




ARTICLE OPEN



CaSSiDI: novel single-cell “Cluster Similarity Scoring and Distinction Index” reveals critical functions for PirB and context-dependent *Cebpb* repression

Robert Nechanitzky^{1,9,10}, Parameswaran Ramachandran^{1,10}, Duygu Nechanitzky¹, Wanda Y. Li², Andrew C. Wakeham¹, Jillian Haight¹, Mary E. Saunders¹, Slava Epelman^{3,4,5,6,7} and Tak W. Mak^{1,2,8}

© The Author(s) 2024

PirB is an inhibitory cell surface receptor particularly prominent on myeloid cells. PirB curtails the phenotypes of activated macrophages during inflammation or tumorigenesis, but its functions in macrophage homeostasis are obscure. To elucidate PirB-related functions in macrophages at steady-state, we generated and compared single-cell RNA-sequencing (scRNAseq) datasets obtained from myeloid cell subsets of wild type (WT) and PirB-deficient knockout (PirB KO) mice. To facilitate this analysis, we developed a novel approach to clustering parameter optimization called “Cluster Similarity Scoring and Distinction Index” (CaSSiDI). We demonstrate that CaSSiDI is an adaptable computational framework that facilitates tandem analysis of two scRNAseq datasets by optimizing clustering parameters. We further show that CaSSiDI offers more advantages than a standard Seurat analysis because it allows direct comparison of two or more independently clustered datasets, thereby alleviating the need for batch-correction while identifying the most similar and different clusters. Using CaSSiDI, we found that PirB is a novel regulator of *Cebpb* expression that controls the generation of Ly6C^{lo} patrolling monocytes and the expansion properties of peritoneal macrophages. PirB's effect on *Cebpb* is tissue-specific since it was not observed in splenic red pulp macrophages (RPMs). However, CaSSiDI revealed a segregation of the WT RPM population into a CD68^{lo}*Irf8*⁺ “neuronal-primed” subset and an CD68^{hi}*Ft1*⁺ “iron-loaded” subset. Our results establish the utility of CaSSiDI for single-cell assay analyses and the determination of optimal clustering parameters. Our application of CaSSiDI in this study has revealed previously unknown roles for PirB in myeloid cell populations. In particular, we have discovered homeostatic functions for PirB that are related to *Cebpb* expression in distinct macrophage subsets.

Cell Death & Differentiation (2024) 31:265–279; <https://doi.org/10.1038/s41418-024-01268-8>

INTRODUCTION


The murine “paired immunoglobulin-like receptor” (Pir)A and PirB molecules are expressed on osteocytes and leukocytes (especially myeloid cells), and transduce activatory and inhibitory signals through “immunoreceptor tyrosine-based activation motifs” (ITAMs) and the corresponding inhibitory motifs (ITIMs), respectively [1–3]. PirB is also expressed on neurons, where it blocks excessive axonal growth and neuronal regeneration [4]. Murine chromosome 7 carries six *PirA* genes and one *PirB* gene. The human orthologs of these genes are referred to as “leukocyte immunoglobulin-like receptors” (*LILR*) and are organized into the activatory *LILRA* family (*LILRA1–6*) and the inhibitory *LILRB* family (*LILRB1–5*).

Pir/LILR-mediated signaling modulates immune responses, and PirB-deficient knockout (PirB KO) mice develop exacerbated graft-versus-host disease [5]. Although the best characterized PirB

ligand is the $\beta 2m$ component of MHC class I, other ligands have been implicated in PirB/LILRB-mediated responses, including non-classical MHC class I molecules such as Qa2 and Cd1d [6], as well as non-MHC molecules such as alarmins and pathogen epitopes [2, 3]. In cancer tissues, abnormal PirB/LILRB expression is linked to enhanced tumor growth and poor patient prognosis [7]. However, MHC class I-LILRB3 interactions can induce the extrusion of precancerous epithelial cells [8]. Therefore, it is unclear precisely how PirB shapes myeloid cell responses.

Single-cell RNA sequencing (scRNAseq) assays provide valuable insights into cellular heterogeneities and functions, but the wide range of parameter choices (such as clustering resolution, number of reduced dimensions) can preclude accurate elucidation of novel activation states hidden within a cluster. In addition, most scRNAseq-based studies either do not describe the systematic methodology used to choose optimal parameter values such as

¹Princess Margaret Cancer Centre, Ontario Cancer Institute, University Health Network, Toronto, ON, Canada. ²Centre for Oncology and Immunology, Hong Kong Science Park, Hong Kong SAR, China. ³Toronto General Hospital Research Institute, University Health Network, Toronto, ON, Canada. ⁴Ted Rogers Centre for Heart Research, Translational Biology and Engineering Program, Toronto, ON, Canada. ⁵Peter Munk Cardiac Centre, UHN, Toronto, ON, Canada. ⁶Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, ON, Canada. ⁷Departments of Immunology and Medical Biophysics, University of Toronto, Toronto, ON, Canada. ⁸Department of Pathology Department of Pathology, School of Clinical Medicine, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong SAR, China. ⁹Present address: Providence Therapeutics Holdings Inc., Calgary, AB, Canada. ¹⁰These authors contributed equally: Robert Nechanitzky, Parameswaran Ramachandran.

email: Robert.Nechanitzky@utoronto.ca; Tak.Mak@uhn.ca

Received: 16 May 2023 Revised: 15 January 2024 Accepted: 7 February 2024

Published online: 21 February 2024

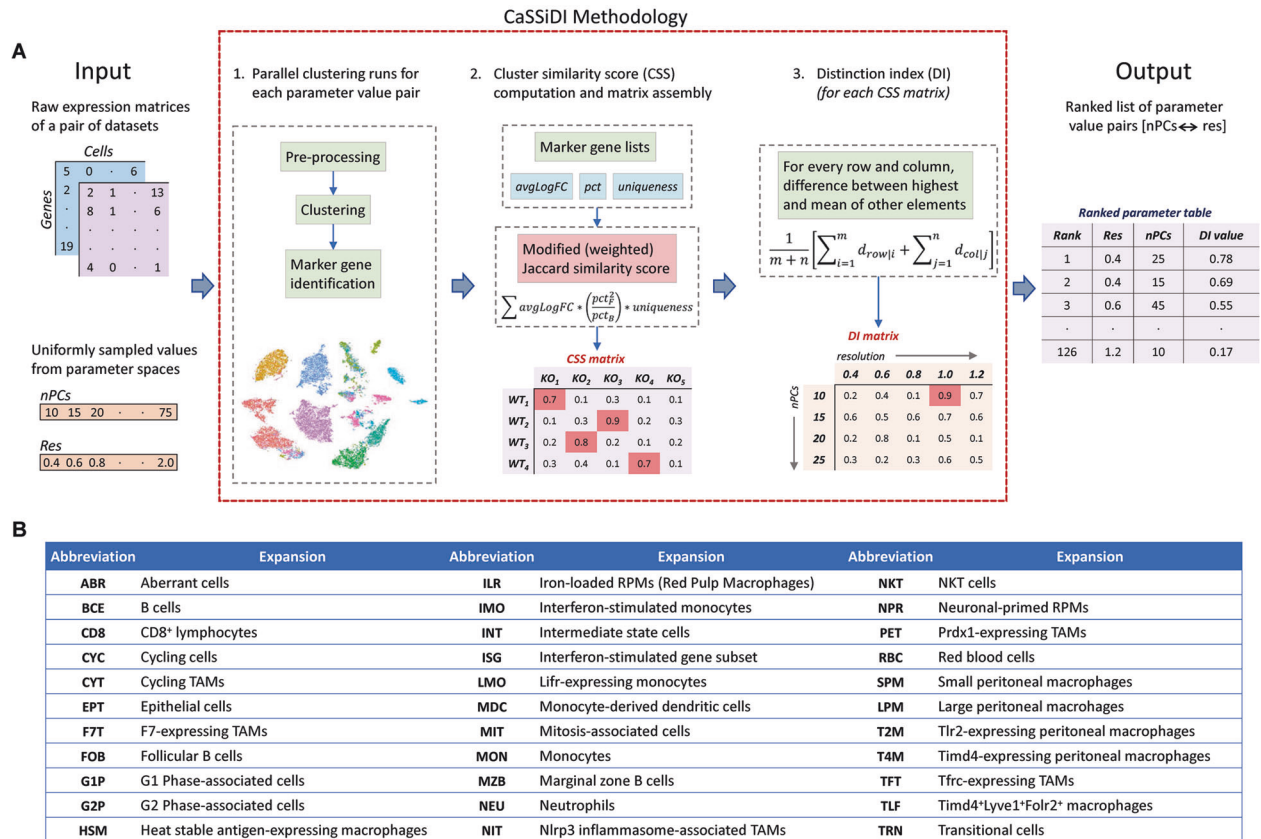


Fig. 1 CaSSiDI framework and cluster identities. **A** Inputs: a pair of raw gene expression matrices for datasets whose clusters are to be matched, and uniformly sampled values from the parameter spaces to be optimized. **1.** Pre-processing, clustering, and marker gene identification steps for all parameter value pairs are conducted in parallel. **2.** For each parameter value pair, an $m \times n$ cluster similarity score (CSS) matrix is assembled containing scores for all possible cluster pairs between the two datasets using modified (weighted) Jaccard similarity. **3.** Finally, a single score for each CSS matrix, called the distinction index (DI), is computed in two steps. Row- and column-level DIs are computed as the differences between the row-wise and column-wise maximums and means. The mean across these $m + n$ DIs yields the matrix level DI. The final output is a ranked list of parameter value pairs from best to worst. **B** A listing of the 3-letter codes used throughout this paper to identify cell subsets and/or clusters, followed by their definitions.

clustering resolution or do not employ one at all. While deep-learning and AI-based approaches can identify novel cellular activation states [9–13], not all researchers have access to advanced bioinformatics analysis pipelines. Here, based on the Jaccard index and known features of cell lineages, we present a data-driven, semi-automated parameter optimization method called “Cluster Similarity Scoring and Distinction Index” (CaSSiDI) (Fig. 1A). Use of CaSSiDI narrows down the clustering-related parameter values to a manageable set of top choices that correspond to meaningful clustering conditions. In the process of developing CaSSiDI, we also constructed useful visualizations such as the “Nebula plot”, which illustrates (in a single plot) the top shared and differentially expressed genes (DEGs) and their expression levels for a given cluster pair within the context of average background expression. In this study, we use these tools to explore the role of PirB in diverse myeloid cell subsets and analyze its effects on their gene expression patterns. We present results on normal macrophage populations in our main text, and additional findings on tumor-associated macrophages (TAMs) in our Supplemental Results.

