1 **Single-cell molecular profiling provides a high-resolution map of basophil and**
2 **mast cell development**
3
4 **Short title: Roadmap of basophil and mast cell differentiation**
5
6 Fiona K. Hamey[1,2], Winnie W.Y. Lau[1], Iwo Kucinski[1], Xiaonan Wang[1], Evangelia
7 Diamanti[1], Nicola K. Wilson[1], Berthold Göttgens[1*], Joakim S. Dahlin[1,3*]
8
9
10 1 Department of Haematology, University of Cambridge, Wellcome–Medical Research
11 Council Cambridge Stem Cell Institute, Cambridge, United Kingdom.
12 2 Present address: JDRF/Wellcome Diabetes and Inflammation Laboratory, Wellcome
13 Centre for Human Genetics, University of Oxford, Oxford, UK.
14 3 Department of Medicine, Karolinska Institutet and Karolinska University Hospital,
15 Stockholm, Sweden.
16
17 *: Corresponding Authors
18 B. Göttgens; Address: University of Cambridge, Department of Haematology,
19 Wellcome–Medical Research Council Cambridge Stem Cell Institute, Jeffrey Cheah
20 Biomedical Centre, Cambridge Biomedical Campus, Cambridge CB2 0AW, UK; E-
21 mail: bg200@cam.ac.uk; Tel. +44-1223-336829; FAX +44-1223-762670
22
23 J. Dahlin; Address: NKS BioClinicum J7:30, Visionsgatan 4, 171 76 Solna, Sweden;
24 E-mail: joakim.dahlin@ki.se; Tel. +46-8-517 761 69
25
26

Title: 100 characters (max 100)

Abstract word count: 169 (max 250)

No of figures: 4

No of tables: 0

Word count: 4008

Reference count: 37

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 and analyzed the Human Cell Atlas data; 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. We also thank the Human Cell Atlas for making the human single-cell transcriptomic data available. J.S.D. is supported by funding from the Swedish Research Council (2015-06322 and 2018-02070), the Swedish Cancer Society, the Åke Wiberg Foundation, the Magnus Bergvall Foundation, the Lars Hierta Memorial Foundation, and Karolinska Institutet. Research in B.G.'s laboratory is supported by Blood Cancer UK (18002), Wellcome (206328/Z/17/Z), CRUK (C1163/A21762), NIH-NIDDK (1 R24 DK106766), MRC (MR/S036113/1) and by core funding from

52 Wellcome and MRC to the Wellcome-MRC Cambridge Stem Cell Institute
53 (203151/Z/16/Z). F.K.H. is funded by a MRC Physical Biology of Stem Cells PhD
54 studentship (MR/K500975/1) and by part of a Wellcome Strategic Award
55 (105031/D/14/Z) awarded to W. Reik, S. Teichmann, J. Nichols, B.D. Simons, T. Voet,
56 S. Srinivas, L. Vallier, B.G. and J.C. Marioni.

57

58 **Conflict of Interest Statement**

59 The authors declare that they have no competing interests.

60

61

Abstract

Background: Basophils and mast cells contribute to the development of allergic reactions. Whereas these mature effector cells are extensively studied, the differentiation trajectories from hematopoietic progenitors to basophils and mast cells are largely uncharted at the single-cell level.

Methods: We performed multicolor flow cytometry, high-coverage single-cell RNA sequencing analyses, and cell fate assays to chart basophil and mast cell differentiation at single-cell resolution in mouse.

Results: Analysis of flow cytometry data reconstructed a detailed map of basophil and mast cell 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. Cell fate assays showed that multicolor flow cytometry and transcriptional profiling successfully predict the bipotent phenotype of a previously uncharacterized population of peritoneal basophil-mast cell progenitors.

Conclusions: A combination of molecular and functional profiling of bone marrow and peritoneal cells provided a detailed roadmap of basophil and mast cell development. ~~We provide a detailed roadmap of basophil and mast cell development through a combination of molecular and functional profiling~~. An interactive web resource was created to enable the wider research community to explore the expression dynamics for any gene of interest.

Keywords: Basophils; differentiation; mast cells; single-cell RNA sequencing; transcriptomics

87

88 **Introduction**

89 Mast cells are sentinel cells that are strategically positioned throughout the body and
90 allow rapid triggering of the immune system upon infections.[1] Mast cell activation also
91 follows IgE-allergen-mediated crosslinking of the FcεRI receptors in atopic
92 individuals, which causes an allergic reaction. Along with basophils, activation of mast
93 cells results in prompt release of proteases and histamine from the cytoplasmic granules
94 as well as synthesis of cytokines and chemokines. These mediators in turn cause
95 inflammation, vasodilation, and leukocyte recruitment to the site of triggering.[1] Thus,
96 the functions of mature basophils and mast cells have been studied in great detail.
97 However, less is known about these cells' development.

98

99 A hierarchical model with distinct megakaryocyte-erythroid, granulocyte-monocyte,
100 and lymphoid branches, was until recently the dominating representation of
101 hematopoiesis.[2] Single-cell RNA sequencing (scRNA-seq) coupled with cell fate
102 assays now reveals that hematopoietic differentiation more likely represents a
103 landscape of cell states with continuous progression from multi- and bipotent
104 progenitors into each respective cell lineage.[3-7] In particular, single-cell transcriptomics
105 of Lin^{-} $c\text{-}Kit^{+}$ mouse bone marrow progenitors uncovers a continuous differentiation
106 from hematopoietic stem cells to bipotent basophil-mast cell progenitors (BMCPs).[4]
107 Microarray analysis of bulk-sorted cells shows distinct gene expression profiles of
108 mature basophils and mast cells.[8] However, investigation of temporal gene expression
109 dynamics during basophil and mast cell specification and maturation is yet to be
110 delineated and requires single-cell resolution.

