

Supplemental Information

Haemopedia: An Expression Atlas of Murine Hematopoietic Cells

Carolyn A. de Graaf, Jarny Choi, Tracey M. Baldwin, Jessica E. Bolden, Kirsten A. Fairfax, Aaron J. Robinson, Christine Biben, Clare Morgan, Kerry Ramsay, Ashley P. Ng, Maria Kauppi, Elizabeth A. Kruse, Tobias J. Sargeant, Nick Seidenman, Angela D'Amico, Marthe C. D'Ombrian, Erin C. Lucas, Sandra Koernig, Adriana Baz Morelli, Michael J. Wilson, Steven K. Dower, Brenda Williams, Shen Y. Heazlewood, Yifang Hu, Susan K. Nilsson, Li Wu, Gordon K. Smyth, Warren S. Alexander, and Douglas J. Hilton

Supplemental Information

Haemopedia: An expression atlas of murine haematopoietic cells

Carolyn A. de Graaf, Jarny Choi, Tracey M. Baldwin, Jessica E. Bolden, Kirsten A. Fairfax, Aaron J. Robinson, Christine Biben, Clare Morgan, Kerry Ramsay, Ashley P. Ng, Maria Kauppi, Elizabeth A. Kruse, Tobias J. Sargeant, Nick Seidenman, Angela D'Amico, Marthe C. D'Ombrain, Erin C. Lucas, Sandra Koernig, Adriana Baz Morelli, Michael J. Wilson, Steven K. Dower, Brenda Williams, Shen Y. Heazlewood, Yifang Hu, Susan K. Nilsson, Li Wu, Gordon K. Smyth, Warren S. Alexander, Douglas J. Hilton.

Supplemental Figures

Figure S1. Related to Figure 1: Features of the cell types included in the atlas.

(A) Sorting strategy plots for progenitor populations included in the atlas (i) Long term haematopoietic stem cells (LTHSC), short term haematopoietic stem cells (STHSC), multi potential progenitors (MPP), megakaryocyte erythrocyte progenitors (MEP), common myeloid progenitors (CMP), granulocyte macrophage progenitors (GMP) and eosinophil progenitors (EoP) (ii) FcγRII/III high CD150+ bipotential progenitors (FcγRBP), CD9 high bipotential progenitors (CD9Hi), bipotential erythroid megakaryocyte progenitors (BEMP) and PreColony forming unit – erythroid (PreCFUE) (iii) Common Lymphoid Progenitors (CLP).

(B) Cytospins of cell populations confirm characteristic morphology of particular cell types. Example populations shown are (i) Common lymphoid progenitors (CLP), (ii) Mast cells, (iii) ortho and poly erythroblasts (EryBIPO) and (iv) neutrophils. (NeutLN). Scale bars indicate 10 microns.

Figure S2. Related to Figure 1: Replicates cluster with similar cell populations

tSNE scaling plot of individual replicate samples for each lineage. Where samples were plotted on top of each other, numbers in brackets show how many samples are there. Colours and abbreviations as given for each lineage in Figure 1.

Figure S3. Related to Figure 1. Comparison of Haemopedia cell types to other murine haematopoietic expression datasets.

tSNE scaling plot of samples included from Haemopedia (filled circles), ImmGen (crosses) (Heng and Painter, 2008) and Gene Expression Commons (empty circles) (Seita et al., 2012). Only a selection of labels is shown for readability, focusing on where cells are not clustering obviously by lineage, or where there are is a large spread of cells within a lineage. Colours as given in Figure 1.

Figure S4. Related to Figure 2: Minimum spanning tree based on surface markers and transcription factors

(A) Heatmap of expression of highly variable probes that were used to generated the MST. Probes that were then used in (B) to make surface marker or transcription factor MST are highlighted.

(B) Probes for (i) surface markers or (ii) transcription factors with an $sd > 2.5$ of a \log_2 scale were selected. Minimum spanning tree based on Euclidean distance measurements was calculated using these probes. Lengths of branches reflect the distance between cell types. Cell type colours as given in Figure 1.

Figure S5: Related to Figure 3. Expression of Haemopedia lineage signatures in other haematopoietic data sets.

Heatmap of lineage signature genes as defined in Figure 3 in the equivalent lineages of other murine haematopoietic expression sets. Cell types from Haemopedia highlighted with pale grey, from Immgen (Heng and Painter, 2008) with dark grey and Gene Expression Commons (Seita et al., 2012) with Black. Cell type abbreviations for the other datasets are as given in the

original paper or found on their web resource. Genes plotted are lineage signature genes that are found on all three platforms, and where there is a representative of the lineage found in at least one of the other datasets. The colour scale is mean normalised expression across each dataset, on a log₂ scale. Black lines separate cells and signatures by lineage, and the grey lines separate cell types by dataset of origin.

