HHS Public Access

Author manuscript

Nat Immunol. Author manuscript; available in PMC 2024 January 17.

Published in final edited form as:

Nat Immunol. 2023 August ; 24(8): 1295–1307. doi:10.1038/s41590-023-01549-3.

ThPOK is a critical multifaceted regulator of myeloid lineage development

Jayati Basu[1,2,#], **Andre Olsson**[3,4,5], **Kyle Ferchen**[3,4], **Elizaveta K. Titerina**[2], **Kashish Chetal**[5], **Emmanuelle Nicolas**[1], **Philip Czyzewicz**[1], **Dmitry Levchenko**[1], **Lu Ge**[1], **Xiang Hua**[1], **H. Leighton Grimes**[3,4,5], **Nathan Salomonis**[3,4,5], **Dietmar J. Kappes**[1,#]

[1]Fox Chase Cancer Center, Philadelphia, PA, USA.

[2]Department of Inflammation and Immunity, Lerner Research Institute, Cleveland Clinic, Cleveland, OH, USA.

[3]Division of Immunobiology, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio 45229, USA

[4]Division of Biomedical Informatics, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio 45229, USA.

[5]Department of Pediatrics, University of Cincinnati, Cincinnati, OH, USA.

Abstract

The transcription factor ThPOK (encoded by *Zbtb7b*) is well known for its role as a master regulator of CD4 lineage commitment in the thymus. Here, we report an unexpected and critical role of ThPOK as a multifaceted regulator of myeloid lineage commitment, differentiation and maturation. Using reporter and knockout mouse models combined with scRNA-Seq, progenitor transfer and colony assays, we show that ThPOK controls monocyte-DC versus granulocyte lineage production during homeostatic differentiation, and serves as a brake for neutrophil maturation in granulocyte lineage-specified cells through transcriptional regulation of lineage-specific transcription factors and RNA via altered mRNA splicing to reprogram intron retention.

[*]Send correspondence to: 215-728-5374 dietmar.kappes@fccc.edu, BASUJ@ccf.org, or 650-576-1646 Nathan.Salomonis@cchmc.org.
[#]contributed equally as senior authors
* # Jayati Basu
Department of Inflammation and Immunity, Lerner Research Institute. Cleveland Clinic, Cleveland, OH, USA.
* Nathan Salomonis
Division of Biomedical Informatics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA
*# Dietmar Kappes
Fox Chase Cancer Center, Philadelphia, PA, USA

Author Contributions Statement: JB and DJK conceived the idea. JB, DJK, and HLG designed the experiments. JB, AO, KF, EKT, KC, EN, PC, DL, LG, XH, NS and DJK performed research. JB, NS, HLG and DJK analyzed data. JB and DJK wrote the MS. NS and HLG reviewed and edited the manuscript.

Competing Interests Statement: The authors declare no competing interests.

Introduction:

Neutrophils and monocytes are key innate immune cells which provide the initial response to pathogens. Inflammatory cytokines and active infection modulates myelopoiesis to enhance the production of specific myeloid subsets (especially neutrophils and monocytes)[1–3 4,5 6].

Hematopoietic lineage choice and differentiation require dynamic changes in gene expression, under tight regulation by lineage-specific transcription factors (TFs), whose interplay can promote or antagonize distinct fates. Our understanding of the TFs that are involved in myelopoeisis and the precise stages at which they function remains rudimentary. Our prior scRNA-Seq studies, coupled with colony assays, identified a novel mixed-lineage intermediate (MultiLin) state in which collapsing hematopoietic stem cell/progenitor (HSCP) gene expression is coupled with opening of early myeloid enhancers by PU.1 and Cebpa. We demonstrated that exit from the MulitLin state can occur via a metastable intermediate exhibiting simultaneous low-level expression of the antagonistic TFs Irf8 and Gfi1, and identified complex cross regulation between these factors and other TFs (Ets1, Per3, Klf4, Zeb2 and Irf5)[7]; however, the complete gene regulatory network (GRN) remains unknown.

The transcription factor ThPOK (encoded by *Zbtb7b*) is well-known for its critical role at multiple stages of T cell development, including development of iNKT cells, selected γδ T cell subsets, and most prominently CD4 T cells[8–11], and has also been implicated as an oncogene during thymopoeisis[12]. However, we recently found that ThPOK expression is not restricted to T cells (ImmGen.org database), so that its role in development and differentiation of other hematpoietic lineages remains an open question. Here we show an unexpected role of ThPOK in myelopoiesis. Using reporter and knockout mouse models combined with in depth single-cell gene expression and splicing analyses, *in vivo* progenitor transfer and clonogenic assays, we demonstrate that ThPOK is a major regulator of homeostatic myeloid lineage commitment, differentiation and maturation, with ThPOK deficiency resulting in homeostatic neutrophilia. Mechanistically, we find that ThPOK regulates crucial TFs required for monocyte/DC vs granulocyte lineage choice as well as RNA binding proteins that control alternate mRNA splicing networks to impart multilayered control over the myeloid lineage differentiation program.

RESULTS

ThPOK deficiency causes homeostatic neutrophilia.

Multiple gene expression databases indicate that ThPOK mRNA is widely expressed among both hematopoietic and non-hematopoietic cell types, including myeloid subsets (Immgen.org; BioGPS.org databases). Utilizing a previously described 15kb ThPOK-GFP reporter transgene[9], containing all known regulatory cis elements from the endogenous *Zbtb7b* locus (Fig. 1A), we detected high GFP expression in all neutrophils, and many Ly6C+ classical and Ly6c- non-classical monocytes from peripheral blood (Fig. 1B).

Next, we asked whether ThPOK expression is necessary for myeloid homeostasis. Indeed, the number and proportion of myeloid subsets in blood, spleen and bone marrow (BM) were significantly altered in ThPOK−/− mice (Fig. 1C, 1D). Strikingly, absolute neutrophil numbers were increased in blood, spleen and BM of ThPOK-deficient compared to WT mice (Fig. 1C, Extended Data Fig. 1A, B). The absolute number of mature (CD101+)[13] neutrophils was also significantly higher in blood, and also tended higher in spleen, lung, BAL, liver and kidney (Extended data Fig. 1b). Recent single cell studies showed that extramedullary neutrophils comprise 3 transcriptomically distinct subsets, namely G5a (IFIT1^{-}CXCR4^{lo}), G5b (IFIT1^{+}) and G5c (IFIT1^{-}CXCR4^{hi})[14]. The absolute number of G5a neutrophils was increased in blood, spleen lung, liver and kidney of ThPOK−/− versus WT mice (statistically significant in blood and spleen). Ly6g+CD11b+ blood neutrophils from ThPOK−/− mice exhibited characteristic mature multi-lobulated morphology (Extended data Fig. 1C), and showed no major differences in surface markers compared to WT neutrophils, except for reduced CD62l expression (Extended data Fig. 2A, Fig.1E). We noted a significant decrease in absolute number of dendritic cells (DC) in BM and spleen of ThPOK−/− mice compared to WT (Extended data Fig. 1E). Also, the absolute number of classical (Ly6c^{hi}) monocytes was increased in BM, but not in blood or spleen of ThPOK−/− versus WT mice (Extended data Fig. 1E), although the absolute cell number of total CD11b+ Ly6g- monocytes was not significantly altered (in BM).

Next, we tested whether ThPOK deficiency alters neutrophil function. Phagocytosis, reactive oxygen species (ROS) production, degranulation (MPO release) and most of the cytokines and chemokines tested in respose to in vitro PMA stimulation were not significantly affected (Extended data Fig. 2A). However, ThPOK-deficient neutrophils, produced significantly less MIP1a ($p<0.001$), MIP1b ($p=0.03$), and MCP1 ($p=0.01$) upon stimulation (Fig. 1F), and released significantly ($p<0.001$) higher total and NET-associated elastase compared to WT neutrophils (Fig. 1G; Extended data Fig. 2A). We next tested neutrophil function *in vivo* in response to pathogen challenge. First, we employed subcutaneous fluorescently labelled *E. coli* bioparticle challenge using the subcutaneous air pouch model[15]. We found that ThPOK−/− and WT neutrophils exhibited similar capacity for homing to airpouch, phagocytosis of bacteria, myeloperoxidase, NET, elastase release and ROS production (Extended data Fig. 2B). Secondly, we used an *in vivo* model of sepsis induced by i.p. injection of cecal bacteria[16]. Strikingly, ThPOK-deficient mice showed substantial protection compared to WT mice in terms of disease progression and longevity (Fig. 1H). Total neutrophil infiltration into the peritoneal cavity was significantly increased in ThPOK−/− compared to WT mice, consisting predominantly of the G5a subtype in ThPOK−/− mice (versus G5b in WT) (Extended data Fig.2C). Peritoneal neutrophils were predominantly immature (CD101−) under both sham and sepsis conditions regardless of genotype. Myeloperoxidase levels, NET, elastase and ROS production were not significant altered in ThPOK−/− versus WT mice under sepsis condition in peritoneal cavity or blood. However, levels of proinflammatory cytokines, IL-1b, IL-6 and TNFa were significantly reduced, while the level of anti-inflammatory IL-10 was increased in peritoneal lavage of ThPOK−/− versus WT mice undergoing sepsis (Extended data Fig 2E). Furthermore, serum level of chemokines, MIP1a, MIP1b and MCP-1 were reduced in ThPOK-deficient compared to WT mice (Extended data Fig. 2E). Finally, ThPOK-deficient mice injected with

cecal bacteria lacked focal areas of acute pleuritis and peritonitis (as evidenced by acute exudate with PMN, fibrin and necrotic materials within peri-kidney area) compared to WT mice (Extended data Fig. 2F). Overall, these results indicate that ThPOK−/− mice show increase number of neutrophils with normal capacity to control bacteria, but diminished inflammatory capacity.

To test whether ThPOK might regulate myeloid development, we evaluated ThPOK-GFP expression in myeloid committed $Lin^{-/low}$ $cKit^{+}$ (LK) progenitors from bone marrow (BM) via flow cytometry. Indeed, ThPOK-GFP expression is already detectable in most MDPs and CMPs, with somewhat higher levels in Flt3+ $CD115^{lo}$ CMP relative to $Flt3^{-}CD115^{-}$ CMP progenitors (Extended data Fig. 3A). Myeloid GMP progenitors show substantially increased GFP levels compared to CMPs (Extended data Fig. 3A). Given that neutrophils arise from progenitors within the GMP gate, and express the highest ThPOK-GFP levels, we dissected ThPOK expression during neutrophil development in greater detail using our prior developed subsetting strategy, based on CITE-Seq (Cellular Indexing of Transcriptomes and Epitopes by Sequencing) analysis[17,18] (Extended data Fig. 3 B, C). This approach previously identified successive stages during normal granulocyte lineage specification and commitment in mouse BM on the basis of their distinct transcriptional programs, and nominated stage-specific surface markers. In particular, we found that granulocytic progenitors undergoing lineage specification are enriched within a Lin-CD16/32+CD34+ gate (“modified” GMP gate) and can be subdivided by relative expression of Ly6c, CD11b, Ly6g, CD117 and VCAM. The earliest granulocytic lineage-specifying progenitors can be identified as CD11b- Ly6g- CD115- CD117+ VCAM+ Ly6c- (proNeu1), whereas those completing specification are defined as CD11b-Ly6g-CD115-CD117+VCAM+ Ly6c+ (proNeu2) (Extended data Fig. 3B). Granulocytic lineage commitment occurs in cells defined as CD11b+CD115-CD117+VCAM+ Ly6c+ with progressive increases in expression of Ly6g marking successive preNeu1, preNeu2/3 and immature Neutrophil (immNeu) stages. Using these markers, we developed a 10-color gating strategy to subset neutrophil precursor populations according to their transcriptional program (Extended data Fig. 3B). Using this cytometry panel[17] we tracked ThPOK-GFP reporter expression in each of these stages of granulopoiesis (Fig. 1I; Extended data Fig. 3B, 3C). Interestingly, we find that ThPOK expression initiates during the earliest HSCP-like progenitor stage of neutrophil specification and then increases with granulocyte commitment and maturation (preNeu1 through immNeu) (Fig. 1I) (note that each subset also contains a GFP- population, which could reflect random silencing of the transgene in a fraction of cells due to position effect variegation). MP and cMoP subsets, which give rise to monocytes, exhibit at least 2-fold lower ThPOK levels than the peak levels in Neutrophil progenitors (Fig. 1I). ThPOK-GFP protein and mRNA expression in granulocyte progenitors correlates well with endogenous ThPOK mRNA expression in sorted subsets (Extended data Fig. 3D), except for the immNeu population (in which GFP protein expression persists after dropoff in ThPOK and GFP mRNA, which may reflect long half-life of GFP protein). As an independent approach, we dissected ThPOK expression in the modified-GMP gate using an unsupervised clustering method (viSNE)[19] based on the same set of surface markers. viSNE analysis identifies 7 distinct ThPOK-expressing populations, pops A-G (Extended data Fig. 3C, upper left panel). Consistent with our previous results, we find several populations that show positive

correlation between elevated ThPOK levels and expression of markers associated with granulopoiesis, i.e. CD11b (pops A, G), Ly6g (pop A), VCAM (pop B,D and E), and c-Kit (pop B , E). In addition, we note that populations C and F exhibit high CD115 expression, characteristic of monocyte/DC specified population. In summary, these data indicate that ThPOK is first significantly induced in immature cKit+ myeloid precursors (HSCP like) and increases with granulocytic commitment and maturation.

Next, to evaluate whether neutrophilia observed in ThPOK deficient animals reflects a cell-intrinsic effect, we transferred equal numbers of CD45.2+ CD45.1+ ThPOK+/+ with CD45.2+ CD45.1- ThPOK+/+ or CD45.2+ CD45.1- ThPOK−/− BM cells into lethally irradiated congenic CD45.2- CD45.1+ recipient mice (Fig. 1J). Post-transplant analysis revealed an enhanced contribution of ThPOK−/− (compared to ThPOK+/+) cells to the whole CD11b+ whole myeloid compartment and neutrophil compartment (CD11b+ Ly6g-) in peripheral blood and in BM (Fig. 1J, Extended data Fig. 3E), indicating that enhanced neutrophil production reflects a BM stem and progenitor cell-intrinsic effect of ThPOK-deficiency. Of note, the above analysis was performed at 6 months post transplantation which reveals most clearly the progeny of long-term hematopoietic stem cell (LT-HSC), because the progenitor pool and progeny of short-term HSC are by then exhausted[20]. We also noted that total ThPOK−/− KO WBCs were relatively reduced in BM chimeras. The basis for this phenomenon is unclear but could suggest that ThPOK KO progenitors engraft more poorly. Finally, to evaluate whether the neutrophil bias in ThPOK-deficient mice might reflect disproportionate proliferation of neutrophil progenitors, rather than an intrinsic differentiation bias, we carried out *in vivo* BrDU labelling. This showed that ThPOK deficiency does not cause selective expansion of early neutrophil progenitors (HSCP1 like and ProNeu1) versus monocyte progenitor (MP-cMoP) (Extended data Fig. 3F).

ThPOK regulates neutrophil versus monocyte lineage choice.

To explore a potential role for ThPOK in myeloid differentiation, we performed detailed FACS analysis of hematopoietic progenitor subsets in the BM from ThPOK−/− versus WT mice. Initial analysis of Lin-Kit+ myeloid progenitor cells revealed that ThPOK deficiency resulted in decreased proportions of CD34− CD16/32− MEP progenitors, but no significant change in CMP and GMP progenitors (Fig. 2A,B, Extended data Fig.4A), and within the CMP gate, the proportions of MDP, Flt3+CMP and Flt3- CMP subsets were unchanged (Extended data Fig.4B). Interestingly, we noted an atypical Kit+ CD34^{lo} CD16^{hi} population (akin to myelocytes) in ThPOK−/− BM (Fig. 2A, blue arrow), further characterized as CD11b^{hi} Ly6c^{hi} Ly6g$^{lo/+}$, but lacking expression of the monocyte/macrophage marker CD115, consistent with a granulocytic rather than monocytic lineage origin (Fig. 2C; Extended data Fig. 4C). This c-Kit+ CD34^{lo} CD16^{hi} population is also detected in haplosufficient ThPOK+/− BM, albeit to a lesser degree, indicating sensitivity to ThPOK gene dosage (Extended data Fig. 4D). Since WT CD16/32^{hi} cells that are downmodulating both cKit and CD34 represent maturing neutrophils, this suggests that limiting ThPOK may sustain cKit expression on maturing neutrophils.

To better delineate the effect of ThPOK deficiency on granulocyte-monocyte lineage development, we analyzed ThPOK−/− BM using our recently described flow panel and

gating strategy[17] (Extended data Fig. 3B). This revealed that ThPOK deficiency greatly increased absolute cell numbers of proNeu1 cells, suggesting enhanced granulocyte specification, as well as preNeu2/3 and immNeu cell frequency/numbers, indicating enhanced neutrophil maturation (Fig. 2 D, E). In contrast, we observed a two-fold decrease in frequency of MP and cMoP monocytic precursors. However, absolute number of MP and cMoP monocytic precursors were not significantly altered, and mature classical monocytes were actually increased in BM (Extended data Fig. 1E). This apparent contradiction may be explained in part by the fact that ThPOK-deficient monocyte progenitor (MP-cMoP) exhibit increased proliferative capacity (Extended data Fig. 3F). Finally, to evaluate whether the effect of ThPOK loss on myeloid development is cell-intrinsic, we employed an *in vitro* colony formation (CFU) assay using flow cytometrically-sorted hematopoeitic progenitors. Indeed, *ThPOK−/−* Lin- progenitors exhibited a substantial increase in CFU-GM myeloid colony production both in primary and particularly evident after secondary secondary plating, revealing cell-intrinsic myeloid bias plating (Extended data Fig. 4D, E). Collectively, these data reveal a multifaceted cell-intrinsic effect of ThPOK loss on myeloid lineage development.

Role of ThPOK in myeloid lineage specification/maturation.

To decipher the role of ThPOK in myeloid differentiation from an unbiased genome-wide perspective, we performed single-cell RNA-Seq analysis on over 46,000 cells (Table S1, Extended data Fig. 3A–F). However, because cKit (CD117) expression was deregulated on maturing ThPOK-deficient neutrophils, it was important to identify potential progenitor marker misexpression and to correctly align progenitor populations via CITE-Seq between ThPOK−/− and wt progenitors (Kit+, CD34+, CD16/34+/−, referred to here as C-GMP: Extended data Fig. 5A). In addition to single-cell transcriptomes, this analysis quantified levels of over 60 different cell surface antigens by virtue of antibody derived tags (ADTs). To appropriately place cell populations within the framework of prior knowledge, we produced and aligned these data to an independent CITE-Seq data set of broad wild-type cKit+ progenitor cells using cellHarmony software[21] (see Methods) (Extended data Fig. 5A–F). CITE-Seq ADT expression based on Flow Cytometry corresponded precisely to these prior defined populations (Extended data Fig. 5G), allowing unambiguous identification of HSCP/MPP (cKiT+/CD38−), multilineage progenitors (mixed lineage surface markers), monocyte and neutrophil subsets[17], with no additional novel populations revealed by unsupervised clustering (Extended data Fig. 5C). While the global cellular and cell-surface marker landscape of these two captures was largely consistent, ThPOK-deficient cells exhibited marked distortions in relative cell state distribution based on their transcriptome (Fig. 3 A,B). In particular, ThPOK deficiency led to an increase in bi-potential monocytic/granulocytic progenitors (IG2), and neutrophil lineage specified proNeu-1 subsets (Extended data Fig. 5F). ThPOK mRNA was expressed most highly and broadly at terminal stages of neutrophil differentiation (preNeu3 and immNeu), while cMoPs, monocyte and early neutrophil progenitors showed broad but low-level expression (also, a few HSCP and Multi-Lineage progenitors showed high expression) (Extended data Fig. 5D, E). To infer cell trajectories we applied the software Velocyto[22] to all genome aligned reads in these scRNA-Seq captures. RNA velocity predictions suggest a block in the differentiation of MDP to DC (Fig. 3A, B) and an increased rate of terminal neutrophil differentiation from pre Neu1

to immNeu, supporting our Flow based data (Fig. 2D, E). Interestingly, this analysis also predicted a redirected trajectory between proNeu-2 and preNeu-1 populations in ThPOK−/− BM (Fig. 3 A, orange arrows).