111

112 Here, we combine multicolor flow cytometry-based index sorting with high-coverage
113 scRNA-seq to investigate the basophil-mast cell bifurcation and the differentiation into
114 each respective lineage. We demonstrate that molecular profiling and pseudotime
115 ordering of single cells highlights genes that are critical for cell differentiation and
116 maturation. The analysis is accompanied with the generation of a user-friendly web
117 resource that allows gene expression to be explored across the single-cell landscape.
118 Finally, we use cell-fate assays to show that single-cell transcriptomics and protein
119 epitope data analysis successfully predict the fate potential of the previously
120 uncharacterized BMCP population in the peritoneal cavity. Taken together, the current
121 resource provides a detailed roadmap of the developmentally related basophils and mast
122 cells, whose activation contributes to allergic diseases.
123

124 **Methods**

125 *Cell isolation and flow cytometry*

126 Experiments involving mice were performed according to the United Kingdom Home
127 Office regulations. PBS with 2 % fetal calf serum (Sigma-Aldrich, St Louis, MO) and
128 1 mM EDTA was injected into the peritoneal cavity of euthanized C57BL/6 mice. The
129 fluid was aspirated following vigorous massage, and the cells were prepared for FACS.
130 Peritoneal lavage samples with excessive blood contamination were discarded before
131 data acquisition. Bone marrow cells were extracted by flushing or crushing the femurs,
132 tibias, and/or ilia. Red blood cells were lysed and the remaining cells were prepared for
133 FACS. The cells were sorted with a BD Influx cell sorter (BD Biosciences, San Jose,
134 CA). Cell doublets were excluded with the width parameters. P1 cells and mast cells
135 were sorted two consecutive times for cell culture experiments. The cells were sorted
136 into Terasaki plates (Greiner Bio-One, Kremsmünster, Austria) or 96-well plate wells.
137 Visual inspection determined colony sizes following culture, and the size was set to 1
138 if no live cells were observed in a particular well. Flow cytometry was typically
139 performed on colonies constituting at least 20 cells, and potential to form a particular
140 cell lineage was based on at least 5 events in a given gate, as described previously.[4]
141 Cultured cells were analyzed with the BD Fortessa flow cytometers (BD Biosciences).
142

143 *Antibodies and cell staining*

144 Primary cells were incubated with the antibodies integrin β7 (clone FIB504), CD34
145 (RAM34), Sca-1 (D7), CD16/32 (93), c-Kit (2B8), FcεRI (MAR-1), IL-33Rα/ST2
146 (DIH9), and/or CD49b (DX5). The Easysep mouse hematopoietic progenitor cell
147 isolation cocktail (STEMCELL Technologies, Vancouver, Canada) stained lineage
148 markers. Cultured cells were stained with c-Kit, FcεRI, CD49b, with or without

149 TER119 (TER119). Fc-block (clone 93) was used where appropriate. The antibodies
150 were from BD Biosciences, Biolegend (San Diego, CA), and Thermo Fisher Scientific
151 (Waltham, MA). DAPI (BD Biosciences) or 7-AAD (Thermo Fisher Scientific) were
152 used to exclude dead cells.
153
154 *Cell culture*
155 The cells were cultured for 6-7 days in IMDM (Sigma-Aldrich) with 20 % heat-
156 inactivated fetal calf serum (Sigma-Aldrich), 100 U/ml penicillin (Sigma-Aldrich), 0.1
157 mg/ml streptomycin (Sigma-Aldrich), 50-200 μM β-mercaptoethanol (Thermo Fisher
158 Scientific). The medium was supplemented with 20 ng/ml IL-3 and 100 ng/ml stem cell
159 factor, or 80 ng/ml stem cell factor, 20 ng/ml IL-3, 50 ng/ml IL-9, and 2 U/ml
160 erythropoietin. All cytokines were recombinant mouse cytokines (Peprotech, Rocky
161 Hill, NJ) except the erythropoietin (Eprex; Janssen-Cilag, High Wycombe, UK), which
162 was human.
163
164 *Flow cytometry analysis*
165 FlowJo v10 (Treestar, Ashland, OR) produced the flow cytometry plots. Diffusion map
166 and principal component analysis (PCA) plots of flow cytometry data were generated
167 using the R programming environment. The flow cytometry events were down-sampled
168 according to the population with the least number of events. Duplicate entries were
169 removed, and the parameters representing fluorescent markers log-transformed.
170 Variables were z-scored and diffusion map plots generated using the *destiny* and
171 *ggplot2* packages. PCA was calculated using the prcomp function. Data projection was
172 performed using the predict function.
173

174 *scRNA-seq data analysis*

175 Primary single-cells were FACS index sorted into lysis buffer, and scRNA-seq was
176 performed based on the Smart-Seq2 protocol.[9] For details of scRNA-seq data
177 processing see Supplementary methods. Analysis was performed using the scanpy v1.4
178 python module[10] and the R programming environment. Interactive websites for plotting
179 gene expression and flow cytometry data are hosted at http://128.232.224.252/bas/ and
180 http://128.232.224.252/per/ for the basophil and mast cell dataset, respectively.

181

182 *Data sharing statement*

183 Protocols and scRNA-seq data generated for this article have been deposited in the
184 Gene Expression Omnibus database (accession numbers GSE128003 and GSE128074).
185 scRNA-seq data of bone marrow BMCPs, analyzed in Dahlin et al[4], are available
186 through GSE106973. Human scRNA-seq data was obtained from the Human Cell
187 Atlas.[11] For other original data, please contact joakim.dahlin@ki.se or
188 bg200@cam.ac.uk.