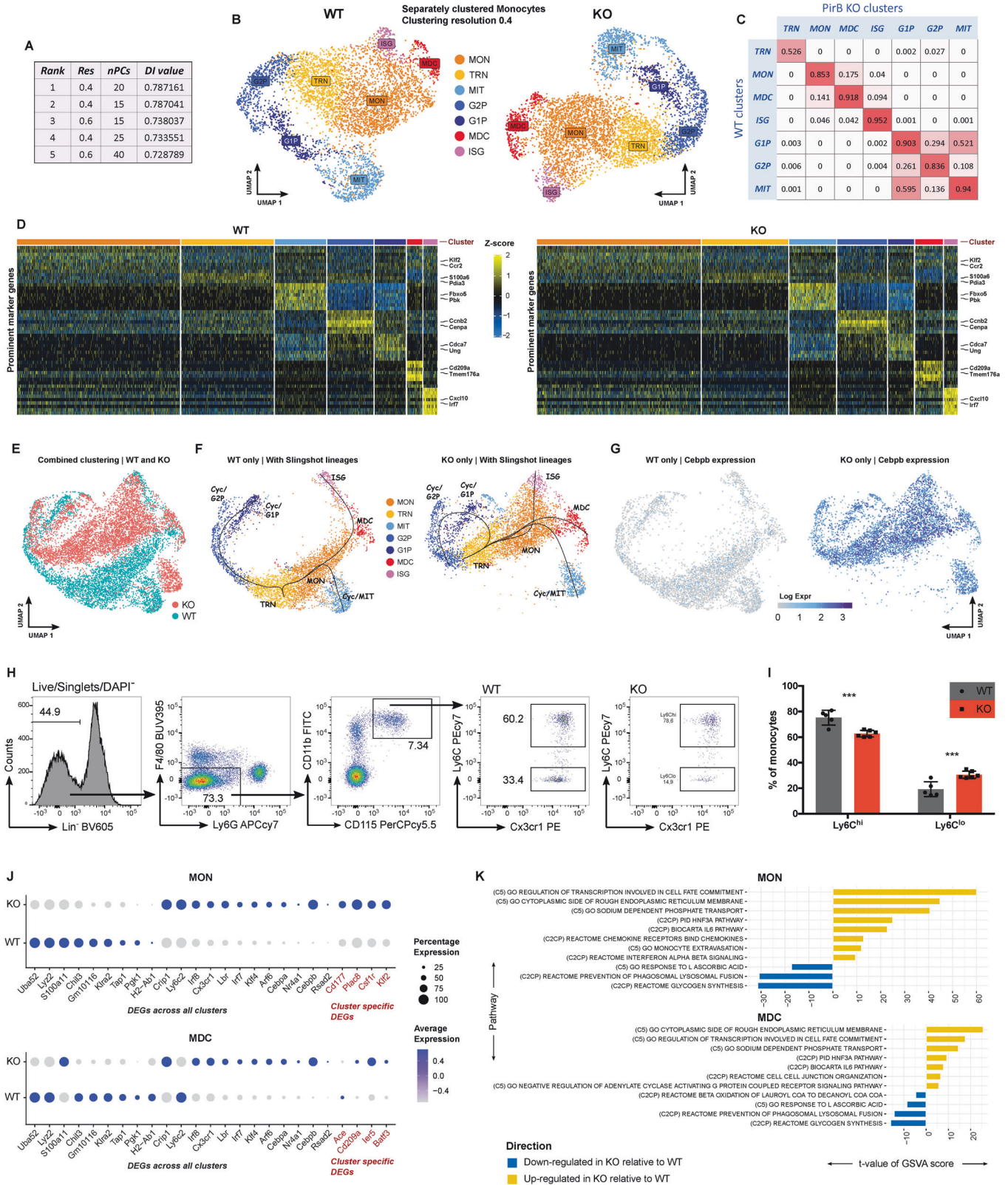
RESULTS

CaSSiDI reveals heterogeneity and *Cebpb* derepression in PirB^{-/-} Ly6C⁺ bone marrow-derived monocytes

To examine PirB’s role in monocyte cell subsets, we used flow cytometry to isolate Lin⁻CD11b⁺CD115⁺Ly6G⁻Ly6C⁺ bone marrow

(BM)-derived monocytes from WT and *Pirb*^{-/-} (PirB KO) female littermate mice and subjected them immediately to scRNAseq. Our CaSSiDI approach (Fig. 1 and see Methods) determined that a resolution (res) of 0.4 and 20 principal components (nPCs) was the optimal set of parameter choices (hereafter, clustering condition) for comparing WT to PirB KO monocytes (Fig. 2A). Seven WT and PirB KO monocyte cluster pairs with high similarity were identified (Fig. 2B, C), with 3 out of 7 pairs representing cycling cells. [Fig. 1B shows a comprehensive legend for all 3-letter cluster identification codes used in this paper.] Examination of cluster-associated marker genes revealed cells expressing *Fbxo5*, *Pbk*, *Cdk1*, *Aurkb* (linked to mitosis; MIT) in both WT and mutant cultures, as well as cells expressing *Cdca7*, *Asf1b*, *Ung* (in G1 phase/S phase; G1P), and those expressing *Ccnb2*, *Birc5*, *Cenpa* (in G2 phase; G2P) (Fig. 2D; Supplementary Tables S1 and S2). Notably, the premonocytic marker *Cxcr4* [14] was expressed by WT and mutant cells in the MIT and G1P clusters (Supplementary Tables S1 and S2).

We sought to elucidate the heterogeneity within specific monocyte subsets from WT and PirB KO mice. In both WT and mutant animals, a cluster termed MON showed high mRNA expression of classical monocyte-associated marker genes, including *Ccr2* (monocyte egress from the BM) [15–17]; the transcription factor (TF) *Klf2* (inhibits proinflammatory monocyte activation) [18]; as well as *Plac8*, *Csf1r* and the classical neutrophil marker *Cd177* (Fig. 2D, J). The monocyte-derived dendritic cell (MDC) cluster showed high mRNA levels of the MHC class II molecules H2-Ab1, H2-Eb1, and H2-Dmb1, as well as the DC-related markers



Cd74, *Cd209a*, *Batf3*, *Tmem176b* [19–21] (Fig. 2D, J) (Supplementary Tables S1 and S2). The monocyte subset expressing high mRNA levels of *Irf7*, *Ifit3*, *Cxcl10* was termed the interferon-stimulated gene (ISG) subset (Fig. 2D, Supplementary Fig. S1A). The transitional cell (TRN) subset showed only a handful of

upregulated genes, including *S100a6*, a calcium-binding protein and critical regulator of myeloid output and mitochondrial metabolism in hematopoietic stem and progenitor cells [22]. However, most genes associated with the TRN subset were downregulated (Fig. 2D, Supplementary Fig. S1A; Supplementary

Fig. 2 Steady-state monocytes: CaSSiDI reveals plausible WT and PirB KO clusters. **A** Top five parameter value pair choices as determined by CaSSiDI (*Res*: clustering resolution, *nPCs*: number of principal components). The top choice of $res = 0.4$, $nPCs = 20$ was used to cluster the WT and PirB KO datasets. **B** UMAP of separately clustered WT and PirB KO BM-derived monocytes. Please see Fig. 1B for definitions of abbreviations. **C** CSS table corresponding to the top-ranking choice in **A**. **D** Prominent marker genes for the indicated (by color) clusters. **E** UMAP of the combined clustering run of the WT and PirB KO monocyte populations. **F** Slingshot trajectory curves originating from the TRN cluster for the separately clustered WT and PirB KO monocyte populations laid out on the combined UMAP space. **G** UMAPs showing *Cebpb* expression for the WT and PirB KO monocyte populations in **F**. **H** Left: Flow cytometric gating strategy to identify blood-resident monocytes. Right: Results of this strategy applied to WT and PirB KO blood monocytes. **I** Quantitation of Ly6C^{hi} and Ly6C^{lo} blood monocytes in the WT and PirB-KO samples in **H** expressed as the percentage of total blood monocytes; WT $n = 5$, KO $n = 6$. $***P < 0.001$ as determined by regression analysis with two-way analysis of variance (ANOVA) followed by Sidak's post hoc multiple comparison test. Data are representative of two independent experiments. **J** Bubble plots showing prominent monocyte-related DEGs (PirB-dependent in black; cluster-specific in red) between WT and PirB KO cells within the indicated clusters as determined by the CSS table in **C**. **K** Top differentially expressed pathways identified by GSEA between WT and PirB KO cells within the indicated clusters. Also shown are t-value scores from a linear model analysis conducted on GSEA scores.

Tables S1 and S2), which is a common feature of progenitor or transitory cell populations [23]. Notably, the PirB KO TRN subset expressed *Pdia3* (Supplementary Table S2), which facilitates hematopoietic progenitor cell anchorage in the BM [24].

The phenotype of the TRN cluster prompted us to use it as the starting point for a Slingshot analysis (Fig. 2E, F). As is typical for progenitor-like cells, the trajectories of the WT and mutant MON, MDC, and ISG subsets indicated that the TRN subset likely possesses self-renewal capacity, with one differentiation path confined to the cycling clusters G2P and G1P and another path leading to the MON state. From there, we observed the MIT state (likely MON amplification) and the emergence of MDC and ISG cells from MON cells. These trajectory-related findings were equivalent in the corresponding WT and PirB KO clusters.

One striking difference between almost all WT and PirB KO monocyte subsets was an increase in mRNA levels of the transcription factor *Cebpb* in the mutant cells (Fig. 2G). We surmised that PirB might be required for *Cebpb* repression in BM-derived monocytes, implicating PirB in regulating the differentiation of Ly6C^{hi} into Ly6C^{lo} monocytes as well as into monocyte-derived macrophages and DCs [25, 26]. Indeed, flow cytometric assessment of blood monocyte subsets revealed a higher proportion of Ly6C^{lo} patrolling monocytes at the expense of Ly6C^{hi} monocytes in PirB KO mice compared to WT animals (Fig. 2H, I).

Although all corresponding WT and mutant monocyte clusters were highly similar (Fig. 2C), an impartial DE analysis and volcano plotting revealed numerous differentially expressed genes (DEGs) (Supplementary Fig. S1B–E). We identified 21 monocytic core genes that were deregulated in the absence of PirB (Fig. 2J and Supplementary Fig. S1A, black gene designations), as well as four DEGs that differed in a cluster-specific manner (Fig. 2J and Supplementary Fig. S1A, red gene designations). Global pathway analysis revealed that PirB deficiency affected numerous pathways but particularly “Response to L-ascorbic acid” and “Regulation of transcription involved in cell fate commitment” (Fig. 2K and Supplementary Fig. S1F). Taken together, these data indicate that PirB restricts *Cebpb* expression in monocytes.

Benchmark validation of CaSSiDI as an approach to improve clustering

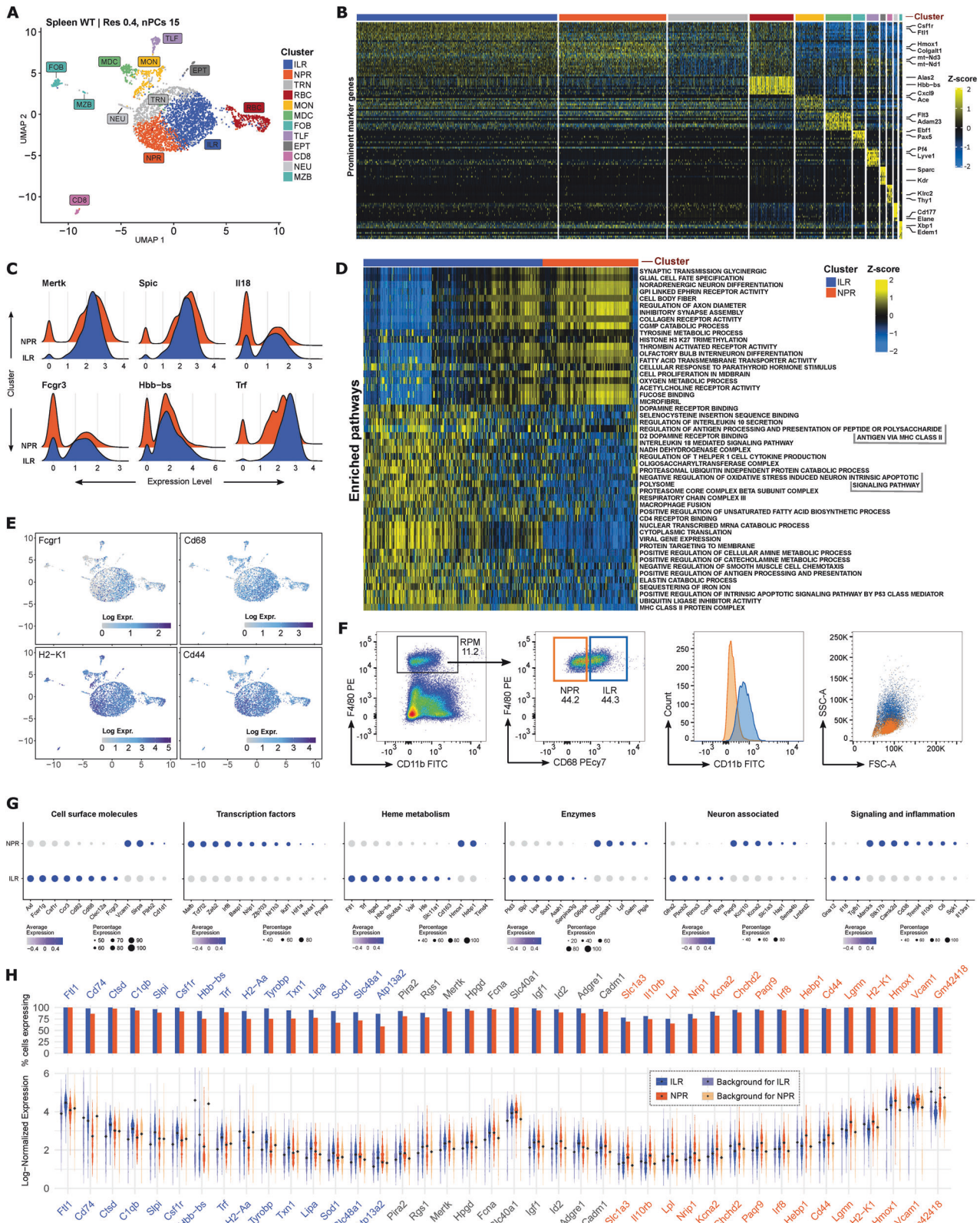
The results above indicated that CaSSiDI determined an optimal clustering condition that reliably predicted a range of established monocyte subsets. To validate our CaSSiDI method, we compared our data outcomes to those produced using a Standard Seurat Pipeline (SSP) procedure for cluster determination (see Supplemental Results for a detailed description). The standard Seurat approach recommends using $res = 0.8$ and determining the number of PCs by elbow plots, plotting *nPCs* against the standard deviation [27]. Accordingly, the optimal *nPC* value can be chosen approximately where the curve takes an elbow turn (Fig. S1G). Our application of SSP to our data above assigned parameters of

