Fundamental properties of unperturbed haematopoiesis from stem cells in vivo – Supplementary Methods…

Article *in* Nature · February 2015

DOI: 10.1038/nature14242

CITATIONS
126

READS
104

9 authors, including:

Katrin Busch
German Cancer Research Center
13 PUBLICATIONS **826** CITATIONS
SEE PROFILE

Kay Klapproth
German Cancer Research Center
18 PUBLICATIONS **1,216** CITATIONS
SEE PROFILE

Susan Mariola Schlenner
Vlaams Instituut voor Biotechnologie and KU …
27 PUBLICATIONS **842** CITATIONS
SEE PROFILE

Thomas Höfer
German Cancer Research Center
195 PUBLICATIONS **4,208** CITATIONS
SEE PROFILE

doi:10.1038/nature14242

Supplementary Methods

Mathematical modelling

Hematopoietic fluxes. Consider two successive cell compartments in hematopoiesis, an 'upstream' and a 'reference' compartment (e.g., HSC and ST-HSC, respectively; Fig. 3a). We assume that the markers used for assigning the cells to these compartments identify homogeneous populations with respect to their differentiation and proliferation properties. The flux of cells from the former to the latter is the number of cells that differentiate per time unit from upstream (U) to reference (R) phenotype. Because this cell flux will be proportional to the number of cells in the upstream compartment, it is appropriate to define their per-capita differentiation rate $\alpha_{\mathrm{U}\to\mathrm{R}}$ so that the cell flux $j_{\mathrm{U}\to\mathrm{R}}$ is given by

$$j_{\mathrm{U}\to\mathrm{R}} = \alpha_{\mathrm{U}\to\mathrm{R}} n_{\mathrm{U}},$$

where n_{U} is the number of cells in the upstream compartment. Experimentally, $1/\alpha_{\mathrm{U}\to\mathrm{R}}$ is the average time an upstream cell takes from its generation by mitosis to differentiation into a reference cell. Because cells in each compartment can proliferate and die, the cell flux is not conserved throughout the hematopoietic system but rather $j_{\mathrm{U}\to\mathrm{R}}$ will take a specific value for each pair of upstream and reference compartments. Thus, the differentiation rates $\alpha_{\mathrm{U}\to\mathrm{R}}$ and the net proliferation rates of the cells within each compartment jointly define the hematopoietic fluxes. The net proliferation rate is the difference of the proliferation and death rates of the cells in a given compartment. Because of proliferation, the outgoing flux from the reference compartment into a downstream compartment $j_{\mathrm{R}\to\mathrm{D}}$ can be larger than the incoming cell flux. We define the amplification of the hematopoietic flux achieved by the reference compartment as the ratio of outgoing to the incoming cell fluxes:

$$\phi_R = \frac{j_{\mathrm{R}\to\mathrm{D}}}{j_{\mathrm{U}\to\mathrm{R}}}$$

Neither net proliferation nor differentiation rates *in vivo* have been accessible from direct measurements. We will now show how these rates can be accurately inferred for steady-state hematopoiesis from two types of measurement: the ratio of compartment sizes and the kinetics of propagation of a heritable label induced in the hematopoietic stem cells *in situ*.

Compartment sizes in steady state. Analogous to the per-capita differentiation rate from the upstream compartment, $\alpha_{\mathrm{U}\to\mathrm{R}}$, we define the rates of cell proliferation, λ_{U}, and cell

death, δ_U, within this compartment. The net proliferation rate is then

$$\beta_U = \lambda_U - \delta_U.$$

The same definitions hold for the recipient compartment (with subscript R). Of course, every compartment except for the stem cells and the final differentiated cells serves both as recipient (reference, R) and donor (upstream, U).