189

190

191 **Results**

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

194 Basophil and mast cell differentiation are closely linked, and the cells share a common
195 bipotent progenitor (Figure 1A). Here, we used multicolor flow cytometry to map these
196 branching trajectories at the single-cell level. Flow cytometry analysis of mouse bone
197 marrow cells captured BMCPs and cells of the basophil differentiation trajectory
198 (Figure 1Bi,ii).[4,12] We performed parallel analysis of Lin^- $c\text{-}Kit^+$ $Fc\varepsilon RI^+$ peritoneal cells
199 in an attempt to capture late mast cell differentiation, which takes place at peripheral
200 sites. Analysis of the Lin^- $c\text{-}Kit^+$ $Fc\varepsilon RI^+$ peritoneal cells distinguished two populations
201 based on integrin β7 expression and cell granularity (measured with the side scatter
202 parameter). A broad gate that included a continuum of prospective mast cell
203 progenitors, intermediate precursors, and mast cells – hereon referred to as population
204 P1 – was set in close proximity to the mast cell gate (Figure 1Biii). To enable 2-
205 dimensional visualization of the flow cytometry single-cell datasets we performed
206 dimensionality reduction using a diffusion map algorithm.[13] This method embeds a
207 dataset by considering the properties of random walks between cells that are close
208 together in the high-dimensional space, and can visualize branching cell differentiation
209 trajectories in single-cell data.[14] Flow cytometry data covered the 5 cell populations
210 recorded with 9 fluorescent and 2 light scatter parameters. The diffusion map revealed
211 a bifurcation at the BMCP stage, establishing the putative entry points to the basophil
212 and mast cell trajectories (Figure 1C). The diffusion map embedding further visualized
213 the progression from BMCP, through basophil progenitors, to basophils. The mast cell
214 trajectory exhibited a similar pattern, with differentiation of BMCPs to mature mast
215 cells.

216

217 Plotting individual surface markers in the diffusion map allowed us to investigate how
218 the proteins are expressed during differentiation. For example, loss of CD34 in
219 combination with downregulation of c-Kit marked the progression from BMCPs to
220 basophils (Figure 1D), and loss of integrin β7 in c-Kit^{+} cells was associated with
221 differentiation along the trajectory from BMCPs to mast cells (Figure 1D). Taken
222 together, the flow cytometry dataset provides a template of basophil-mast cell
223 differentiation at single-cell level and highlights the bifurcation towards the two
224 lineages.

225

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

228 Analysis by flow cytometry suggested that the flow cytometry gating strategies we used
229 could be capturing a continuum of differentiation towards basophils and mast cells. To
230 first identify changes in gene expression programs during basophil differentiation, we
231 performed scRNA-seq of primary basophil progenitor (BaP) cells and basophil (Ba)
232 cells from mouse bone marrow. Both PCA and diffusion maps showed separation
233 between the majority of cells from the two sorting gates (Figure 2A, Figure S1A). To
234 investigate which genes were driving this separation, we performed differential
235 expression analysis, identifying 212 upregulated and 833 downregulated genes in Ba
236 cells compared to BaPs (adjusted p-value < 0.01, t-test with Benjamini-Hochberg
237 correction) (Table S1). Enrichment analysis of these gene lists revealed that upregulated
238 genes were enriched for granulocyte immune response terms (Table S2, Figure S1B).
239 Downregulated genes were enriched for cell cycle related terms (Table S2, Figure 2B),
240 suggesting a difference in cell cycle behavior throughout the differentiation process.

241 This observation is in line with other hematopoietic differentiation pathways, where
242 progenitors commonly loose proliferative capacity as they mature into the fully
243 differentiated cell types.
244
245 To further explore this, we then performed analysis to computationally assign cell cycle
246 state to the single-cell profiles.[15] Consistent with the gene list enrichment analysis, the
247 majority of cells in the BaP gate were assigned to S and G2M states (69%), whereas
248 87% of cells in the Ba gate were assigned to G1 state (Figure 2C, D). The effect of cell
249 cycle status was clear in the diffusion map dimensionality reduction (Figure S1C),
250 confounding attempts to order cells using pseudotime algorithms. Instead,
251 downregulation of progenitor marker genes such as *Cd34* and *Kit* indicated that
252 ordering cells along PC1 could be used to arrange cells in pseudotime (Figure S1D).
253 Visualization of index sorting data also showed clear dynamics of the different surface
254 markers along PC1 (Figure 2E). As expected, CD34 and c-Kit protein expression
255 showed a negative correlation with pseudotime (compare Figure 1D and 2E), which
256 indicates their downregulation during basophil differentiation. In addition, the basophil
257 marker CD49b (DX5) showed a positive correlation with pseudotime ordering (Figure
258 2E).
259
260 Using the PC1 pseudotime ordering, we then identified genes that dynamically changed
261 during differentiation (Figure 2F). Clustering sorted these dynamic genes into two
262 groups: one increasing and one decreasing with differentiation (Table S3). Basophil
263 differentiation was associated with upregulation of *Hdc*, which is associated with
264 histamine synthesis, and increased expression of the basophil gene E-cadherin (*Cdh1*).
265 We further observed downregulation of the proteases *Mcpt8*, *Prss34* and *Ctsg* and

266 upregulation of the transcription factors *Cebpa*, *Stat5b*, and *Spi1* (Figure 2G). To
267 validate the full lists of dynamically regulated genes, we compared these to mast cell
268 and basophil signature genes identified using bulk microarray analysis.[8] Genes
269 upregulated during basophil differentiation exhibited a significant overlap with the
270 previously described basophil signature genes ($p = 4.0 \times 10^{-29}$, hypergeometric test,
271 Figure S1Ei), whereas genes that were downregulated during differentiation had
272 significant overlap with the previously described mast cell signature genes ($p = 1.3 \times$
273 10^{-4}, hypergeometric test, Figure S1Eii).
274
275 Together, this analysis offers a description of the dynamics of gene expression during
276 basophil differentiation and highlights changes in cell cycle activity as one of the major
277 occurrences during this maturation process.
278
279 *Single-cell gene expression analysis suggests a continuum of mast cell differentiation*
280 *in the peritoneal cavity*
281 After exploring the basophil progenitors, we next decided to focus on mast cell
282 differentiation in the peritoneal cavity. The flow cytometry data suggested the existence
283 of both peritoneal BMCPs and mast cells (Figure 1), so we performed single-cell RNA-
284 sequencing on these primary cell populations to characterize them based on gene
285 expression. A subset of the P1 cells clustered separately from the mast cells in the
286 diffusion map plot, demonstrating a difference between the transcriptome of these cells
287 and the peritoneal mast cells (Figure 3A). In previous work we characterized bone
288 marrow BMCPs at the single-cell gene expression level.[4] To examine the similarity of
289 these bone marrow progenitors to the peritoneal mast cell differentiation, single-cell
290 bone marrow BMCP profiles from Dahlin et al[4] were projected onto the peritoneal