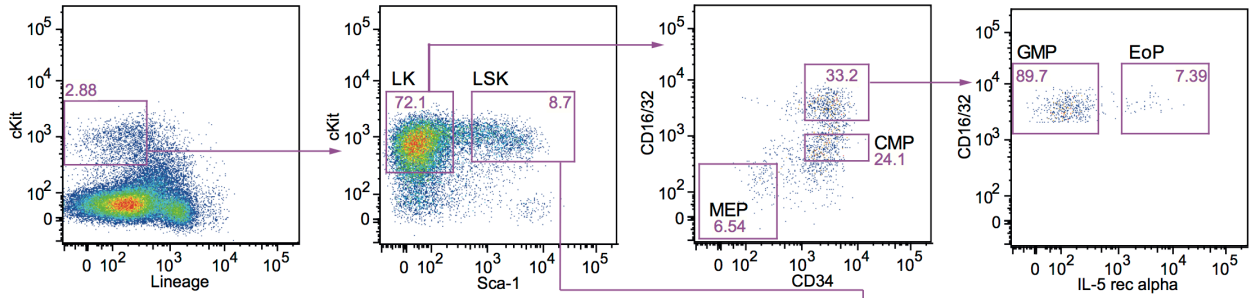
Figure S6. Related to Figure 4: MDS of mouse cell types vs human cell types

Multidimensional plot of mouse and human cell types after mean normalization of expression of one-to-one orthologues between species. Mouse cells are marked with squares, human cells are marked with crosses. Only a selection of labels is shown for readability. Mouse cell type labels are as before with an “m” prefix, human labels are as given in original publication of data (Novershtern et al., 2011), with an “h” prefix. Colours used are for the lineages as given in Figure 1. Distances on the plot represent leading fold change, the average absolute log₂-fold change of the top 500 genes that distinguish each pair of cell types. A)