To understand the molecular basis for the altered cellular trajectory due to ThPOK deficiency, we examined differential gene expression in ThPOK−/− compared to matching wild-type progenitor populations using cellHarmony[21] (Table S2). CellHarmony organizes gene expression changes into predominant patterns of regulation across cell populations. This analysis revealed that up-regulated differentially expressed genes (DEGs) tend to coincide with genes positively associated with neutrophil specification (e.g., *Elane, Ctsg, Mif, Prtn3, Vcam*) or commitment (e.g., *Cebpe, Hmgb2, S100a11, Hp, Birc5*)[17], while downregulated DEGs tend to impact broad steps in hematopoiesis, including lineage specification and commitment or are negatively associated with neutrophil progenitor commitment (preNeu-1 through 3) (e.g., *Irf8, Zeb2, Flt3, Runx1, Foxp1*) (Fig. 3C–E, Table S2, S3). This analysis further highlighted that ThPOK-deficient MDPs also induce genes associated with MultiLin-2 and neutrophil specification (proNeu-1) programs, while ThPOK-deficient proNeu-1 cells prematurely induce a preNeu-1 neutrophil commitment program (Fig. 3E). Intersecting ThPOK ChIPseq (GSE116506) with our scRNAseq analysis showed that 82% of DEGs are direct targets of ThPOK (Table S4). Next, to decipher the molecular basis for the block in MDP to DC trajectory due to ThPOK loss, we compared the gene expression profile of ThPOK−/− MDP population to matching WT counterparts. This revealed that ThPOK loss resulted in reduction in the expression of critical TFs essential for monocyte and DC development[23] (e.g. *Irf8* and *Runx1)* (Fig. 3D). Gene Ontology enrichment analyses also revealed downregulation of genes related to phagocytosis and antigen presentation (Fig. 3F). As expected, ThPOK loss strongly perturbed the transcriptional profile in early-stage myeloid specified progenitors, i.e. mixed-lineage MultiLin and metastable (IG2) states. Interestingly, we noticed downregulation of TFs characteristic of stemness, e.g. *Runx1*, *Meis1* and *Bcl11a*[24], in MultiLin stage, suggestive of collapsing stemness program in absence of ThPOK. At IG2 we noticed downregulation of monocyte specification genes including *Irf8*, and *Zeb2*. Upregulated DEGs in ThPOK−/− progenitors include several transcription factors required for neutrophil specification, including *Id1* (upregulated in myeloproliferative disease), *Cebpb* and *Cebpe* (required for granulopoiesis). The leukemia transcriptional co-repressor *Cbfa2t3*[25,26] is also upregulated in ThPOK−/− Multilin progenitors. We conclude that ThPOK deficiency results in widespread transcriptional reprogramming beginning in early stage of myeloid specified progenitors and causes pronounced skewing towards neutrophil specification and commitment programs.

ThPOK loss profoundly impacts the RNA regulon.

Previous studies show that splicing factors and RBPs play a central role in post-transcriptional gene regulation to shape the stage-specific transcriptome during normal hematopoiesis[27,28]. In particular, alternate intron-retention contributes to regulation of normal granulopoeisis[29], concordant with the fact that downregulation of splicing regulators is linked to granulocyte differentiation and myeloid leukemia[30]. In addition to confirming downregulation of genes associated with RNA regulation (mRNA splicing, RNA stability

and degradation) in normal neutrophil commitment (proNeu and preNeu), we find substantial differential expression of these factors in ThPOK−/− progenitors (up- and downregulated) (Fig. 4A, B; Extended data Fig. 5 G–N and Table S2). In particular, we observed downregulation of numerous splicing silencers including HNRNPs[27], PTBPs[28], and core spliceosomal components, such as Srsf1, Srsf2, Sf3b1, Ptbp1 (Fig. 4B). These genes also included splicing factors implicated in hematological malignancies, such as Hnrnpk and Mbnl1, the latter of which was shown to suppress intron retention[30]. Therefore, we next examined whether ThPOK deficiency causes altered RNA splicing. Since most high-throughput nanodroplet technologies (e.g.10x Genomics Chromium 3' platform) do not readily enable the detection of alternative splice variants, we applied a Smart-Seq2 single-cell protocol (Fluidigm C1), which provides relatively unbiased deeper sequencing of isoform fragments and meaningful splice-event quantification using our previously applied MultiPath-PSI workflow in AltAnalyze. We first identified unique splicing events associated with specific stages of normal granulopoiesis. After projecting labels from the 10x Genomics cKit data onto the Fluidigm transcriptome of WT cells (Extended data Fig. 6 A–B), we observed a progressive increase in alternative splicing in proNeu-1 through preNeu-1 stages. In agreement with prior reports[29], we noted a marked change in intron inclusion between preNeu3 and immNeu stages (>70%) (Fig. 4C, Table S5). Comparison of alternative splicing events (ASEs) between ThPOK−/− and WT progenitors during neutrophil specification and commitment revealed that ThPOK deficiency causes a premature increase in intron retention at the MultiLin state, which increases progressively with granulocyte specification (Fig. 4D, Table S6). While <33% of ASEs in WT proNeu1 reflected retained introns, this frequency was remarkably increased in ThPOK-deficient cells (Fig. 4D). Next, we assessed how altered splicing in ThPOK-deficient progenitors impinged on the differentiation process. We found that ThPOK loss in proNeu1 progenitors promoted two atypical splicing programs, that were respectively associated with accelerated maturation or regression to a stem/multipotent progenitor program (Fig. 4E–G; Extended data Fig. 6 C, D). Interestingly, some of the major ASEs impacted were key transcriptional regulators, including Hmga1 and Ezh2, both implicated in normal stem-progenitor functions and malignant hematopoiesis[31,32] (Fig. 4 E–H: Extended data Fig. 6F). Notably, a Hmga1 ASE enriched in ThPOK−/− proNeu1 cells is predicted by AltAnalyze to result in a gain of a phosphoserine binding site in the protein, facilitating its interaction with Hipk2 and *Cdc2*. Further, an *Ezh2* exon-inclusion event enriched in ThPOK−/− proNeu1 cells is predicted to result in enhanced nonsense-mediated decay of Ezh2 mRNA (note that potential enrichment for this *Ezh2* exon-inclusion event in other ThPOK−/− subsets cannot be directly assessed due to insufficient cell numbers). In support of predicted Ezh2 degradation, we find decreased Ezh2 protein expression in ThPOK-deficient monocytic progenitors and proNeu1 cells, and strong downregulation of the H3K27Me3 mark (Fig.4H, Extended data Fig. 6G). Focussing on Ezh2 target genes (ChIP-seq analysis, GSE181873) that are upregulated (in light of its role in transcriptional repression) in ThPOK−/− vs. WT BM progenitors for any cell-population, we identified several genes implicated in neutrophil differentiation including Cebpa, Slpi and Rgcc (Extended data Fig. 6H). Importantly, biological processes affected by these alterations in transcript usage in ThPOK−/− progenitors include transcription co-factor activity, nonsense mediated decay, cell migration, as well as RNA splicing itself (Extended data Fig. 6 C–D), suggesting a novel role of ThPOK in physiological granulopoiesis via

maintenance of the RNA regulon. Deregulation of these processes due to ThPOK loss impacts transcriptional and cell cycle differentiation velocity, in line with observed changes in relative expression of ASE products.

ThPOK-deficient GMPs are biased towards the granulocytic lineage.

To directly assess the effect of ThPOK on neutrophil versus monocyte lineage potential we performed a bilineage colony formation assay on flow sorted GMP-gated progenitors (Lin- Kit+ CD34hi CD16hi) (Extended data Fig. 7A) from ThPOK-deficient versus WT mice. Strikingly, ThPOK$^{-/-}$ GMP progenitors produced a 10-fold increase in granulocytic colonies (CFU-G) compared to WT GMPs (60% versus 5% of total colonies), but a 3-fold decrease in granulocyte/macrophage colonies (CFU-GM) (Fig. 5A, B, C), supporting that ThPOK deficiency promotes neutrophil versus macrophage lineage differentiation. This granulocytic bias was not seen for similar cultures of sorted ThPOK-deficient CMP-gated (Flt3+CD115-) progenitors (Extended data Fig. 8A). Lineage choice and specification of progenitors *in vivo* depends on diverse ligand-receptor interactions, which are not necessarily reproduced under defined cytokine and ligand conditions provided *in vitro*. To test whether ThPOK-deficient GMPs undergo similar neutrophil lineage skewing *in vivo*, we adoptively transferred ThPOK-deficient and WT GMPs into non-irradiated recipient mice. Indeed, ThPOK-deficient progenitors gave rise to a 2-fold higher number of Ly6G+ neutrophils *in vivo*, and could not give rise to non-classical monocytes (Fig. 5D, Extended data Fig. 7B). However, unlike our *in vitro* cultures, ThPOK$^{-/-}$ but not WT progenitors gave rise to an unusual population of F4–80^{hi} CD11b+ Ly6g- CD115+ CD43+ macrophages (Fig. 5E, Extended data Fig. 7B middle and lower panels).

ThPOK-deficient MDP produce neutrophils *in vitro and in vivo*.

WT MDPs can differentiate to monocytes and dendritic cells but do not frequently produce neutrophils (when strictly gated on c-Kit+CD34+CD16/32-, Flt3+, CD115+). Our CITE seq analysis reveals profound downregulation of the monocytic lineage TFs Runx1, Irf8 and Zeb2, as well as upmodulation of the granulocytic TF Cebpe and other granulocyte-specific proteins (e.g. Cd63 and Nkg7) in ThPOK−/− MDPs (Fig. 3C, Extended data Fig. 6E). To directly test lineage potential of ThPOK-deficient MDPs, we cultured FACS-sorted MDPs (Extended data Fig. 7A) in methycellulose medium with supporting cytokines in which WT MDP only differentiate to monocyte or macrophage lineages. While WT MDPs produced only these 2 lineages, ThPOK−/− MDPs instead produced massive numbers of granulocytic colonies with far fewer monocytic colonies, similar to ThPOK−/− GMPs (Fig. 6A, B and Extended data Fig. 8B). FACS analysis of cells from MDP-derived colonies confirms that in contrast to WT MDP, ThPOK-deficient MDPs give rise predominantly to neutrophils and F4–80^{hi} macrophages, with reduced numbers of monocytes (Fig. 6C). Of note, the CD11b-Ly6g- population is strongly reduced in ThPOK−/− MDP-derived cultures, suggesting that ThPOK deficiency in MDPs imparts bias towards granulocyte differentiation. Next we tested the effect of ThPOK deficiency on myeloid differentiation under trilineage differentiation conditions, in which WT MDP can differentiate to DC in addition to monocytes and macrophages. Strikingly, ThPOK-deficient MDPs continued to produce massive numbers of granulocytic colonies with decreased monocytic DC-like colonies (Fig. 6D), and FACS analysis confirmed predominance of neutrophils in ThPOK-deficient versus WT MDPs

cultures (Fig. 6E). The few DCs that arose from ThPOK-deficient MDPs failed to fully upmodulate MHC II compared to DCs similarly cultured from WT MDPs (Fig. 6E). Overall, our *in vitro* lineage differentiation assays indicate that absence of ThPOK strongly promotes neutrophil differentiation from MDP, while severely impairing DC differentiation.

Next, we evaluated *in vivo* lineage choice of adoptively transfered MDP subsets. Consistent with our *in vitro* results, ThPOK-deficient MDPs differentiated predominantly towards neutrophils *in vivo* (Fig. 6F, left upper panels; Extended data Fig. 8), while failing to efficiently differentiate into either cDC or CD11b+ DC (Fig. 6F, left lower panels; Extended data Fig. 8C). Interestingly, ThPOK-deficient MDPs were able to give rise to classical Ly6c+ monocytes, while non-classical Ly6c-CD43+ and Ly6c-CD43- monocytes were reduced (Fig 6F, right center panels; Extended data Fig. 8C). Moreover, ThPOK-deficient MDPs gave rise to increased number of F4/80+ macrophages (Fig 6F, right center panels; Extended data Fig. 8C).

We previously showed that Irf8 and Zeb2 are crucial for monocyte/ granulocyte lineage choice[7] and here we find that ThPOK loss causes downregulation of Irf8 and Zeb2 (Fig. 3C, Extended data Fig. 6E). Therefore, we performed ChIP using Flow sorted CMP and GMP progenitor subsets from heterozygous Flag-tagged-ThPOK knock-in mice to ascertain whether ThPOK may directly regulate promoter or enhancers of *Irf8* and *Zeb2* loci in these progenitors. Indeed, we find that ThPOK binds to multiple *Irf8* regulatory elements, as well as the *Zeb2* promoter in CMP progenitors, and Enh2 of *Irf8* in GMP progenitors (Fig. 6G, center panels).

Collectively our data suggest that ThPOK is an important activator of the DC/Monocyte gene regulatory network (Fig. 6H), which conversely restricts neutrophil granulocyte potential, such that ThPOK deficiency leads to lineage infidelity and neutrophilia.

Discussion:

Using reporter and knockout mouse models combined with scRNA-Seq analysis and clonogenic assays, we have identified a pivotal role for the transcription factor ThPOK in myeloid lineage fidelity and homeostasis. In particular, the presence of ThPOK is required at: a) The dendritic cell-monocyte-granulocyte tri lineage decision point to control homeostatic lineage choice, and b) at late granulocyte differentiation stage to maintain quantal maturation of preNeu to immNeu.

While stage- and lineage-specific functions of ThPOK were previously reported during T cell development in the thymus, where it plays a critical role in directing/supporting development to multiple T cell lineages, its role in development of other hematopoietic lineages has not so far been addressed. Here, using ThPOK-GFP reporter mice, we show for the first time that ThPOK is expressed in all developmental stages during myelopoiesis, albeit with markedly different levels among different lineages/stages. We further show using ThPOK-deficient mice that ThPOK controls multiple myeloid developmental transitions. Of note, relative expression level of ThPOK is not a good predictor of its functional importance. Thus while very high levels of ThPOK during granulopoiesis are implicated for regulating

the velocity of neutrophil maturation, much lower levels at the MultiLin, IG2 and MDP stage are important for promoting the DC-monocytic fate.

Myelopoiesis is regulated by combinatorial TFs that instruct the precisely orchestrated gene expression programs controlling alternate lineage and differentiation fates. By intersecting single cell transcriptomic and ThPOK ChIP seq analyses we determined that ThPOK regulates lineage- and stage-specific developmental gene expression at the levels of transcription and alternative RNA splicing. Firstly, ThPOK is involved in activating the monocyte-DC lineage TFs Irf8 and Zeb2 and inhibiting granulocyte lineage determining TFs Cebpb, Cebpe at MultiLin2, IG2 and MDP stages, thus orchestrating homeostatic commitment of DC/monocyte versus neutrophil lineage. Secondly, ThPOK likely impacts RNA splicing through the transcriptional regulation of multiple RNA binding proteins. Recent studies find that developmental cell states can be distinguished based on RBP expression, indicating a key role in defining cellular identity[27,28]. RBPs are the central players in post-transcriptional regulation and are involved in physiological regulation of myelopoiesis[27,28,33]. The ThPOK-mediated gene regulatory network accordingly includes important RNA regulon components previously implicated in hematopoiesis[28,34] e.g. Ptbps, Srsf, Upf2 Msi2, Zfp36 family, Mttl3 etc.[34]. Interestingly, our comparison of alternative splicing events (ASEs) during normal neutrophil lineage specification and commitment finds a sequential increase in alternative splicing during proNeu1 to preNeu1 transition (specification to commitment), with the highest frequency of intron-retention observed at the immNeu stage (>70% intron-inclusion in immNeu versus preNeu3). In accordance, we previously showed that the majority of mRNA transcripts in human neutrophils, unlike other cell types, show evidence of retained introns from scRNA-Seq[17]. Surprisingly, ThPOK deficiency leads to premature initiation of alternative spliced transcripts at the proNeu1 stage, which are largely predicted to undergo nonsense-mediated decay. In line with the bioinformatic prediction we find increased ASEs of Ezh2, correlating with decrease in EZH2 protein and strong down modulation of the H3K27Me3 mark in monocytic progenitors and proNeu1 cells. Deregulated expression of several RBPs (Magoh, Zfp36l2, Upf2 etc.) upon ThPOK loss, which are linked with NMD, may add to the premature onset of mRNA decay in ThPOK-deficient proNeu1 progenitors. Our future studies will explore whether ThPOK is directly involved in mRNA splicing. Collectively, our results provide strong evidence that ThPOK-mediated transcriptional and splicing regulatory feedback loops velocity of cell state traversal during granulopoiesis, such that ThPOK acts as a physiological brake to maintain neutrophil homeostasis.

The use of next-generation sequencing technologies has increasingly highlighted the importance of RBPs in the pathogenesis of specific hematological malignancies. Our study reveals for the first time the unexpected role of ThPOK in multifaceted regulation of myeloid lineage commitment, differentiation and maturation via transcriptional regulation of TF and RBPs, making ThPOK an attractive target to develop novel AML therapeutic strategies. On the other hand, ThPOK appears to limit neutrophil production under stress and also promote a proinflammatory effect of neutrophils. Thus, ablation of ThPOK not only increases the functional neutrophil output upon sepsis induction, and also offers major protection against sepsis via down-modulation of proinflammatory cytokines, such that

targeting ThPOK may provide a novel strategy to combat sepsis, which is major health problem worldwide.

Methods:

Mice.