291 dataset (Figure 3B). This demonstrated that the P1 peritoneal cells furthest from the
292 peritoneal MCs were most similar to the bone marrow BMCPs, supporting that these
293 were the most immature cells in the dataset.
294

295 To understand expression changes during mast cell maturation, we then performed
296 pseudotime ordering of the peritoneal cells (Figure 3C). As expected, interrogation of
297 cell surface markers along pseudotime showed a strong downregulation of integrin β7
298 and strong upregulation of markers such as Sca1 and ST2 (compare Figure 1D and 3D).
299 Genes exhibiting dynamic expression patterns were identified and clustered as for the
300 basophil trajectory (Table S4, Figure 3E). Annotation from the Panther database[16] was
301 used to interrogate the two gene clusters for overlap with specific annotated gene sets
302 such as proteases. Protease genes downregulated during mast cell differentiation
303 included *Mcpt8* and *Gzmb*, whereas *Cpa3*, *Cma1*, *Mcpt1*, *Mcpt4*, *Tpsb2,* and *Tpsab1*
304 increased with differentiation (Figure 3F). To investigate the temporal induction and
305 loss of protease genes, we changed visualization method and scaled the gene expression
306 according to the cell with maximum expression (instead of z-scoring genes across the
307 dataset). Early onset proteases included *Cpa3*, followed by *Tpsb2*, and finally *Tpsab1*,
308 indicating that the protease induction occurs in stages (Figure 3G, raw values for
309 individual genes shown in Figure S2C).
310

311 To validate the full lists of dynamically regulated genes in the peritoneal mast cell
312 dataset, we compared these to mast cell and basophil signatures identified in Dwyer et
313 al[8]. The upregulated genes significantly overlapped with the mast cell signature genes
314 ($p = 3.7 \times 10^{-65}$, hypergeometric test, Figure S2Di), including *Ndst2* and *Meis2* ~~and *Hdc*~~
315 (Figure 3H). Some genes showed expression enrichment mainly in the mast cells

316 (*Meis2*), whereas others were expressed more evenly across the trajectory save for
317 lower expression at the beginning of pseudotime (*Ndst2*). Similar to basophil
318 differentiation, mast cell differentiation was associated with *Hdc* upregulation (Figure
319 3H). There was also a small overlap between the downregulated genes and basophil
320 signature genes ($p = 2.5 \times 10^{-5}$, hypergeometric test, Figure S2Dii). To investigate the
321 link between gene and protein expression we also interrogated the expression of *Itga4*
322 and *Itgb7*, which encode subunits of integrin β7. *Itga4* was significantly downregulated
323 with a similar expression pattern to integrin β7 in the flow cytometry data whereas *Itgb7*
324 was not significantly changing in pseudotime (Figure 3D, H).
325
326 *P1 cells in the peritoneal cavity exhibit basophil and mast cell-forming potential*
327 The flow cytometry-based and transcriptional analyses revealed an immature cell
328 population with BMCP-like characteristics in the peritoneal cavity. However, a
329 population of bipotent peritoneal BMCPs has not previously been described at this site.
330 We therefore explored whether the protein and transcriptional analyses successfully
331 predicted the developmental state of the peritoneal P1 cells and mast cells.
332 Fluorescence-activated cell sorting (FACS) isolated P1 cells and mast cells were
333 cytochemically stained with May-Grünwald Giemsa. Primary P1 cells displayed little
334 cytoplasm that contained no or few granules, consistent with the morphology of blasts
335 (Figure 4A). In contrast, primary mast cells were filled with numerous metachromatic
336 granules, in agreement with a mature morphology (Figure 4A).
337
338 We cultured the peritoneal cells to investigate whether the P1 cell population exhibited
339 capacity to generate basophils and mast cells. P1 cells cultured with IL-3 and stem cell

340 factor generated c-Kit$^-$ FcεRI$^+$ CD49b$^+$ basophils and c-Kit$^+$ FcεRI$^+$ mast cells, whereas
341 primary mast cells only displayed mast cell-forming capacity (Figure 4B-D).
342
343 By contrast to bulk cultured cells, only cell-fate assays performed at the single-cell level
344 have the potential to reveal whether the P1 population consists of bipotent progenitors.
345 Therefore, single P1 cells and mast cells were index sorted into individual wells, the
346 resulting colony sizes were measured, and the colonies were subjected to flow
347 cytometry analysis and cytochemical staining. To visualize the cell culture data, we
348 first performed PCA of the flow cytometry data presented in Figure 1C, henceforth
349 referred to as the reference dataset (Figure 4E). We then projected the FACS index sort
350 data onto the principal component space of the reference dataset, and plotted colony
351 size and colony type data (Figure 4F). Analysis of colony sizes showed that colonies
352 derived from P1 cells were large, whereas cells along the mast cell trajectory exhibited
353 reduced proliferation rate (Figure 4F, S3). Notably, the cell-fate assays revealed that
354 primary P1 cells formed pure basophil colonies, pure mast cell colonies or mixed
355 basophil-mast cell colonies (Figure 4F, S3A). Colonies derived from single mast cells
356 were too small to analyze with flow cytometry. However, mast cells cultured in bulk
357 remained mast cells as expected (Figure 4C-D, S3B). Further analysis of the FACS
358 index sort data revealed that primary cells that formed large colonies comprising
359 basophils and/or mast cells were mainly integrin β7$^{+/hi}$ P1 progenitors (Figure S4). This
360 observation agrees with the pseudotime ordering of the single-cell transcriptomics data,
361 which showed that loss of integrin β7 is associated with differentiation. We also
362 cultured the P1 peritoneal cells in erythroid-promoting conditions, as the early basophil-
363 mast cell differentiation is closely linked to the erythrocyte trajectory.[3] However, no

