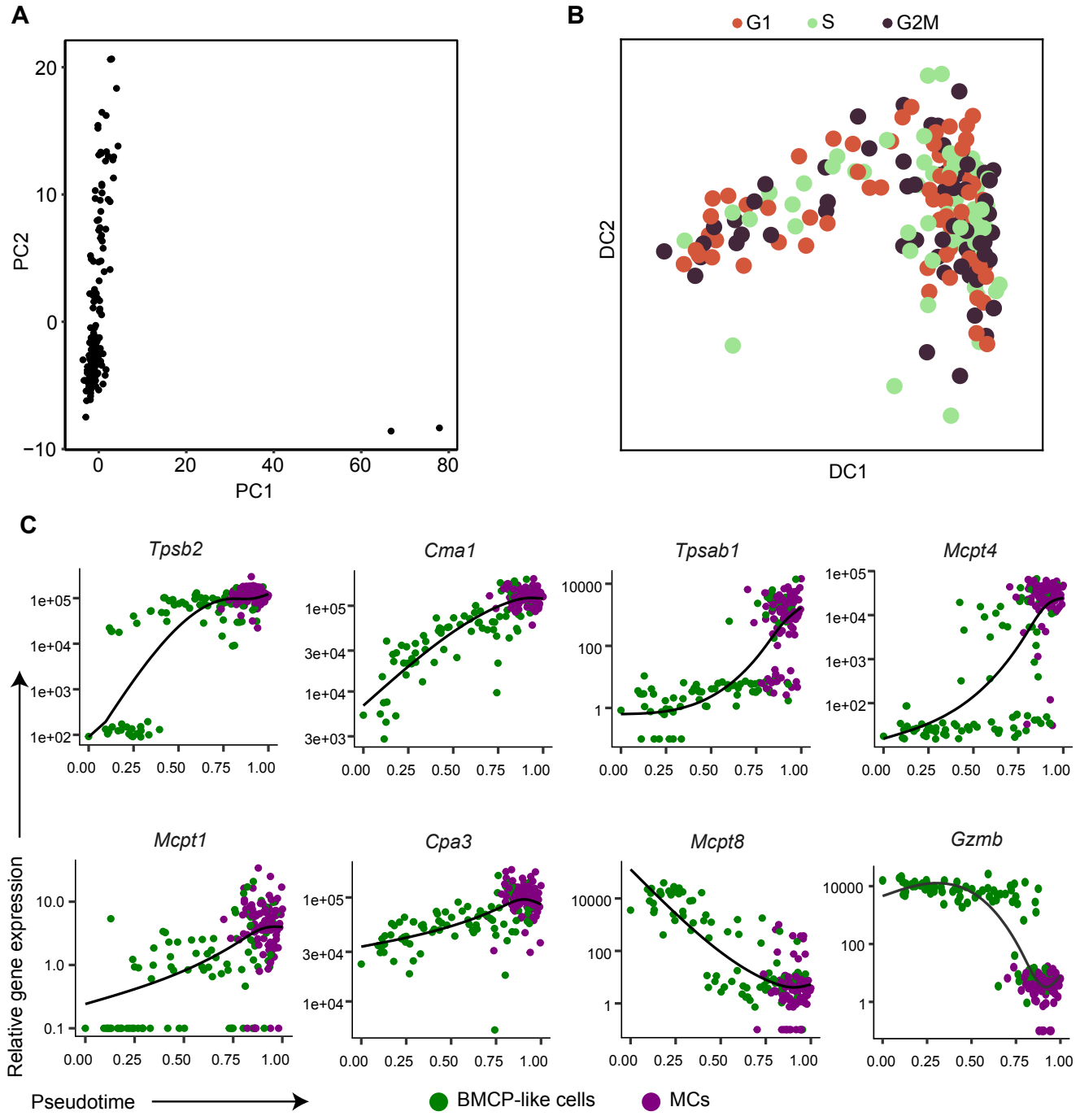


Figure S1