$res = 0.8$ and $nPC = 10$, generating nine WT and eight PirB KO monocyte clusters (Fig. S1H). Inspection of the marker genes associated with these clusters revealed the following insights, which highlight the utility of CaSSiDI. 1) For the WT sample, the MON, MDC, and the three cell cycle-associated clusters (G1P, G2P, MIT) showed similar arrays of marker genes when analyzed by SSP or CaSSiDI (Supplementary Table S3). However, SSP also assigned the largest cluster (ART) to cells that did not display positively regulated genes (Supplementary Table S3; Supplementary Fig. S1H). This outcome is not biologically meaningful and thus can only be interpreted as an artifact stemming from a suboptimal data analysis pipeline. 2) SSP assigned two clusters (TRN and TRN2) to cells containing only 41 or 12 positively regulated genes, respectively. These genes were predominantly linked to progenitor- and cell cycle-associated features, making it difficult to derive distinct biological functions (Supplementary Table S3). 3) The last two WT monocyte clusters (ISG and ISG2) identified by SSP shared a similar ISG-related gene expression signature that included *Irf7*, *Stat1*, *Oasl2*, *Irf47*, which is likely also due to over-clustering (Supplementary Table S3). 4) For the PirB KO monocyte dataset, both SSP and CaSSiDI identified MON, MDC, ISG, G1P, G2P, and MIT clusters with similar gene signatures. However, SSP generated an additional cluster with a vast number of cell cycle-related genes (CYC), as well as two transitional cell clusters (TRN and TRN2) with fewer than 12 positively regulated marker genes each (Supplementary Table S4). These clusters were functionally hard to interpret, making the SSP-derived outcome less accurate and harder to work with than those obtained with CaSSiDI-determined clustering.

Our comparative analysis against SSP establishes that CaSSiDI-based determination of optimal clustering conditions simplifies the detailed comparison of two scRNAseq datasets, and produces biologically more relevant outcomes. We repeated this CaSSiDI versus SSP comparison for splenic macrophages and peritoneal macrophages and found the same improvement in clustering using the former over the latter (see below and Supplemental Results).

CaSSiDI reveals two functional subsets of WT red pulp macrophages

Having established that CaSSiDI could reliably identify known subsets among BM-derived monocytes, we applied our approach to sorted splenic red pulp macrophages (RPMs) from WT mice. The primary function of RPMs is homeostatic scavenging of senescent erythrocytes and recycling of their heme and iron [28, 29], but RPMs may also control *P. falciparum* infections [28, 30]. It is unclear if these differing functions are executed by two distinct RPM subsets [31, 32]. We therefore isolated the Lin[−]F4/80⁺CD11b^{int}MHCII^{hi} RPM population from digested spleens of WT female littermate mice. CaSSiDI determined that the most suitable clustering condition in this case was $res = 0.4$ and $nPCs = 15$ (Supplementary Fig. S2A). As expected, the steady-state WT RPM population showed



contamination by other cell populations (Fig. 3A). Inspection of cluster-associated marker genes showed no over- or under-clustering, meaning that CaSSiDl yielded biologically meaningful clusters with a non-redundant distribution of marker genes. (Fig. 3B; Supplementary Table S5). A small macrophage cluster expressed

Lyve1, *Folr2*, *Fcrls*, *Pf4*, likely representing a yolk sac-derived TLF subset [33]. However, the pan-macrophage marker *Mertk* [34, 35] and the bona fide RPM marker *Spi-C* [36, 37] were confined to two enormous clusters (orange and blue) in the UMAP plot (Fig. 3A; Supplementary Table S5). Another cluster (gray) showed elevated

Fig. 3 Spleen: WT NPR vs. WT ILR clusters. A UMAP for clusters of WT splenic red pulp macrophages (RPMs) obtained using the indicated top-ranked CSS parameter values. **B** Prominent marker genes for the clusters (indicated by color) in **A**. **C** Ridge plots showing expression levels of the indicated DEGs between WT ILRs (blue) and NPRs (orange). **D** Top differentially expressed pathways identified by GSEA between WT ILRs (blue) and NPRs (orange). **E** UMAPs showing expression levels of the indicated genes in WT splenic RPMs. **F** Left: Flow cytometric gating strategy to identify ILRs and NPRs among WT RPMs. CD3⁺, CD19⁺ and NK1.1⁺ cells were excluded (please see Methods). ILRs (blue) and NPRs (orange) were distinguished by their CD68 expression. Right: Flow cytometric analysis to show differences in CD11b expression and cell morphology between WT ILRs and NPRs. **G** Bubble plots organized by the indicated gene groups illustrating DEGs between WT ILRs and NPRs. **H** Nebula plot showing prominent ILR-specific (blue, left), shared (gray, middle), and NPR-specific (orange, right) marker genes. Please see Methods for a full explanation of this plot type.

expression levels (as compared with other clusters) of only 16 genes, with 13 being associated with mitochondria (Fig. 3B). The remaining upregulated genes, namely *Lgmn*, *Siglece*, *Hebp1*, suggested that these cells were in a transitory state (TRN, Supplementary Table S5) [23].

Cluster and pathway analyses of the blue and orange clusters (ILR and NPR, respectively, abbreviations defined below) revealed two RPM subsets (Fig. 3A–E). Orange cluster cells expressed classical RPM genes like *Hmox1*, *Vcam1*, *Hebp1*. Because this cluster was linked to neuronal-associated pathways such as “Synaptic transmission glycinergic”, “Glial cell fate specification”, and “Noradrenergic neuron differentiation”, it was deemed to contain “Neuronal-primed RPMs” (NPRs). Blue cluster cells expressed a different set of classical RPM genes involved in either iron storage or antigen presentation, including *Csf1r*, *Ftl1*, *Ctsd*, *H2-Ab1*, *CD74* (Fig. 3A–E, G, Supplementary Fig. S2B). We termed the cells in this cluster “Iron-loaded RPMs” (ILRs). ILRs also showed greater expression of *Mertk*, *Spic*, *Fcgr3*, *Irf18* as well as erythrocyte-associated hemoglobin (*Hbb-bs*) and serotransferrin (*Trf*) (Fig. 3C, G), suggesting that ILRs contain more iron than NPRs. The higher iron load and IL-18 expression in ILRs may better equip these cells to eliminate bacteria. Accordingly, GSEA pathway analysis of the blue cluster highlighted the global gene expression programs “MHC class II protein complex”, “Sequestering of iron ion”, and “Positive regulation of antigen processing and presentation” (Fig. 3D).

Next, we performed a COMET analysis [38] of WT RPMs to determine the most suitable cell surface antigen combinations for differential isolation of NPRs vs. ILRs (Fig. 3E). Unexpectedly, only CD68 showed a sufficient difference in expression to facilitate separation of ILRs and NPRs by flow cytometry (Supplementary Fig. S2C). Separating Lin⁺F4/80⁺CD11b^{int} into CD68^{lo} and CD68^{hi} cells (NPRs and ILRs, respectively) revealed a positive correlation between CD68 and CD11b expression (Fig. 3F). Thus, NPRs express low levels of CD68 and CD11b, whereas ILRs express elevated levels of both of these surface markers. Strikingly, this correlation was also reflected in cell morphology, with ILRs exhibiting increased cell size (FSC-A) and granularity (SSC-A) compared to NPRs (Fig. 3F). These differences may reflect a maturation process whereby macrophages that can engage in phagocytosis (NPRs) become those that can no longer do so (ILRs) and instead rely on iron-based mechanisms to combat pathogens.

Among DEGs (Fig. 3G), ILRs showed increased *Ftl1*, *Trf*, *Slc48a1*, *Hfe*, *Sod1*, which are involved in heme/iron biology [29, 39]. In contrast, the transcription factors *Zeb2*, *Irf*, *Ikzf1* were more highly expressed in NPRs than ILRs. *Zeb2* safeguards tissue-specific macrophage states, while *Irf8* increases macroautophagy [40], and *Ikzf1* is associated with an anti-inflammatory macrophage phenotype [41, 42].

To account for background gene expression derived from clusters outside a population of interest, we custom-designed a visualization that we call the “Nebula plot” and applied it to a comparison of NPRs and ILRs (Fig. 3H). Genes in blue are the top 15 ILR-related genes, followed by the top 10 genes shared between ILRs and NPRs in black, and the top 15 NPR-related genes in orange. Close to 100% of ILRs and NPRs expressed *Slc40a1*

(ferroportin) [43] and *Adgre1* (F4/80). Other shared surface receptor genes were *Mertk* and *Pira2* (co-expressed with PirB) [34, 35]. Genes expressed primarily by ILRs included *Ftl1*, *Trf*, *Hbb-bs*. In NPRs, the most highly expressed genes were *Hmox1*, *Vcam1*, *GM42418*.

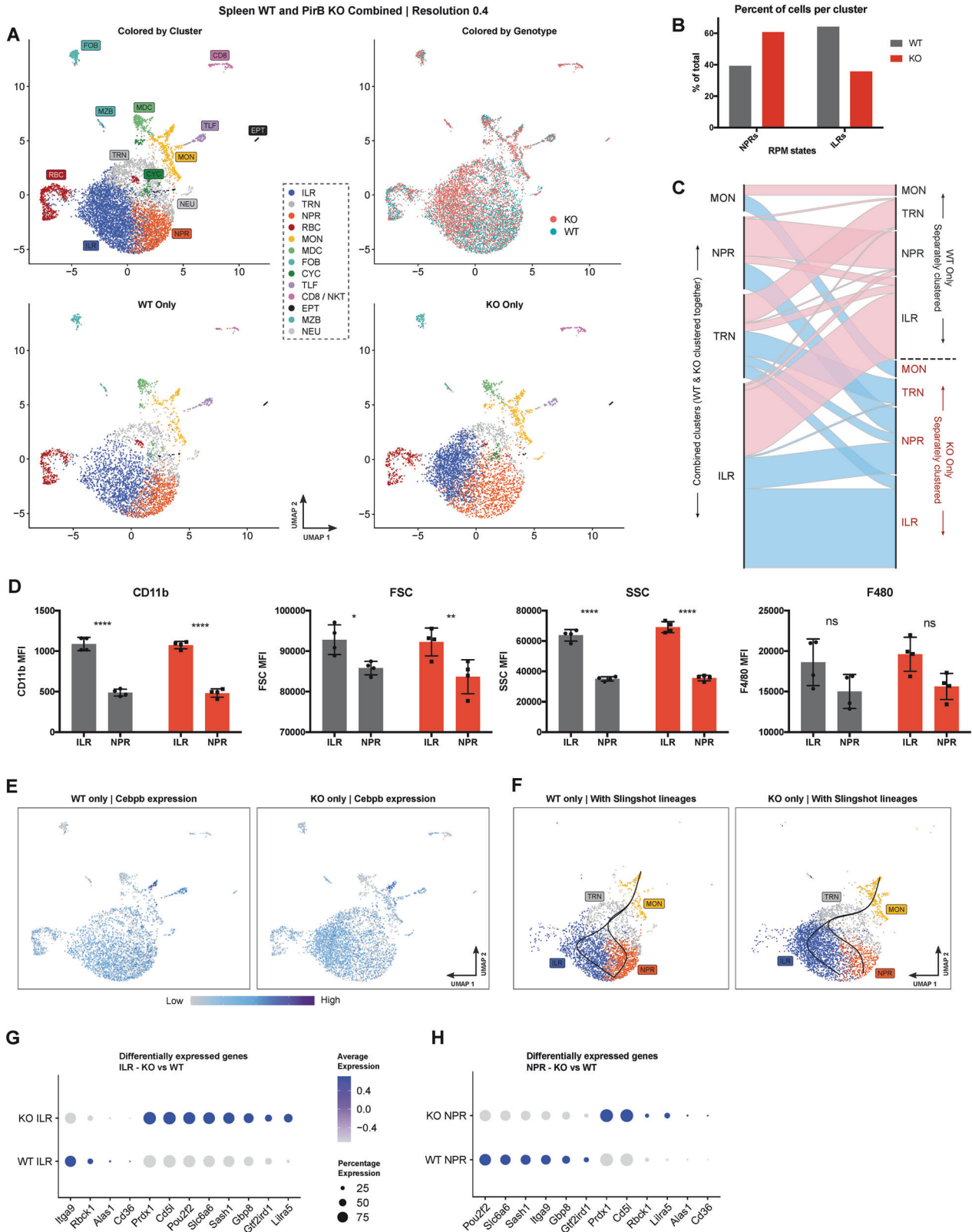
We sought to validate candidate genes identified by our scRNA-seq DEG analysis pipeline. To this end, we sorted WT splenic Lin⁺F4/80⁺CD11b^{int} macrophages into the CD68^{lo} and CD68^{hi} subsets, prepared cDNAs, and examined expression levels of selected candidate genes by qPCR (Supplementary Fig. S2D). As expected, there was no difference in *Slc40a1* transcript levels between ILRs and NPRs, confirming our original finding shown in Fig. 3H. ILRs showed increased *Ctsd* mRNA levels, whereas *Paqr9* and *Cd44* mRNAs were reduced in ILRs compared with NPRs (Supplementary Fig. S2D). *Hmox1* and *Vcam1* showed a trend towards decreased expression in ILRs compared to NPRs but these differences were not statistically significant.