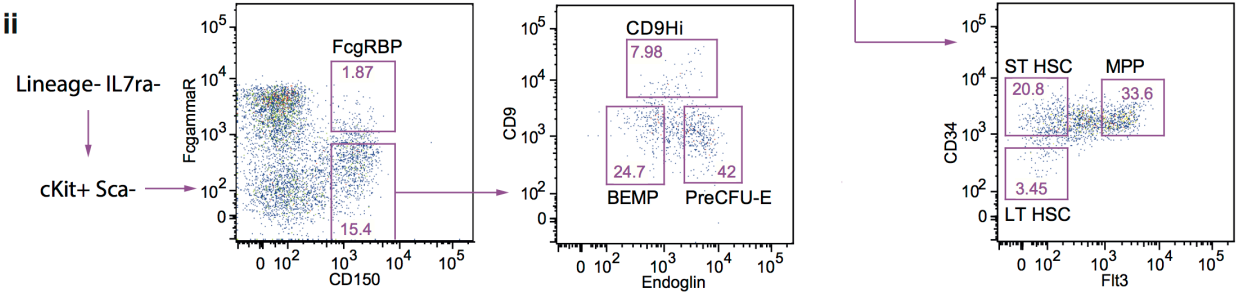
Figure S1

A

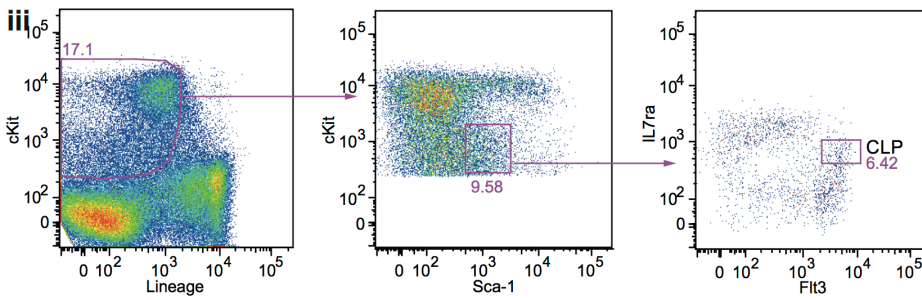
i



ii

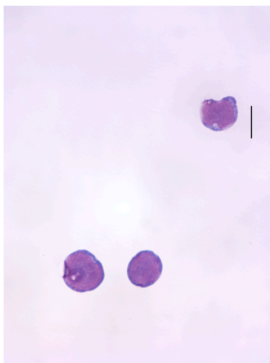


iii

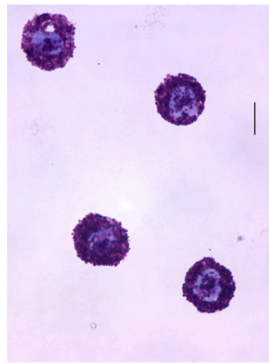


B

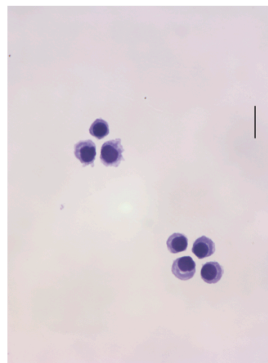