Wild-type C57BL/6J mice (CD45.2) and congenic CD45.1 mice (B6.SJL-*PtprcaPepcb*/BoyJ), were purchased from Jackson Laboratories. ThPOK-GFP transgenic mice[9] and ThPOK-GFP reporter knock-in mice[35] were described previously, and are on the C57BL/6J background. In these studies ThPOK reporter lines were validated by FACS and qPCR. Thus, reporter expression correlated precisely with ThPOK mRNA expression levels, including in CD4+CD8+ and CD4-CD8+ thymocytes known to lack ThPOK mRNA expression by qPCR[9]. FlagTagged-ThPOK mice were provided by I. Taniuchi (RIKEN, Japan). To generate ThPOK-GFP knockin mice, a Zinc Finger Nuclease (ZFN)-mediated genome engineering approach was used. In short, a pair of ZFN RNAs that recognize a target site upstream of Zbtb7b coding exon 1 were designed and generated by Millipore-Sigma (Genome Editing division). mRNAs encoding site-specific ZFNs (50 ng/μl) were introduced along with a homologous donor construct (2 ng/μl) containing the GFP cassette fused with the Cd3d 5'UTR or Flag tagged *ThPOK* exon1 and flanked by 1.5-kb 3′ and 0.8-kb 5′ homologous arms into one-cell mouse oocytes by standard pronuclear injection approach. Positive founder pups were identified using mutation-specific primers and mated to C57BL/6 mice to generate stable heritable knock-in lines. ThPOK GFP KI mice were genotyped for site specific eGFP integration using F: 5'-GCCTAAACCA GGACTCCAGT-3' and GFPR: 5'-CGGCGAGCTGCACGCTGCCGTCCTC-3' primer pairs and presence of WT allele was ascertained using F: 5'-ACATGAAAGGTGGTTTGGGA-3' and R: 5'-AGCTTGGAGTACAGCCGGCA-3' primer pairs. Homozygous ThPOK-GFP knockin mice lacked CD4 T cells in peripheral blood, thymus, lymph nodes and spleen, similar to other ThPOK-deficient lines. Flag tagged FhThPOK allele was genotyped by Fh F: 5'-CGATGACGACAAGGACTACAAA-3', R: 5'-TTGAAGTAGTGGCTACAGGC-3'. ThPOK-GFP transgenic mice was typed using GFPF: 5'-CCGCGGGGGCCTCTGGAAGAGCAGTC-3', GFPR: 5'-TTAATTAAGTCGAGGAATTCCGATCATATTC-3'. Mice were maintained at either Fox Chase Cancer Center, Cincinnati Children's Hospital Medical Center or Lerner Research Institute. All procedures were performed with IACUC approval (FCCC protocols 92–14, and 20–14; LRI 00002896). Animals used in all experiments were aged 6–12 weeks, and males and females were used in equal proportions as far possible (no difference was noted between males and females in any experiment). Animal care was in accordance with the National Institutes of Health (NIH) guidelines. Mice were maintained on a 12 h light:dark cycle, at 24 °C and 50% humidity.

Immune cell isolation/processing from tissues.

Broncho alveolar lavage: After euthanasia, skin from abdomen to neck was incised to expose thoracic cage and neck. Trachea was exposed by gently removing the muscle around the neck. ~10 cm-long nylon string was passed under the trachea. To expose the heart and the lungs by cutting the ribs carefully. 22 G × 1" Exel Safelet catheter was inserted into the

trachea and tied the catheter with trachea firmly with a nylon string[36]. Using 1ml syringe 1 ml of HBSS was injected and aspirated 5 times, ~ 3–4ml of fluid was collected. For all the mouse, BAL fluid volume was equalized to 5ml. Immune cells were collected by centrifugation at 500 × g for 10 min at 4 °C and total cell pellet was used for flow analyses.

Lungs: Lungs collected from the euthanized mouse after broncho alveolar lavage, washed in 5% FBS HBSS in Petri dish and minced with scissors. Minced organ was digested with 0.5mg/ml Collagenase type IV (Sigma, C5138) and 25U/ml DNase I (Sigma, D4263) at 37°C shaker incubator for 45 min. After incubation cells were filtered through 70μm cell strainer.

Liver: Part of the lower left lobe of the liver were collected and minced and digested with enzyme mix [0.5mg/ml Collagenase type IV (Sigma, C5138) and 25U/ml DNase I (Sigma, D4263)]) at 37°C shaker incubator for 45 min. After incubation cells were passed through 70μm cell strainer and spun at 400g for 5 min. The pellet was resuspended in 40% percoll and layered on top of 80% Percoll gradient (40%/80%). Separation was performed by centrifugation at 1100 × g for 25 min with slow acceleration and 0 deceleration. Immune cells were collected from the interphase and used for flowcytometry.

Kidney: Left mouse kidney was removed after euthanasia. 1/3 of the kidney was minced with razorblade in Petri dish and transferred into the tube with 10ml RPMI and 300μl of digestion mix containing 20mg/ml Collagenase type II (Sigma, C6885) and 50mg/ml DNase I (Sigma, D4263). Samples were incubated in 37°C bath for 45 min with shaking every 10 min. After incubation samples were vortexed and filtered through 70μm cell strainer. Single-cell suspension was used for flow cytometry analysis.

Peritoneal exudate: After euthanasia, abdomen was soaked with 70% alcohol and small incision was made along the midline with sterile scissors. Abdominal skin was retracted manually to expose the intact peritoneal wall. 10 ml of HBSS was injected in the peritoneum using 24-G needle, and masseged gently for 15 seconds. Using the same syringe and 22G needle, aspirated the fluid from peritoneum. Approximately 8 ml fluid recovered per mouse[37]. The immune cells were obtained by centrifugation at 400g for 5 min and the supernatant was frozen at −80°C for measurement of MPO, Elastase, NET, cytokine and chemokine.

Blood neutrophil isolation.—Highly purified neutrophils from peripheral blood were isolated by immunomagnetic negative selection using EasySep™ Mouse Neutrophil Enrichment Kit (Stemcell technology, Cat:19762) according to the manufacturer's protocol. Purity of CD11b+Ly-6G+ neutrophils was assessed by flow cytometry and found to be 88–90% pure.

Subcutaneous air pouch model

The air pouch model was described and validated previously[15]. Briefly, Mice were anesthetized with isoflurane and 3 ml of air was injected subcutaneously to create a dorsal air pouch with a top-up of air 3 d later. After 5 days of the creation of the air pouch, mice

were injected with 1ml of PBS (control) or challenged with 100μg of pHrodo conjugated E. coli BioParticles (Invitrogen, Cat# P35366) directly into the pouch. Challenged mice were culled 14 h after injection. Recruited cells within the air pouch were harvested by carefully washing the pouch with 3ml of HBSS. Immune cells were pelleted by centrifugation at 400g for 5min, used for phagocytosis and neutrophil characterization using flowcytometry. Serum and plasma were preserved for other analyses.

Cell staining for flow cytometry and FACS sorting.—The following fluorochrome-conjugated antibodies were used for flow cytometric analysis and FACS sorting: CD34 (clone RAM34), FcgR (CD16 and CD32; clone Ab93); Sca-1 (clone 108113), c-Kit (CD117; clone 2B8), Ly6C (clone HK1.4), Ly6G (clone 1A8), F4/80 (clone BM8), CD115 (clone AFS98), CD317 (clone 927), CD11c (clone N418), CD86 (clone GL-1), B220 (clone RA3–6B2), I-Ab (clone AF6–120.1), CD8a (clone 53–6.7), CD45.1 (clone A20), CD45.2 (clone 104) from BioLegend; CD43 (clone S7), CD11b (clone M1/70), CD135 (clone A2F10.1) from BD Biosciences. Where possible, non-specific antibody binding was prevented by prior incubation with Fc block (anti-CD16 and anti-CD32). Fc blocking was not possible for identification of progenitors by FcgR expression, so cells were stained for FcgR prior to staining with other antibodies. An LSRFortessa instrument (BD Biosciences) was used for flow cytometry and data were analyzed with FlowJo.

IFIT1 intracellular staining for Flow Cytometry.—Cells were incubated with fluorochrome-conjugated antibodies for surface staining with (Ly6g, CD11b, CD101, CXCR4) for 10 min at 20°C in the dark and washed twice with PBS with 5% FBS. Intracellular staing was performed using Foxp3 / Transcription Factor Staining Buffer Set (eBioscience, cat #00-5523-00). Cells were fixed and permeabilized for 20 min at room temperature in the dark and washed twice. Cells were stained with rabbit polyconal anti-IFIT1/p56 antibody (ABF117, Sigma-Aldrich, Cat# 621832) 1:75 dilution for 30 min at 25° C in 200uL buffer. Cells were washed twice and then resuspended in 200uL buffer containing 1:500 secondary goat anti-rabbit-Alexa Flour 647 antibody (Invitrogen Cat # A-21246) for 30 min. Cells were washed twice, resuspended in 200uL buffer and analyzed on LSRFortessa Flow Cytometer.

Cell preparation for Flow cytometry and cell sorting.—Peripheral blood was collected into EDTA-coated tubes. For BM isolation, hind limb bones (femurs, tibias, and the iliac crest) were harvested immediately after euthanasia and stored in cold FACS buffer (1% FBS, in DPBS) under sterile conditions. Bones were flushed using a syringe for transplantation,or crushed using a mortar and pestle for all other applications, then passed through a 40 μM cell strainer (BD Biosciences) to obtain single cell suspensions for downstream applications. To enrich for murine stem/progenitor populations, freshly isolated bone marrow (BM) cells were incubated with CD117 Microbeads or with components of the Mouse Lineage Depletion kit and separated on an AutoMACS Pro separator (Miltenyi) according to manufacturer specifications.

Whole Peripheral blood, spleen and BM analysis.: Prior to analytical flow of blood, spleen and BM cells, erythrocytes were lysed using ACK buffer (Gibco) and washed

in FACS buffer. Peripheral blood cells were stained in FACS buffer with a mix of antibodies: CD16/CD32 (clone 2.4G2, Becton, Dickinson and Company), CD3 (clone 145–2C11, Becton, Dickinson and Company), CD45R (clone RA3–6B2, Becton, Dickinson and Company), CD11b (clone M1/70, BioLegend), and Ly6g (clone 1A8, BioLegend), F4/80 (clone BM8), CD115 (clone AFS98), CD317 (clone 927), CD11c (clone N418), Ly6C (clone HK1.4), CD43 (clone S7). For analysis of transplant chimaeras, cells were also stained with CD45.1-BV605 (clone A20, Biolegend) and CD45.2-AlexaFluor700 (clone 104, Biolegend).

Myeloid progenitor analysis.: Miltenyi AutoMacs CD117-enriched bone marrow cells were stained for 1 h at 20°C protected from light with a mix of biotin-conjugated antibodies (including anti-CD3e (clone 145–2C11, BioLegend), anti-CD4 (clone RM4–5, Thermo Fisher), anti-CD8 (clone 53–6.7), anti-CD19 (clone 6.D5, BioLegend), anti-CD127 (clone B12–1, BD), anti-B220 (clone RA3–6B2, BioLegend), and anti-Ter119 (clone TER-119, BioLegend)), as well as PE-Cy7-conjugated anti-SCA-1 (clone D7, BD), Brilliant Violet 785-conjugated anti-LY6C (clone HK1.4, BioLegend), PerCp Cy5.5-conjugated anti-LY6G (clone 1A8, BioLegend), Brilliant Violet 421-conjugated anti-CD34 (clone RAM34, BD), APC-conjugated anti-CD117 (clone 2B8, BioLegend), and Brilliant Violet 605-conjugated anti-CD115 (clone T38–320, BD), PE- Cy5 conjugated CD135 and PerCP-eFluor710-conjugated anti-CD16/CD32 (clone 93, Thermo Fisher). Cells were then washed twice with FACS buffer and incubated for 15 min at 20°C with streptavidin-APC-Cy7 (BD). Cells were washed once in FACS buffer, and resuspended in FACS buffer for analyses.

Cell sorting.—AutoMacs CD117-enriched BM cells were stained for 1 h at 20°C protected from light with a mix of biotin-conjugated antibodies (including anti-CD3e (clone 145–2C11, BioLegend), anti-CD4 (clone RM4–5, Thermo Fisher), anti-CD8 (clone 53–6.7), anti-CD19 (clone 6.D5, BioLegend), anti-CD127 (clone B12–1, BD), anti-B220 (clone RA3–6B2, BioLegend), and anti-Ter119 (clone TER-119, BioLegend)), together with PE-Cy7-conjugated anti-SCA-1 (clone D7, BD), Brilliant Violet 785-conjugated anti-LY6C (clone HK1.4, BioLegend), PerCp Cy5.5-conjugated anti-LY6G (clone 1A8, BioLegend), Brilliant Violet 421-conjugated anti-CD34 (clone RAM34, BD), APC-conjugated anti-CD117 (clone 2B8, BioLegend) and Brilliant Violet 605-conjugated anti-CD115 (clone T38–320, BD), PE- Cy5 conjugated CD135 and BUV395-conjugated anti-CD16/CD32 (clone 93, Thermo Fisher). MDPs (Lin-, KIT+, SCA-, CD16/32-, CD34hi, CD135+, CD115+), GMP(Lin-, KIT+, SCA-, CD16/32+,CD34hi) and C-GMP (Lin-, KIT+, SCA-, CD16/32+/−,CD34hi) (Fig. 3A), were sorted using FACSAria II (Becton, Dickenson, and Company) and cells were collected in a solution of DPBS + 50% FBS (Atlanta Biologicals). Other granulocytic and monocytic lineage progenitors stages were sorted based on the gating strategy shown in Extended data Fig.1E.

Colony-forming assays.—To evaluate the lineage potential of hematopoietic progenitors, 10,000 cells were plated in triplicate [10,000 cells (sorted from 3 mice) /well] in MethoCult GFM3434 (STEMCELL Technologies; components include insulin, transferrin, stem cell factor, IL-3, IL-6, erythropoietin). For the trilineage assay, we also included 10ng/ml Flt3L, and 10ng/ml GMCSF. Colonies were identified and counted 7 days later.

For flow cytometry, methyl cellulose medium was solubilized by addition of 10ml of warm (37°C) RPMI + 10% FBS, and cells pelleted and then washed 2x with RPMI + FBS, and then stained with fluorescently labeled antibodies as usual.

To evaluate the myeloid and erythroid lineage potential, we flow sorted lineage (CD11b, CD127, B220, CD19, Gr1, NK1.1, Ter-119, CD3) negative cells, 5,000 cells were plated in triplicate (5000 cells /well) in MethoCult GFM3434 (STEMCELL Technologies; components include insulin, transferrin, stem cell factor, IL-3, IL-6, erythropoietin). Colonies were identified and counted 7 days later. For secondary plating and flowcytometry, methyl cellulose medium was solubilized by addition of 10ml of warm (37°C) RPMI + 10% FBS, and cells were pelleted and then washed 2x with 15 ml of RPMI + FBS, and counted. 10, 000 cells were plated in MethoCult GFM3434 (STEMCELL Technologies; components include insulin, transferrin, stem cell factor, IL-3, IL-6, erythropoietin). 7–10 days later colonies were counted.

Cytospin.—Freshly prepared cells were diluted in 200 μl of FACS buffer and spun onto VWR VistaVision Histobond slides (VWR) for 3 min at 900 rpm using a Cytospin 4 apparatus (Thermo Scientific). Slides were then fixed and stained using the Camco stain pak (Cambridge Diagnostic Products), and once dry, sealed using Cytoseal 60 (Thermo Scientific) and microscope cover glass (Globe Scientific).

***In vivo* progenitor differentiation.—**Wild-type CD45.2 GMP or MDP progenitors were intravenously injected (50×10^3 cells/mouse in 100 μL PBS) into non-irradiated CD45.1 recipient mice on day 0. Mice were sacrificed at d5 post-transfer; bone marrow (femurs and tibias), spleens and blood (500μl) were harvested and single cell suspensions were prepared. Erythrocytes were lysed with ammonium chloride. Cells were stained for CD45.1, CD45.2, CD11b, Ly6C, Ly6G, F4/80, CD115, CD43, MHCII and CD11c for neutrophil, monocyte and dendritic cell identification. Of note, only the blood provided useful information, as donor-derived cells in BM and spleen proved too rare to analyse.

scRNA-Seq experimental protocol.

CITE-Seq: We applied the same CITE-Seq protocol previously described[13], using an expanded collection of 65 TotalSeq antibodies to characterize diverse hematopoietic lineages. In brief, before FACS-sorting, samples that were previously stained with fluorochrome-conjugated antibodies were stained with the pool of TotalSeq antibodies (BioLegend) in staining buffer (2% BSA/0.02% Tween20, PBS) for 30 minutes at 4°C . Cells were washed three times in staining buffer, centrifuging at 400 *g* for 5 minutes at 4°C in between each wash. Cells were FACS-sorted, diluted to 1,000 cells/μL, and 10,000 cells from 3 littermate male or female mice (wild-type, 4 TotalSeq antibodies) or 1 mouse, were loaded in each lane of a Chromium Controller Instrument (10x Genomics) according to the manufacturer's instructions, until the cDNA amplification step. The Antibody Derived Tags (ADTs) were further processed according to the CITE-Seq protocol (New York Genome Center, https://cite-seq.com/protocol/). The final pooled library was comprised of ~90% mRNA-derived library and ~10% ADT library. Sequencing was performed on the Illumina HiSeq-4000 platform (Novogene).

Fluidigm: FACS-sorted ThPOK$^{-/-}$ cells purified from murine bone marrow were captured using the C1 Single-Cell Auto Prep System (Fluidigm) according to the manufacturer's instructions. In brief, sorted cells were counted and resuspended at a concentration of 333,333 cells per 1 ml DPBS then loaded onto a primed 5–10 μm C1 Single-Cell Auto Prep Integrated Fluidic Chip. Cell separation was visually scored to identify wells with only a single live cell, with two replicate captures performed. Cells were lysed on the chip and reverse transcription was performed using the Clontech SMARTer® Kit with the mRNA Seq RT + Amp (1771x). After the reverse transcription, cDNAs were quantified using the Quant-iT™ PicoGreen® dsDNA Assay Kit (Life Technologies, Grand Island, NY) and Agilent High Sensitivity DNA Kit (Agilent Technologies (Santa Clara, CA). Libraries for 96 cells were prepared using the Nextera XT DNA Library Preparation Kit (Illumina Inc., Santa Clara, CA). In each single-cell library preparation, a total of 125pg cDNA was tagmented at 55°C for 20 minutes. Libraries were pooled and purified on the AMPure® bead-based magnetic separation Kit prior to a quality control check using both the Qubit® dsDNA HS Assay Kit (Life Technologies, Grand Island, NY) and Agilent High Sensitivity DNA Kit. All single-cell RNA-Seq libraries were finally sequenced on a HiSeq 2500 gel.

scRNA-Seq analysis.—The Fluidigm and CITE-Seq analyses are described below,

CITE-Seq: The 10x Genomics libraries were aligned to the mouse genome (mm10), with unique molecular index quantified (UMI) gen counts obtained using the Cell Ranger workflow (version 3.1.0) using the mm10–2.1.0 reference transcriptome or GFP sequence. Each sample was sequenced to a depth of >890 million reads, with over 9,600 called cell barcodes from each library, with >70,000 UMI per cell on average (default filtering options). Using the default filtered cellular barcodes, the associated sparse-filtered HDF5 files were processed in AltAnalyze and ICGS2 (Iterative Clustering and Guide-gene selection version 2) to obtain normalized gene counts (counts per ten thousands (CPTT) UMIs) and initial predicted cell populations. ICGS2 was run using the default options (cosine clustering) in addition to conservative exclusion of cell-cycle effects (--excludeCellCycle)[38] and exclusion with greater than 500 genes expressed (CPTT>1), were retained for further analysis (>90% of cell barcodes for all captures) (see summary metrics in Table S1). Reference cell states were defined in the cKit progenitor capture, by first defining new cell populations from ICGS2 (e.g., progressive Erythroid progenitor cell-states), followed by supervised classification of cell-states using the software cellHarmony and our prior defined 10× 3' v2 cluster centroids[17]. Ambient RNAs were excluded to minimize batch effects using the SoupX package (https://github.com/constantAmateur/SoupX) from the cKit defined cell populations (top two markers from each cluster with a MarkerFinder rho> 0.7 to estimate a contamination fraction = 0.1). Subsequent C-GMP captures cell-states were defined using these expanded cKit population centroids, using cellHarmony. Differential expression analyses were performed using these alignment results in an additional cellHarmony run using the options: referenceType None, eBayes p<0.05 FDR adjusted and fold>1.2. To infer lineage trajectories, we applied the Velocyto toolkit, which examines RNA velocity of spliced versus unspliced transcripts using the SeuratWrappers framework[22]. Predictions from this workflow were projected onto the cellHarmony UMAP embeddings using the

CreateDimReducObject function. To obtain ADT UMIs for each cellular barcode, each ADT barcode was supplied as input in the CITE-Seq-Counts python program.