Thus, our dissection of steady-state WT RPMs by CaSSiDI and the Nebula plot has revealed two subsets with substantial bona fide gene expression differences. NPRs display a neuronal-associated gene expression program and express high levels of *Hmox1* and *Hebp1* to recycle senescent erythrocytes. ILRs may be fully loaded with iron and therefore prepared to fight bacteria efficiently.

PirB deficiency alters gene expression patterns and subset proportions among steady-state RPMs

To dissect PirB’s functions in RPMs, we compared RPM samples from WT and PirB KO female littermate mice using a combined clustering approach and detected the same ILR and NPR clusters as revealed by CaSSiDI-based separate clustering (Fig. 4A, Supplementary Fig. S3A, B). Visual inspection of UMAP plots suggested an increase in NPRs at the expense of ILRs in the mutant. Indeed, enumeration of WT and PirB KO cells per cluster confirmed this trend, suggesting a role for PirB in RPM differentiation or subset-specific proliferation (Fig. 4B). We next performed an alluvial plot analysis confirming the close match between the separate and combined clustering approaches (Fig. 4C). The greatest difference was observed in the TRN cluster, where most cells were correctly assigned to the separate PirB KO TRN cluster (457 cells), but another 231 and 173 cells were allocated to the separate PirB KO ILR and NPR clusters, respectively (Supplementary Table S6). Flow cytometric differences in CD11b expression, cell size (FSC-A), and granularity (SSC-A) differed to the same degree between ILRs and NPRs of WT and PirB KO mice. F4/80 expression was comparable across WT and mutant ILRs and NPRs (Fig. 4D). Strikingly, unlike PirB KO monocytes, PirB KO RPMs did not show elevated *Cebpb* (Fig. 4E).

A Slingshot trajectory analysis confirmed that both WT and PirB KO steady-state monocytes become TRN RPMs that differentiate into NPRs or ILRs (Fig. 4F). When we investigated DEGs in WT vs. PirB KO ILRs, we found that mutant ILRs showed drastic decreases in *Itga9*, *Rbck1*, *Alas1*, *Cd36*, whereas *Pou2f2*, *Slc6a6*, *Sash1*, *Gbp8*, *Gtf2ird1*, *Lilra5* were elevated in these cells (Fig. 4G, Supplementary Fig. S3C). In comparing WT and PirB KO NPRs, we found that *Pou2f2*, *Slc6a6*, *Sash1*, *Itga9*, *Gbp8*, *Gtf2ird1*



transcripts were less abundant in PirB KO NPRs, but that *Rbck1*, *Lilra5*, *Alas1*, *CD36* mRNAs were all elevated (Fig. 4H, Supplementary Fig. S3D). *Prdx1* and *Cd51* (inflammatory mediators) were significantly increased in both the ILRs and NPRs of the mutant compared to the WT (Fig. 4G, H). GSVA comparing WT vs.

PirB KO NPRs and ILRs showed increased expression of pathways involving secretion and regulation of IL-10, IL-12, and IL-13 in both mutant subsets (Supplementary Fig. S3E). Thus, a lack of PirB impinges on global RPM and subset-specific gene expression patterns at steady-state.

Fig. 4 Spleen: WT NPR and ILR vs. PirB KO NPR and ILR clusters. A UMAPs showing combined clustering results of WT and PirB KO splenic RPMs, with cells grouped as indicated: by combined cluster identities, by genotype, and by identities obtained from clustering the WT and KO cells separately. **B** Quantitation of percentages of NPRs and ILRs among WT and PirB KO RPMs in **A**. **C** Alluvial diagram linking the combined clusters and the separate clusters, showing the flow of cells and the regroupings between the combined and separate clustering strategies. **D** Quantitation of the mean fluorescent intensities (MFI) of the indicated features discriminating between NPRs and ILRs isolated from spleens of WT and PirB KO mice (WT $n = 4$, KO $n = 4$) using the gating strategy shown in Fig. 3F. * $P < 0.05$, ** $P < 0.01$, **** $P < 0.0001$ as determined by regression analysis with two-way ANOVA followed by Sidak's post hoc multiple comparison test. Data are representative of two independent experiments. **E** UMAPs showing *Cebpb* mRNA levels in WT and PirB KO RPMs. **F** Slingshot trajectory curves originating from the MON cluster for the separately clustered WT and PirB KO RPM populations laid out on the combined UMAP space. **G, H** Bubble plots comparing prominent DEGs in WT vs. PirB KO ILRs and WT vs. PirB KO NPRs.

Lastly, as noted above, we applied SSP to the above splenic macrophage dataset and obtained inferior clustering outcomes (Supplementary Fig. S8A, B; Supplemental Results). This result bolsters the validity of our CaSSiDI method.

Validation of the CaSSiDI approach by analysis of Zeb2 in splenic myeloid cells

The TF *Zeb2* drives expression of essential RPM-related genes. To validate our CaSSiDI approach in a biological context, we chose a published study by Scott et al. that contained scRNAseq data obtained from steady-state WT and *Zeb2* KO RPMs [44]. The combined clustering analysis carried out therein focused on WT vs. *Zeb2* KO differences and did not pinpoint cluster-associated features or identify RPM subsets. We processed the Scott data through our CaSSiDI pipeline and compared the Scott results with data from our scRNAseq analysis of steady-state WT RPMs.

When we compared clusters of steady-state PirB WT and *Zeb2* WT samples, both were found to contain the ILR and NPR RPM subsets (Supplementary Fig. S4A). The TRN cluster was absent from both WT populations, but a myeloid cell cluster with an ISG signature (*Cxcl9*, *Stat1*, *Ifit2*, *Marco*, *Ifi47*, *Fcgr4*, *Cd274*, *Icam1*) was present in both samples (Supplementary Fig. S4A, Supplementary Tables S7 and S8). Importantly, NPRs and ILRs were readily identified in the independent biological replicates represented by the PirB WT and *Zeb2* WT samples, despite their differences in quality and quantity of contaminating non-macrophage cell types (Supplementary Fig. S4A, B). Thus, CaSSiDI operates in a manner analogous to data integration pipelines. Moreover, the top clustering condition chosen by CaSSiDI in Supplementary Fig. S4C ($res = 0.4$, $nPCs = 35$) differs from that in comparing PirB WT and PirB KO RPMs in Fig. 4 ($res = 0.4$, $nPCs = 15$). This outcome indicates that CaSSiDI is sensitive to the nature of the input data.

We then used CaSSiDI to compare steady-state *Zeb2* WT and *Zeb2* KO clusters. Once again, the ILR and NPR RPM subsets were identified, ISG cells appeared, and the TRN cluster was absent (Fig. 5A, B). Each WT RPM cluster was highly similar to the corresponding *Zeb2* KO cluster (Fig. 5B), but a previously unidentified cluster (ABR, red) appeared in *Zeb2* KO RPMs (Fig. 5A, B). In line with *Zeb2*'s control of classical DC development [45–47], these red cluster cells co-expressed genes typical of not only macrophages but also DCs and mast cells, in particular *Itgax* (CD11c), *Zbtb46*, *Ly6e*, *Sipi*, *Gm2a* (Fig. 5C, Supplementary Fig. S4D). Compared to PirB deficiency, loss of *Zeb2* had a greater skewing effect on the proportions of ILR vs. NPR cells among RPMs (Fig. 5D). These results confirmed the vital function of *Zeb2* in RPM subset differentiation and bolstered our confidence in our CaSSiDI method.

PirB deficiency alters gene expression patterns of peritoneal macrophage subsets

RPMs, alveolar macrophages (AMs), and peritoneal macrophages (PMs) are tissue-resident macrophages (TRMs) that maintain their numbers by slow and local self-renewal [33]. In contrast, skin- and dermis-associated TRMs are constantly replaced by circulating monocytes [48, 49]. The factors controlling local self-renewal versus monocyte-dependent replacement are not fully

understood but may include sex-dependent differences and elements of the tissue-specific microenvironment [50, 51].

There are two PM subsets: large PMs (LPMs) and small PMs (SPMs). SPMs are 10x less abundant than LPMs and express *F4/80^{lo}* (*Adgre1*), *Cd226*, *Retnla*. *Irf4* is indispensable for SPM generation [50, 52, 53]. In contrast, LPM differentiation is governed by *Gata6*, *Cebpb*, retinoic acid, and omentum-derived factors [54, 55]. It is mainly LPMs that conduct immunosurveillance of the peritoneal cavity and adjacent viscera [56, 57]. To determine if PirB contributes to PM heterogeneity and function, we compared scRNAseq datasets from steady-state WT and PirB KO PMs. CaSSiDI established that the optimal clustering condition was $res = 0.4$ and $nPCs = 15$ (Fig. 6A), which yielded 8 WT and 6 PirB KO total PM cell clusters (Fig. 6B, C). Among WT PMs, we identified three clusters (G1P, G2P, and MIT) that each contained small numbers of cells showing an overrepresentation of cell cycle-related genes. Only the G1P and G2P clusters appeared in the KO PM population (Fig. 6B, D) (Supplementary Tables S9 and S10). It is possible that an MIT cluster was not detected due to the low cell numbers in that cluster and/or an overlap in cell cycle genes with other clusters. Notably, PirB deficiency caused global derepression of *Cebpb* in PMs as it did in monocytes (Fig. 6E).

The WT PM population also contained a cluster (yellow) that likely represented an intermediate state (INT) and was lacking in the KO population (Fig. 6B–D). The INT cluster exhibited few upregulated genes but did include *Prtn3* (a neutrophil serine protease; NSP), which contributes to the control of LPM vs. SPM differentiation [54]. NSPs are repressed during the monocyte-macrophage transition [58]. *Prtn3* limits the self-renewal of hematopoietic progenitors [59] and may do the same for self-renewing LPMs. INT cells also expressed *Sipi*, an anti-apoptotic regulator that controls NSP activities and safeguards the granulocyte lineage [60–62]. Additional INT cluster-linked genes were *Ypel3* (p53 target) [63, 64]; *Rgs2* (negative regulator) [65]; and *Cirbp* (mediates hemorrhagic shock and sepsis) [66] (Fig. 6C, D; Supplementary Table S9). Combined clustering confirmed the absence of the INT cluster in PirB KO PMs (Fig. 6E). Notably, cells of the PirB KO SPM cluster (green), which expressed the SPM signature gene *Retnla* (Fig. 6D), spread into the region of the WT INT cluster (Fig. 6E). This observation suggests that the WT INT cluster is stabilized in a prospective SPM state; that is, it contains monocytes destined to undergo SPM differentiation. The absence of an equivalent INT cluster in the PirB KO PM population reinforces the reported effect of PirB deficiency on BM-derived monocytes, which rapidly adopt a macrophage-like state upon BM egress [67, 68]. Interestingly, our PirB KO SPMs expressed the WT INT gene *Sipi* (Fig. 6F, G, Supplementary Table S10).

Consistent with previous work [51], our CaSSiDI pipeline elucidated three subsets of WT mature functional LPMs: T2M (*Tlr2*-expressing macrophages), T4M (*Timd4*-expressing macrophages), and HSM (*Heat stable antigen* (*CD24*)-expressing macrophages). For details on the phenotypes of these subsets, please see Supplemental Results.