i



ii



iii



iv

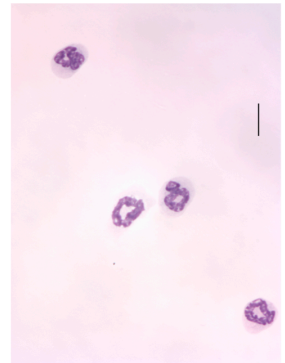


Figure S2

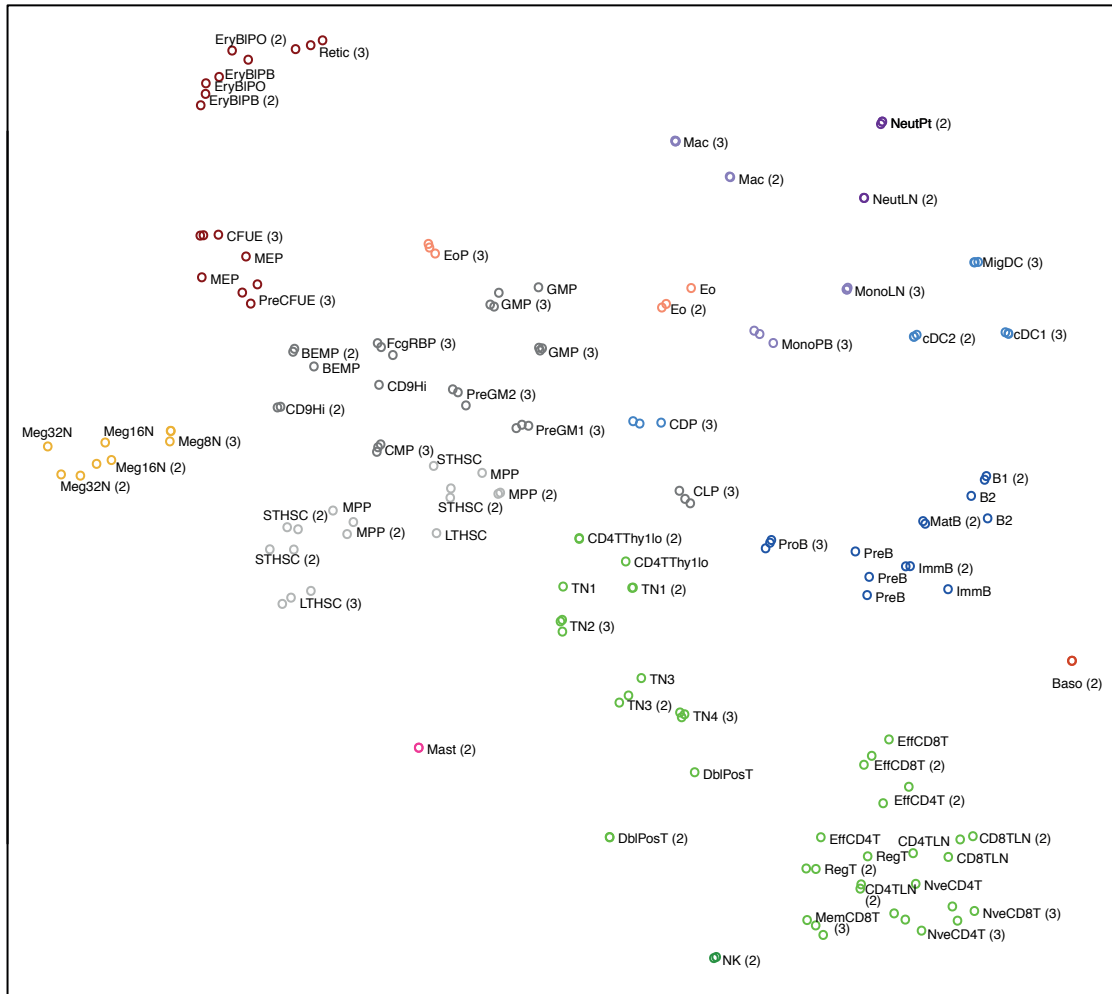


Figure S3

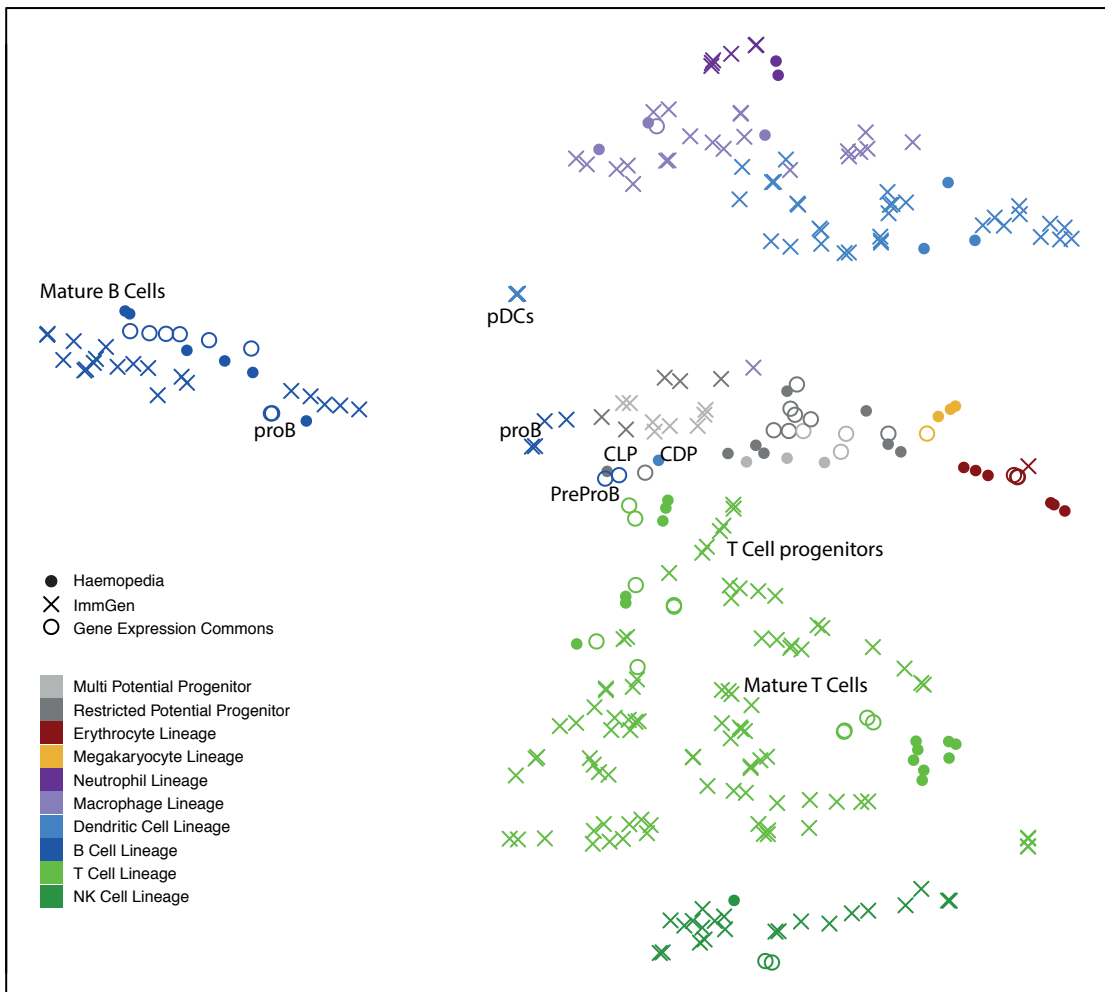
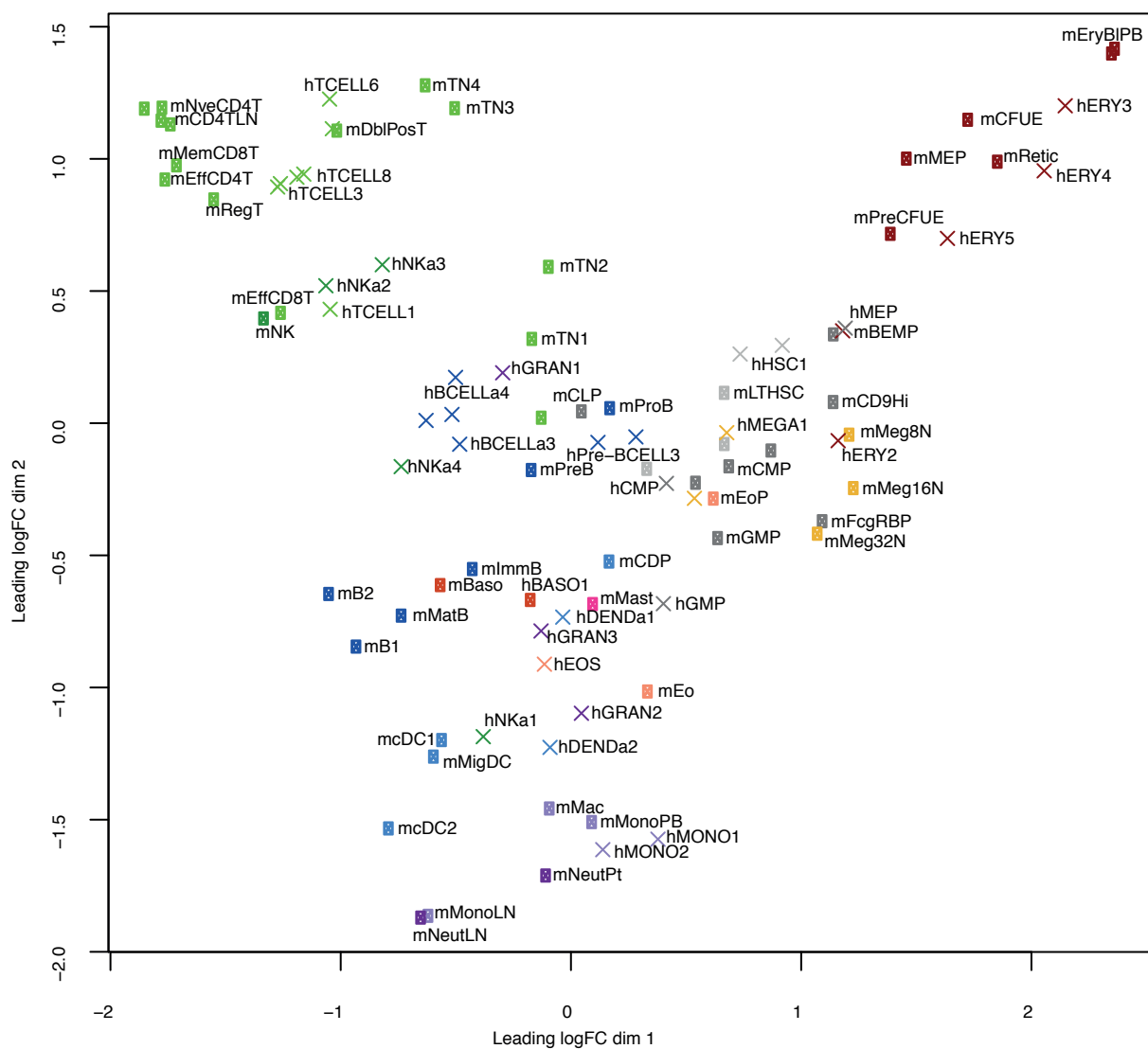