Fluidigm: The primary Fluidigm C1 single-cell RNA-Seq analysis for all captured libraries (wild-type and mutant) was performed as previously described[7]. In brief, RSEM was used to obtain transcript per million (TPM) estimates following read alignment to the mm9 reference transcriptome (RefSeq). Only libraries with > 300,000 paired-end reads were retained for further analysis. Wild-type libraries were combined with our prior described mouse Fluidigm captures (LSK, CMP, GMP, LK CD34+). UMAP analysis indicates that the new captures intermixed with the old captures, without clear evidence of batch effects. Cell populations were derived in both the WT and ThPOK$^{-/-}$ through reference-based aligned to the cKit defined cell populations with cellHarmony. Differential gene expression analysis between matching the WT and ThPOK$^{-/-}$ cell populations was using the options --referenceType None, eBayes $p<0.05$ and fold>1.2 and matched to analogous 10x Genomics differentials in the same comparisons with the same fold direction using a custom python script. Alternative splicing within these single-cell populations and between ThPOK$^{-/-}$ and WT cells, was performed as previously described[39]. In brief, the MultiPath-PSI workflow in AltAnalyze was used to quantify exon-exon and exon-intron junction expression from STAR aligned BAM files (mm10 genome/transcriptome reference), requiring detection of splicing in at least 20% of the single-cell libraries of each cellHarmony assigned cell populations (dPSI > 0.1, eBayes t-test $p<0.05$). To determine the concordance of ASEs in ThPOK$^{-/-}$ versus WT and ASEs observed in normal hematopoiesis, we applied a previously described comparison analysis[30].

Gene ontology analysis: A gene set of interest was processed using GO-Elite analysis in AltAnalyze to identify enriched biological processes. The resulting list was filtered to remove those terms with fewer than 4 genes in the gene set (fewer than 3 genes for Fluidigm C1 splicing data) and plotted to visualize Z-scores and Fisher-Exact p-values (visualized using GraphPad Prism)

Sepsis assay: Sepsis induction was carried out as previously described[16]. In brief, each mouse was given an intra- peritoneal (i.p.) injection of 0.4 mL (~10^9 strict anaerobe bacteria) of cecal slurry solution (200mg/ml) using a 27G needle. Sham-treated mice were injected with sterile normal saline. Analgesia was provided by a subcutaneous injection of buprenorphine (0.1 mg/kg). Monitoring of the health of the animals was conducted every 2 hours after the induction of sepsis for 12 hours, and then every hour thereafter. Severity of disease progression was quantified based on 7 distinct criteria[16]. Functional assays and neutrophil subset analysis were carried out at 14hrs after sepsis induction.

RNA isolation, reverse transcription, and RT-PCR: Total RNA was prepared from FACS sorted progenitor populations with the RNA Extraction RNeasy Plus Mini-kit (QIAGEN) and RNA was reverse-transcribed with the HighCapacity cDNA Reverse Transcription kit (Applied Biosystems; Carlsbad, CA) according to the manufacturer's protocol. Real-time quantitative PCR (SYBR-green, Applied Biosystems) assays were performed with an Applied

Biosystems 7900HT Fast Real-Time PCR System sequencer detector. Expression was normalized to the expression of the *Actb* housekeeping gene. *ThPOK:* F 5'-ACCCAACGGCTGAAAGGA-3', R 5'-GCTGCTGTGGTCTGGCAAT-3'. *Irf8*: F 5'-GACACACACCATTCAGCTTTC-3', R: 5'-CATCCGGCCCATACAACTTA-3'. *Zeb2:* F 5'-CGCCACGAGAAGAATGAAGA-3', R 5'-GGTTAGCATTTGGTGCTGATC-3', *Cebpe:* F 5'-GCAGCCACTTGAGTTCTCAGG-3', R 5'-GATGTAGGCGGAGAGGTCGAT-3', *Ezh2* (ref transcript): F 5'-TGTTACCAGCATCTGCCAC-3', R 5'-AACTCCTTAGCTCCCTCCTT-3', *Ezh2 (ASE)*: F 5'-GTGTTACCAGCATCTGGAGG-3', R 5'-GTCTGCTACTGTTATTCGGAAGT-3', *Hmga1* (ref. transcript) F 5'-TGGTCGGGAGTCAGAAAGA-3'. R 5'-CTTCTCCAGTTTCTTGGGTCTG-3'. *Hmga1*ASE: F 5'-CAAGCAGCCTCCGAAAGA-3'. R 5'-TTCCTCCCTGGAGCTGT-3'. For the rest we used Taqman qPCR assays (Applied Biosystem). Runx1-Mm01213405_m1, NKg7-Mm00452524_g1, Pde4d-Mm00456879_m1, Ddx3x-Mm04207948_gH, Cd63-Mm01966817_g1 Srsf5-Mm00833629_g1and used Actb: Mm01205647_g1 as normalizer.

ChIP.—CMP and GMP progenitors were purified by flow cytometry from compound heterozygous Flag tagged ThPOK (ThPOK^{FH}/+) mice. Chromatin crosslinking and immunoprecipitation were performed using the iDeal ChIP–seq kit for Transcription Factors (Diagenode) according to the manufacturer's protocol. The anti-Flag (Diagenode, C15200209) antibody was used for immunoprecipitation and purified DNA sequences were analyzed by qPCR using following regulatory elements. *Irf8* Promoter: F-CAAGGGAACCGATAATGCG, R- CGTCTGCCTAAGGTCAAGGA. *Irf8* Enhancer1: F-TTCTGAGTGTGTGGCTCCCC, R- TTCAGAGCACGCCTGCAAGA. *Irf8* Enhancer2: F-GAAATGTCAGCCATGCGTGT, R- AAAACTGAGGTCCAGCAGGC. *Irf8* Enhancer3: F-ACACAGGGTGGACTGACAAA, R- CCCTGAAGGACCCAGGACAT. *Irf8* Enhancer4: F-AGGGCTGTCCTCTCACTCTT, R- GGTACGCTGAAGAGTCCTGG. *Zeb2* promoter: F-CCTCAGAAAAGTCCTCAGCC, R- GTTGTTTGTTACTTTTTTCCCCTC.

BrdU pulse-labelling.—For *in vivo* assays, mice were injected intraperitoneally with 2mg 5-bromo-20-deoxyuridine (BrdU; Sigma-Aldrich) at indicated time points. To detect BrdU incorporation into neutrophil subsets, cells were stained with Propidium iodide, surface-stained, fixed, permeabilized, and subjected to intracellular staining with PECy7-conjugated anti-BrdU antibody, according to the manufacturer's protocol (BrdU Flow kit; BD) prior to analysis by flow cytometry.

Intracellular ROS measurement.—Intracellular ROS was measured using a FACS-based method. Neutrophils were incubated with 2.5 μg ml−1 (7 μM) CellROX™ Deep Red Reagent (Invitrogen, Cat# C10422) in complete RPMI 1640 medium and stimulated with 50 nM of phorbol 12-myristate 13-actetate (PMA) (Sigma-Aldrich) for 20 min at 37 °C. Cells were subsequently washed with PBS and fluorescence intensity per cell measured by flow cytometry.

Bone marrow transplantation.—A single-cell suspension of 1×10^6 unfractionated bone marrow cells from 6-week-old CD45.2+ WT or ThPOK−/− mice together with

CD45.1/CD45.2 heterozygous (competitor cells) was injected into the tail veins of lethally irradiated (10 Gy) CD45.1 congenic recipients.

Phagocytosis.—The phagocytic capacity of neutrophils was measured by FACS using fluorescent labelled pHrodo™ red *E. coli (Invitrogen, Cat: P35361)* and pHrodo™ red *S. aureus (Invitrogen, Cat:* A10010*)*. Purified neutrophils were incubated with fluorescent *E. coli or S. aureus* at a multiplicity of infection of 10 in complete RPMI 1640 medium for 15 min at 37 °C. Neutrophils were subsequently washed with PBS and the fluorescence intensity was measured by flow cytometry.

Myeloperoxidase, Elastase, and NETosis assay.: Blood neutrophil (suspended at 10^6/ml) were stimulated with 5 nM of PMA for 4h. Myeloperoxidase release was assayed using a commercial Neutrophil Myeloperoxidase Activity Assay Kit (Cayman chemical, Cat: 600620). Both soluble Elastase and NET associated elastase were assayed using a NETosis Assay Kit (Cayman chemical, Cat: 601010). Citrullinated H3 (CitH3), which is released from neutrophils during NETosis, was measured by Citrullinated Histone H3 (Clone 11D3) ELISA Kit (Cayman chemical, Cat: 501620).

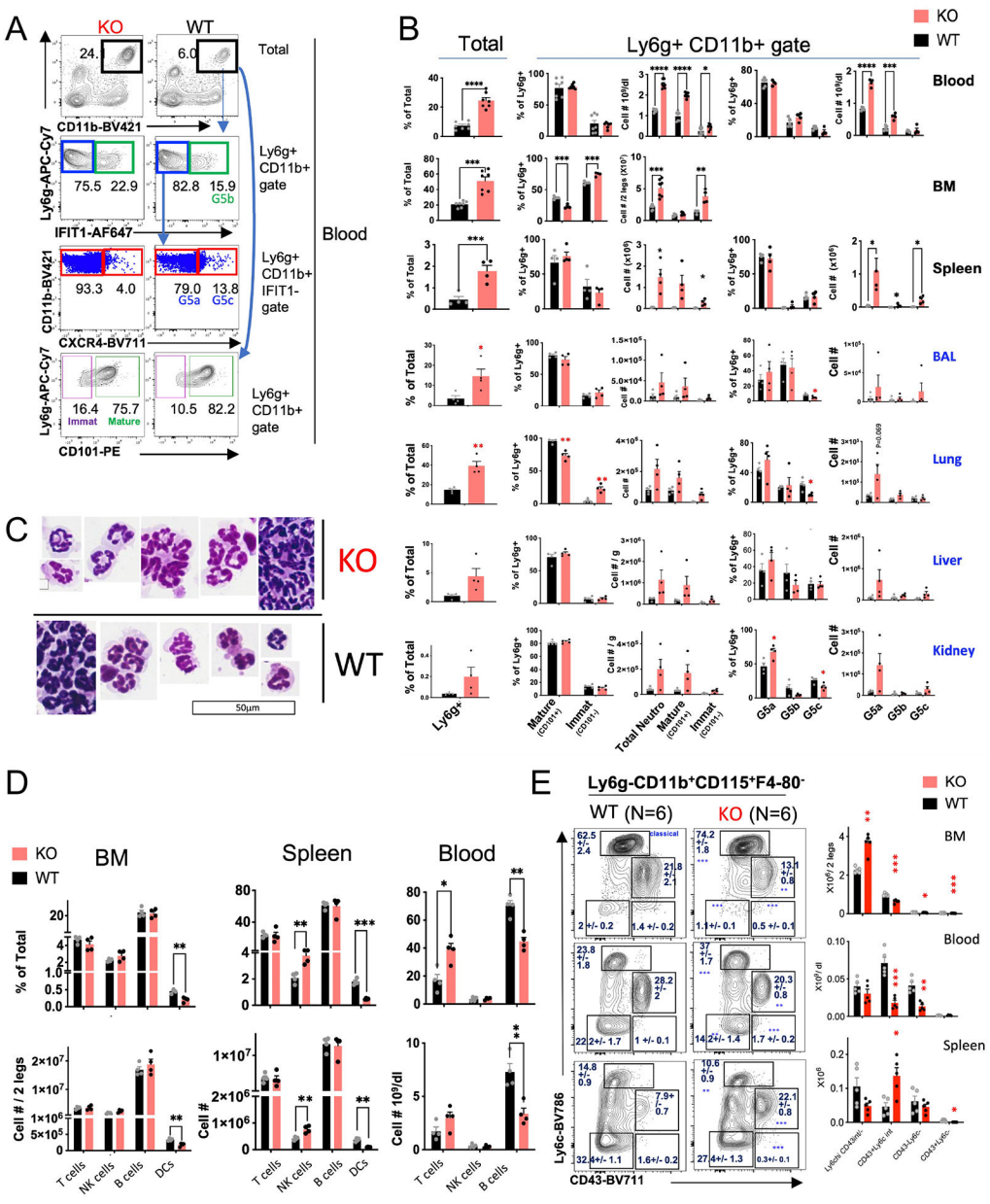
Fluorescence microscopy for EZH2 and H3K27Me3.: FACS-sorted WT and ThPOK−/− BM progenitors were plated on PolyD lysine coated 96 well plates. Intracellular staining was performed using a Transcription Factor Buffer Set kit (BD Biosciences, Cat: 562574). Primary antibody staining was carried out with 5ug/ml of EZH2 Rabbit polyclonal antibody (Cell signaling: Cat: 5246S) and mouse monoclonal Histone H3K27me3 antibody (Active motif, Cat: 61017), followed by Alexa Fluor™ 633 conjugated Goat anti-Rabbit IgG (H+L) cross-adsorbed secondary antibody and Alexa Fluor 546 conjugated Goat anti-Mouse IgG (H+L) cross-adsorbed secondary antibody (Invitrogen), to measure EZH2 protein and histone H3K27Me3 in the same cells. Finally, cells were stained with DAPI for visualization of the nucleus. In each well of a 96 well plate, eight separate image fields in each wavelength were captured at 20X with 5 z-steps separated by 5um with an automated microscope (ImageXpress Confocal; Molecular Devices, Sunnyvale, CA) driven by MetaXpress software. Two-dimensional projection images were analyzed utilizing the Multi-Wavelength Scoring module (MetaXpress, Molecular Devices) to segment the images for cell morphology parameters and intensity parameters in each channel.

Chemokine and cytokine quantification.—Chemokines and cytokines were assayed in Serum samples after 14h of sepsis induction and culture supernatant of blood neutrophil (10^5/100µl medium) with or with stimulation of 5 nM PMA for 5h. LEGENDplex mouse proinflammatory chemokine Panels (Biolegend, Cat: #740451, #741068 and #740007) were used to measure Chemokines. Cytokines were measured using LEGENDplex™ Cytokine Panel (Cat: 741043), Mouse M2 macrophage panel (740446), Mouse inflammation panel (Cat:740446), Mouse Cytokine Panel 2 (Cat: 740134), and Mouse HSC Panel (Cat: 740677), all from Biolegend.

Statistics and reproducibility.—No statistical method was used to determine sample size, but our sample sizes are similar to those reported in previous publications[40,41]. Sample sizes were rationalized by weighing sufficient replication (to determine the extent

of biological variation) with a reduction of total animals used. Data were excluded only for technical reasons, such as low cell viability. In particular, bioinformatically, single cell RNA-Seq data of low quality was excluded according to pre-established criteria to ensure robust and accurate downstream analyses of the data as a whole. With regard to replication, all *in vivo* analyses were performed on a total of three to six animals per genotype (across at least three separate experiments). All attempts at replication were successful. Randomization was not used; assignment to experimental groups was based on genotype. To exclude physiological and environmental covariates, mice of different genotypes were derived from the same litters as control mice (as much as possible), or co-housed before analysis. Blinding was not possible because genotyping was necessary for all experiments. The investigators were not blinded to allocation during experiments and outcome assessment. Statistical analysis for nonsequencing data was performed using GraphPad Prism9 software. Data were analyzed by applying unpaired, two-tailed Student's t-tests, and one-way analysis of variance (ANOVA) with Bonferroni's correction. A P value < 0.05 was considered to be significant. *P < 0.05, **P < 0.01, ***P < 0.001. All statistical analyses performed for sequencing data are mentioned elsewhere in Methods.

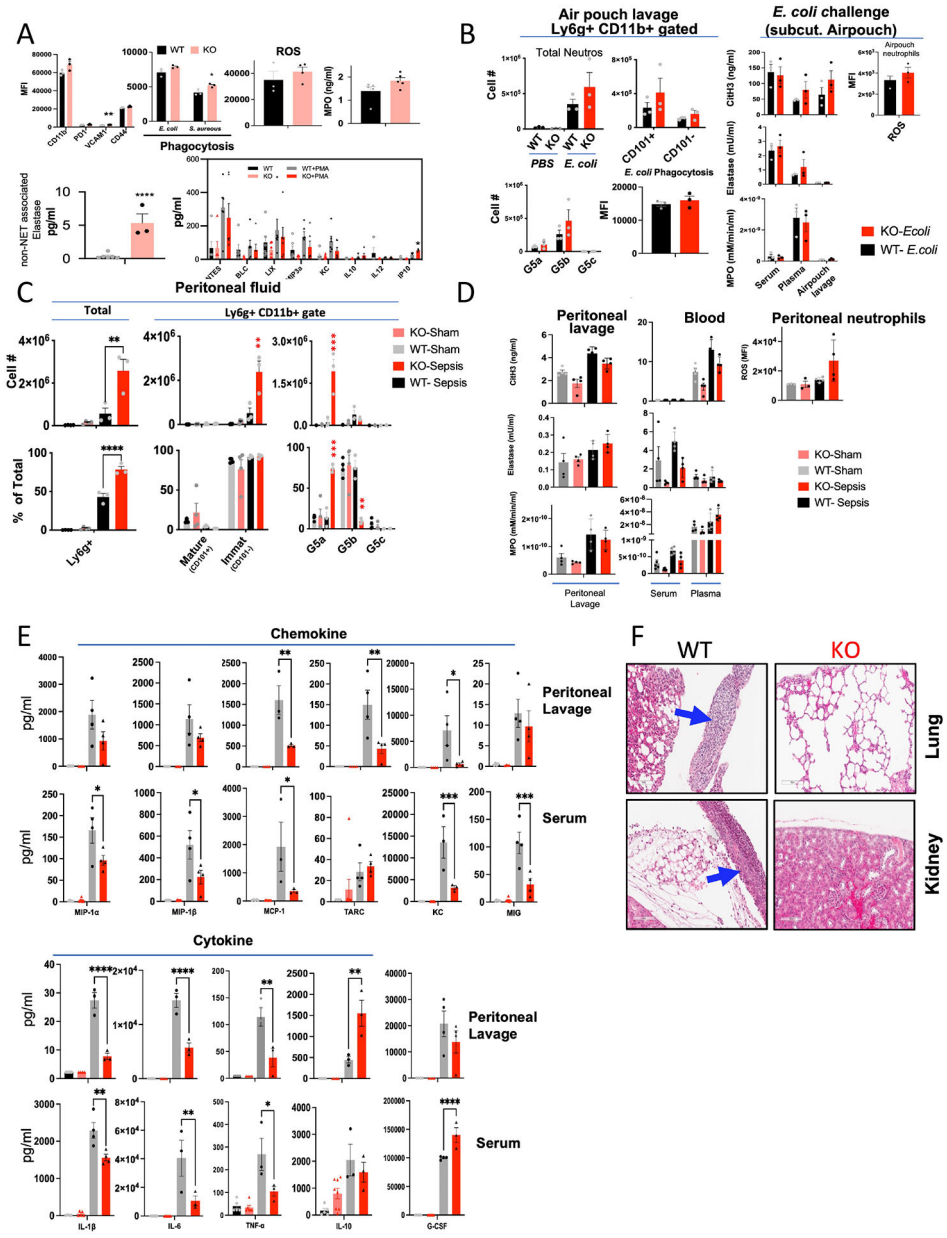
Extended Data



Extended data Fig. 1. ThPOK deficiency alters neutrophil and monocyte subset distribution.