We next set out to test potential functional differences between WT and PirB KO PMs. Of note, *Cebpb* expression is known to be a prerequisite for LPM generation [55]. We adoptively transferred

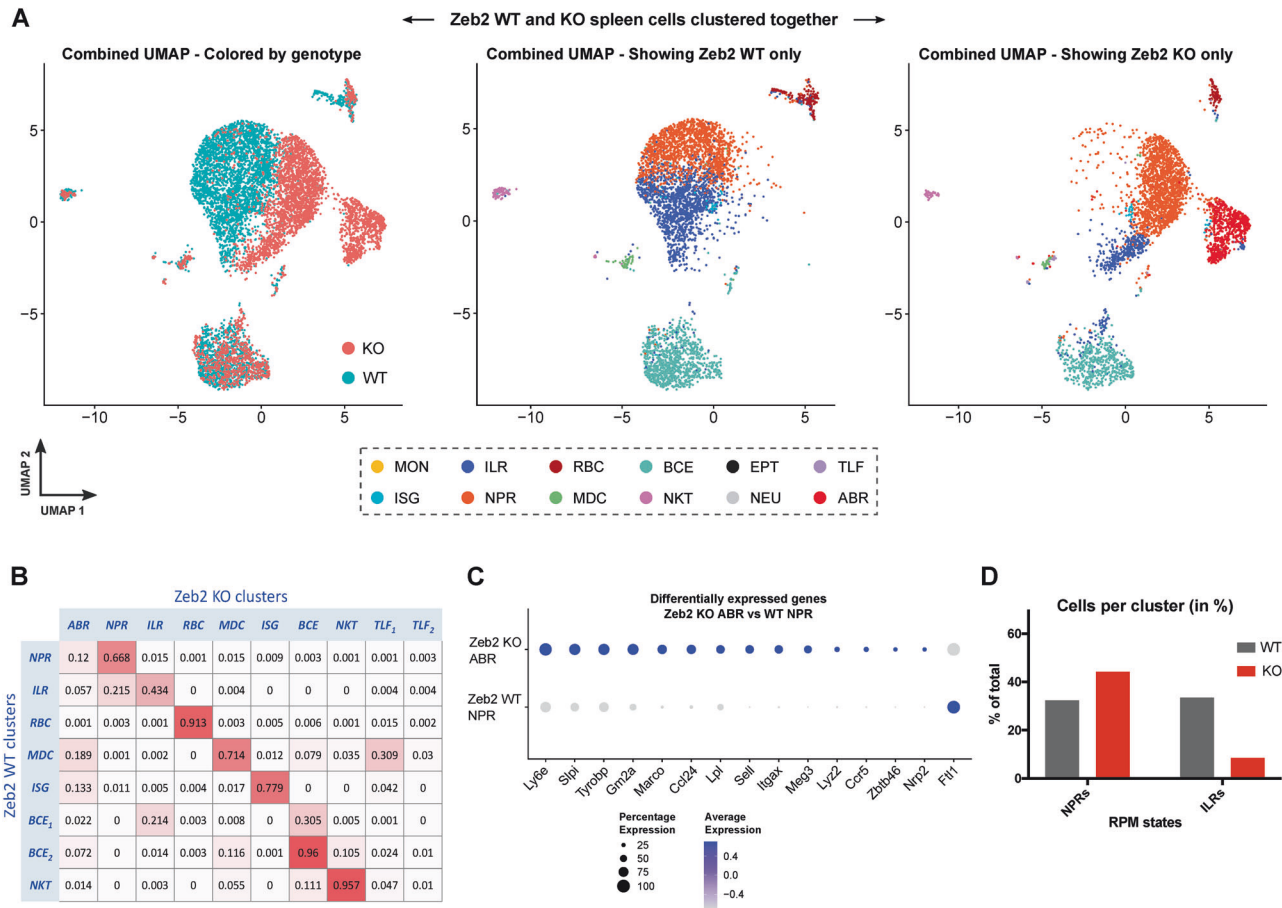


Fig. 5 Validation of CaSSiDl utilizing scRNA-seq data from Zeb2 KO RPMs. **A** UMAPs showing combined clustering results of WT and Zeb2 KO RPMs, with cells grouped as indicated: by genotype, and by identities obtained from clustering the WT and KO samples separately. **B** CSS table comparing clusters obtained using CaSSiDl on data from splenic WT and Zeb2 KO samples presented in ref. [44]. **C** Bubble plots illustrating prominent DEGs between the Zeb2 WT NPR and Zeb2 KO ABR populations of ref. [44]. **D** Quantitation of percentages of NPRs and ILRs among the Zeb2 WT and Zeb2 KO RPM populations of ref. [44].

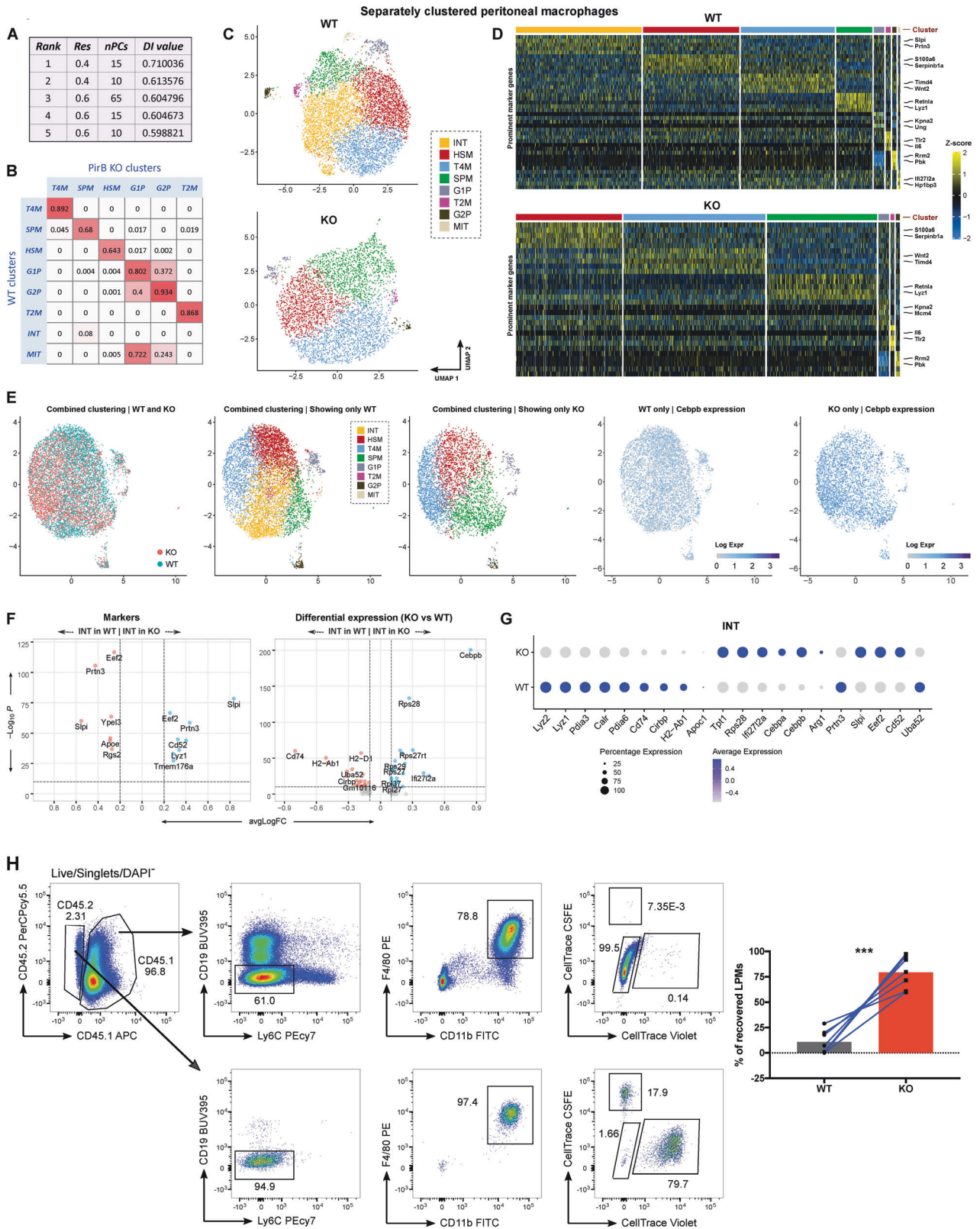
WT or PirB KO CD3⁺CD19⁺CD11c⁺Ly6C^{lo}CD11b^{hi}F4/80^{hi} LPMs (CD45.2⁺) into B6 recipient mice harboring the congenic marker CD45.1. Consistent with the elevated *Cebpb* mRNA levels observed in PirB KO PMs, we recovered greater than four-fold more CD45.2⁺, donor-derived PirB KO LPMs from recipients than WT LPMs at 21 days post-transplantation (Supplementary Fig. S5A). To overcome potential biases due to individual and/or host-dependent transplantation factors, we performed a competitive adoptive transfer experiment in which equal numbers of purified and dye-labeled CD45.2⁺ WT (CellTrace CFSE) and CD45.2⁺ PirB KO (CellTrace Violet) LPMs were co-transplanted into the same recipient mouse bearing unlabeled CD45.1⁺ WT LPMs. By 12 days after co-transplantation (the manufacturer-indicated guarantee of dye stability), significantly more CellTrace Violet⁺ PirB KO LPMs than CellTrace CFSE⁺ WT LPMs were recovered from recipients (Fig. 6H). These data indicate that PirB-deficient LPMs have a competitive edge over WT LPMs following adoptive transfer.

Taken together, these data identify PirB as a molecule that represses *Cebpb* expression to limit the differentiation or expansion of PMs in the peritoneal cavity. However, as was true for RPMs, loss of PirB has only mild effects on non-differentiating PM subsets at steady-state.

Finally, as noted above, we applied SSP to the above peritoneal macrophage dataset and obtained inferior clustering outcomes (Supplementary Fig. S8C, D; Supplemental Results). Once again, this result demonstrates the beneficial qualities of our CaSSiDl approach.

DISCUSSION

In this study, we have devised a novel single-cell cluster matching and optimization approach called CaSSiDl, and have used it to elucidate several cell type-associated differences in gene expression patterns between WT and PirB KO myeloid cell subsets. CaSSiDl proved to have several important advantages compared to a Standard Seurat Pipeline (SSP) in producing interpretable clustering outcomes. The objective of CaSSiDl is to find similar clusters in any two independently clustered scRNAseq datasets. To this end, CaSSiDl utilizes a marker gene-based quantitative score to carry out side-by-side comparisons of clusters across multiple resolutions and numbers of principal components in a grid-search fashion (See Methods for further details). Importantly, CaSSiDl considers only positive marker genes (genes overexpressed in a cluster compared to the background population) to determine cluster similarities. Negative markers (genes underexpressed in a cluster compared to the background population) were often found to be redundant across clusters and therefore unreliable for use in cluster identification. Moreover, CaSSiDl prevents the distortion of cell cluster allocations stemming from genetically divergent samples (e.g., WT vs. KO) that can occur when a classical combined clustering approach requiring batch correction is utilized. That being said, joint application of the separate and combined clustering approaches can help reveal biological relevance and correlations between two scRNAseq samples of interest. In this study, we have illustrated this utility by our analyses of murine RPMs, PMs, and TAMs (see Supplemental Results).



In addition to circumventing the need for batch correction, CaSSiDI's rank and score tables highlight cluster similarities in a graded fashion and thus also indicate potential cluster trajectories in a single glance. Notably, we used CaSSiDI to straightforwardly identify known monocyte and PM subsets in our WT samples. Based on this

success, we were able to use CaSSiDI to explore heterogeneities within TRM populations that are less well-described in the literature. Our data have thus expanded knowledge in this field.

CaSSiDI's enhanced ability to refine complex single-cell data analysis should be useful in several scenarios. First, as

Fig. 6 Peritoneum: WT vs. PirB KO PMs. **A** Top five parameter value pair choices as determined by CaSSiDI. The top choice of $res = 0.4$, $nPCs = 15$ was used to cluster the WT and PirB KO PM datasets. **B** CSS table corresponding to the top-ranking choice in **A**. **C** UMAPs of separately clustered WT and PirB KO PM populations. **D** Prominent marker genes for the indicated (by color) clusters. **E** UMAPs showing combined clustering results of WT and PirB KO PMs, with cells grouped as indicated: by genotype, and by identities obtained from clustering the WT and KO samples separately. The far-right plots are color-coded by *Cebpb* expression level. Volcano (**F**) and bubble (**G**) plots comparing markers and DEGs expressed by WT vs. PirB KO cells that occupied the INT cluster as identified using the combined clustering approach. **H** Left: Flow cytometric gating strategy for the identification of unlabeled CD45.1⁺ recipient LPMs or donor CD45.2⁺ LPMs, which were either labeled WT (CellTrace CFSE) or PirB KO (CellTrace Violet) LPMs. Right: Quantitation of WT and PirB KO LPMs recovered from CD45.1 recipients (WT $n = 7$, KO $n = 7$) at day 12 after co-transplantation. *** $P < 0.001$ as determined by two-tailed, paired t-test. Data are representative of two independent experiments.

demonstrated for the scRNAseq data in this study, CaSSiDI can be applied to optimize parameters for a combined clustering run by pitting the combined dataset (WT + KO) against the WT or KO samples individually. Depending on how much the WT and KO datasets differ, running CaSSiDI twice, once with WT and once with KO, may yield either the same best parameter value pair or two different choices. In the latter situation, the final decision on which of the two choices must be accepted for the combined clustering can be made by carefully considering the biological relevance and contexts associated with the clusters.