Figure S6



Supplemental Tables

S1 Table. Related to Figure 1: Replicate information for samples included in Haemopedia

S2 Table Related to Figure 1: Sorting criteria for cells included in Haemopedia

References for cell populations and surface markers are given in the table (Akashi et al., 2000; Böiers et al., 2010; Chen et al., 2009; Gordon and Taylor, 2005; Guilliams et al., 2014; Hardy et al., 2000; Heazlewood et al., 2013; Iwasaki et al., 2005; Jakubzick et al., 2013; Kaech et al., 2002; Kondo et al., 1997; Linch et al., 2012; Malbec et al., 2007; Naik et al., 2007; Ng et al., 2011, 2014; Nutt and Kee, 2007; Osawa et al., 1996; Overgaard et al., 2015; Pronk et al., 2007; Wang et al., 2014; Wu, 2006; Yang et al., 2005; Yoshimoto and Nakanishi, 2012; Zhang et al., 2001).

S3 Table. Related to Figure 1: Antibodies used to sort cells

S4 Table. Related to Figure 2: Genes used to create transcription factor and surface marker MSTs

Table S5. Related to Figure 3: Details of lineage specific genes

Table S6. Related to Figure 3: Ontologies and gene signatures associated with lineage specific genes

Table S7. Related to Figure 5: Murine lineage signature genes with maximal expression in human orthologous lineage

Table S8. Related to Figure 6: Genes with highest positive correlation to Ccr3

Supplemental Experimental Procedures

Generation of Mx-GFP mice

A targeting vector was constructed for generation of a modified *Mx* allele via homologous recombination in embryonic stem (ES) cells in which sequence encoding a biotinylation recognition site, MASSLRQILDSQKMEWRSNAGGS, was inserted immediately upstream of the of the stop codon in exon 7, followed by an internal ribosome entry site (IRES)-green fluorescent protein (GFP)-FRT-PGK-neo-FRT cassette. Chimeric mice were generated from an

ES cell clone bearing the targeted locus and heterozygous offspring established on a C57BL/6 background. The PGK-neo sequence was removed by intercrossing with transgenic mice ubiquitously expressing Flp recombinase (Farley et al., 2000). The resultant allele acts as a GFP reporter for expression from the *Mkx* locus; experiments were performed using mice homozygous for this allele. All procedures were approved by The Walter and Eliza Hall Institute of Medical Research Animal Ethics Committee.

Populations were analysed on a flow cytometer using the population definitions as outlined in Table S2.

Analysis of *Mkx*-GFP mice

Haematopoietic cells were isolated from the bone marrow, spleen, blood and peritoneal cavity of *Mkx*^{GFP/GFP} and C57BL/6 mice for flow cytometry as follows: Bone marrow was flushed from femurs and tibias and made into a single cell suspension by passing through a 23 gauge needle. Splenic cells were isolated by mashing tissue through a 40µm filter. Peripheral blood was collected from the retro-orbital sinus, and peritoneal cells obtained by peritoneal lavage. Red cells were removed from the bone marrow, spleen and blood samples by hypotonic lysis.