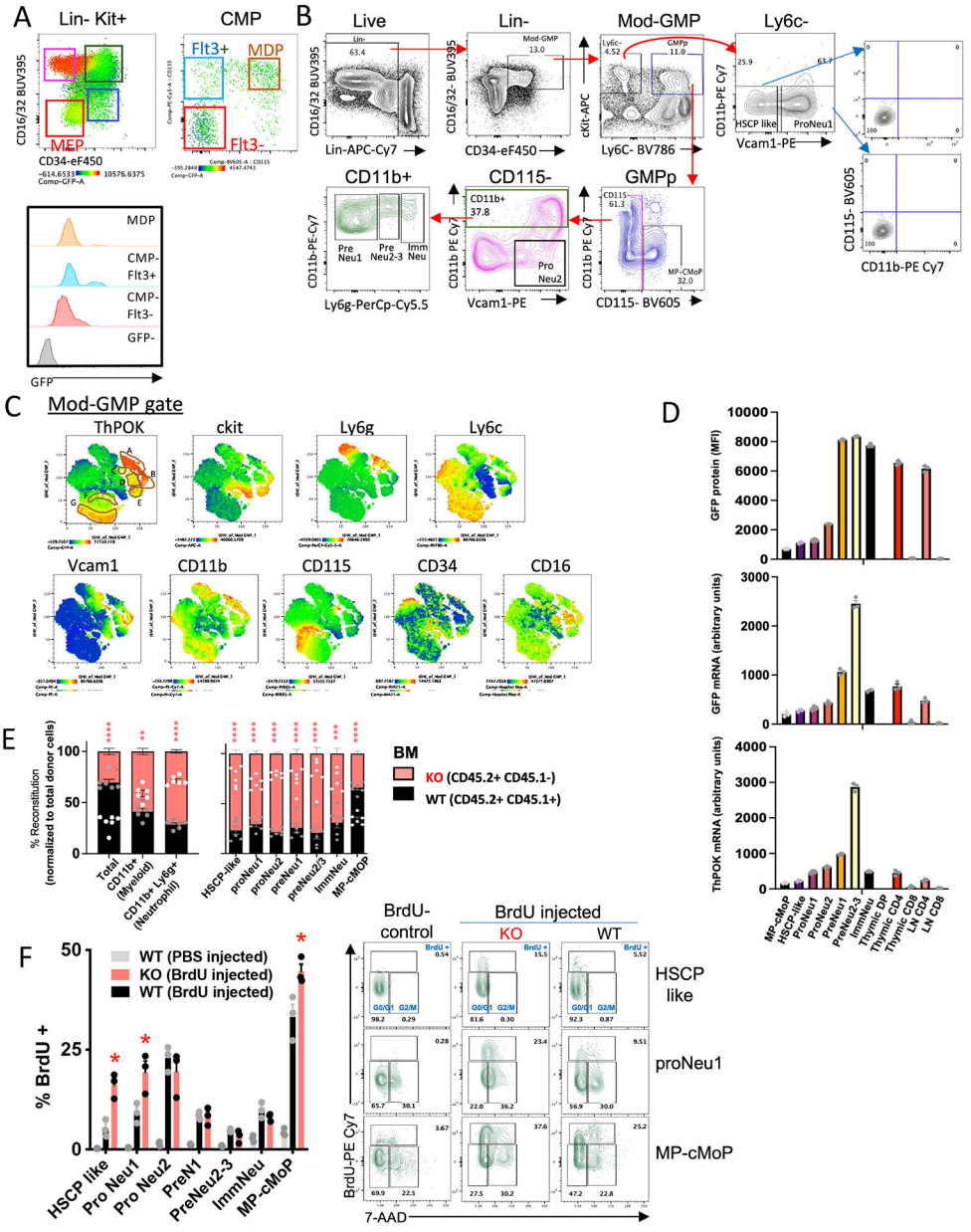
a) Gating strategy to reveal neutrophil maturation and differentiation status, according to indicated markers. **b)** Cell-population percentages and absolute cell numbers of indicated neutrophil subsets from indicated organs of wt versus ThPOK-deficient mice (based on gating strategy in panel a). Data are displayed as mean ± SEM. *P < 0.05; **P < 0.01; ***P < 0.001; ****P < 0.0001. Significant differences were determined by two-sided unpaired t test. **c)** Histological examination of flow sorted (Ly6g+CD11b+) blood neutrophils. Cytospins of neutrophils from WT or ThPOK−/− mice were treated with Wright's stain and examined with a light microscope. The experiment was repeated 3 times independently with similar results. **d)** FACS analysis of CD11b+ Ly6g-CD115+F4/80- cells from indicated

organs from WT and ThPOK−/− mice (left panels). Note increase in Ly6c+ classical monocytes in ThPOK-deficient compared to WT mice in BM. Plots showing total number of indicated gated populations in indicated organ in ThPOK-deficient compared to WT mice (right panels; same mice as in FACS panels). Error bars represent standard deviations. **e)** Plots showing frequency (top panels) or cell number (bottom panels) of T cells, NK cells, B cells, and DCs in BM, blood and spleen of ThPOK-deficient or WT mice. Significant differences between ThPOK−/− and WT mice were determined by unpaired T test with Welchs correction, and indicated by asterisks (* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$).



Extended data Fig. 2. ThPOK deficiency perturbs selective neutrophil functions.

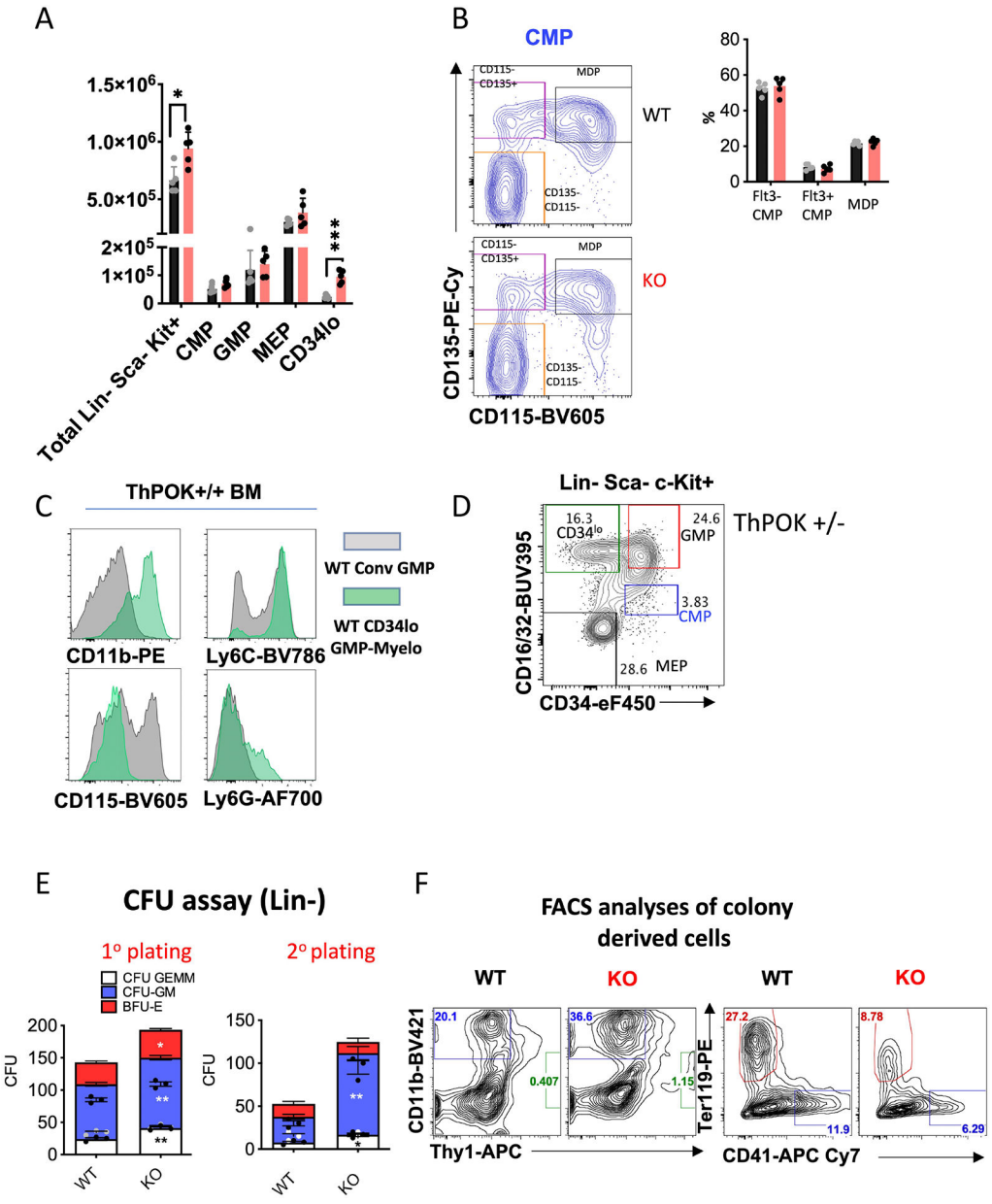
a) Neutrophil surface marker expression (n=3) (Top 1st panel from left), phagocytosis [data expressed as geometric Mean Fluorescent Intensity (MFI) ± SEM (n=3)] (Top 2nd panel), PMA-mediated ROS production using dihydrorhodamine 123 (DHR) [data are shown as geometric means ± SEM (n = 4)] (Top 3rd panel), Myeloperoxidase release (n=5) (Top 4th panel), NET associated elastase release (n=5 for wt and n=3 for ThPOK−/− neutrophil, respectively) (Bottom left panel), and cytokine/chemokine release upon *in vitro* stimulation (N=6) (Bottom right panel). Data are representative of three independent experiments. Significance was calculated by two-sided unpaired T-test with Welch's correction (*p < 0.05; ***p < 0.001). **b)** Absolute numbers of total neutrophils or indicated subsets from airpouch lavage of wt versus ThPOK-deficient mice, 14hrs after injection of *E. coli* or PBS into airpouch (top left, top 2nd from left, and bottom left panels). Same mice were used to assess phagocytosis by pHrodo™ green *E. coli* incorporation [geometric MFI ± SEM (n=3)] (bottom 2nd from left panel), ROS production [geometric means ± SEM (n = 3)] (top right panel), Myeloperoxidase release (n=3) (bottom right panel), NET associated elastase release (n=3) (center right panel), and Citrullinated H3 (CitH3) release (Top 2nd from right panel) for airpouch lavage, serum and plasma samples from the same mice. Significance was calculated by two-sided unpaired T-test with Welch's correction (*p < 0.05; **p < 0.01; ***p < 0.001). **c)** Frequency (bottom panels) or cell number (top panels) of indicated subsets in WT or ThPOK-deficient mice 16hrs after sepsis induction or normal saline administration. Significant differences were determined by one way ANOVA posthoc Bonferroni's multiple comparison test (*p < 0.05; **p < 0.01; ***p < 0.001). **d)** Same mice as in panel c were assessed for ROS production [geometric means ± SEM (n = 4)] (top right panel), Myeloperoxidase release (n=4) (bottom left panel), NET associated elastase release (n=4) (center left panel), and Citrullinated H3 (CitH3) release (top left panel). Assays were performed for blood serum and plasma samples from the same mice (top, middle and bottom panels in center). **e)** Cytokine/chemokine levels in peritoneal lavage, blood serum and plasma of ThPOK−/− and WT mice 14hrs after sepsis induction (N=4). Significance was calculated by one way ANOVA posthoc Bonferroni's multiple comparisons test (* p < 0.05; **p < 0.01; ***p < 0.001; **** p < 0.0001). **f)** H&E staining of tissue sections from indicated organ 16hrs after sepsis induction. Images shown are representative of 6 mice/genotype. The experiment was repeated 3 times with similar results.



Extended data Fig. 3. ThPOK expression and effect of ThPOK deficiency on gene expression.

a) GFP expression by indicated subsets from ThPOK-GFP reporter mice. GFP expression level is color-coded according to heat-scale below each plot. **b)** Gating strategy to reveal granulocyte-monocyte lineage development (according to REF 17). Red arrows indicate sequential gates. **c)** Unsupervised clustering (viSNE) of cells in the modified-GMP gate (upper left). Other panels show expression of indicated surface marker superimposed on the same clusters. **d)** GFP protein (top), GFP mRNA (center), and endogenous ThPOK mRNA levels, in indicated progenitor subsets sorted according to gating strategy in (b), as well as indicated control T cell populations from transgenic ThPOK-GFP reporter mice. **e)** Competitive reconstitution assay with WT and ThPOK−/− BM cells (same experiment as depicted in Fig. 1J) (n = 6, per genotype). Relative contribution to indicated BM populations

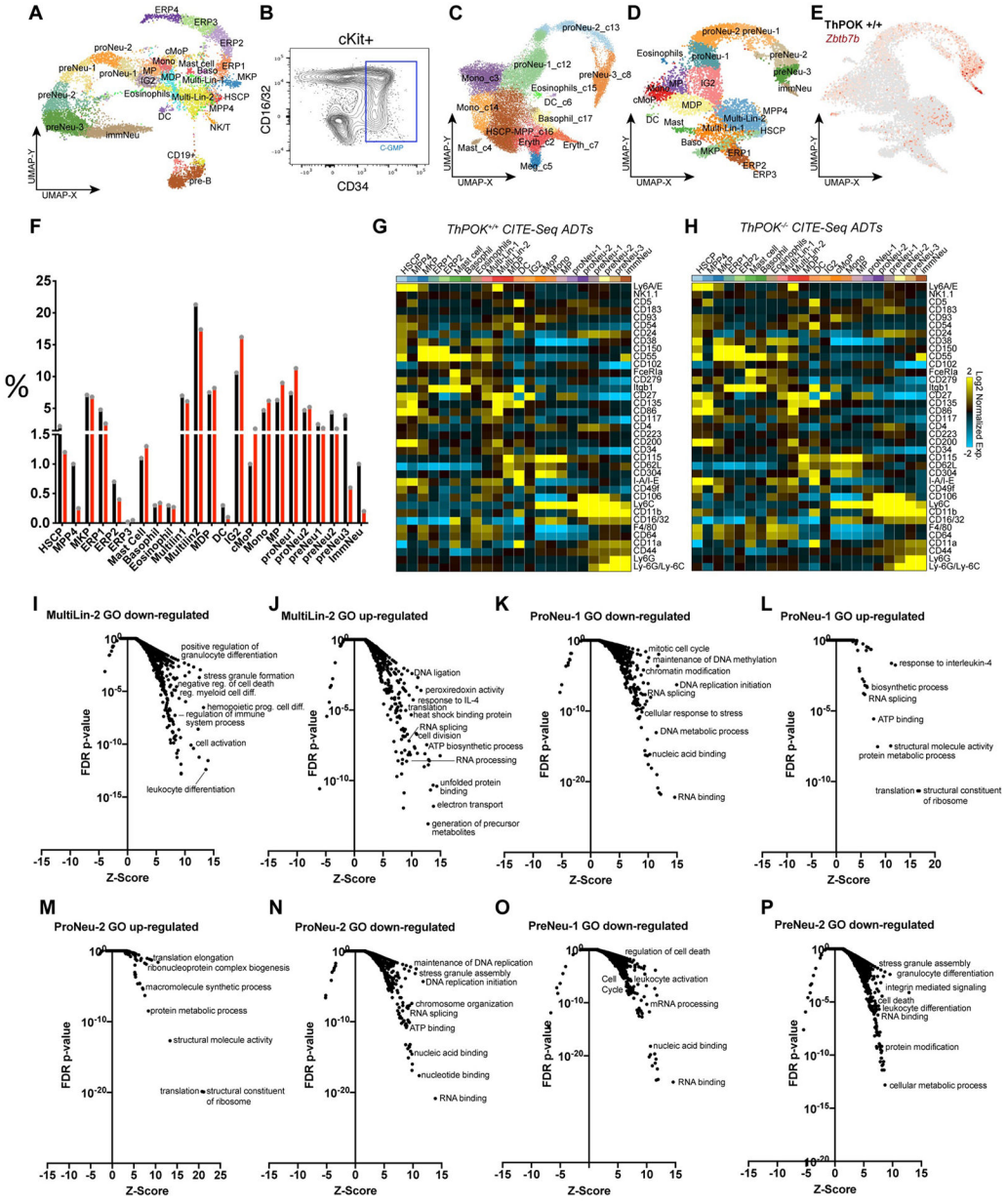
was assessed at 6 months post-transplant. Data are normalized for only donor cells, so that the sum of WT and KO donor cells equals 100%. Two sided unpaired T test with Welch's Correction was performed. Data are displayed as mean ± SEM (*P < 0.05; **P < 0.01; ***P <0.001; ****P < 0.0001). **f)** Cell cycle analyses of indicated gated subsets in WT and ThPOK −/− BM 20h after a single in vivo BrdU pulse. Representative FACS plots of BM subsets from PBS-injected WT controls (BrdU-), or WT and ThPOK−/− mice that received BrdU, as indicated (right panels). Cell cycle stages were defined by flow cytometry based on BrDU intensity and DNA content (7-AAD staining), as indicated. Bar graphs at left summarize the results for indicated progenitor populations, expressed as means ± SEM (n=3 independent animals per genotype) and are representative of two experiments. Two sided unpaired T-test with Welch's Correction was performed (*p < 0.05; **p < 0.01; *** p < 0.001).