364 erythroid output was observed (Figure S5), indicating that the P1 cells indeed consisted
365 of bipotent basophil-mast cell progenitors.
366
367 After investigating the bifurcation of bipotent BMCPs in mouse, we decided to explore
368 single-cell transcriptomics data in human to see if it supports a similar relationship
369 between basophil and mast cell differentiation. We processed data of human bone
370 marrow cells from the Human Cell Atlas.[11,17] By subsetting the data for pertinent
371 progenitor populations we identified a distinct differentiation trajectory with a gene
372 expression profile characteristic of basophils and mast cells (*HDC* and *MS4A2*), which
373 was separate from other myelo-erythroid lineages (Figure 4G-H, S6). Observed
374 expression of basophil (*CLC* and *CEBPA*) and mast cell (*TPSB2* and *TPSAB1*) genes
375 highlighted the gradual differentiation and entry points of the respective lineages
376 (Figure 4H, S6C). The observation of neighboring entry points indicated a close
377 developmental relationship between human basophils and mast cells, whereas
378 erythrocyte development progressed on a separate trajectory.
379
380 Taken together, the cell culture assays revealed that the protein and gene expression
381 analyses successfully predicted the differentiation state of the P1 cell population in the
382 peritoneal cavity.
383
384 **Discussion**
385 Single-cell transcriptomics coupled with index sorting of thousands of bone marrow
386 HSPCs has previously been used to chart erythrocyte and granulocyte-monocyte
387 differentiation.[18,19] BMCPs represent a minor fraction of the bone marrow HSPCs, and
388 capturing the early basophil-mast cell axis therefore requires analysis of tens of

389 thousands of HSPCs.[4] The early differentiation of progenitors with mast cell-forming
390 capacity occurs in the bone marrow.[20] However, full mast cell differentiation and
391 maturation takes place at peripheral sites,[20] and we therefore specifically sorted Lin^- c-
392 Kit^+ $Fc\varepsilon RI^+$ cells extracted from the peritoneal cavity to capture this process. Cell
393 isolation from peritoneum does not require enzymatic digestion, thus minimizing
394 external stimuli during cell processing. Basophil differentiation takes place in bone
395 marrow, and we therefore analyzed basophils and their progenitors from this site.
396

397 The single-cell transcriptomics data presented here capture a continuum of cells from
398 peritoneal BMCPs to mast cells. Recent studies have explored whether bone marrow
399 HSPCs constitute the primary source of peritoneal mast cells. Transferred bone marrow
400 cells contribute little to the peritoneal mast cell numbers unless the local pool of mature
401 mast cells are depleted first.[21,22] A known feedback mechanism, in which mast cells
402 inhibit recruitment and differentiation of their progenitors, provides a likely explanation
403 to these observations.[23,24] The peritoneal mast cells in adult mice emerge from
404 definitive hematopoiesis.[21,22] However, this observation does not necessarily imply that
405 bone marrow HSPCs are the main source of mast cells. Further studies exploring the
406 relationship between bone marrow HSPCs and peritoneal mast cell differentiation are
407 therefore needed. It will also be important to generate reference maps of mast cell
408 differentiation at alternative compartments and in the prenatal setting, given the
409 heterogeneity of the mast cell population.
410

411 We reveal the existence of a progenitor with dual basophil-mast cell-forming potential
412 in the peritoneal cavity. BMCPs have previously been described in the mouse spleen
413 and bone marrow,[4,12,25] and the presence of a bipotent progenitor population indicates

414 that there is a close association between the basophil and mast cell differentiation
415 trajectories. One study has questioned the bipotent nature of splenic BMCPs,[26] as only
416 mast cell colonies were observed following culture. The failure to detect basophils in
417 that study is yet to be explained.
418
419 Recent data suggest that the erythroid axis is coupled with the basophil and/or mast cell
420 fates.[3,5,27-29] However, we did not observe erythrocyte-forming potential among P1 cells
421 in the peritoneum. In agreement with this, BMCPs in the spleen and bone marrow are
422 unable to generate erythrocytes,[4,12] altogether suggesting that loss of erythrocyte-
423 forming potential is an early event along the differentiation trajectory from
424 hematopoietic stem cells to basophil and mast cells. Similarly, the human single-cell
425 transcriptional landscape presented here reveals a unique trajectory of cells diverging
426 into basophils and mast cells, separate from the erythroid trajectory. This observation
427 is in agreement with recent studies that demonstrate the presence of human progenitor
428 populations that produce basophils and mast cells.[5,27] The low proliferation capacity of
429 mast cells complicates culture-based approaches to determine whether the cell
430 populations harbor bipotent basophil-mast cell progenitors.[30] ~~Further investigation of~~
431 ~~the basophil-mast cell fate decision in human is still awaited.~~ Previous culture
432 experiments suggest that the human basophil and eosinophil differentiation trajectories
433 are adjacent to each other.[31,32] Simultaneous existence of bipotent basophil-mast cell
434 progenitors and basophil-eosinophil progenitors is in line with the landscape model of
435 hematopoiesis.[20] However, the exact association between the mast cell, basophil, and
436 eosinophil fates in mouse and human is still to be deciphered.
437

438 Temporal ordering of the cells in the transcriptomic datasets allows exploration and
439 verification of molecular processes in differentiating basophils and mast cells. We show
440 that *Ndst2* (encoding *N*-deacetylase/*N*-sulphotransferase-2) is upregulated during
441 differentiation from BMCPs to mature mast cells, and this was also associated with the
442 appearance of numerous densely stained granules. In agreement with these findings,
443 dense May-Grünwald Giemsa staining of the peritoneal mast cell granules requires
444 sulphated heparin, which is dependent on *Ndst2* expression.[33] Mast cells and basophils
445 are major producers of histamine, which ~~Histamine~~ is quickly released upon ~~basophil~~
446 ~~and mast~~ cell activation.[34] Here, we ~~show~~ verified that the expression of the enzyme
447 that catalyzes the histamine synthesis, histidine decarboxylase (*Hdc*), increased upon
448 differentiation of both basophils and mast cells. Analysis of the single-cell
449 transcriptomics data can also give insights into more complex regulatory processes. For
450 example, integrin β7 expression on progenitors with mast cell-forming potential is
451 important for cell migration into the lungs in a mouse model of allergic airway
452 inflammation.[35] Downregulation of integrin β7 is a hallmark of terminal mast cell
453 differentiation.[36] However, we did not observe downregulation of *Itgb7* gene
454 expression during the transition from BMCPs to mast cells, despite downregulation of
455 the surface protein. Integrins constitute αβ heterodimers when localized to the cell
456 surface, and further investigation into the gene expression profile revealed decreased
457 expression of *Itga4*, the binding partner of the integrin β7 subunit, upon differentiation.
458 Thus, the loss of integrin α4 gene expression likely explains the downregulation of
459 integrin β7 protein expression on the cell surface.