Cells were stained at a concentration of 2×10^7 cells/mL in phosphate buffered saline (PBS) containing 2% fetal bovine serum (FBS) and the following antibodies: B220-Alexa-Fluor A700, CD3-biotin, CD125-PE, SiglecF-AlexaFluor 647, Ly6C-e-450, Ly6G-PeCy7 and CD16/32. Secondary staining was performed with Streptavidin-PerCpCy5.5 in PBS+2% FBS. Antibodies as in Table S3. Cells were resuspended PBS + 2%FCS containing 2mM EDTA and propidium iodide (Invitrogen), and analysed using an LSRII flow cytometer (Becton Dickinson). Subsequent data analyses were performed with FlowJo software (FlowJo) and plots generated using Prism (GraphPad Software).

Bioinformatic Analyses

Transcription Factor and Surface Marker Gene Sets

Transcription factor genes were identified as genes associated with any one of the following gene ontologies: GO:0005667 (transcription factor complex), GO:0009299 (mRNA transcription), GO:0003677 (DNA binding) (The Gene Ontology Consortium, 2015). Surface marker genes were selected as genes associated with the following gene ontologies: GO:0005886 (plasma membrane) (The Gene Ontology Consortium, 2015). Any genes that were found in both groups, such as *Notch* (Andersson et al., 2011), were removed.

Gene set enrichment analysis

Gene sets were downloaded from the Molecular Signatures Database (www.broadinstitute.org/gsea/msigdb) (Liberzon et al., 2011). Human gene identifiers were mapped to mouse orthologues using the Jackson Laboratory Human and Mouse Orthology Report (downloaded 12 December 2012), and the resulting mouse signature sets were uploaded to <http://bioinf.wehi.edu.au/software/MSigDB>. Lineage signatures were compared to curated genes sets (C2) and motif genes sets (C3). Curated gene sets include chemical and genetics perturbations, canonical pathways, BioCarta pathways, KEGG gene sets and the Reactome pathway. Motif gene sets contain genes that share a cis-regulatory motif that is conserved across the human, mouse, rat and dog genomes. Gene sets that were enriched in the lineage specific datasets were identified with Fisher's exact test, and p values were adjusted with Benjamini-Hochberg correction. A cutoff of adjusted p value of 0.05 was used to denote significance.

Comparison to different datasets.

Transcriptional profiles of ImmGen (Heng and Painter, 2008) were obtained from their data portal at (<http://rstats.immgen.org/DataRequest/>) and Gene Expression Commons (Seita et al., 2012) samples were obtained by downloading the published data from the Gene Expression Omnibus (GEO) with accession number GSE34723. We quantile normalized the data and

where there were replicates, samples were averaged. Samples were then associated with a particular lineage according to our scheme, and non haematopoietic samples were removed. Probes were then associated with Ensembl gene ids by BioMart (www.biomart.org). We then reduced the data sets down to Ensembl Ids which were represented in all three datasets, leaving 15126 genes. The data for each gene was then mean centered at zero separately for each data set.

For the comparison of cell types, we removed the basophils, mast cells and eosinophils from Haemopedia as these cell types were not in either of the two other datasets. We then made a tSNE plot in R using standard parameters from the package *Rtsne*.

To examine the expression of our lineage specific gene signatures across the different datasets, we again only looked at those that had a probe with a matching Ensembl Id in both other datasets, and used the mean centered data. For this plot we only showed the mature cells of each lineage and multipotential progenitors, as these were the most equivalent to the cell types that we originally defined the lineage signatures in. The exceptions to this were the erythrocyte and megakaryocyte lineages where we have included the progenitors from the Gene Expression Commons dataset as these are the only cell types of these lineages available in other data sets. We also removed the plasmacytoid dendritic cells from the ImmGen data set, as Haemopedia does not include equivalent cells.

Haemosphere Methods

Dataset processing for Haemosphere

The datasets included in Haemosphere have been chosen based on the following key criteria:

1. The dataset should represent haematopoietic expression data at the whole genome level.
2. The dataset should be primarily composed of a range of wildtype (for mouse) or healthy (for human) haematopoietic cell types, so that it can serve as a reference dataset. Other cell types may be included if they

serve as controls or help to interpret the primary data on steady state haematopoiesis.

Each dataset in Haemosphere is manually curated to ensure that a set of appropriately normalized probe intensities is used for analysis and visualization. Illumina probes in Haemopedia have been filtered based on the annotation by Barbosa-Morais et al (Barbosa-Morais et al., 2009) (those with “bad” or “no match” values according to this annotation were removed).