Extended data Fig. 4. ThPOK loss does not affect MDP development:

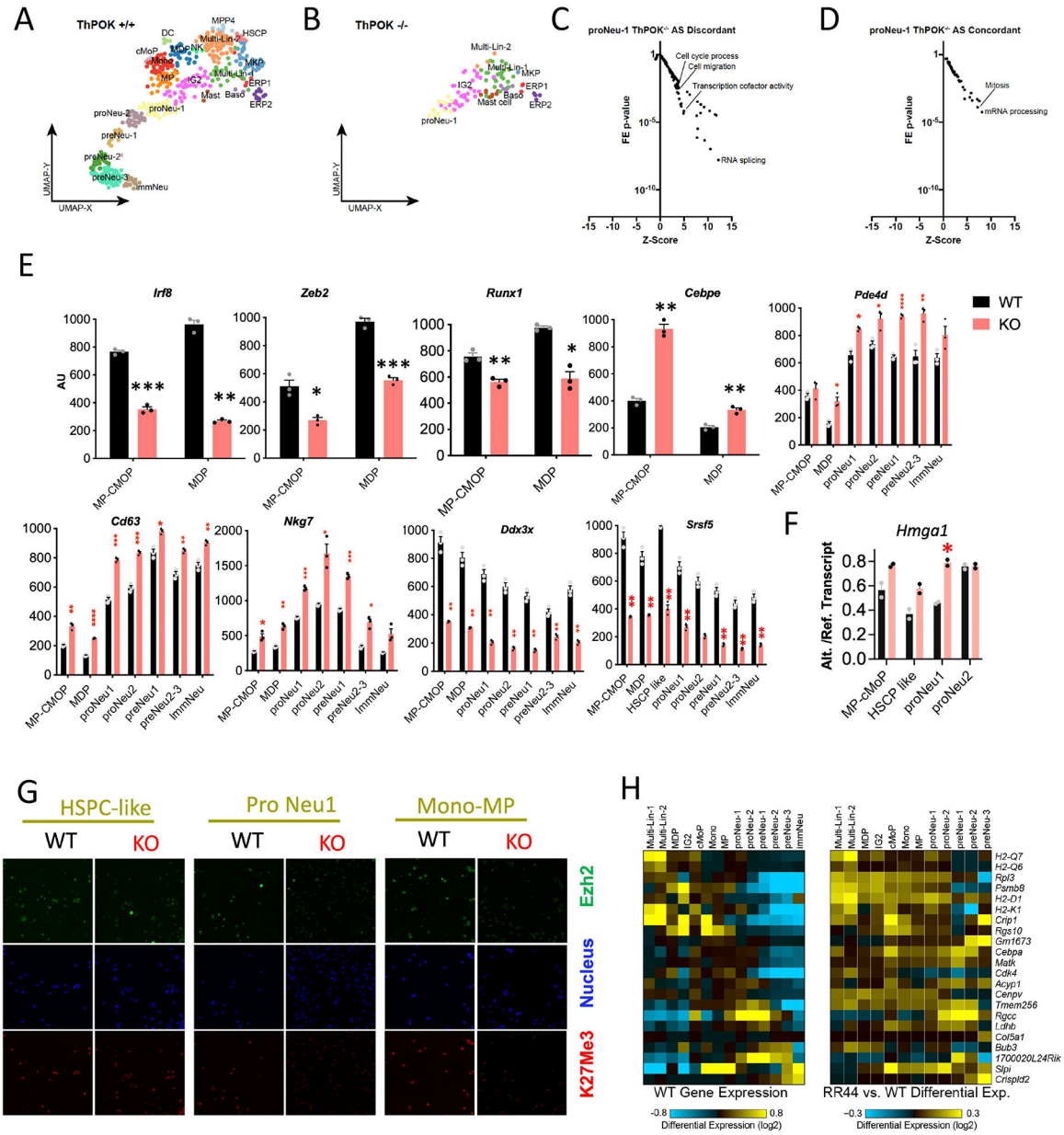
a) Bar graphs showing absolute cell numbers for CMP, GMP, MEP and CD34^{lo} subsets among gated cKit+ Sca1- Lin- BM cells from ThPOK-deficient or WT mice (same mice as Fig. 2 A, B) (n = 5 independent animals per genotype). Error bars represent SEM. Significant differences between ThPOK−/− and WT mice were determined by two-sided unpaired T test with Welch's correction, and indicated by asterisks (* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$). **b)** FACS analysis of CD135 and CD115 expression by gated CMPs from WT and ThPOK−/− mice (left panels). Bar graph at right shows % of indicated gated subsets (n = 5 independent animals per genotype). **c)** FACS analysis of CD11b, CD115, Ly6c, and Ly6g expression by gated cGMP (conventional GMP) (grey histograms) and c-Kit+ CD34lo CD16hi Lin- (green histograms) BM subsets from WT mice. **d)** FACS analysis

of CD16/32 and CD34 expression by gated Lin- c-Kit+ Sca1- BM cells from ThPOPK+/− mice. CMP, GMP, MEP and CD34^{lo} subsets are marked. **e)** CFU assay of FACS-sorted Lin- BM cells, from indicated WT or ThPOK-deficient mice. Error bars represent SEM. Significant differences between ThPOK−/− and WT mice were determined by two-sided multiple multiple unpaired T with Welch's correction, and indicated by asterisks (* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$). Note that *ThPOK−/−* Lin- progenitors exhibited a substantial (** $p < 0.01$) increase in CFU-GM myeloid colony production even after secondary plating. **f)** FACS analysis of CD11b, Thy1, CD41 and Ter119 lineage marker expression by WT and ThPOK-deficient cells after secondary CFU assay (same experiment as in panel d).



Extended data Figure 5. Single-cell transcriptome analysis of ThPOK−/− and WT BM populations.

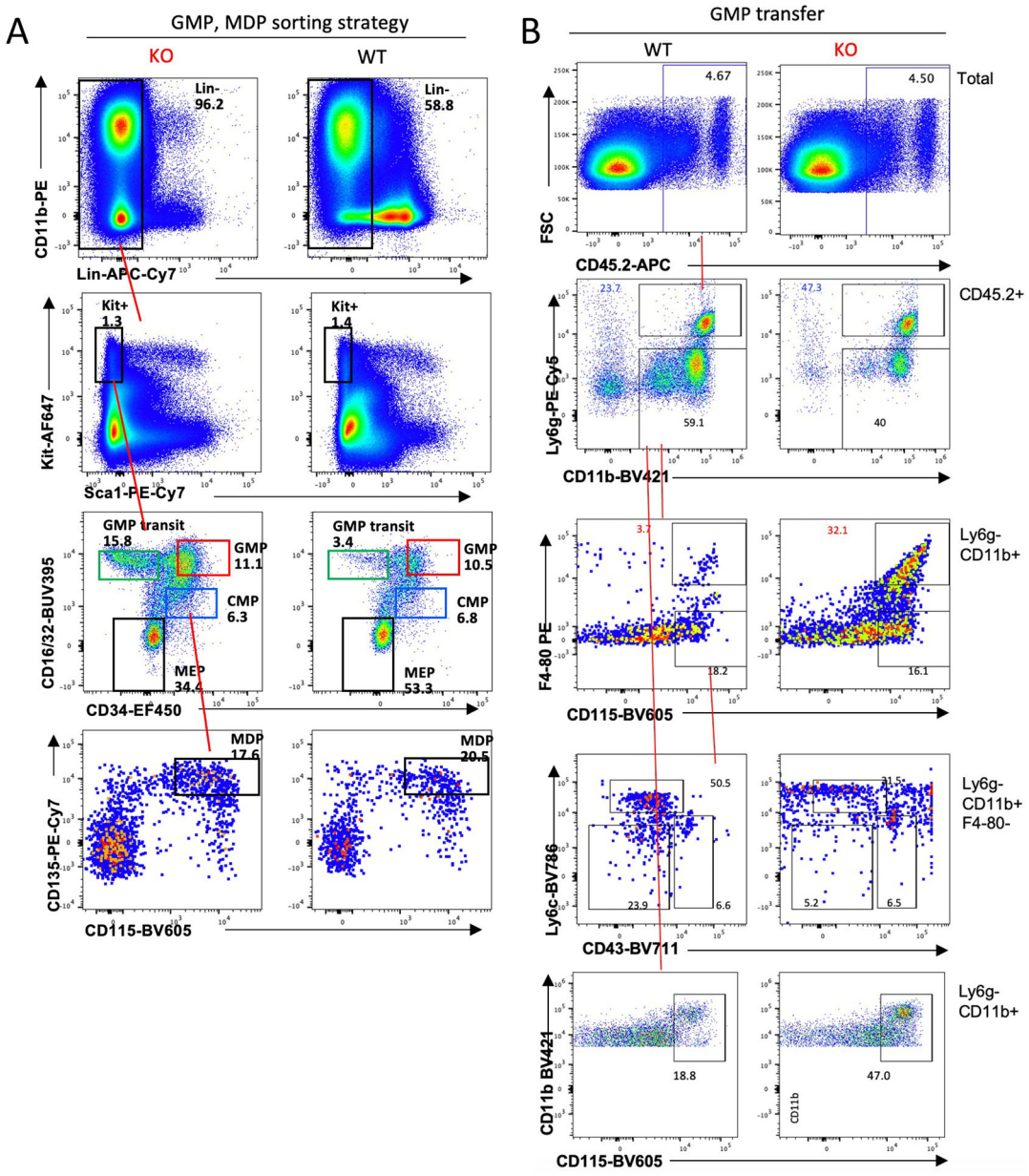
a) UMAP plot of curated cKit+ progenitor scRNA-Seq cell populations that serves as the reference for cellHarmony alignment analyses in these studies (see Methods), **b**) Flow cytometry selection of c-Kit+ BM C-GMP progenitors (gated cells denoted in blue). **c**) Unsupervised clustering UMAP plot of ~26,000 combined C-GMP CITE-Seq captured mRNA profiles (WT and ThPOK−/−) following analysis with the software ICGS2. Indicated cell-population labels are those automatically assigned by ICGS2 (AltAnalyze BioMarker database). **d**) UMAP plot of all WT and ThPOK−/− cells from panel c aligned to the reference described in panel a. **e**) UMAP plot from panel d showing the expression levels of ThPOK mRNA in the wild-type CITE-Seq populations. **f**) Cell-population percentage of distinct BM progenitor populations detected by CITE-Seq from WT and ThPOK−/− mice. **g–h**) Heatmap of relative normalized (TotalVI) CITE-Seq antibody derived-tag (ADT) intensities for cellHarmony cKit+ aligned cell populations. Panels (**g)** and **(h)** displays cells from WT and ThPOK−/− BM progenitors, respectively. **i–p**) Gene Ontology enrichment analyses from the software GO-Elite, for each of the indicated cell population differential expression analysis comparisons (all ThPOK−/− vs. WT).



Extended data Fig. 6. ThPOK−/− impacts global splicing and cellular process networks:

(**a–b)** UMAP plot of single-cell Fluidigm RNA-Seq analysis of prior profiled **(a)** wild-type hematopoietic progenitors and **(b)** Ly6c-ThPOK−/− GMPs aligned to cKit+ CITE-Seq. Consistent alignment of cell populations in the UMAP space indicates a lack of apparent batch effects. **c)** Gene Ontology enrichment analysis of ThPOK−/− dependent alternative splicing events in proNeu-1 cells, for splicing events with the opposite pattern of exon inclusion in proNeu-1 versus MultiLin (discordant = more immature splicing) or **(d)** with the same pattern (concordant = promoting neutrophil specification associated splicing). **(e−f)** Validation of deregulated gene expression in ThPOK−/− versus WT progenitors for important transcripts as observed in CITE seq data: **e)** qRTPCR analyses of indicated flow-sorted progenitors (pooled from 4 mice/genotype) for monocyte/DC genes (*Irf8, Zeb2*,

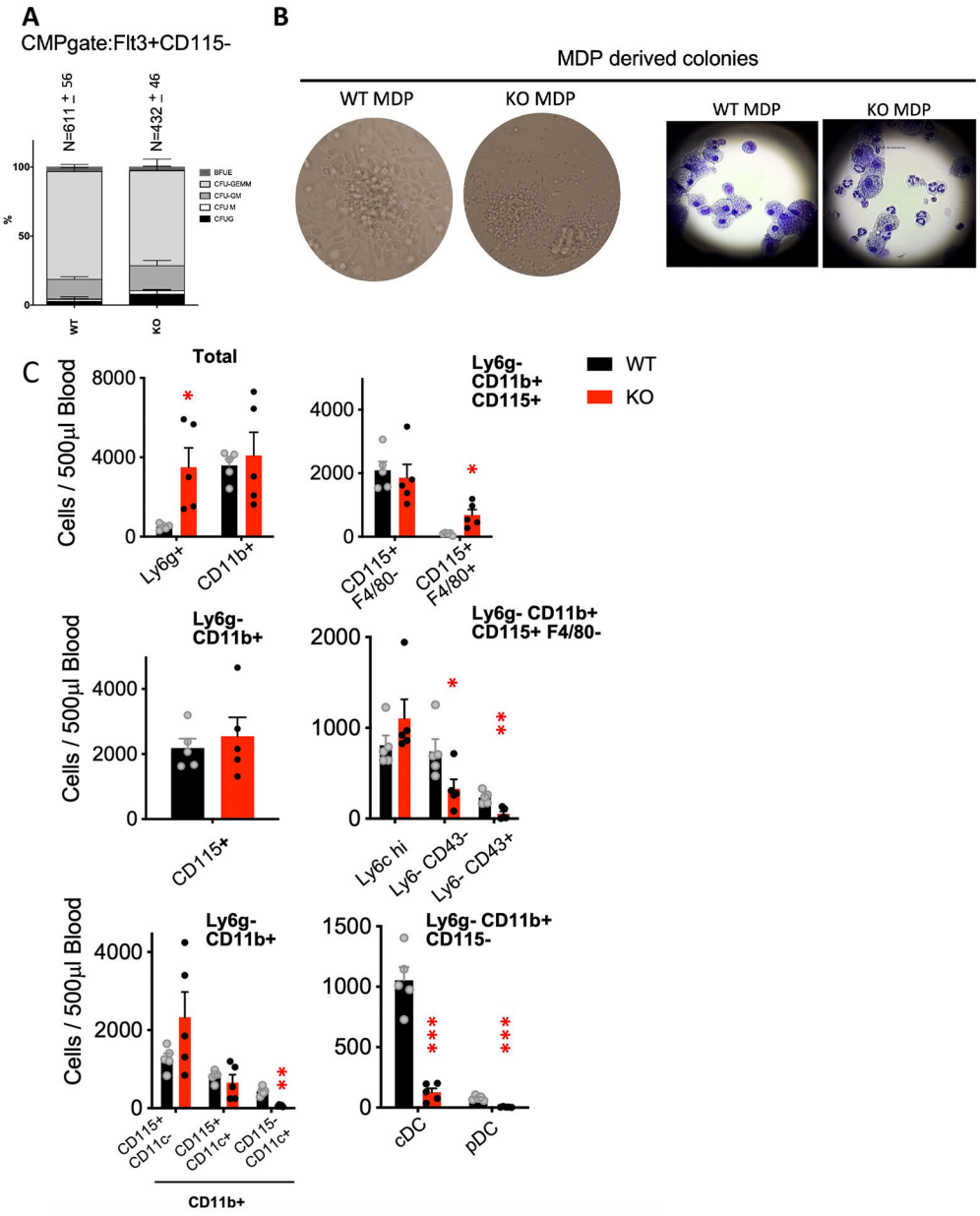
Runx1), granulocyte lineage genes (*Cebpe, Pde4d, Cd63 and Nkg7*) and RNA binding protein genes (*Ddx3x* and *Srsf5*). The experiment was repeated 3 times independently with similar results. **f)** Expression ratio of *Hmga1* alternate transcript versus reference transcript in indicated progenitor populations (pooled from 4 mice/genotype). **g)** Representative confocal micrographs of EZH2 protein (top), DAPI (middle), and H3K27Me3 (bottom) staining in indicated subsets. Each experiment was reproduced twice and significant differences between ThPOK−/− and WT mice were determined by two-sided unpaired T test with Welch's correction, and indicated by asterisks (* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$). **h)** Heat maps showing relative mRNA expression of indicated genes in WT progenitors (left panel), or in ThPOK−/− versus WT progenitor subsets (right panel) for genes that are relatively up-regulated in any ThPOK−/− progenitor population and that are known Ezh2 targets in WT immune cells by ChIP-seq (GSE181873). Note that in WT mice most of these genes show marked stage-specific regulation.



Extended data Fig. 7. ThPOK−/− progenitor transfer assays.

a) Gating strategy for isolation of GMPs and MDPs for *in vitro* and *in vivo* transfer studies.

b) FACS analysis of Ly6g, CD11b, F4–80, CD115, Ly6c, and CD43 in blood of mice reconstituted with sorted GMPs from BM of WT or ThPOK−/− mice. Note substantial population of F4–80^{hi} CD11b+ Ly6g- CD115+ macrophages in ThPOK-deficient versus WT mice (middle panels).



Extended data Fig. 8. ThPOK−/− MDP exhibit neutrophil differentiation bias *in vitro*:
a) CFU assay of FACS-sorted CMP (lin- Kit+ Sca-CD34+ CD16/32-Flt3+CD115-) cells from WT and ThPOK KO mice, as indicated. Centre line refers to mean. **b)** Representative images of colony morphology after culture of WT or ThPOK KO MDPs in methyl cellulose (left), or of dissociated cells after Giemsa staining (right). Note increased neutrophil frequency in ThPOK −/− versus WT MDP cultures. **c)** Bar graphs indicate absolute cell number of indicated donor-derived gated myeloid subsets (n = 5 independent animals) in blood after *in vivo* MDP transfer (gated as in Fig. 6F). Error bars represent standard deviations (centre line refers to mean). Significant differences between ThPOK−/− and WT mice were determined by two-sided unpaired t test with Welch's correction, and indicated by asterisks (* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements:

This work was supported by NIH grants R01 AI068907 (DJK), R01 GM107179 (DJK), R21 CA195356 (DJK), R21AI164333 (JB), R01 HL122661 (to HLG.), P30 CA006927 (FCCC Comprehensive Cancer Center Core Grant), and R50 CA211479 (MBE). We are thankful for the assistance of Margret B Einarson in confocal imaging. We acknowledge the assistance of the following core facilities of the Fox Chase Cancer Center: Transgenic Mouse, Flow Cytometry, Cell Culture, DNA Sequencing, Genomic and Laboratory Animal facilities. We are grateful to D.L. Wiest, G. Rall, E. Koltsova for critical comments and suggestions on the manuscript.We thank Dr W. Baldwin III, for providing expertise with analysis of neutrophil morphology. We thank Dr C. E. Canty for logistical support during the revision.

Data availability.

scRNA-Seq and CITE-Seq raw and processed data/metadata are deposited as GEO accession number GSE163844. There are no data restrictions.