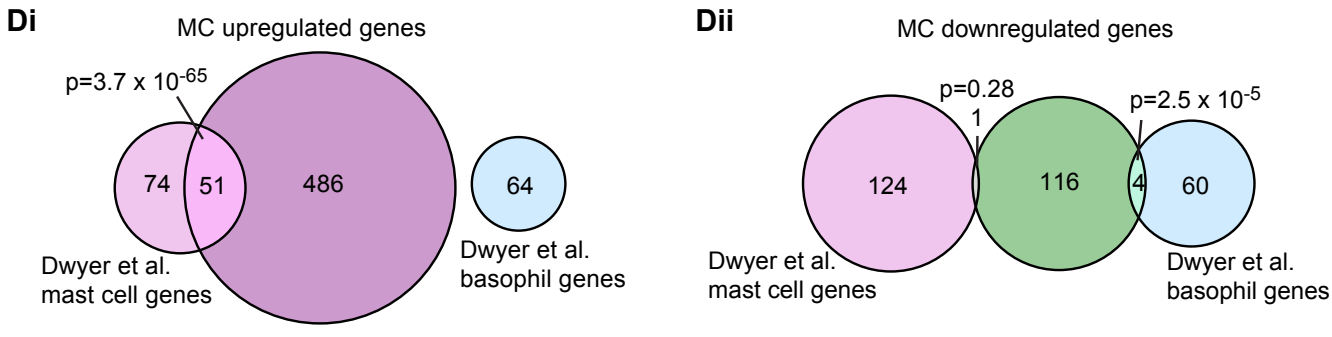


Figure S2

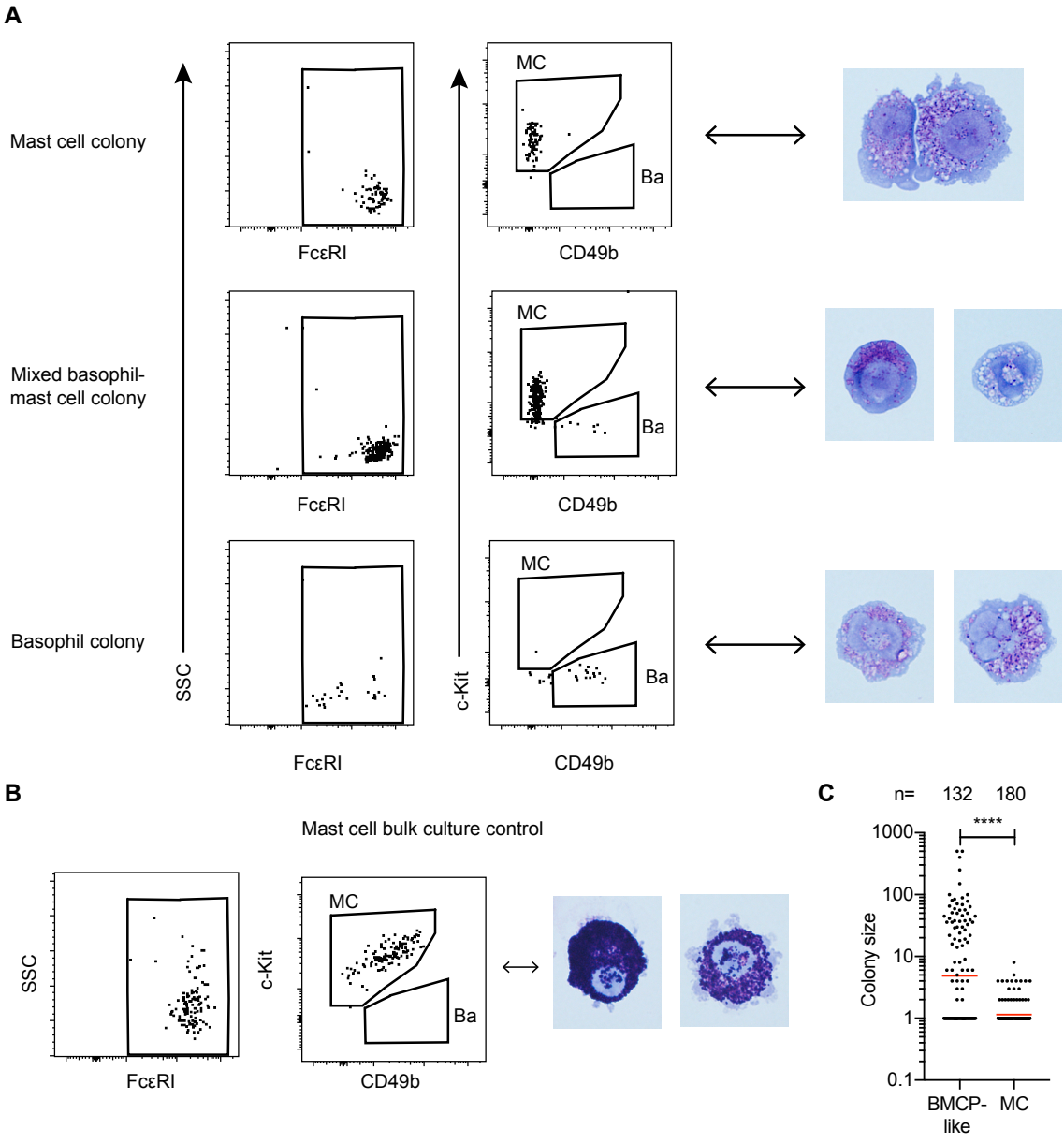