Gene definition

To provide optimal usefulness to the user, Haemosphere uses gene annotation from a combination of sources. Ensembl gene ids are used as unique gene identifiers in the system, but each Ensembl gene was also carefully matched against Entrez genes in order to obtain synonym information. Then MGI was used to obtain gene orthologues (ftp://ftp.informatics.jax.org/pub/reports/HOM_AllOrganism.rpt). Details of particular versions of these data sources used in Haemosphere are available on the website under the “about” section, as well as the ability to download the complete list of genes used by the system as a text file.

Code base

Haemosphere has been written in Python and JavaScript, using popular free-to-use packages and frameworks including pyramid, pandas, rpy2, angularjs and d3.

Supplemental References

Akashi, K., Traver, D., Miyamoto, T., and Weissman, I.L. (2000). A clonogenic common myeloid progenitor that gives rise to all myeloid lineages. *Nature* *404*, 193–197.

Andersson, E.R., Sandberg, R., and Lendahl, U. (2011). Notch signaling: simplicity in design, versatility in function. *Development* *138*, 3593–3612.

Barbosa-Morais, N.L., Dunning, M.J., Samarajiwa, S.A., Darot, J.F.J., Ritchie, M.E., Lynch, A.G., and Tavaré, S. (2009). A re-annotation pipeline for Illumina

BeadArrays: Improving the interpretation of gene expression data. *Nucleic Acids Res.* **38**, e17.

Böiers, C., Buza-Vidas, N., Jensen, C.T., Pronk, C.J.H., Kharazi, S., Wittmann, L., Sitnicka, E., Hultquist, A., and Jacobsen, S.E.W. (2010). Expression and role of FLT3 in regulation of the earliest stage of normal granulocyte-monocyte progenitor development. *Blood* **115**, 5061–5068.

Chen, K., Liu, J., Heck, S., Chasis, J. a, An, X., and Mohandas, N. (2009). Resolving the distinct stages in erythroid differentiation based on dynamic changes in membrane protein expression during erythropoiesis. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 17413–17418.

Farley, F.W., Soriano, P., Steffen, L.S., and Dymecki, S.M. (2000). Widespread recombinase expression using FLPeR (flipper) mice. *Genesis* **28**, 106–110.

Gordon, S., and Taylor, P.R. (2005). Monocyte and macrophage heterogeneity. *Nat. Rev. Immunol.* **5**, 953–964.

Guilliams, M., Ginhoux, F., Jakubzick, C., Naik, S.H., Onai, N., Schraml, B.U., Segura, E., Tussiwand, R., and Yona, S. (2014). Dendritic cells, monocytes and macrophages: a unified nomenclature based on ontogeny. *Nat. Rev. Immunol.* **14**, 571–578.

Hardy, R.R., Li, Y.S., Allman, D., Asano, M., Gui, M., and Hayakawa, K. (2000). B-cell commitment, development and selection. *Immunol. Rev.* **175**, 23–32.

Heazlewood, S.Y., Williams, B., Storan, M.J., and Nilsson, S.K. (2013). The Prospective Isolation of Viable, High Ploidy Megakaryocytes from Adult Murine Bone Marrow by Fluorescence Activated Cell Sorting. In *Stem Cell Niche: Methods and Protocols*, pp. 121–133.

Heng, T.S.P., and Painter, M.W. (2008). The Immunological Genome Project: networks of gene expression in immune cells. *Nat. Immunol.* **9**, 1091–1094.

Iwasaki, H., Mizuno, S., Mayfield, R., Shigematsu, H., Arinobu, Y., Seed, B., Gurish, M.F., Takatsu, K., and Akashi, K. (2005). Identification of eosinophil lineage-committed progenitors in the murine bone marrow. *J. Exp. Med.* **201**, 1891–1897.

Jakubzick, C., Gautier, E.L., Gibbings, S.L., Sojka, D.K., Schlitzer, A., Johnson, T.E., Ivanov, S., Duan, Q., Bala, S., Condon, T., et al. (2013). Minimal Differentiation of Classical Monocytes as They Survey Steady-State Tissues and Transport Antigen to Lymph Nodes. *Immunity* **39**, 599–610.

Kaech, S.M., Wherry, E.J., and Ahmed, R. (2002). Effector and memory T-cell

differentiation: implications for vaccine development. *Nat. Rev. Immunol.* 2, 251–262.

Kondo, M., Weissman, I.L., and Akashi, K. (1997). Identification of clonogenic common lymphoid progenitors in mouse bone marrow. *Cell* 91, 661–672.

Liberzon, A., Subramanian, A., Pinchback, R., Thorvaldsdóttir, H., Tamayo, P., and Mesirov, J.P. (2011). Molecular signatures database (MSigDB) 3.0. *Bioinformatics* 27, 1739–1740.