References:

1. Netherby CS, et al. The Granulocyte Progenitor Stage Is a Key Target of IRF8-Mediated Regulation of Myeloid-Derived Suppressor Cell Production. J Immunol 198, 4129–4139 (2017). [PubMed: 28356386]
2. Yanez A, Goodridge HS, Gozalbo D & Gil ML TLRs control hematopoiesis during infection. Eur J Immunol 43, 2526–2533 (2013). [PubMed: 24122753]
3. Manz MG & Boettcher S Emergency granulopoiesis. Nat Rev Immunol 14, 302–314 (2014). [PubMed: 24751955]
4. Yanez A, Megias J, O'Connor JE, Gozalbo D & Gil ML Candida albicans induces selective development of macrophages and monocyte derived dendritic cells by a TLR2 dependent signalling. PLoS One 6, e24761 (2011). [PubMed: 21935459]
5. Basu S, et al. "Emergency" granulopoiesis in G-CSF-deficient mice in response to Candida albicans infection. Blood 95, 3725–3733 (2000). [PubMed: 10845903]
6. Serbina NV, Hohl TM, Cherny M & Pamer EG Selective expansion of the monocytic lineage directed by bacterial infection. J Immunol 183, 1900–1910 (2009). [PubMed: 19596996]
7. Olsson A, et al. Single-cell analysis of mixed-lineage states leading to a binary cell fate choice. Nature 537, 698–702 (2016). [PubMed: 27580035]
8. He X, et al. The zinc finger transcription factor Th-POK regulates CD4 versus CD8 T-cell lineage commitment. Nature 433, 826–833 (2005). [PubMed: 15729333]
9. He X, et al. CD4-CD8 lineage commitment is regulated by a silencer element at the ThPOK transcription-factor locus. Immunity 28, 346–358 (2008). [PubMed: 18342007]
10. Park K, et al. TCR-mediated ThPOK induction promotes development of mature (CD24-) gammadelta thymocytes. EMBO J 29, 2329–2341 (2010). [PubMed: 20551904]
11. Engel I, Zhao M, Kappes D, Taniuchi I & Kronenberg M The transcription factor Th-POK negatively regulates Th17 differentiation in Valpha14i NKT cells. Blood 120, 4524–4532 (2012). [PubMed: 23034280]
12. Lee HO, et al. Disregulated expression of the transcription factor ThPOK during T-cell development leads to high incidence of T-cell lymphomas. Proc Natl Acad Sci U S A 112, 7773–7778 (2015). [PubMed: 26056302]
13. Evrard M, et al. Developmental Analysis of Bone Marrow Neutrophils Reveals Populations Specialized in Expansion, Trafficking, and Effector Functions. Immunity 48, 364–379 (2018). [PubMed: 29466759]

14. Xie X, et al. Single-cell transcriptome profiling reveals neutrophil heterogeneity in homeostasis and infection. Nat Immunol, 21, 1119–1133 (2020). [PubMed: 32719519]
15. Blazek K, et al. IFN-λ resolves inflammation via suppression of neutrophil infiltration and IL-1β production. J Exp Med. 212, 845–53 (2015). [PubMed: 25941255]
16. Shrum B, et al. A robust scoring system to evaluate sepsis severity in an animal model. BMC Res Notes 7, 233 (2014). [PubMed: 24725742]
17. Muench DE, et al. Mouse models of neutropenia reveal progenitor-stage-specific defects. Nature 582, 109–114 (2020). [PubMed: 32494068]
18. Kwok I, et al. Combinatorial Single-Cell Analyses of Granulocyte-Monocyte Progenitor Heterogeneity Reveals an Early Uni-potent Neutrophil Progenitor. Immunity (2020).
19. Amir el AD, et al. viSNE enables visualization of high dimensional single-cell data and reveals phenotypic heterogeneity of leukemia. Nat Biotechnol 31, 545–552 (2013). [PubMed: 23685480]
20. Yamamoto R, et al. Clonal analysis unveils self-renewing lineage-restricted progenitors generated directly from hematopoietic stem cells. Cell. 154, 1112–1126 (2013). [PubMed: 23993099]
21. DePasquale EAK, et al. cellHarmony: cell-level matching and holistic comparison of single-cell transcriptomes. Nucleic Acids Res 47, e138 (2019). [PubMed: 31529053]
22. La Manno G, et al. RNA velocity of single cells. Nature, 494 (2018).
23. Nutt SL, Chopin M Transcriptional Networks Driving Dendritic Cell Differentiation and Function. Immunity16, 942 (2020).
24. Sunami Y, et al. BCL11A promotes myeloid leukemogenesis by repressing PU.1 target genes. Blood Adv. 6, 1827–1843 (2022). [PubMed: 34714913]
25. Gruber TA, et al. An Inv(16)(p13.3q24.3)-encoded CBFA2T3-GLIS2 fusion protein defines an aggressive subtype of pediatric acute megakaryoblastic leukemia. Cancer Cell 22, 683–697 (2012). [PubMed: 23153540]
26. Masetti R, Bertuccio SN, Pession A & Locatelli F CBFA2T3-GLIS2-positive acute myeloid leukaemia. A peculiar paediatric entity. Br J Haematol 184, 337–347 (2019). [PubMed: 30592296]
27. Saha S, et al. Transcriptomic Analysis Identifies RNA Binding Proteins as Putative Regulators of Myelopoiesis and Leukemia. Front Oncol 9, 692 (2019). [PubMed: 31448224]
28. Diaz-Munoz MD & Turner M Uncovering the Role of RNA-Binding Proteins in Gene Expression in the Immune System. Front Immunol 9, 1094 (2018). [PubMed: 29875770]
29. Wong JJ, et al. Orchestrated intron retention regulates normal granulocyte differentiation. Cell 154, 583–595 (2013). [PubMed: 23911323]
30. Itskovich SS, et al. MBNL1 regulates essential alternative RNA splicing patterns in MLL-rearranged leukemia. Nat Commun 11, 2369 (2020). [PubMed: 32398749]
31. Skoda RC & Schwaller J Dual roles of EZH2 in acute myeloid leukemia. J Exp Med 216, 725–727 (2019). [PubMed: 30890555]
32. Sumter TF, et al. The High Mobility Group A1 (HMGA1) Transcriptome in Cancer and Development. Curr Mol Med 16, 353–393 (2016). [PubMed: 26980699]
33. Iervolino A, et al. hnRNP A1 nucleocytoplasmic shuttling activity is required for normal myelopoiesis and BCR/ABL leukemogenesis. Mol Cell Biol 22, 2255–2266 (2002). [PubMed: 11884611]
34. Hodson DJ, Screen M & Turner M RNA-binding proteins in hematopoiesis and hematological malignancy. Blood 133, 2365–2373 (2019). [PubMed: 30967369]
35. Basu J, et al. Essential role of a ThPOK autoregulatory loop in the maintenance of mature $CD4^+$ T cell identity and function. Nat Immunol 22, 969–982 (2021). [PubMed: 34312548]
36. Sun F, Xiao G, Qu Z Murine Bronchoalveolar Lavage. Bio Protoc. 7: e2287 (2017).
37. Gonçalves R and Mosser DM The isolation and characterization of murine macrophages. Curr. Protoc. Immunol 111: 14.1.1–14.1.16 (2015).
38. Venkatasubramanian M, Chetal K, Schnell DJ, Atluri G, Salomonis N, Resolving single-cell heterogeneity from hundreds of thousands of cells through sequential hybrid clustering and NMF. Bioinformatics 36: 3773 (2020). [PubMed: 32207533]
39. Muench DE, et al. SKI controls MDS-associated chronic TGF-beta signaling, aberrant splicing, and stem cell fitness. Blood 132: e24–e34 (2018). [PubMed: 30249787]

40. Muller PA, et al. Microbiota modulate sympathetic neurons via a gut-brain circuit. Nature. 583: 441–446 (2020). [PubMed: 32641826]

41. Esterházy D, et al. Compartmentalized gut lymph node drainage dictates adaptive immune responses. Nature. 569: 126–130 (2019). [PubMed: 30988509]

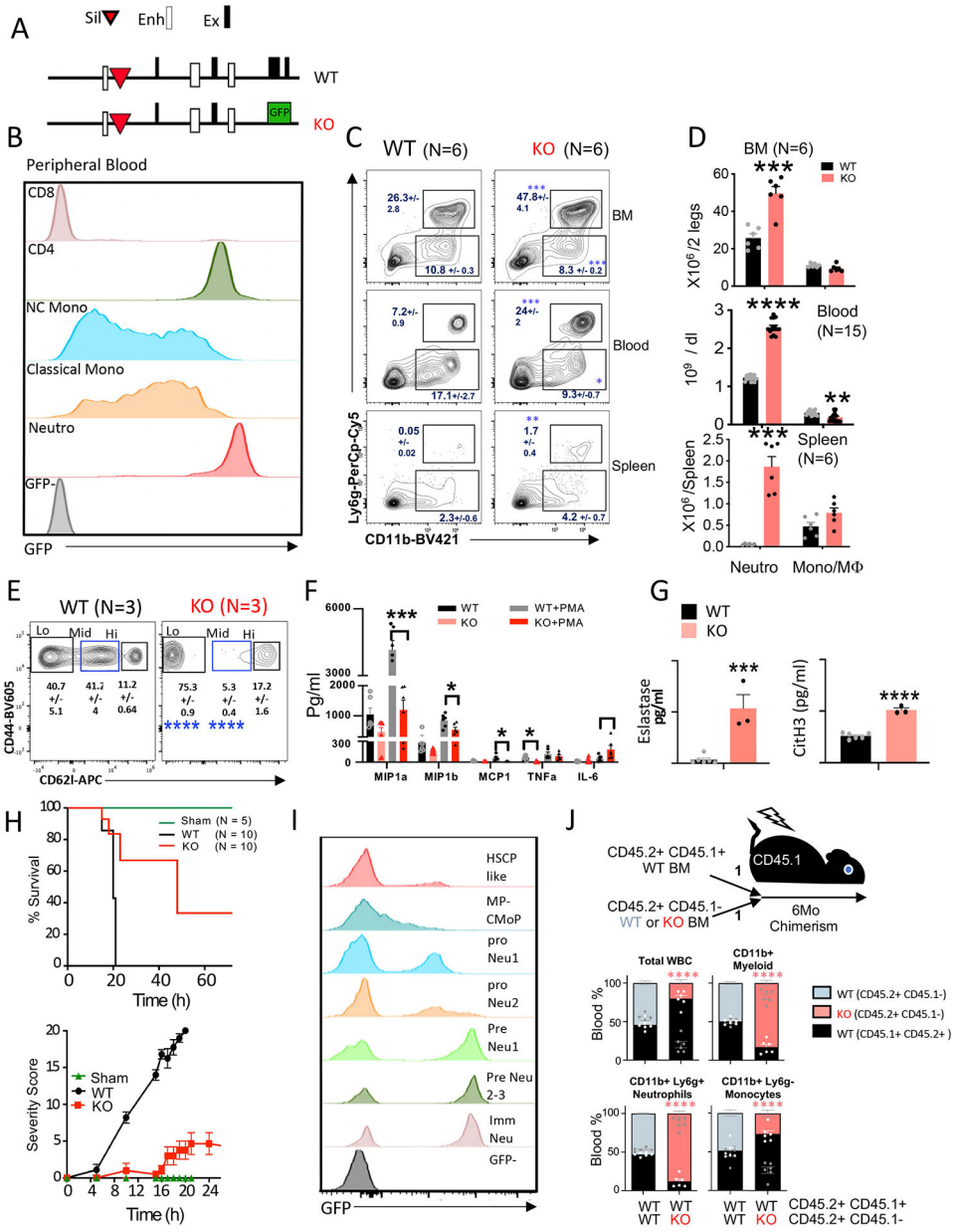


Fig. 1. ThPOK knockout causes homeostatic neutrophilia.

a) Schematic of ThPOK gene organization in wt mice (top row), or ThPOK-GFP reporter transgenic construct (bottom row). Black boxes indicate exons. Enhancers and silencer elements are shown as white boxes, or red triangle, respectively. **b)** FACS analysis of ThPOK-GFP reporter expression by indicated gated peripheral blood subsets. **c)** FACS analysis of CD11b and Ly6g expression by total cells from indicated organs. Note increase in $CD11b^{+}$ $Ly6g^{hi}$ neutrophils in ThPOK-deficient compared to WT mice. **d)** Plots showing total number of neutrophils or monocytes/macrophages per indicated organ in ThPOK-deficient compared to WT mice (same mice as panel c). Error bars represent standard deviations. Two sided unpaired T test was performed. **e)** FACS analysis of CD62l and CD44 expression by ThPOK−/− and WT neutrophils in peripheral blood. **f)** cytokine/chemokine

release by ThPOK−/− and WT neutrophils upon *in vitro* stimulation. Two sided unpaired T test was performed. **g)** Plots showing Elastase (right) or NET (left) upon *in vitro* stimulation. **h)** Plots of survival (left) or disease severity score (right) in an *in vivo* model of sepsis, comparing ThPOK-deficient or WT mice treated with coliform bacteria, or sham-infected, as indicated. Note enhanced protection of ThPOK-deficient versus WT mice. **i)** FACS analysis of ThPOK-GFP reporter expression by indicated gated bone marrow subsets. **j)** Competitive reconstitution assay to compare ability of WT and ThPOK−/− BM cells to contribute to myeloid compartment. Equal numbers of CD45.2+ CD45.1- ThPOK+/+ or ThPOK−/− BM cells were cotransferred with CD45.1+ CD45.2+ ThPOK+/+ competitor BM cells into lethally irradiated congenic CD45.1+ recipient mice. Relative contribution to the myeloid compartment in peripheral blood was assessed at 6 months post-transplant. Data are normalized for only donor cells, so that the sum of WT and KO donor cells equals 100% of the total (n = 6, per genotype). Data are displayed as mean ± SEM. *P < 0.05; **P < 0.01; ***P < 0.001; ****P < 0.0001. Two sided unpaired T test with Welch's Correction was performed.

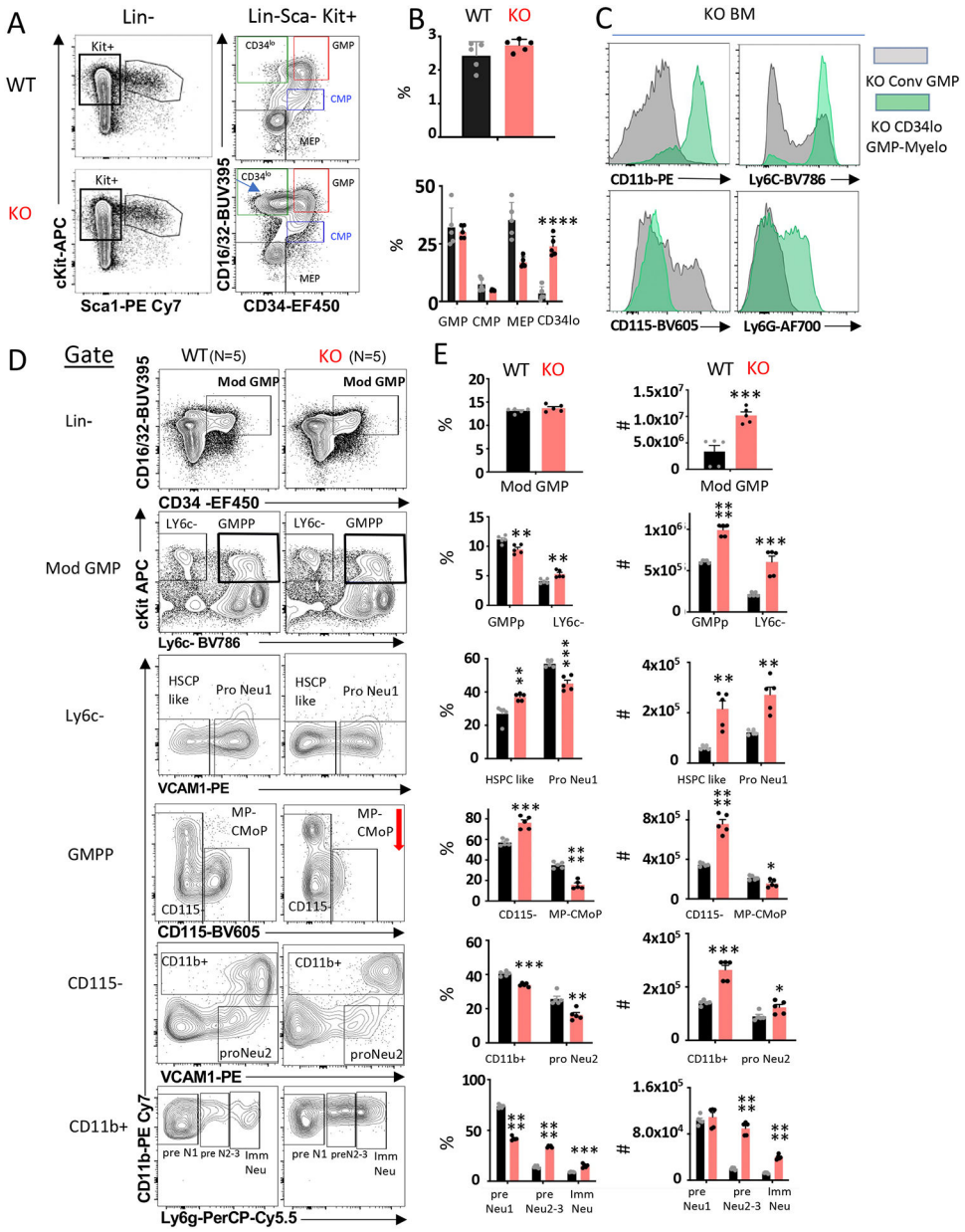


Fig. 2. ThPOK expression is necessary for neutrophil versus monocyte lineage choice and homeostatic maturation of neutrophils.

a) FACS analysis of cKit, Sca1, CD16/32 and CD34 expression by gated total lineage negative (Lin-) or Lin- cKit+ Sca1- BM cells, as indicated. CMP, GMP, MEP and CD34^{lo} gates are shown in right panels. **b)** Plots showing frequency of Lin-Sca-cKit+ cells in total BM and of CMP, GMP, MEP and CD34^{lo} subsets in gated cKit+ Sca1- Lin- BM cells from ThPOK-deficient or WT mice (same gates as panel a) (n = 5 biologically independent animals per genotype). Error bars represent standard deviations (centre refers to mean). Significant differences between ThPOK−/− and WT mice were determined by two sided unpaired T test with Welch's correction, and indicated by asterisks (* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$). **c)** FACS comparison of CD11b, CD115, Ly6c, and Ly6g expression by gated cGMP (conventional GMP) (grey histograms) and atypical c-Kit+

$CD34^{lo}$ $CD16^{hi}$ Lin- (green histograms) BM subsets, from ThPOK-deficient mice. **d)** FACS analysis of CD11b, Ly6g, Vcam1, CD115, Ly6c, CD16/32, cKit and CD34 expression by indicated gated BM subsets. Gates for preNeu1-3 (preN1-3), proNeu1/2 and other precursor populations are shown. **e)** Plots showing frequency (left panels) or cell number (right panels) of indicated gated precursor populations in ThPOK-deficient or WT mice (same mice as panel d). Data are presented as mean values +/− SEM. Significant differences between ThPOK−/− and WT mice were determined by two sided unpaired T test, and indicated by asterisks (* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$). n = 5 biologically independent animals.

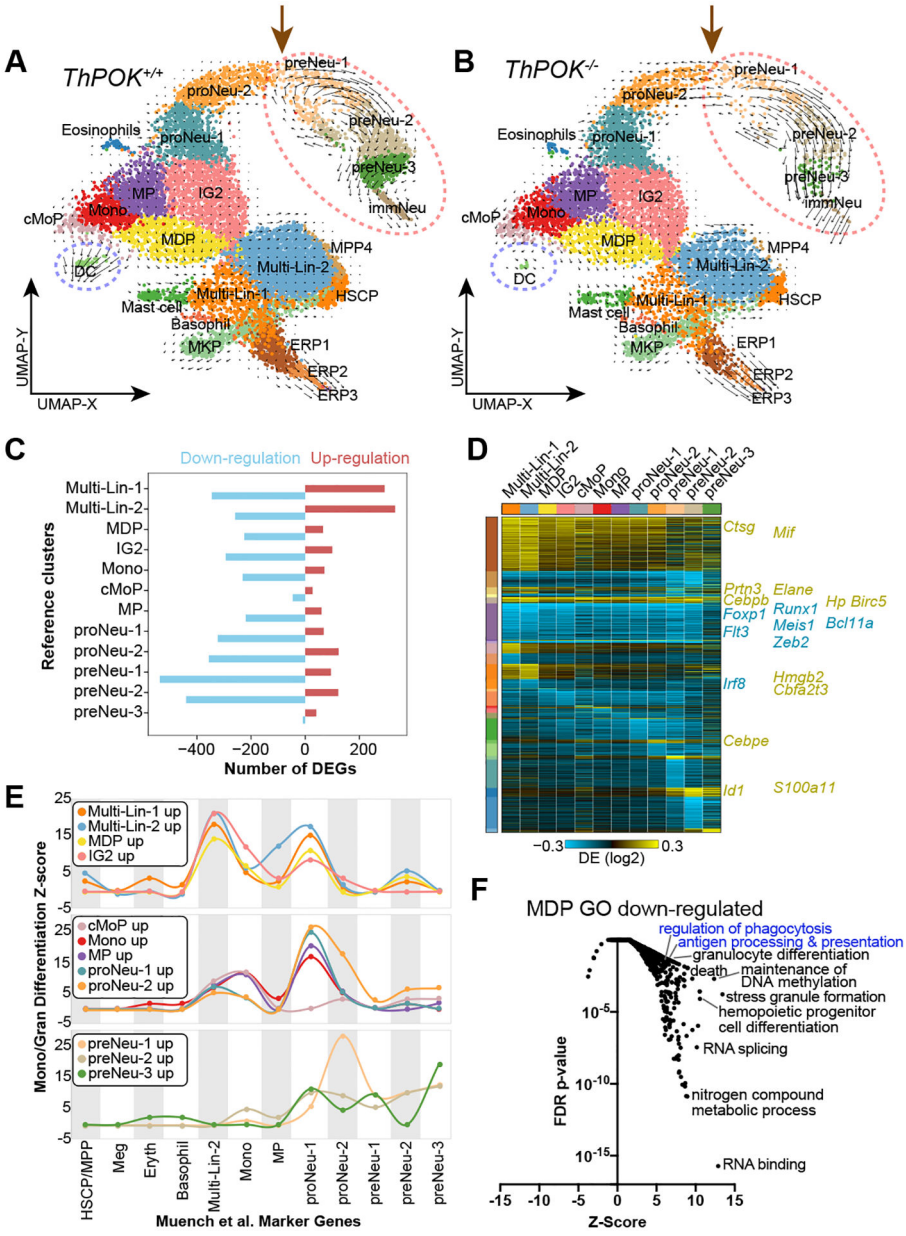


Figure 3. ThPOK controls neutrophil differentiation.

a-b) UMAP plots of (**a**) wild- type (left) and (b) ThPOK−/− (right) C-GMP myeloid progenitors (Lin-, Kit+, Sca1-, CD34hi, CD16/32+/−, gated as in Extended data Fig.5b). Shown cell populations were defined based on reference alignment to a prior curated murine cKit+ CITE-Seq dataset (Extended data Fig.5a). Overlaid arrows indicate predicted RNA velocities derived from spliced versus unspliced scRNA-Seq reads. Areas with the greatest predicted observed trajectory differences are denoted by purple or red circles. **c**) Number of differentially expressed genes (DEGs) that are up- or downregulated in ThPOK−/− versus wild-type cells by cellHarmony analysis for indicated cell populations. **d**) cellHarmony organized heatmap of dynamically regulated DEGs in ThPOK−/− and wt for the most frequency detected cell populations (n=1,685 genes, fold > 1.1 and empirical Bayes

moderated t-test p<0.05, FDR corrected). Yellow = upregulated gene, blue = downregulated gene. Genes noted in the text are called out to the right of the heatmap. **e**) Relative statistical enrichment (GO-Elite Z-score) of ThPOK−/− versus wild-type up-regulated genes against all prior defined hematopoietic differentiation markers[17], indicates altered differentiation programs in ThPOK−/− mice, for lineage priming (top), lineage specification (middle) and neutrophil commitment (bottom). **f**) Gene Ontology enrichment analysis (GO-Elite) of down- (right) and up-regulated genes in ThPOK−/− versus wt MDP cells, with example terms highlighted.