The rate of change of cells in the reference compartment is given by the balance of influx through differentiation from the upstream compartment, net proliferation and efflux through onward differentiation to downstream compartment(s) 'D':

$$\frac{dn_R}{dt} = \alpha_{U\to R} n_U - (\alpha_{R\to D} - \beta_R) n_R \tag{1}$$

There will be homeostatic and feedback regulation of the compartment sizes, which, in general, will make the proliferation and differentiation rates dependent on cell density. However, in steady state the compartment size n_R is constant and hence $\frac{dn_R}{dt} = 0$, implying

$$\frac{\bar{n}_R}{\bar{n}_U} = \frac{\alpha_{U\to R}}{\alpha_{R\to D} - \beta_R}. \tag{2}$$

Thus the ratio of reference and upstream compartment sizes in steady state, $\bar{n}_R$ and $\bar{n}_U$, contains information on the differentiation rates into ($\alpha_{U\to R}$) and out of ($\alpha_{R\to D}$) the reference compartment as well as the net proliferation rate in the reference compartment (β_R). Importantly, these rates are characteristic of the steady state of the hematopoietic system; upon challenge (e.g., infection or irradiation), the rates may change due to feedback or homeostatic regulation. Moreover, the rates could depend on the age of the cells. However, we assume for the fitting of the model to experimental data that the rates do not vary over the time periods of measurement, which was consistent with the experimental observations.

Label progression measures degree of self-renewal in compartments downstream of hematopoietic stem cells. In contrast to the total compartment sizes, a heritable label switched on in the hematopoietic stem cells will initially not be in steady state. The dynamics of labelled cells (denoted by an asterisk) is given by

$$\frac{dn_R^*}{dt} = \alpha_{U\to R} n_U^* - (\alpha_{R\to D} - \beta_R) n_R^* \tag{3}$$

As show experimentally for the $Tie2^{MCM/+}Rosa^{YFP}$ cells, the labelling does not alter the biological properties of HSC. Hence, the appearance of labelled cells does not perturb the steady state of the hematopoietic system: labelled cells coming into a compartment replace unlabelled cells that leave it. Therefore, the labelled cells will have the same steady-state differentiation and proliferation rates as their unlabelled counterparts.

It is practical to measure the frequency of labelled cells in a compartment, instead of

their total number:

$$f_R = \frac{n_R^*}{\bar{n}_R}.$$

Dividing Equation 3 by $\bar{n}_R$ and taking into account Equation 2, we find for the dynamics of the label frequency the simple law:

$$\frac{df_R}{dt} = (\alpha_{R\to D} - \beta_R)(f_U - f_R) \tag{4}$$

This is the key equation for interpreting the label progression data; it shows that label progression measures the degree of self-renewal in a compartment. To understand this property, first notice that Equation 4 implies that, in the long run, the labelling frequency in the reference compartment will equilibrate with the labelling frequency in the upstream compartment: $f_R = f_U$ is the steady state. The characteristic time taken to reach this steady state is

$$\tau_R = \frac{1}{\alpha_{R\to D} - \beta_R} \tag{5}$$

This is the time needed to reduce the size of the reference compartment to $1/e = 37\%$ of its steady state value when the influx from the upstream compartment is switched off (cf. Equation 1). Hence τ_R measures how long cells and their progeny remain in the compartment before being lost by differentiation and death; we call τ_R the residence time of the compartment.

In a perfectly self-renewing compartment, cells that differentiate will be fully replaced by proliferation. Therefore $\alpha_{R\to D} = \beta_R$ and the residence time τ_R will (theoretically) be infinite. The observation that the label frequency in the HSC does not decline over time is consistent with perfect self-renewal.

For all compartments downstream of the hematopoietic stem cells, onward differentiation will exceed net proliferation. Compartments that operate close to self-renewal will have a small positive difference $\alpha_{R\to D} - \beta_R$ and hence a large residence time, whereas for 'transit' compartments this difference will be larger and the residence time correspondingly smaller. Based on Equation 4, a mathematical model for label progression through the hematopoietic compartments can be written down. This model contains as the only free parameters the residence times of the different compartments, τ_i ($i = \text{ST-HSC}, \text{MPP}, \ldots$; see below). Therefore, fitting this model to label progression data for the compartments downstream of HSC will identify their residence times and, thus, their degree of self-renewal.