Second, CaSSiDI can effectively identify cell types in an uncharacterized dataset using a dataset for which the number and types of clusters have already been characterized. A CaSSiDI run with the characterized and uncharacterized datasets as inputs will systematically look for the best cluster matches at optimal resolution and nPCs and, in the process, characterize or label the clusters in the new dataset.

Third, CaSSiDI is not restricted to just RNAseq data or optimizing the resolution and nPCs parameters. It can be applied to data from other types of single-cell assays such as CYTOF and CITE-seq, and will work with any parameter pair that significantly influences clustering outcomes in a setting where one-to-one matches are expected to exist across cluster sets originating from a pair of datasets. The framework can also be easily expanded to include more than two parameters at a time for optimization, as long as computational resources are adequate. If, for example, we introduce a third parameter with 10 sampled values in addition to our usual set of res and $nPCs$, we would then have to run the module 1260 times, yielding as many CSS matrices. Since the runs are entirely independent, they can be done in parallel using a multicore computer cluster or a cloud-based computing service, both of which are ubiquitous and cheap to access at present. Once the clustering runs are complete and the CSS matrices are prepared, the DI computations are simple enough to execute on a standard laptop computer.

Fourth, depending on available computational resources, the number of samples from each parameter space can be increased to improve the power and accuracy of the search. For example, we have chosen our nPCs values to range from 10 to 75 in steps of 5, yielding 14 values. We can increase this to 33 values by reducing the step size to 3. Similarly, we can increase the resolution space by reducing the step size to 0.1 from 0.2. Consequently, many more parallel runs of the modules would be required. Nevertheless, since the runs are independent and can be executed in parallel, they should not consume any additional computational time if a sufficient number of parallel cores are available.

CaSSiDI does have a few limitations. First, CaSSiDI employs a grid-search style parameter optimization approach, which confines the method's output to the predefined grid points and ranges for each parameter search. If the actual optimal values fall outside the specified ranges or reside between the grid values, CaSSiDI may yield near-approximations instead. This hurdle is a fundamental challenge associated with grid searches in general, and can be mitigated to some extent by widening the parameter ranges and employing denser search spaces (provided sufficient computational resources are available). Second, while CaSSiDI can hint at possible cluster trajectories, it is not designed as a primary trajectory inference method and should not be employed as such.

It may overlook intricate trajectory structures and complex networks, which can only be unveiled through dedicated methods specifically designed for these purposes. Third, CaSSiDI is not an independent clustering method. Its performance relies on the quality of the initial clustering approach used to create clusters at different resolutions. Hence, selecting the most appropriate clustering method for the specific data type is paramount, as it forms the foundation upon which CaSSiDI operates effectively.

Turning to our biological results, perhaps this study's most critical conclusion is our unexpected finding that PirB is important for repressing *Cebpb* in some myeloid cell subsets at steady-state. *Cebpb* induces the differentiation of classical Ly6C^{hi} monocytes [26] and is also required for proper macrophage differentiation in normal BM, peritoneal cavity, and lung [26, 55, 69]. However, this engagement appears to be tissue-specific since, although PirB KO PMs showed elevated *Cebpb* mRNA levels and a precocious monocyte-to-macrophage transition, PirB KO RPMs did not differ from the WT in their *Cebpb* abundance.

Another factor to consider is that the *Cebpb* mRNA encodes two activating TFs called "Liver-enriched activator proteins" (LAPs) and one repressive TF called "Liver-enriched inhibitory protein" (LIP); the expression of these TFs is known to be differentially regulated [68, 70]. Furthermore, the translation of LAPs vs. LIP may be context- and cell type-dependent, and also cross-regulated by other cell signaling pathways [71]. Thus, the differential and antagonistic functions of LAPs and LIP, as well as their multifaceted pre- and post-translational control, makes the dissection of *Cebpb* regulation challenging. We found that loss of PirB upregulated some genes in certain macrophage subsets while downregulating others, suggesting that signaling through PirB controls gene expression in ways other than through *Cebpb* repression. Additional study is required to resolve precisely how PirB-mediated signaling affects *Cebpb*-derived TFs during the generation of various myeloid cell subsets. These findings also highlight the need to elucidate the natural PirB ligands in various tissues and contexts that are required to invoke *Cebpb* expression and/or repression.

Lastly, our analyses have uncovered a previously unknown RPM differentiation/maturation process that generates NPRs and ILRs. These subsets differ in cell size, morphology, and surface expression of CD11b and CD68. Our scRNA-based analyses suggest that NPRs are *Irf8*⁺ and represent a macrophage state that is 'ready to eat', whereas ILRs are 'done eating', generate IL-18, and bear an increased iron content that may allow them an alternative means of combatting pathogens. Future studies will be needed to determine the factors that govern PM maturation.

In conclusion, we believe that our novel CaSSiDI method makes a valuable contribution to the field of single-cell assay analysis, and that we have established its utility by demonstrating a role for PirB in the repression of *Cebpb* in certain subsets of myeloid lineage cells.

METHODS

Mice

PirB knockout mice (Pirb^{-/-} 129/B6 mice; [72]) were purchased from The Jackson Laboratory. Female littermates, age 11-13 weeks, were used for all

scRNAseq experiments. All animals were maintained in fully accredited facilities at the Princess Margaret Cancer Centre. No particular means for animal randomization or blinding investigators were implemented during experimentation.

Cell preparations, cell sorting, and flow cytometry

Cell preparations and flow cytometric analyses were performed essentially as described [73]. In brief, mice were humanely euthanized utilizing CO₂ displacement and cervical dislocation. Bone marrow-derived monocytes were recovered by flushing the bone marrow of femur plus tibia with ice-cold FACS buffer (PBS^{-/-} supplemented with 2% heat-inactivated FCS and 2 mM EDTA). To obtain splenocyte single-cell suspensions, spleens were crushed through a 70 µm filter using a syringe and suspended in 4 ml ice-cold FACS buffer. For peritoneal lavage, 5–7 ml ice-cold FACS buffer was injected into the peritoneal cavity followed by abdominal massage. Buffer containing peritoneal cells was aspirated by syringe equipped with a 20 G needle. All cell suspensions were subjected to red blood cell (RBC) lysis using a commercially available buffer (Sigma). Cells were pelleted by centrifugation (300 × g, 5 min, 4 °C), resuspended in ice-cold FACS buffer, and subjected to cell count determinations.

All single-cell suspensions were treated with Fc-blocking antibody (anti-CD16/32; 2.4G2; Tonbo) before staining with titrated antibody dilutions. The following antibodies were used for cell surface staining. Lin⁻ staining: anti-CD3 (145-2c11), anti-CD19 (1D3), anti-NK1.1 (PK136), anti-Ter119 (Ter119), anti-TCRgd (GL3). Target cell identification: anti-CD11b (M1/70), anti-CD11c (N418), anti-F4/80 (BM8), anti-CD115 (AFS98), anti-CD45.1 (A20), anti-CD45.2 (104), anti-CD68 (FA-11), anti-Ly6C (HK1.4), anti-Ly6G (1A8), anti-Cx3cr1 (SAO11F11), anti-MerTK (108928), anti-CD64 (X54-5/7.1). All antibodies were purchased from BD Biosciences, BioLegend, or Thermo Fisher Scientific. Dead cells were stained and excluded using 12.5 ng/ml DAPI (Sigma).

Adoptive transfer of peritoneal macrophages

Peritoneal cells were harvested by lavage of the peritoneal cavity with 7 ml cold PBS. Cells were washed using FACS buffer (PBS with 2% heat-inactivated FBS) and passed through a 70 µm filter before being stained with the appropriate antibodies as indicated in the Results section, the Figures, and the Figure legends. LPMs were sorted as indicated using a BD FACSAria Fusion cell sorter. WT and PirB KO LPMs were counted and labeled with CellTrace CFSE or CellTrace Violet, respectively, according to the manufacturer's instructions (Thermo Fisher Scientific). Equal numbers of purified and dye-labeled CD45.2⁺ WT and PirB KO LPMs were co-transplanted into the same recipient mouse bearing unlabeled CD45.1⁺ WT LPMs.

Single-cell RNA sequencing

Processing of murine cells for scRNAseq was performed essentially as previously described [33]. In brief, BM monocytes, splenic macrophages, peritoneal macrophages, or B16 tumor cells were isolated by flow cytometric cell sorting at low pressure. Single-cell suspensions were prepared according to the Chromium Single Cell 3' Reagent Kits User Guide (v2 Chemistry). Samples were loaded onto a 10x Chromium instrument to produce sequencing libraries, which were processed according to methods provided by 10x Genomics. Cells were sequenced and processed to generate expression matrices using Cell Ranger (10x Genomics). Raw base call (BCL) files from a HiSeq2500 sequencer were demultiplexed into FASTQ files, which were aligned (STAR) and filtered, followed by barcode and UMI counting to generate the counts table. The scRNAseq package Seurat v3.1.0 [74] was used for all downstream analyses using R 3.6.1. To take advantage of some improved plotting features, Seurat v3.2.3 in R v4.0.3 was also selectively used.

CaSSiDI: cluster similarity scoring methodology

The rationale and derivation of CaSSiDI are presented in detail in Supplemental Methods. In brief, one of the main challenges we faced in analyzing our scRNAseq data was in arriving at the most meaningful set of values for the computational parameters that wield a significant influence on biological inference: the clustering resolution (res) and the number of principal components (nPC) that are input for clustering. We also had the unique requirement of comparing and contrasting the PirB KO datasets with their corresponding WT counterparts. Therefore, it was necessary to find a set of parameter values that, on the one hand, revealed as many direct one-to-one cluster mappings as possible between WT and KO (thus identifying the same or similar cell types/states), while on the other hand

identified any stand-alone clusters that were unique to either the WT or KO population.

To achieve these objectives, we devised a multi-step strategy that: 1) sampled from the space defined by res and nPCs; 2) computed the clusters corresponding to each pair of parameter values chosen from the space; 3) evaluated clustering quality using a score; and 4) used the scores to determine the best pair of parameter values satisfying predefined criteria. The score characterizing clustering quality was designed such that it rewards parameter value pairs that identify clear one-to-one matches between WT and KO clusters along with any clusters unique to one genotype, but penalizes parameter value pairs that yield partial or unclear matches or too many clusters (over-clustering). In other words, we sought to find the parameter value pair that brought out the most striking one-to-one correspondences between WT and KO cluster sets while, at the same time, kept the number of clusters as small as possible to avoid over-clustering.

The first step in our strategy involves choosing the set of values for the res and the nPC parameters that will represent our sampled space of search. For all our analyses, we set nPC to range from 10 to 75 in steps of 5 and res to range from 0.4 to 2 in steps of 0.2. This translates into 14 nPC values and 9 res values, or a total of 126 value pairs. We ran Seurat's clustering module for each of these pairs for both the WT and KO datasets and identified markers for the clusters. Only markers satisfying an adjusted p-value threshold of 0.05 were retained. These markers were then used to compute a cluster similarity score (CSS) between each WT cluster and KO cluster using a weighted and enhanced similarity metric derived from the popular Jaccard similarity coefficient (<https://www.jstor.org/stable/2427226>). For two clusters with marker gene sets M₁ and M₂, a CSS can be defined using the original Jaccard similarity coefficient as

$$CSS_{basic} = \frac{C_{12}}{C_{12} + U_1 + U_2}$$

where C₁₂ is the number of common markers between clusters 1 and 2, U₁ is the number of markers unique to cluster 1, and U₂ is the number of markers unique to cluster 2. The score ranges between 0 and 1, with 1 representing identical clusters with all markers being shared and 0 representing the most dissimilar clusters with no common marker genes. While the CSS_{basic} formulation is an acceptable measure of similarity, it is too simple in this context as it is solely based on gene names and does not take into consideration other essential measurements such as the proportion of cells expressing each marker, the average expression strength, or how exclusively the gene serves as a marker for a cluster. All of these additional measurements are important indicators of the effectiveness of a marker gene for a cluster. We reasoned that incorporating these measures into the CSS formulation would enhance the score's sensitivity in defining similar and different clusters. The proportions of cells expressing a marker in a cluster (pct1) compared to the background (pct2), and the average relative expression level compared to the background (avgLogFC), are generated by Seurat in the marker lists and so can be extracted straightforwardly.