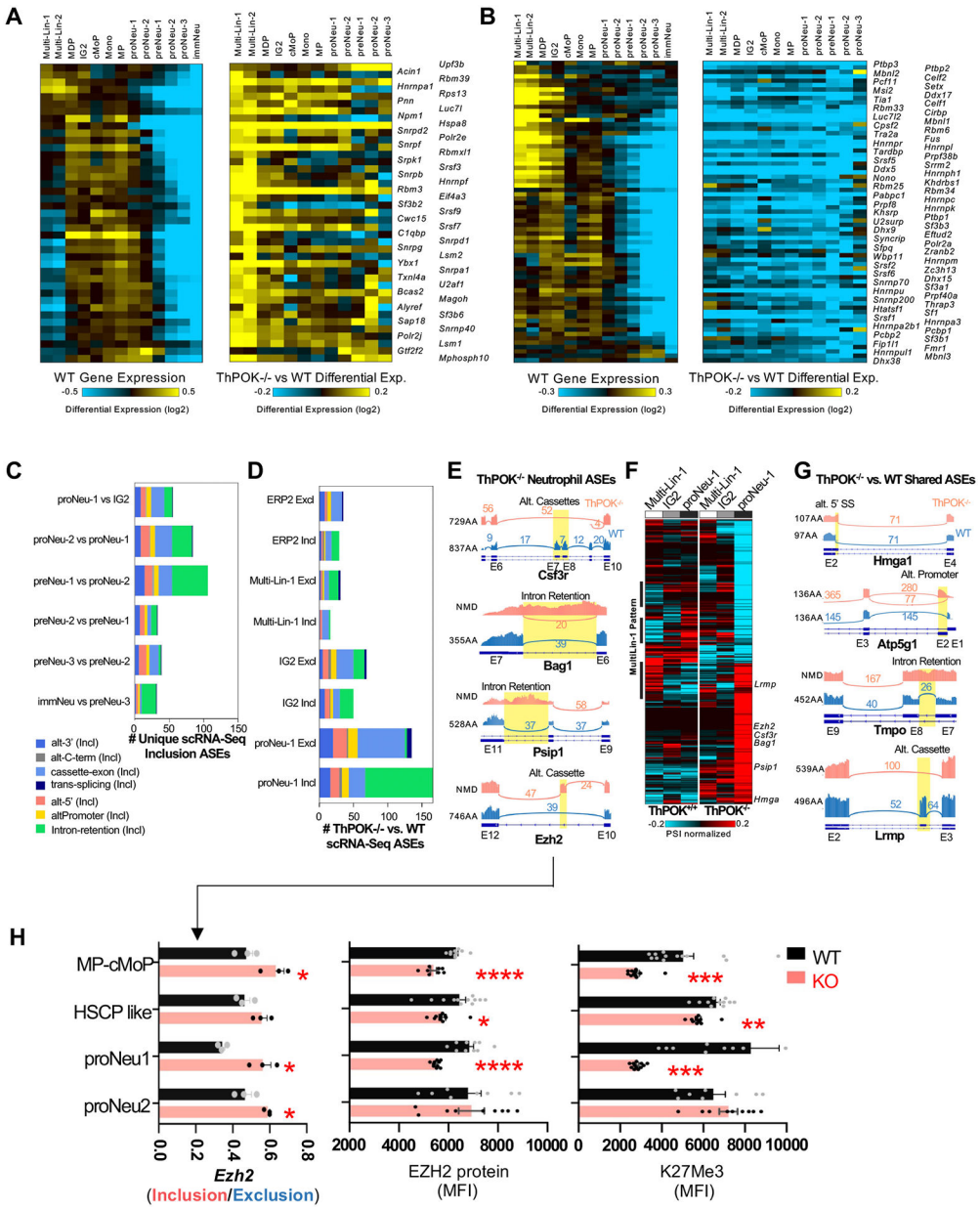


Fig. 4. ThPOK alters normal neutrophil specification mRNA splicing, particularly by control of intron-retention.

a-b) Heat maps showing relative mRNA expression of indicated mRNA binding protein genes in ThPOK−/− versus WT progenitor subsets (right panels), for genes that are relatively up- **(a)** or downregulated **(b)** in ThPOK−/− cells. Note that in WT mice most/all of these genes are strongly downregulated during granulopoiesis (**a,b**, left panels). **c)** Differential splicing analysis (MultiPath-PSI algorithm) of WT BM cell progenitor subsets identifies frequent intron retention in granulopoiesis (only inclusion Alternative Splicing Events (ASEs) are shown). **d)** Differential splicing analysis of ThPOK−/− versus WT BM progenitors showing unique inclusion (incl) and exclusion (excl) ASEs for indicated cell populations. **e)** SashimiPlot visualization of example novel ASEs from ThPOK−/− progenitors (pink) versus predominant ASEs from WT progenitors (blue). Yellow regions

indicate the location of alternative exon/intron regions (exon numbers indicated below). The outcome of the ASE is indicated at the left (i.e. size of novel protein generated, or nonsense-mediated decay [NMD], as appropriate). **f**) Heat map showing relative expression of ASE products for indicated ThPOK −/− and WT progenitor subsets. **g**) SashimiPlot visualization of example ASEs from ThPOK−/− cells (pink) versus ASEs observed in normal myelopoiesis (blue). Yellow regions indicate the location of alternative exon/intron regions (exon numbers indicated below), the outcome of the ASE is indicated at the left. **h)** Expression of Ezh2 alternate transcript (ratio compared to reference Ezh2 transcript; left panel), EZH2 protein (middle panel), or H3K27Me3 mark (right panel), in indicated progenitor populations from WT and ThPOK−/− mice. mRNA was pooled from 4 mice for each genotype. EZH2 protein and H3K27Me3 levels were assessed by confocal microscopy (mean fluorescent intensities, MFI/per cell, calculated from 8 different fields containing >7 cells/field). Note that increase in alternate Ezh2 transcript correlates with decrease in EZH2 protein expression and H3K27Me3 mark in monocytic progenitors as well as proNeu1 cells. Data are presented as mean values +/− SEM. Significant differences between ThPOK−/− and WT mice were determined by two-sided unpaired t test with Welch's correction, and indicated by asterisks (* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$). For left panel, n = 3 technical replicates, each derived from pooled samples from 5 mice. For center and right panels, n = 8–12 microscopic field scans, each containing at least 8 cells.

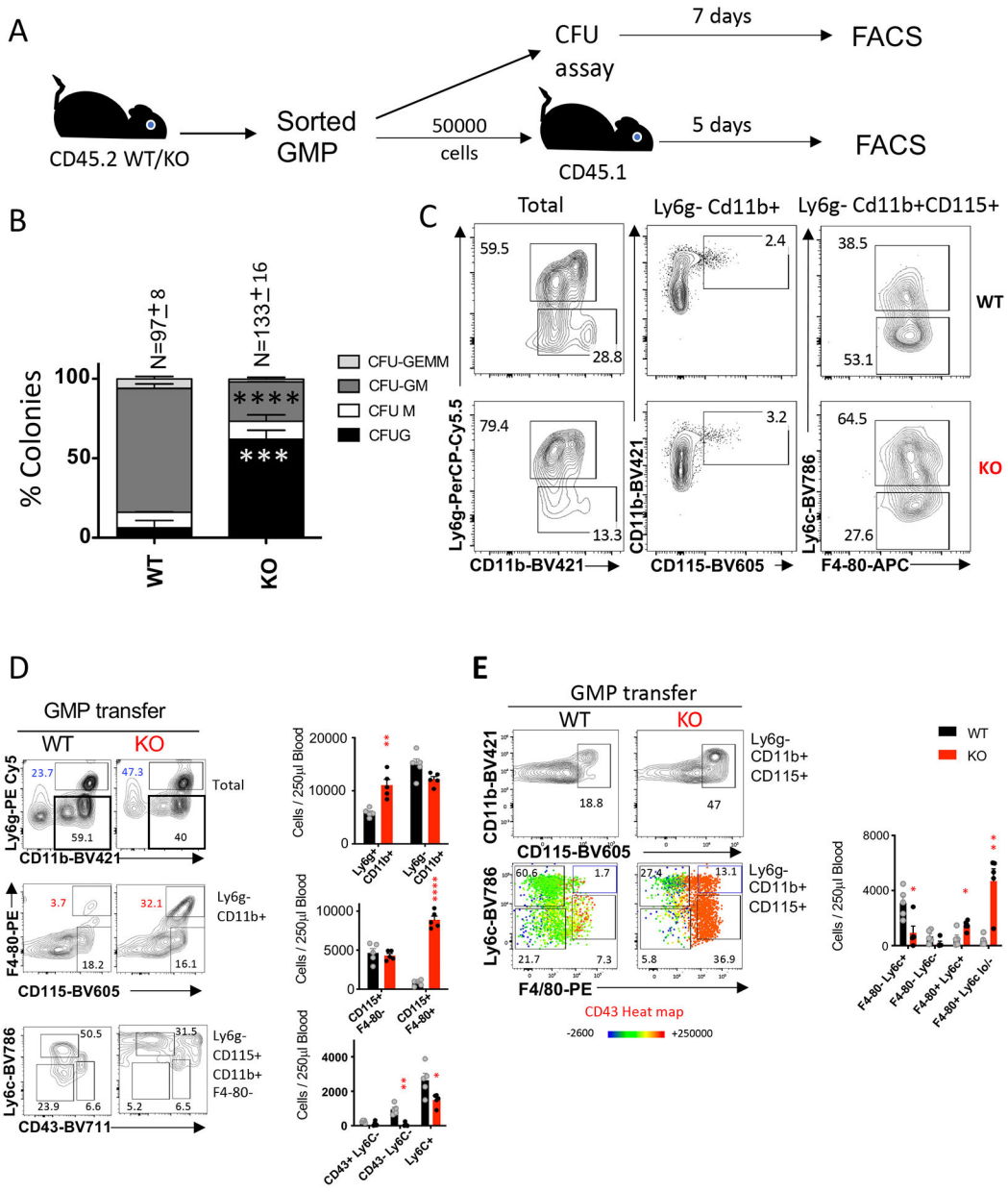


Fig. 5. ThPOK-deficient GMPs are biased towards the granulocytic lineage.

a) Schematic of in vitro and in vivo assays of GMP lineage potential. **b)** Frequency of morphologically identified colonies from *in vitro* bilineage colony formation assay on flow-sorted GMP-gate progenitors (Lin- Kit+ CD34hi CD16hi) from ThPOK-deficient versus WT mice. Note, greatly increased number of granulocytic colonies produced by *ThPOK*$^{-/-}$ GMPs (GEMM = granulocyte, erythrocyte, macrophage, megakaryocyte; GM = granulocyte, macrophage; G = granulocyte; M = macrophage). Significant differences were determined by two-sided unpaired t test with Welch's correction. **c)** Flow cytometric analysis of cells from same *in vitro* bilineage colony formation assay as in panel b. **d), e)** Flow cytometric analysis of blood cells derived from *in vivo* transfer of flow sorted GMP-gate progenitors (Lin- Kit+ CD34hi CD16hi) from ThPOK-deficient versus WT mice transferred into

non-irradiated recipient mice. Note that ThPOK-deficient progenitors show predominant differentiation towards neutrophils. Bar graphs indicate absolute cell number of indicated gated myeloid subsets (n = 5 biologically independent animals per genotype). Error bars represent standard deviations. Significant differences between ThPOK−/− and WT mice were determined by two-sided unpaired t test with Welch's correction, and indicated by asterisks (* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$).

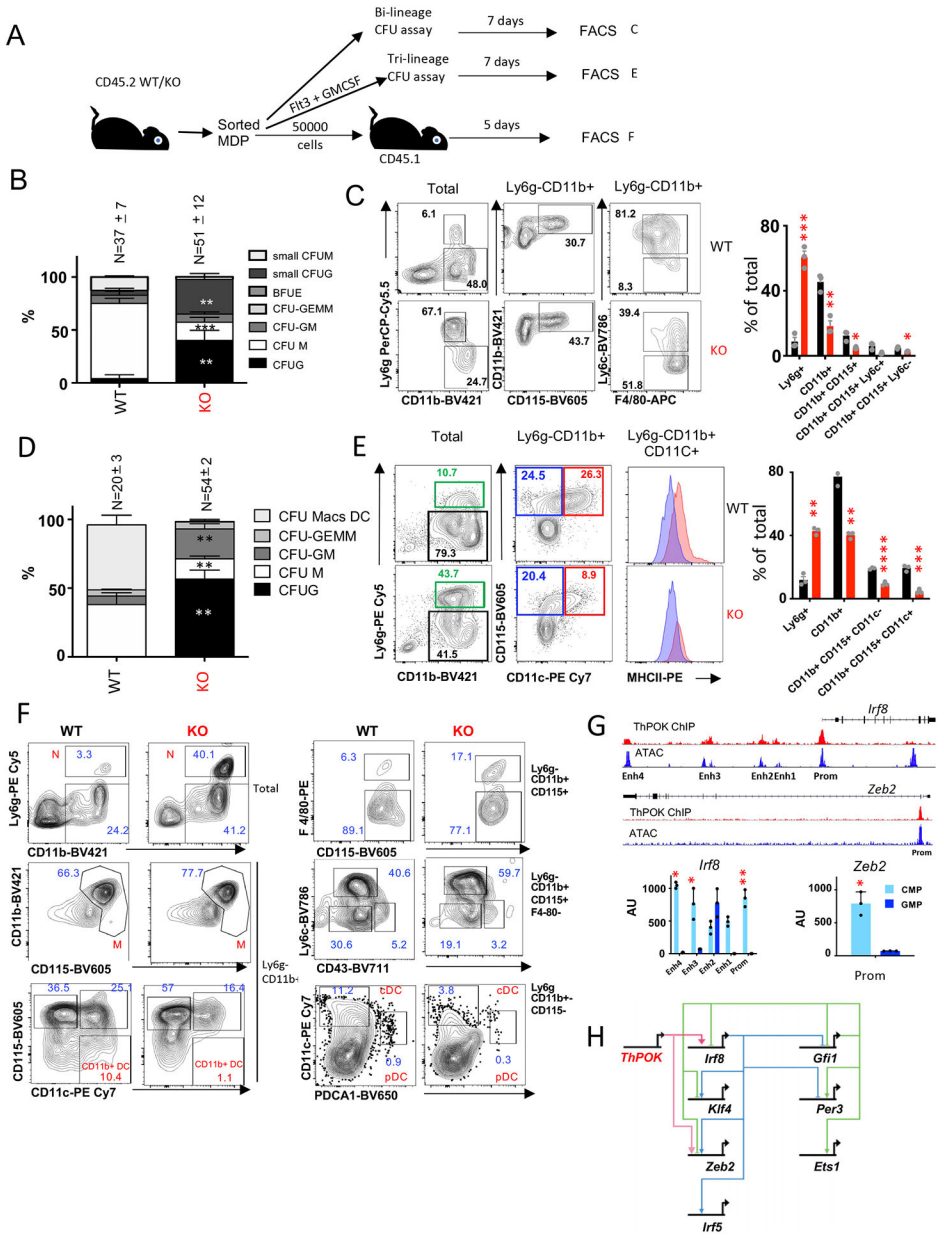


Fig. 6. ThPOK-deficient MDP produce neutrophils *in vitro and in vivo*.

a) Schematic of *in vitro* and *in vivo* assays of MDP lineage potential. **b)** Frequency of morphologically identified colonies from *in vitro* bilineage colony formation assay on flow sorted MDP-gated (Kit+CD34+CD16/32-, Flt3+, CD115+) progenitors from ThPOK-deficient versus WT mice. Note, greatly increased number of granulocytic colonies produced by *ThPOK*$^{-/-}$ MDPs. N = 37 and N = 51 refer to average number of colonies per well (averaged for 3 wells). Data are presented as mean values +/− SEM. **c)** Flow cytometric analysis of cells from same *in vitro* bilineage colony formation assay as in panel b. Bar graphs at right represent percentages of indicated subsets/ well according to gating strategy shown at left (n = 3). Single data point represents pulled MDP population from 3 mice. Data are displayed as mean ± SEM. *P < .05; **P < .01; ***P < .001; ****P < .0001. **d)**

Frequency of morphologically identified colonies from *in vitro* trilineage (DC-permissive) colony formation assay on flow sorted MDP-gate progenitors from ThPOK-deficient versus WT mice. n = average colony number per well. Data are presented as mean values +/− SEM. **e)** Flow cytometric analysis of cells from same *in vitro* bilineage colony formation assay as in panel D. Bar graphs at right represent percentages of indicated subsets according to gating strategy shown at left (n = 3). Data are displayed as mean ± SEM. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$. Single data point represents pulled MDP population from 3 mice. **f)** Flow cytometric analysis of blood cells derived from *in vivo* transfer of flow sorted MDP-gate progenitors from ThPOK-deficient versus WT mice transferred into non-irradiated recipient mice. **g)** Anti-ThPOK Chip-seq (red; GSE116506) and ATAC-seq (blue; GSE168772.) tracks for Irf8 and Zeb2 genes (top), ThPOK Chip analyses of FACS-sorted CMP or GMP subsets from heterozygous Flag-tagged-ThPOK knock-in mouse, showing differential ThPOK binding at *Irf8* and *Zeb2* loci (bottom). Data are presented as mean values +/− SD. **h)** Schematic of regulation of myeloid gene regulatory network by ThPOK. n = 3 technical replicates, each derived from pooled samples from 5 mice.