Rates of cell differentiation and net proliferation can be identified from data. Combining the knowledge about the residence times of the hematopoietic compartments gained from label progression with the measurement of the compartment size ratios at steady state, Equation 2 can be rearranged as follows:

$$\alpha_{U\to R} = \frac{1}{\tau_R}\frac{\bar{n}_R}{\bar{n}_U}. \tag{6}$$

The right-hand side of this equation contains only measured quantities, and hence the differentiation rate from upstream to recipient compartment is now known.

Fitting the label-progression model based on Equation 4 to the corresponding data will also establish the residence time of the upstream compartment, τ_U. Knowing $\alpha_{U\to R}$ and τ_U, we can infer the net proliferation rate in the upstream compartment β_U (Equation 5):

$$\beta_U = \alpha_{U\to R} - \frac{1}{\tau_U}. \tag{7}$$

Every intermediate compartment serves both as recipient and donor. Equations 6 and 7 imply that label progression kinetics and ratio of compartment sizes are sufficient to infer the differentiation and net proliferation for all these compartments. In the following section we will describe the details of this inference.

Parameter inference from the experimental data

Succession of stem and progenitor cell compartments. We begin by assuming the following model of hematopoiesis: HSC are the earliest progenitors, with constant number under unperturbed conditions. HSC differentiate into short-term reconstituting hematopoietic stem cells (ST-HSC), which then give rise to multi-potent progenitors (MPP). From MPP onward the development branches into the two main lineages, lymphoid and myeloid, giving rise to common lymphoid progenitors (CLP) and common myeloid progenitors (CMP). This model is studied in Figure 3. The continuation of differentiation in the bone marrow from CMP into granulocyte-macrophage progenitors (GMP) and megakaryocyte-erythroid progenitors (MEP) as well as from CLP into pro B cells is considered in Extended Data Fig. 6. We assume that all differentiation steps are irreversible.

Residence times from label progression. The frequencies of labelled cells follow the dynamics given by the differential equations:

$$\begin{aligned}
\frac{df_{\text{ST-HSC}}}{dt} &= \frac{1}{\tau_{\text{ST-HSC}}}(f_{\text{HSC}} - f_{\text{ST-HSC}}) \\
\frac{df_{\text{MPP}}}{dt} &= \frac{1}{\tau_{\text{MPP}}}(f_{\text{ST-HSC}} - f_{\text{MPP}}) \\
\frac{df_{\text{CMP}}}{dt} &= \frac{1}{\tau_{\text{CMP}}}(f_{\text{MPP}} - f_{\text{CMP}}) \\
\frac{df_{\text{CLP}}}{dt} &= \frac{1}{\tau_{\text{CLP}}}(f_{\text{MPP}} - f_{\text{CLP}})
\end{aligned} \tag{8}$$

with $f_{\text{HSC}} =$ constant and the initial conditions: $f_{\text{ST-HSC}} = f_{\text{MPP}} = f_{\text{CMP}} = f_{\text{CLP}} = 0$. These equations can be solved explicitly and fitted to the label progression data, establishing the residence times $\tau_{\text{ST-HSC}}$, τ_{MPP}, τ_{CMP} and τ_{CLP}. To to this, we used data of 110 adult mice treated with tamoxifen at age ≥ 6 weeks and measured between 1 and 34 weeks after label