460

461 *Mcpt1* expression is typically associated with mucosal mast cells. Nevertheless, *Mcpt1*
462 was upregulated during differentiation of serosal-type peritoneal mast cells. However,

463 the levels detected were several orders of magnitude lower than the levels of *Tpsb2*,

464 *Cma1*, and *Mcpt4*.

465

466 During basophil differentiation, the transcription factors *Stat5b* and *Cebpa* are

467 upregulated along the progression of pseudotime. The expression of C/EBPα is STAT5-

468 dependent, and both genes are required for basophil formation.[12,25] Dynamic expression

469 of transcription factors with currently unknown functions in basophil and mast cell

470 differentiation was also recognized. For example, *Spi1*, which encodes PU.1, is

471 upregulated during late basophil differentiation. It is known to be involved in neutrophil

472 granulocyte maturation,[37,38] but the role of PU.1 in basophil differentiation is yet to be

473 delineated. During mast cell differentiation, we describe the increase of the

474 transcription factor *Meis2*. Primary mast cells from human skin express this

475 transcription factor,[39] but the potential function during mast cell differentiation is yet

476 to be described.

477

478 Microarray and RNA sequencing analyses reported previously provide detailed gene

479 expression patterns of mature hematopoietic cell populations, including bulk-sorted

480 mature basophils and mast cells.[8,40] We observed that differentiation into basophils and

481 mast cells involves activation of mutually exclusive lineage programs. However, a

482 small subset of the previously reported signature genes is not unique to mature cells,

483 but can also be observed in bipotent progenitors. For example, we show that *Mcpt8*

484 expression is not restricted to basophils but is also expressed by BMCPs. Indirect

485 evidence also supports the validity of this observation.[41,42] ~~This~~ Transient *Mcpt8*

486 expression at the BMCPs stage in fact provides an explanation to a major conundrum

487 in the field. Basophils, identified as *Mcpt8*-expressing cells, have been reported to

488 exhibit potential to transdifferentiate into mast cells.[43] Our results show that a more
489 likely scenario is that a subset of the previously reported *Mcpt8*-expressing cells
490 constitutes bipotent BMCPs that can give rise to mast cells.
491
492 In summary, here we have reported the generation of a high-resolution single-cell map
493 of the BMCP bifurcation and mast cell and basophil differentiation. A user-friendly
494 interactive website has been created for the wider community to enable further
495 exploration of the data.
496

497 **References**

498

499 1. Nathan C. Points of control in inflammation. *Nature.* 2002;420:846-852.

500 2. Akashi K, Traver D, Miyamoto T, Weissman IL. A clonogenic common
501 myeloid progenitor that gives rise to all myeloid lineages. *Nature.* 2000;404:193-197.

502 3. Tusi BK, Wolock SL, Weinreb C, et al. Population snapshots predict early
503 haematopoietic and erythroid hierarchies. *Nature.* 2018;555:54-60.

504 4. Dahlin JS, Hamey FK, Pijuan-Sala B, et al. A single-cell hematopoietic
505 landscape resolves 8 lineage trajectories and defects in Kit mutant mice. *Blood.*
506 2018;131:e1-e11.

507 5. Drissen R, Thongjuea S, Theilgaard-Monch K, Nerlov C. Identification of two
508 distinct pathways of human myelopoiesis. *Sci Immunol.* 2019;4.

509 6. Pellin D, Loperfido M, Baricordi C, et al. A comprehensive single cell
510 transcriptional landscape of human hematopoietic progenitors. *Nat Commun.*
511 2019;10:2395.

512 7. Velten L, Haas SF, Raffel S, et al. Human haematopoietic stem cell lineage
513 commitment is a continuous process. *Nat Cell Biol.* 2017;19:271-281.

514 8. Dwyer DF, Barrett NA, Austen KF, Immunological Genome Project C.
515 Expression profiling of constitutive mast cells reveals a unique identity within the
516 immune system. *Nat Immunol.* 2016;17:878-887.

517 9. Picelli S, Faridani OR, Bjorklund AK, Winberg G, Sagasser S, Sandberg R.
518 Full-length RNA-seq from single cells using Smart-seq2. *Nat Protoc.* 2014;9:171-
519 181.

520 10. Wolf FA, Angerer P, Theis FJ. SCANPY: large-scale single-cell gene
521 expression data analysis. *Genome Biol.* 2018;19:15.

522 11. Human Cell Atlas: Census of Immune Cells.
523 https://data.humancellatlas.org/explore/projects/cc95ff89-2e68-4a08-a234-
524 480eca21ce79.
525 12. Arinobu Y, Iwasaki H, Gurish MF, et al. Developmental checkpoints of the
526 basophil/mast cell lineages in adult murine hematopoiesis. *Proc Natl Acad Sci U S A.*
527 2005;102:18105-18110.
528 13. Coifman RR, Lafon S, Lee AB, et al. Geometric diffusions as a tool for
529 harmonic analysis and structure definition of data: Diffusion maps. *Proceedings of the*
530 *National Academy of Sciences.* 2005;102:7426-7431.
531 14. Haghverdi L, Buettner F, Theis FJ. Diffusion maps for high-dimensional
532 single-cell analysis of differentiation data. *Bioinformatics.* 2015;31:2989-2998.
533 15. Macosko EZ, Basu A, Satija R, et al. Highly Parallel Genome-wide
534 Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell.*
535 2015;161:1202-1214.
536 16. Mi H, Muruganujan A, Ebert D, Huang X, Thomas PD. PANTHER version
537 14: more genomes, a new PANTHER GO-slim and improvements in enrichment
538 analysis tools. *Nucleic Acids Res.* 2019;47:D419-D426.
539 17. Regev A, Teichmann SA, Lander ES, et al. The Human Cell Atlas. *Elife.*
540 2017;6.
541 18. Nestorowa S, Hamey FK, Pijuan Sala B, et al. A single-cell resolution map of
542 mouse hematopoietic stem and progenitor cell differentiation. *Blood.* 2016;128:e20-
543 31.
544 19. Paul F, Arkin Y, Giladi A, et al. Transcriptional Heterogeneity and Lineage
545 Commitment in Myeloid Progenitors. *Cell.* 2015;163:1663-1677.