Linch, S.N., Danielson, E.T., Kelly, A.M., Tamakawa, R. a, Lee, J.J., and Gold, J. a (2012). Interleukin 5 is protective during sepsis in an eosinophil-independent manner. *Am. J. Respir. Crit. Care Med.* 186, 246–254.

Malbec, O., Roget, K., Schiffer, C., Iannascoli, B., Dumas, A.R., Arock, M., and Daeron, M. (2007). Peritoneal Cell-Derived Mast Cells: An In Vitro Model of Mature Serosal-Type Mouse Mast Cells. *J. Immunol.* 178, 6465–6475.

Naik, S.H., Sathe, P., Park, H.-Y., Metcalf, D., Proietto, A.I., Dakic, A., Carotta, S., O’Keeffe, M., Bahlo, M., Papenfuss, A., et al. (2007). Development of plasmacytoid and conventional dendritic cell subtypes from single precursor cells derived in vitro and in vivo. *Nat. Immunol.* 8, 1217–1226.

Ng, A.P., Kauppi, M., Metcalf, D., Di, L., Hyland, C.D., and Alexander, W.S. (2011). Characterization of thrombopoietin (TPO)-responsive progenitor cells in adult mouse bone marrow with in vivo megakaryocyte and erythroid potential. *Proc. Natl. Acad. Sci. U. S. A.* 109, 2364–2369.

Ng, A.P., Kauppi, M., Metcalf, D., Hyland, C.D., Josefsson, E.C., Lebois, M., Zhang, J.-G., Baldwin, T.M., Di Rago, L., Hilton, D.J., et al. (2014). Mpl expression on megakaryocytes and platelets is dispensable for thrombopoiesis but essential to prevent myeloproliferation. *Proc. Natl. Acad. Sci. U. S. A.* 111, 5884–5889.

Novershtern, N., Subramanian, A., Lawton, L.N., Mak, R.H., Haining, W.N., McConkey, M.E., Habib, N., Yosef, N., Chang, C.Y., Shay, T., et al. (2011). Densely interconnected transcriptional circuits control cell states in human hematopoiesis. *Cell* 144, 296–309.

Nutt, S.L., and Kee, B.L. (2007). The transcriptional regulation of B cell lineage commitment. *Immunity* 26, 715–725.

Osawa, M., Hanada, K., Hamada, H., and Nakauchi, H. (1996). Long-term lymphohematopoietic reconstitution by a single CD34-low/negative hematopoietic stem cell. *Science* 273, 242–245.

Overgaard, N.H., Jung, J.-W., Steptoe, R.J., and Wells, J.W. (2015).

CD4⁺/CD8⁺ double-positive T cells: more than just a developmental stage? *J. Leukoc. Biol.* **97**, 31–38.

Pronk, C.J.H., Rossi, D.J., Månsson, R., Attema, J.L., Norddahl, G.L., Chan, C.K.F., Sigvardsson, M., Weissman, I.L., and Bryder, D. (2007). Elucidation of the Phenotypic, Functional, and Molecular Topography of a Myeloerythroid Progenitor Cell Hierarchy. *Cell Stem Cell* **1**, 428–442.

Seita, J., Sahoo, D., Rossi, D.J., Bhattacharya, D., Serwold, T., Inlay, M.A., Ehrlich, L.I.R., Fathman, J.W., Dill, D.L., and Weissman, I.L. (2012). Gene expression commons: An open platform for absolute gene expression profiling. *PLoS One* **7**, 1–11.

The Gene Ontology Consortium (2015). Gene Ontology Consortium: going forward. *Nucleic Acids Res.* **43**, D1049–D1056.

Wang, H., Yan, M., Sun, J., Jain, S., Yoshimi, R., Abolfath, S.M., Ozato, K., Coleman, W.G., Ng, A.P., Metcalf, D., et al. (2014). A Reporter Mouse Reveals Lineage-Specific and Heterogeneous Expression of IRF8 during Lymphoid and Myeloid Cell Differentiation. *J. Immunol.* **193**, 1766–1777.

Wu, L. (2006). T lineage progenitors: The earliest steps en route to T lymphocytes. *Curr. Opin. Immunol.* **18**, 121–126.

Yang, L., Bryder, D., Adolfsson, J., Nygren, J., Månsson, R., Sigvardsson, M., and Jacobsen, S.E.W. (2005). Identification of Lin⁻Sca1⁺kit⁺CD34⁺Flt3⁻ short-term hematopoietic stem cells capable of rapidly reconstituting and rescuing myeloablated transplant recipients. *Blood* **105**, 2717–2723.

Yoshimoto, T., and Nakanishi, K. (2012). Generation and characterization of mouse basophils from bone marrow and purification of basophils from spleen. *Curr. Protoc. Immunol.* 1–16.

Zhang, X., Izikson, L., Liu, L., and Weiner, H.L. (2001). Activation of CD25⁺CD4⁺ regulatory T cells by oral antigen administration. *J. Immunol.* **167**, 4245–4253.