induction (except for the CMP compartment, for which only data measured between 6 and 34 weeks after label induction were included; Extended Data Fig. 4). As detailed below, this number of mice was chosen to cover this time period evenly and to estimate mean values with sufficient accuracy for subsequent model parameterisation. In order to compare the data for different mice, the frequencies of labelled cells in each compartment were normalised to the label frequency measured in HSC. The thus normalised label frequencies in compartments downstream of HSC vary between 0 (at time of label induction in HSC) and 1 (full equilibration of label) and show, for each mouse, how close label at time t (time point of sacrificing the mouse for measurement) has come to its eventual equilibrium value. The data were measured at 28 time points (1 week $\geq t \geq$ 34 weeks), with 3-4 mice per time point. To obtain a consistent estimate of the measurement error, we pooled the variances for all the data to calculate the standard error of the mean. Moreover, to limit noisy variation in the mean for the fit of the model to the data, we pooled nearby time points as follows: 7-28 days post label induction ($n = 19$), 35 – 52 days ($n = 18$), 57 – 77 days ($n = 19$), 87 – 119 days ($n = 13$), 126 – 140 days ($n = 14$), 154 – 168 days ($n = 12$), 182 – 190 days ($n = 8$) and 210 – 238 days ($n = 7$). These are the blue data points in Fig. 2l and Fig. 3c. For completeness, we also give the averaging for the data after embryonic HSC labelling at day E10.5 (red data points in Fig. 2l) that have not been used for model fitting: 2.5 days ($n = 9$), 5.5 – 9.5 days ($n = 40$) and 11.5 – 34.5 days ($n = 17$).

The model was then fitted to the labelling data for adult mice, using the lsqnonlin MATLAB function for nonlinear optimization. Pooling of the data into 16 instead of 8 time points gave practically identical results for the residence times of the compartments; Fig. 3c shows the result for the 8-point pooling of the data specified above. This model predicts labelling frequencies at time points beyond 34 weeks that were measure after model fitting, as shown in Fig. 3c and g.

Rates of differentiation and net proliferation. To determine these rates for HSC, ST-HSC and MPP, we rewrite Equation 8 as follows

$$\begin{aligned}
\frac{df_{\text{ST-HSC}}}{dt} &= (\alpha_{\text{ST}\to\text{MPP}} - \beta_{\text{ST}})(f_{\text{HSC}} - f_{\text{ST-HSC}}) \\
\frac{df_{\text{MPP}}}{dt} &= (\alpha_{\text{MPP}\to\text{CMP}} + \alpha_{\text{MPP}\to\text{CLP}} - \beta_{\text{MPP}})(f_{\text{ST}} - f_{\text{MPP}}) \\
\frac{df_{\text{CMP}}}{dt} &= \frac{1}{\tau_{\text{CMP}}}(f_{\text{MPP}} - f_{\text{CMP}}) \\
\frac{df_{\text{CLP}}}{dt} &= \frac{1}{\tau_{\text{CLP}}}(f_{\text{MPP}} - f_{\text{CLP}})
\end{aligned} \tag{9}$$

and consider the ratios between the compartment sizes:

$$\begin{aligned}
\frac{\bar{n}_{\text{ST-HSC}}}{\bar{n}_{\text{LT}}} &= \frac{\alpha_{\text{HSC}\to\text{ST-HSC}}}{\alpha_{\text{ST-HSC}\to\text{MPP}} - \beta_{\text{ST-HSC}}} \\
\frac{\bar{n}_{\text{MPP}}}{\bar{n}_{\text{ST-HSC}}} &= \frac{\alpha_{\text{ST-HSC}\to\text{MPP}}}{\alpha_{\text{MPP}\to\text{CMP}} + \alpha_{\text{MPP}\to\text{CLP}} - \beta_{\text{MPP}}} \\
\frac{\bar{n}_{\text{CMP}}}{\bar{n}_{\text{MPP}}} &= \alpha_{\text{MPP}\to\text{CMP}}\,\tau_{\text{CMP}} \\
\frac{\bar{n}_{\text{CLP}}}{\bar{n}_{\text{MPP}}} &= \alpha_{\text{MPP}\to\text{CLP}}\,\tau_{\text{CLP}}
\end{aligned} \tag{10}$$