546 20. Grootens J, Ungerstedt JS, Nilsson G, Dahlin JS. Deciphering the
547 differentiation trajectory from hematopoietic stem cells to mast cells. *Blood Adv.*
548 2018;2:2273-2281.
549 21. Gentek R, Ghigo C, Hoeffel G, et al. Hemogenic Endothelial Fate Mapping
550 Reveals Dual Developmental Origin of Mast Cells. *Immunity.* 2018;48:1160-1171.
551 22. Li Z, Liu S, Xu J, et al. Adult Connective Tissue-Resident Mast Cells
552 Originate from Late Erythro-Myeloid Progenitors. *Immunity.* 2018;49:640-653 e645.
553 23. Kanakura Y, Kuriu A, Waki N, et al. Changes in numbers and types of mast
554 cell colony-forming cells in the peritoneal cavity of mice after injection of distilled
555 water: evidence that mast cells suppress differentiation of bone marrow-derived
556 precursors. *Blood.* 1988;71:573-580.
557 24. Nilsson G, Dahlin JS. New insights into the origin of mast cells. *Allergy.*
558 2019;74:844-845.
559 25. Qi X, Hong J, Chaves L, et al. Antagonistic regulation by the transcription
560 factors C/EBPalpha and MITF specifies basophil and mast cell fates. *Immunity.*
561 2013;39:97-110.
562 26. Mukai K, BenBarak MJ, Tachibana M, et al. Critical role of P1-Runx1 in
563 mouse basophil development. *Blood.* 2012;120:76-85.
564 27. Grootens J, Ungerstedt JS, Wu C, Hamberg Levedahl K, Nilsson G, Dahlin
565 JS. CD203c distinguishes the erythroid and mast cell-basophil differentiation
566 trajectories among human FcepsilonRI(+) bone marrow progenitors. *Allergy.*
567 2020;75:211-214.
568 28. Zheng S, Papalexi E, Butler A, Stephenson W, Satija R. Molecular transitions
569 in early progenitors during human cord blood hematopoiesis. *Mol Syst Biol.*
570 2018;14:e8041.

571 29. Inclan-Rico JM, Hernandez CM, Henry EK, et al. Trichinella spiralis-induced
572 mastocytosis and erythropoiesis are simultaneously supported by a bipotent mast
573 cell/erythrocyte precursor cell. *PLoS Pathog.* 2020;16:e1008579.
574 30. Dahlin JS, Malinovschi A, Ohrvik H, et al. Lin- CD34hi CD117int/hi
575 FcepsilonRI+ cells in human blood constitute a rare population of mast cell
576 progenitors. *Blood.* 2016;127:383-391.
577 31. Grundstrom J, Reimer JM, Magnusson SE, Nilsson G, Wernersson S, Hellman
578 L. Human cord blood derived immature basophils show dual characteristics,
579 expressing both basophil and eosinophil associated proteins. *PLoS One.*
580 2012;7:e48308.
581 32. Denburg JA, Telizyn S, Messner H, et al. Heterogeneity of human peripheral
582 blood eosinophil-type colonies: evidence for a common basophil-eosinophil
583 progenitor. *Blood.* 1985;66:312-318.
584 33. Forsberg E, Pejler G, Ringvall M, et al. Abnormal mast cells in mice deficient
585 in a heparin-synthesizing enzyme. *Nature.* 1999;400:773-776.
586 34. Voehringer D. Protective and pathological roles of mast cells and basophils.
587 *Nat Rev Immunol.* 2013;13:362-375.
588 35. Abonia JP, Hallgren J, Jones T, et al. Alpha-4 integrins and VCAM-1, but not
589 MAdCAM-1, are essential for recruitment of mast cell progenitors to the inflamed
590 lung. *Blood.* 2006;108:1588-1594.
591 36. Dahlin JS, Ding Z, Hallgren J. Distinguishing Mast Cell Progenitors from
592 Mature Mast Cells in Mice. *Stem Cells Dev.* 2015;24:1703-1711.
593 37. Scott EW, Simon MC, Anastasi J, Singh H. Requirement of transcription
594 factor PU.1 in the development of multiple hematopoietic lineages. *Science.*
595 1994;265:1573-1577.

596 38. Anderson KL, Smith KA, Pio F, Torbett BE, Maki RA. Neutrophils deficient
597 in PU.1 do not terminally differentiate or become functionally competent. *Blood.*
598 1998;92:1576-1585.
599 39. Motakis E, Guhl S, Ishizu Y, et al. Redefinition of the human mast cell
600 transcriptome by deep-CAGE sequencing. *Blood.* 2014;123:e58-67.
601 40. Akula S, Paivandy A, Fu Z, Thorpe M, Pejler G, Hellman L. Quantitative In-
602 Depth Analysis of the Mouse Mast Cell Transcriptome Reveals Organ-Specific Mast
603 Cell Heterogeneity. *Cells.* 2020;9.
604 41. Lunderius C, Xiang Z, Nilsson G, Hellman L. Murine mast cell lines as
605 indicators of early events in mast cell and basophil development. *Eur J Immunol.*
606 2000;30:3396-3402.
607 42. Ohnmacht C, Schwartz C, Panzer M, Schiedewitz I, Naumann R, Voehringer
608 D. Basophils orchestrate chronic allergic dermatitis and protective immunity against
609 helminths. *Immunity.* 2010;33:364-374.
610 43. Metcalf D, Ng AP, Baldwin TM, Di Rago L, Mifsud S. Concordant mast cell
611 and basophil production by individual hematopoietic blast colony-forming cells. *Proc
612 Natl Acad Sci U S A.* 2013;110:9031-9035.
613
614
615