To quantify the degree to which a marker is unique to a particular cluster, we define a *marker uniqueness score* inspired by the work of Carson et al. [75] as $u_i = 1 - \sqrt{\frac{L_i}{L_n}}$ where u_i is the uniqueness score, L_i is the number of clusters associated with each marker gene, and L_n is the total number of clusters in the WT and KO sets combined. Upon incorporating these additional measurements into CSS_{basic}, we get an enhanced formulation CSS_{enh} as

$$CSS_{enh} = \frac{\sum_{i=1}^{C_{12}} w_{j1} \cdot u_i + \sum_{i=1}^{C_{12}} w_{i2} \cdot u_i}{\sum_{i=1}^{C_{12}} w_{i1} \cdot u_i + \sum_{i=1}^{C_{12}} w_{i2} \cdot u_i + \sum_{j=1}^{U_1} w_{j1} \cdot u_j + \sum_{k=1}^{U_2} w_{k2} \cdot u_k}$$

where C₁₂, U₁, and U₂ carry the same meanings as in CSS_{basic}. Variables u_i , u_j , and u_k are the uniqueness scores and w_{j1} , w_{i2} , w_{j1} , and w_{k2} are the weights for genes i , j , and k as indicated in the first subscript and corresponding to the clusters indicated in the second subscript. The weight w_{j1} for gene i in cluster 1 is defined as $avgLogFC_{i1} \left(\frac{pctF_{i1}^2}{pctB_{i1}} \right)$, where $avgLogFC_{i1}$ is the average log fold change, $pctF_{i1}$ is the proportion of cells expressing gene i in cluster 1, and $pctB_{i1}$ is the proportion of cells expressing gene i in the appropriate background population for cluster 1. All other weights, w_{i2} , w_{j1} , and w_{k2} are defined in the same manner.

Before arriving at the above final formulation, six other variations of this score, including the introduction of only one measurement at a time or combinations thereof, were tested and the results analyzed. In the end, the

above formulation, which simultaneously incorporates the uniqueness, strength of expression, and extent of expression (i.e., the percentage of cells that express the gene), proved to be most effective in improving the sensitivity of the score in defining similar and different clusters. A gene that is close to being uniquely expressed, is more strongly expressed, and is more widely expressed is awarded a higher weight than a gene that is shared, weakly expressed, and narrowly expressed. Just like CSS_{basic} , CSS_{enh} values range between 0 and 1.

For a given pair of res and nPCs values, if there are m WT clusters and n KO clusters, there would be a total of $m \times n$ CSS_{enh} values. These can be assembled into a structure we call the CSS matrix. The CSS matrix provides a snapshot of the one-to-one relationships between the WT and KO clusters. It lists the pairwise similarity scores, comparing each WT cluster with every KO cluster to quantify how similar or different they are. For example, for our spleen data, the WT-KO dataset pair generated a total of 126 CSS matrices, one for every combination of res and nPCs.

The goal of the second major step of our strategy is to compute a single metric representing each CSS matrix that aims to quantify the overall clustering quality of the res and nPCs parameter value pair associated with the CSS matrix. To achieve this goal, we first compute, for each row and column, a *distinction index* (DI) equal to the difference between the maximum element and the mean of the remaining elements. This yields a total of m row-level and n column-level distinction indices. We express the mean value of this set of $m \times n$ indices as $\frac{1}{mn} \left[\sum_{i=1}^m d_{row/i} + \sum_{j=1}^n d_{col/j} \right]$, which we refer to as the distinction index for the entire CSS matrix (CaSSiDI). Thus, CaSSiDI represents a quantification of the clustering quality associated with a given res/nPCs parameter pair choice. Clustering quality can have many different meanings depending on the context. Here, we interpret the DI of a CSS matrix and, by extension, the corresponding res/nPCs parameter pair choice, as a metric that quantifies how well the parameter pair choice has been able to identify matching WT and KO clusters while also identifying and segregating any clusters in either genotype that do not have a good match in the other genotype. For our spleen data, the set of 126 CaSSiDI values for a pair of WT-KO datasets can then be directly used to rank the parameter pair choices from best to worst, with the highest CaSSiDI value corresponding to the best parameter pair choice.

Nebula plot

In Fig. 3H, for each gene, four violin plots are shown. The inner two violins correspond to the ILR (left) and NPR (right) subsets, while the outer two violins correspond to their respective background populations used by Seurat for marker identification. Black diamonds represent the average expression level used by Seurat. Genes are ordered as follows: from left to right, those labeled in blue are in decreasing order of inner blue violin diamond level, while those labeled in orange are in increasing order of inner orange violin diamond level. Those labeled in gray are ordered based on the mean of the inner blue and orange violin diamond levels, with highest mean at the center and falling off on either side. The bar plots at the top show the proportion of cells in the cluster (ILR or NPR) that expresses the given gene.

Thus, through the Nebula plot, it is possible to get an overview of the most important genes characterizing a given cluster in terms of differential expression compared to background. Moreover, it is possible to compare two clusters at a time and get an idea of how widely each gene is expressed in a population through the bar plots.

DATA AVAILABILITY

The 10X single-cell sequencing raw data files and processed matrices associated with this study have been deposited in the Gene Expression Omnibus database under the accession code [GSE252466](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE252466).

CODE AVAILABILITY

A software implementation of CaSSiDI along with critical scripts used to analyze the scRNAseq data presented in the paper can be found at <https://github.com/CodelnTheSkies/CaSSiDI-PirB-Macrophage-Paper>.

REFERENCES

1. Takai T. Paired immunoglobulin-like receptors and their MHC class I recognition. *Immunology*. 2005;115:433–40.

- Takeda K, Nakamura A. Regulation of immune and neural function via leukocyte Ig-like receptors. *J Biochem*. 2017;162:73–80.
- van der Touw W, Chen HM, Pan PY, Chen SH. LILRB receptor-mediated regulation of myeloid cell maturation and function. *Cancer Immunol Immunother*. 2017;66:1079–87.
- Atwal JK, Pinkston-Gosse J, Syken J, Stawicki S, Wu Y, Shatz C, et al. PirB is a functional receptor for myelin inhibitors of axonal regeneration. *Science*. 2008;322:967–70.
- Nakamura A, Kobayashi E, Takai T. Exacerbated graft-versus-host disease in PirB $-/-$ mice. *Nat Immunol*. 2004;5:623–9.
- Li D, Wang L, Yu L, Freundt EC, Jin B, Screaton GR, et al. Ig-like transcript 4 inhibits lipid antigen presentation through direct CD1d interaction. *J Immunol*. 2009;182:1033–40.
- Kang X, Kim J, Deng M, John S, Chen H, Wu G, et al. Inhibitory leukocyte immunoglobulin-like receptors: Immune checkpoint proteins and tumor sustaining factors. *Cell Cycle*. 2016;15:25–40.
- Ayukawa S, Kamoshita N, Nakayama J, Teramoto R, Pisheshia N, Ohba K, et al. Epithelial cells remove precancerous cells by cell competition via MHC class I-LILRB3 interaction. *Nat Immunol*. 2021;22:1391–402.
- Li X, Wang K, Lyu Y, Pan H, Zhang J, Stambolian D, et al. Deep learning enables accurate clustering with batch effect removal in single-cell RNA-seq analysis. *Nat Commun*. 2020;11:2338.
- Pasquini G, Rojo Arias JE, Schafer P, Busskamp V. Automated methods for cell type annotation on scRNA-seq data. *Comput Struct Biotechnol J*. 2021;19:961–9.
- Bej S, Galow AM, David R, Wolfen M, Wolkenhauer O. Automated annotation of rare-cell types from single-cell RNA-sequencing data through synthetic oversampling. *BMC Bioinforma*. 2021;22:557.
- Menden K, Marouf M, Oller S, Dalmia A, Magruder DS, Kloiber K, et al. Deep learning-based cell composition analysis from tissue expression profiles. *Sci Adv*. 2020;6:eaba2619.
- Bassler K, Schulte-Schrepping J, Warnat-Herresthal S, Aschenbrenner AC, Schultze JL. The Myeloid Cell Compartment-Cell by Cell. *Annu Rev Immunol*. 2019;37:269–93.
- Chong SZ, Evrard M, Devi S, Chen J, Lim JY, See P, et al. CXCR4 identifies transitional bone marrow premonocytes that replenish the mature monocyte pool for peripheral responses. *J Exp Med*. 2016;213:2293–314.
- Shi C, Pamer EG. Monocyte recruitment during infection and inflammation. *Nat Rev Immunol*. 2011;11:762–74.
- Ginhoux F, Jung S. Monocytes and macrophages: developmental pathways and tissue homeostasis. *Nat Rev Immunol*. 2014;14:392–404.
- Winkler CW, Woods TA, Robertson SJ, McNally KL, Carmody AB, Best SM, et al. Cutting Edge: CCR2 Is Not Required for Ly6C(hi) Monocyte Egress from the Bone Marrow but Is Necessary for Migration within the Brain in La Crosse Virus Encephalitis. *J Immunol*. 2018;200:471–6.
- Das M, Lu J, Joseph M, Aggarwal R, Kanji S, McMichael BK, et al. Kruppel-like factor 2 (KLF2) regulates monocyte differentiation and functions in mBSA and IL-1beta-induced arthritis. *Curr Mol Med*. 2012;12:113–25.
- Cheong C, Matos I, Choi JH, Dandamudi DB, Shrestha E, Longhi MP, et al. Microbial stimulation fully differentiates monocytes to DC-SIGN/CD209(+) dendritic cells for immune T cell areas. *Cell*. 2010;143:416–29.
- Tussiwand R, Lee WL, Murphy TL, Mashayekhi M, Kc W, Albring JC, et al. Compensatory dendritic cell development mediated by BATF-IRF interactions. *Nature*. 2012;490:502–7.
- Picotto G, Morse LR, Nguyen N, Saltzman J, Battaglini R. TMEM176A and TMEM176B Are Candidate Regulators of Inhibition of Dendritic Cell Maturation and Function after Chronic Spinal Cord Injury. *J Neurotrauma*. 2020;37:528–33.
- Grahn THM, Niroula A, Vegvari A, Oburoglu L, Pertesi M, Warsi S, et al. S100A6 is a critical regulator of hematopoietic stem cells. *Leukemia*. 2020;34:3323–37.
- Sun J, Ramos A, Chapman B, Johnnidis JB, Le L, Ho YJ, et al. Clonal dynamics of native haematopoiesis. *Nature*. 2014;514:322–7.
- Santana-Codina N, Carretero R, Sanz-Pamplona R, Cabrera T, Guney E, Oliva B, et al. A transcriptome-proteome integrated network identifies endoplasmic reticulum thiol oxidoreductase (ERp57) as a hub that mediates bone metastasis. *Mol Cell Proteom*. 2013;12:2111–25.
- Jaitin DA, Weiner A, Yofe I, Lara-Astiaso D, Keren-Shaul H, David E, et al. Dissecting Immune Circuits by Linking CRISPR-Pooled Screens with Single-Cell RNA-Seq. *Cell*. 2016;167:1883–96.e1815.
- Mildner A, Schonheit J, Giladi A, David E, Lara-Astiaso D, Lorenzo-Vivas E, et al. Genomic Characterization of Murine Monocytes Reveals C/EBPbeta Transcription Factor Dependence of Ly6C(-) Cells. *Immunity*. 2017;46:849–62.e847.
- Satija R. Quickly Pick Relevant Dimensions. Available from: <https://satijalab.org/seurat/reference/elbowplot> (2019).
- Borges da Silva H, Fonseca R, Pereira RM, Cassado Ados A, Alvarez JM, D'Imperio Lima MR. Splenic Macrophage Subsets and Their Function during Blood-Borne Infections. *Front Immunol*. 2015;6:480.