Figure S3

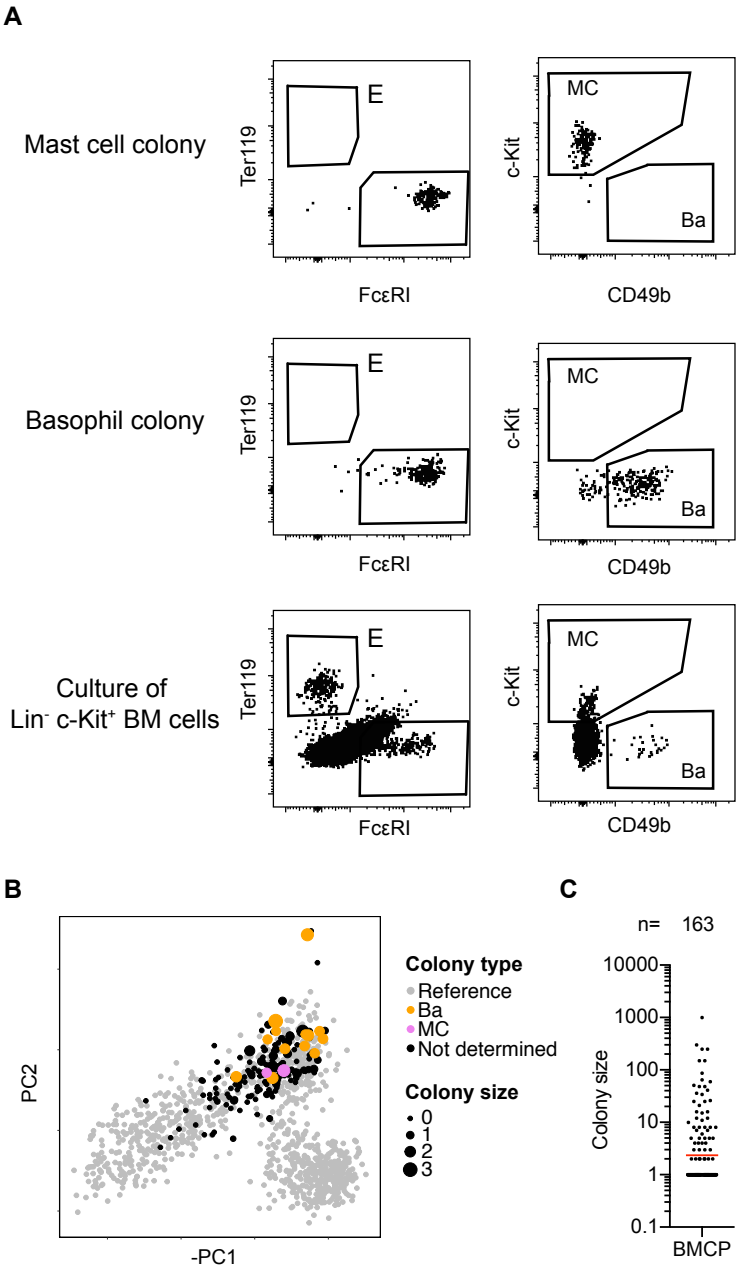


Figure S4