To estimate the parameters, we fitted the model to the experimental data in a standard way by minimizing the sum of the weighted squared residuals:

$$c = \sum_j \sum_{i=1}^{N} \left(\frac{F_j(t_i) - F_j^{(o)}(t_i)}{\sigma_{F_{j,i}}} \right)^2 + \sum_j \left(\frac{R_j - R_j^{(0)}}{\sigma_{R_j}} \right)^2 \tag{11}$$

where:

- $j \in \{\text{ST-HSC,MPP,CMP,CLP}\}$;
- $i = 1,\ldots,N$ index the time points, t_i, at which the data are pooled;
- $F_j(t_i)$ is the computed normalised frequency of labelled cells in compartment j at time t_i: $F_j(t_i) = \frac{f_j(t_i)}{f_{\text{HSC}}(t_i)}$;
- $F_j^{(o)}(t_i)$ is the same as above, but observed instead of computed;
- $R_j = \frac{\bar{n}_{\text{j}}}{\bar{n}_{\text{j-precursor}}}$ is the computed ratio of the size of compartment j to the size of its precursor;
- $R_j^{(o)}$ is the same as above, but observed instead of computed;
- $\sigma_{F_{j,i}}$ and σ_{R_j} are the respective standard errors of the mean.

In addition to the data described in Section 2.2, the compartment size ratios $R_j^{(o)}$ have been measured for 10 mice (Extended Data Fig. 6a). The parameter values resulting from optimization (using MATLAB lsqnonlin) are depicted in Fig. 3 d,e. Moreover, we also fitted this model to data from older mice (332 – 802 days after label induction), finding a decline in lymphoid differentiation (Fig. 3g, h).

Confidence bounds for the parameterized models. To establish confidence bounds, we used non-parametric bootstrapping of the data[35] and the profile-likelihood method[36]. All 95% confidence bounds for the kinetic predictions of the model (grey shaded areas in Fig. 3

c, g; Extended Data Fig. 6d and 7d) were computed from refitting the model to 10,000 bootstrap samples of the original data. All 95% confidence bounds on the residence times obtained with the model (8) and the rates of differentiation and net proliferation obtained with model (9) (Fig. 3 d, e and h) were established through computing profile likelihoods (Extended Data Fig. 6b, c). The model parameters and confidence bounds are summarised in Supplementary Table 1.

Extensions of the model. This basic model has been extended in two ways. First, we included measurements for MEP and GMP compartments, which were placed downstream of CMP, as well as pro B cells downstream of CLP (Extended Data Fig. 6d). From Eq. 9, the differential equations for CMP and CLP were modified to account for the fact that data for the subsequent compartments are now considered explicitly and differential equations for GMP, MEP and pro B cells were added, as follows:

$$\begin{aligned}
\frac{df_{\mathrm{CMP}}}{dt} &= (\alpha_{\mathrm{CMP}\to\mathrm{GMP}} + \alpha_{\mathrm{CMP}\to\mathrm{MEP}} - \beta_{\mathrm{CMP}})(f_{\mathrm{MPP}} - f_{\mathrm{CMP}}) \\
\frac{df_{\mathrm{CLP}}}{dt} &= (\alpha_{\mathrm{CLP}\to\mathrm{proB}} + \alpha_{\mathrm{CLP}\to\mathrm{DN\ thymocytes}} - \beta_{\mathrm{CLP}})(f_{\mathrm{MPP}} - f_{\mathrm{CLP}}) \\
\frac{df_{\mathrm{GMP}}}{dt} &= \frac{1}{\tau_{\mathrm{GMP}}}(f_{\mathrm{CMP}} - f_{\mathrm{GMP}}) \\
\frac{df_{\mathrm{MEP}}}{dt} &= \frac{1}{\tau_{\mathrm{MEP}}}(f_{\mathrm{CMP}} - f_{\mathrm{MEP}}) \\
\frac{df_{\mathrm{proB}}}{dt} &= \frac{1}{\tau_{\mathrm{proB}}}(f_{\mathrm{CLP}} - f_{\mathrm{proB}})
\end{aligned} \tag{12}$$