616 **Figure legends**

617 **Figure 1. Flow cytometry analysis reveals differentiation trajectories from**
618 **bipotent basophil-mast cell progenitors to basophils and mast cells.**
619 (A) Illustration outlining the basophil and mast cell differentiation trajectories. (B)
620 Flow cytometry-based gating strategies of (Bi) bipotent basophil-mast cell progenitors
621 (BMCPs) from bone marrow, (Bii) basophil progenitors (BaP) and basophils (Ba) from
622 bone marrow, and (Biii) P1 cells and mast cells from peritoneal cavity. Lineage markers
623 include 7-4, CD5, CD11b, CD19, CD45R/B220, Ly6G/C (Gr-1), and TER119. (C)
624 Diffusion map visualization of the flow cytometry data colored by cell type. (D)
625 Diffusion map visualization of the flow cytometry data colored by protein expression
626 or light scatter parameters. The surface expression parameters and light scatter
627 parameters are visualized on log-transformed and linear scales, respectively.
628 Expression of lineage markers and viability staining are not shown. The data are
629 representative of 4 independent experiments.
630
631 **Figure 2. Bone marrow basophil progenitors downregulate cell cycle genes during**
632 **differentiation.** (A) PCA of scRNA-seq profiles colored by cell surface marker
633 phenotype. PC, principal component. (B) Top 5 GO Biological Process terms
634 associated with the genes significantly upregulated in BaP cells compared to Ba cells,
635 ranked by adjusted p-value. Benjamini-Hochberg correction for multiple hypotheses
636 testing. Genes upregulated in Ba compared to BaP are presented in Figure S1B. (C)
637 Proportion of scRNA-seq profiles from each phenotype computationally assigned to
638 G1, S or G2M cell cycle states based on gene expression using the scanpy
639 *score_genes_cell_cycle* function. (D) PCA colored by cell cycle state. (E) Levels of
640 cell surface markers for cells ordered by PC1 pseudotime. Index data values were log-

641 transformed, smoothed along pseudotime by using a sliding window of size 20 and
642 scaled between 0 and 1 for each marker. Correlation values indicate the pearson
643 correlation coefficient between pseudotime and the unsmoothed expression values for
644 each surface marker. Colorbar at the top indicates the phenotypic cell type proportions
645 within each window. Blue corresponds to entirely BaPs and orange to Ba cells. (F)
646 Heatmap displaying the expression of genes dynamically expressed along the PC1
647 pseudotime ordering. The top colorbar indicates the cell type proportion in each
648 window. Expression is smoothed along a sliding window and z-scored for each gene,
649 and genes were clustered using Louvain clustering into groups showing different
650 dynamics. Dynamic genes defined as adjusted p-value < 0.01 in permutation test,
651 details in supplementary methods. (G) PCA colored by z-scored expression of specific
652 genes. The data represents cells pooled from 3 individual mice.
653
654 **Figure 3. Transcriptional profiling of peritoneal mast cell progenitors captures a**
655 **differentiation continuum.** (A) Diffusion map dimensionality reduction of scRNA-
656 seq profiles colored by cell phenotype. DC, diffusion component. (B) Bone marrow
657 BMCP cells from Dahlin et al[4] were projected into the PCA space of the peritoneal cells
658 and the 10 closest peritoneal neighbors of each bone marrow cell were identified in
659 these co-ordinates. Cells are colored by a score representing how frequently each
660 peritoneal cell was the nearest neighbor of a bone marrow BMCP. (C) Diffusion map
661 colored by pseudotime ordering of cells. DPT, diffusion pseudotime. (D) Levels of cell
662 surface markers for pseudotime ordered cells. Index data values were log-transformed,
663 smoothed along pseudotime by using a sliding window of size 20 and scaled between
664 0 and 1 for each marker. Correlation values indicate the pearson correlation coefficient
665 between pseudotime and the unsmoothed expression values for each surface marker.

666 Colorbar at the top indicates the phenotypic cell type proportions within each window.
667 Green corresponds to entirely P1 cells and purple to MCs. (E) Heatmap displaying the
668 expression of genes dynamically expressed along the pseudotime ordering. The top
669 colorbar indicates the proportion of cell type in each window. Expression is smoothed
670 along a sliding window and z-scored for each gene, and genes were clustered using
671 Louvain clustering into groups showing different dynamics. Dynamic genes defined as
672 adjusted p-value < 0.01 in permutation test, details in supplementary methods. (F)
673 Heatmap of dynamically regulated proteases showing z-scored gene expression along
674 pseudotime. Genes were ordered using the hierarchical clustering indicated by the
675 dendrogram. Colorbar indicates the Louvain cluster from (E) for each gene. (G)
676 Expression trends of specific genes along pseudotime. Genes are scaled by their
677 maximum expression value rather than z-scoring as in the heatmap. (H) Diffusion map
678 colored by z-score scaled expression of specific genes. The data represents cells pooled
679 from 4 individual mice.
680
681 **Figure 4. P1 peritoneal cells exhibit potential to form basophils and mast cells.** (A-
682 B) May-Grünwald Giemsa staining of primary and in vitro cultured P1 cells and mast
683 cells extracted from the peritoneal cavity. Ba, basophil; MC, mast cell. Two or seven
684 independent experiments revealed the morphology of primary P1 cells and mast cells,
685 respectively. (C) Flow cytometry gating strategy to identify basophils and mast cells
686 cultured from primary P1 cells and mast cells. (D) Quantification of cell type output
687 following bulk-culture and flow cytometry analysis of P1 cells and mast cells. Pooled
688 data from 4 independent experiments per population are shown. The means and SEMs
689 are shown. (E) Principal component analysis of the flow cytometry reference dataset,
690 provided in Figure 1C, colored by cell type. (F) Projection of index-sorted cells into the

691 principal component space of the reference dataset. The point size represents log_{10}-
692 transformed colony size and the colors represent colony type following cell culture.
693 Panel F shows data pooled from 2 independent experiments. The cells were cultured
694 with IL-3 and stem cell factor. (G) UMAP visualization of the Human Cell Atlas single-
695 cell transcriptomics data colored by cell type. Identity of the clusters was assigned
696 based on established marker genes. (H) UMAP visualization of the basophil-mast cell
697 differentiation trajectory colored by cluster or expression of different genes.