29. A-Gonzalez N, Castrillo A. Origin and specialization of splenic macrophages. *Cell Immunol.* 2018;330:151–8.
30. Lai SM, Sheng J, Gupta P, Renia L, Duan K, Zolezzi F, et al. Organ-Specific Fate, Recruitment, and Refilling Dynamics of Tissue-Resident Macrophages during Blood-Stage Malaria. *Cell Rep.* 2018;25:3099–109.e3093.
31. A-Gonzalez N, Quintana JA, Garcia-Silva S, Mazariegos M, Gonzalez de la Aleja A, Nicolas-Avila JA, et al. Phagocytosis imprints heterogeneity in tissue-resident macrophages. *J Exp Med.* 2017;214:1281–96.
32. Sukhbaatar N, Weichhart T. Iron Regulation: Macrophages in Control. *Pharmaceuticals.* 2018;11:137.
33. Dick SA, Wong A, Hamidzada H, Nejat S, Nechanitzky R, Vohra S, et al. Three tissue resident macrophage subsets coexist across organs with conserved origins and life cycles. *Sci Immunol.* 2022;7:eabf7777.
34. Gautier EL, Shay T, Miller J, Greter M, Jakubczik C, Ivanov S, et al. Gene-expression profiles and transcriptional regulatory pathways that underlie the identity and diversity of mouse tissue macrophages. *Nat Immunol.* 2012;13:1118–28.
35. Lavin Y, Winter D, Blecher-Gonen R, David E, Keren-Shaul H, Merad M, et al. Tissue-resident macrophage enhancer landscapes are shaped by the local microenvironment. *Cell.* 2014;159:1312–26.
36. Kohyama M, Ise W, Edelson BT, Wilker PR, Hildner K, Mejia C, et al. Role for Spi-C in the development of red pulp macrophages and splenic iron homeostasis. *Nature.* 2009;457:318–21.
37. Haldar M, Kohyama M, So AY, Kc W, Wu X, Briseno CG, et al. Heme-mediated SPI-C induction promotes monocyte differentiation into iron-recycling macrophages. *Cell.* 2014;156:1223–34.
38. Delaney C, Schnell A, Cammarata LV, Yao-Smith A, Regev A, Kuchroo VK, et al. Combinatorial prediction of marker panels from single-cell transcriptomic data. *Mol Syst Biol.* 2019;15:e9005.
39. Danzeisen R, Achsel T, Bederke U, Cozzolino M, Crosio C, Ferri A, et al. Superoxide dismutase 1 modulates expression of transferrin receptor. *J Biol Inorg Chem.* 2006;11:489–98.
40. Agod Z, Pazmandi K, Bencze D, Vereb G, Biro T, Szabo A, et al. Signaling Lymphocyte Activation Molecule Family 5 Enhances Autophagy and Fine-Tunes Cytokine Response in Monocyte-Derived Dendritic Cells via Stabilization of Interferon Regulatory Factor 8. *Front Immunol.* 2018;9:62.
41. Mougialakos D, Bach C, Bottcher M, Beier F, Rohner L, Stoll A, et al. The IKZF1-IRF4/IRF5 Axis Controls Polarization of Myeloma-Associated Macrophages. *Cancer Immunol Res.* 2021;9:265–78.
42. Oh KS, Gottschalk RA, Lounsbury NW, Sun J, Dorrington MG, Baek S, et al. Dual Roles for Ikaros in Regulation of Macrophage Chromatin State and Inflammatory Gene Expression. *J Immunol.* 2018;201:757–71.
43. Donovan A, Lima CA, Pinkus JL, Pinkus GS, Zon LI, Robine S, et al. The iron exporter ferroportin/Slc40a1 is essential for iron homeostasis. *Cell Metab.* 2005;1:191–200.
44. Scott CL, T'Jonck W, Martens L, Todorov H, Sichien D, Soen B, et al. The Transcription Factor ZEB2 Is Required to Maintain the Tissue-Specific Identities of Macrophages. *Immunity.* 2018;49:312–325.e315.
45. Bagadia P, Huang X, Liu TT, Durai V, Grajales-Reyes GE, Nitschke M, et al. An Nfil3-Zeb2-Id2 pathway imposes Irf8 enhancer switching during cDC1 development. *Nat Immunol.* 2019;20:1174–85.
46. Huang X, Ferris ST, Kim S, Choudhary MNK, Belk JA, Fan C, et al. Differential usage of transcriptional repressor Zeb2 enhancers distinguishes adult and embryonic hematopoiesis. *Immunity.* 2021;54:1417–32.e1417.
47. Wu X, Briseno CG, Grajales-Reyes GE, Haldar M, Iwata A, Kretzer NM, et al. Transcription factor Zeb2 regulates commitment to plasmacytoid dendritic cell and monocyte fate. *Proc Natl Acad Sci USA.* 2016;113:14775–80.
48. Liu Z, Gu Y, Chakarov S, Bleriot C, Kwok I, Chen X, et al. Fate Mapping via Ms4a3-Expression History Traces Monocyte-Derived Cells. *Cell.* 2019;178:1509–25.e1519.
49. Bain CC, Bravo-Blas A, Scott CL, Perdiguero EG, Geissmann F, Henri S, et al. Constant replenishment from circulating monocytes maintains the macrophage pool in the intestine of adult mice. *Nat Immunol.* 2014;15:929–37.
50. Bain CC, Hawley CA, Garner H, Scott CL, Schridde A, Steers NJ, et al. Long-lived self-renewing bone marrow-derived macrophages displace embryo-derived cells to inhabit adult serous cavities. *Nat Commun.* 2016;7:ncomms11852.
51. Bain CC, Gibson DA, Steers NJ, Boufeia K, Louwe PA, Doherty C, et al. Rate of replenishment and microenvironment contribute to the sexually dimorphic phenotype and function of peritoneal macrophages. *Sci Immunol.* 2020;5:eabc4466.
52. Kim KW, Williams JW, Wang YT, Ivanov S, Gilfillan S, Colonna M, et al. MHC II+ resident peritoneal and pleural macrophages rely on IRF4 for development from circulating monocytes. *J Exp Med.* 2016;213:1951–9.
53. Accarias S, Genthon C, Rengel D, Boullier S, Foucras G, Tabouret G. Single-cell analysis reveals new subset markers of murine peritoneal macrophages and highlights macrophage dynamics upon *Staphylococcus aureus* peritonitis. *Innate Immun.* 2016;22:382–92.
54. Okabe Y, Medzhitov R. Tissue-specific signals control reversible program of localization and functional polarization of macrophages. *Cell.* 2014;157:832–44.
55. Cain DW, O'Koren EG, Kan MJ, Womble M, Sempowski GD, Hopper K, et al. Identification of a tissue-specific, C/EBPbeta-dependent pathway of differentiation for murine peritoneal macrophages. *J Immunol.* 2013;191:4665–75.
56. Bain CC, Jenkins SJ. The biology of serous cavity macrophages. *Cell Immunol.* 2018;330:126–35.
57. Ansel KM, Harris RB, Cyster JG. CXCL13 is required for B1 cell homing, natural antibody production, and body cavity immunity. *Immunity.* 2002;16:67–76.
58. Cheung P, Schaffert S, Chang SE, Dvorak M, Donato M, Macaubas C, et al. Repression of CTSG, ELANE and PRTN3-mediated histone H3 proteolytic cleavage promotes monocyte-to-macrophage differentiation. *Nat Immunol.* 2021;22:711–22.
59. Karatepe K, Zhu H, Zhang X, Guo R, Kambara H, Loison F, et al. Proteinase 3 Limits the Number of Hematopoietic Stem and Progenitor Cells in Murine Bone Marrow. *Stem Cell Rep.* 2018;11:1092–105.
60. Klimenkova O, Ellerbeck W, Klimiankou M, Unalan M, Kandabarau S, Gigina A, et al. A lack of secretory leukocyte protease inhibitor (SLPI) causes defects in granulocytic differentiation. *Blood.* 2014;123:1239–49.
61. McGarry N, Greene CM, McElvaney NG, Weldon S, Taggart CC. The Ability of Secretory Leukocyte Protease Inhibitor to Inhibit Apoptosis in Monocytes Is Independent of Its Antiprotease Activity. *J Immunol Res.* 2015;2015:507315.
62. McNeely TB, Dealy M, Dripps DJ, Orenstein JM, Eisenberg SP, Wahl SM. Secretory leukocyte protease inhibitor: a human saliva protein exhibiting anti-human immunodeficiency virus 1 activity in vitro. *J Clin Invest.* 1995;96:456–64.
63. Baker SJ. Small unstable apoptotic protein, an apoptosis-associated protein, suppresses proliferation of myeloid cells. *Cancer Res.* 2003;63:705–12.
64. Kelley KD, Miller KR, Todd A, Kelley AR, Tuttle R, Berberich SJ. YPEL3, a p53-regulated gene that induces cellular senescence. *Cancer Res.* 2010;70:3566–75.
65. Lee HK, Park DW, Bae JH, Kim HJ, Shin DG, Park JS, et al. RGS2 is a negative regulator of STAT3-mediated Nox1 expression. *Cell Signal.* 2012;24:803–9.
66. Qiang X, Yang WL, Wu R, Zhou M, Jacob A, Dong W, et al. Cold-inducible RNA-binding protein (CIRP) triggers inflammatory responses in hemorrhagic shock and sepsis. *Nat Med.* 2013;19:1489–95.
67. Ma G, Pan PY, Eisenstein S, Divino CM, Lowell CA, Takai T, et al. Paired immunoglobulin-like receptor-B regulates the suppressive function and fate of myeloid-derived suppressor cells. *Immunity.* 2011;34:385–95.
68. Marigo I, Bosio E, Solito S, Mesa C, Fernandez A, Dolcetti L, et al. Tumor-induced tolerance and immune suppression depend on the C/EBPbeta transcription factor. *Immunity.* 2010;32:790–802.
69. Liu SS, Lv XX, Liu C, Qi J, Li YX, Wei XP, et al. Targeting Degradation of the Transcription Factor C/EBPbeta Reduces Lung Fibrosis by Restoring Activity of the Ubiquitin-Editing Enzyme A20 in Macrophages. *Immunity.* 2019;51:522–34.e527.
70. Begay V, Baumeier C, Zimmermann K, Heuser A, Leutz A. The C/EBPbeta LIP isoform rescues loss of C/EBPbeta function in the mouse. *Sci Rep.* 2018;8:8417.
71. Spike AJ, Rosen JM. C/EBP β Isoform Specific Gene Regulation: It's a Lot more Complicated than you Think! *J Mammary Gland Biol Neoplasia.* 2020;25:1–12.
72. Ujiike A, Takeda K, Nakamura A, Ebihara S, Akiyama K, Takai T. Impaired dendritic cell maturation and increased T(H)2 responses in PIR-B(–/–) mice. *Nat Immunol.* 2002;3:542–8.
73. Nechanitzky R, Akbas D, Scherer S, Gyory I, Hoyler T, Ramamoorthy S, et al. Transcription factor EBF1 is essential for the maintenance of B cell identity and prevention of alternative fates in committed cells. *Nat Immunol.* 2013;14:867–75.
74. Butler A, Hoffman P, Smibert P, Papalexis E, Satija R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol.* 2018;36:411–20.
75. Carson MB, Liu C, Lu Y, Jia C, Lu H. A disease similarity matrix based on the uniqueness of shared genes. *BMC Med Genomics.* 2017;10:26.

ACKNOWLEDGEMENTS

We thank the Princess Margaret Genomics Centre and the UHN Bioinformatics and HPC Core for RNA sequencing; Shawn P. Kubli for the B16 subcutaneous transplantation protocol; and Ryan Harrietha, and Kyle T. Gill for support with animal breeding and genotyping. We are also grateful to the Genotyping Facility and the Animal Resource Centre at the Princess Margaret Cancer Centre (Toronto). Lastly, we thank Drs. Tracy McGaha, Jeffrey Pollard and Florent Petitprez for careful reading of the manuscript.

AUTHOR CONTRIBUTIONS

R.N. and T.W.M. conceived the biological study and obtained funding. R.N. conceived CaSSiDI with input from P.R. P.R. conceived, developed, and designed the mathematical framework for CaSSiDI with input from R.N. P.R. performed all bioinformatic analyses including the implementation of CaSSiDI. R.N. and D.N. designed experiments and analyzed data. R.N. and D.N. performed experiments with support from W.L., A.C.W., and J.H. R.N. and P.R. wrote the original draft of the

manuscript with input from D.N. and S.E. M.E.S. provided scientific editing advice. T.W.M supervised the research. All authors have reviewed and approved the manuscript in its current form.

FUNDING

This work was supported by a grant to TWM from the Canadian Institutes of Health Research (CIHR) and a postdoctoral fellowship to RN from EMBO and Marie Curie (ALTF 725-2015).

COMPETING INTERESTS

The authors declare that they have no competing financial interests concerning the work described in this manuscript. TWM owns equity in Treadwell Therapeutics and Agios Pharmaceuticals, and is a consultant for AstraZeneca and Tessa Therapeutics.

ETHICS APPROVAL

The Ontario Cancer Institute Animal Care Committee approved all animal research and mouse experiments used in this study.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41418-024-01268-8>.

Correspondence and requests for materials should be addressed to Robert Nechanitzky or Tak W. Mak.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024