In modification and extension of Eq. 10, we included the following relations for the cell number ratios:

$$\begin{aligned}
\frac{\bar{n}_{\mathrm{CMP}}}{\bar{n}_{\mathrm{MPP}}} &= \frac{\alpha_{\mathrm{MPP}\to\mathrm{CMP}}}{\alpha_{\mathrm{CMP}\to\mathrm{GMP}} + \alpha_{\mathrm{CMP}\to\mathrm{MEP}} - \beta_{\mathrm{CMP}}} \\
\frac{\bar{n}_{\mathrm{CLP}}}{\bar{n}_{\mathrm{MPP}}} &= \frac{\alpha_{\mathrm{MPP}\to\mathrm{CLP}}}{\alpha_{\mathrm{CLP}\to\mathrm{proB}} + \alpha_{\mathrm{CLP}\to\mathrm{DN\ thymocytes}} - \beta_{\mathrm{CLP}}} \\
\frac{\bar{n}_{\mathrm{GMP}}}{\bar{n}_{\mathrm{CMP}}} &= \alpha_{\mathrm{CMP}\to\mathrm{GMP}}\tau_{\mathrm{GMP}} \\
\frac{\bar{n}_{\mathrm{MEP}}}{\bar{n}_{\mathrm{CMP}}} &= \alpha_{\mathrm{CMP}\to\mathrm{MEP}}\tau_{\mathrm{MEP}} \\
\frac{\bar{n}_{\mathrm{proB}}}{\bar{n}_{\mathrm{CLP}}} &= \alpha_{\mathrm{CLP}\to\mathrm{proB}}\tau_{\mathrm{proB}}
\end{aligned} \tag{13}$$

For the model that included the subdivision of ST-HSC according to CD229 expression (Extended Data Fig. 7), we split the ST-HSC cell number into $\mathrm{CD229}^-$ and $\mathrm{CD229}^+$ variables and implemented a branch into CLP and MPP as shown. The two extended models were fit to the data and analysed in exactly the same way as described above for the basic model.

Supplementary Table 1

Summary of parameter estimation from experimental data

Residence time	Best fit value [d]	95% c.i. [d]
$\tau_{\text{ST-HSC}}$	330	[250, 450]
τ_{MPP}	70	[40, 110]
τ_{CMP}	1.1	[0.6, 4.1]
τ_{CLP}	60	[30, 170]
Differentiation rate	**Best fit value [d^{-1}]**	**95% c.i. [d^{-1}]**
$\alpha_{\text{HSC}\to\text{ST-HSC}}$	0.009	[0.006, 0.012]
$\alpha_{\text{ST-HSC}\to\text{MPP}}$	0.045	[0.025, 0.1]
$\alpha_{\text{MPP}\to\text{CMP}}$	3.992	[0.272, 4]
$\alpha_{\text{MPP}\to\text{CLP}}$	0.022	[0.008, 0.059]
$\alpha^{\text{old}}_{\text{MPP}\to\text{CMP}}$	4	[0.08, 4]
$\alpha^{\text{old}}_{\text{MPP}\to\text{CLP}}$	0.005	[0.003, 0.011]
Net proliferation rate	**Best fit value [d^{-1}]**	**95% c.i. [d^{-1}]**
β_{HSC}	0.009	[0.006, 0.012]
$\beta_{\text{ST-HSC}}$	0.042	[0.021, 0.1]
β_{MPP}	4	[0.3, 4]
Flux amplification	**Best fit value**	**95% c.i.**
ϕ_{ST}	15	[7, 40]
ϕ_{MPP}	280	[50, 450]
Myeloid bias	**Best fit value**	**95% c.i.**
$\alpha_{\text{MPP}\to\text{CMP}}/\alpha_{\text{MPP}\to\text{CLP}}$	180	[20, 610]
Compartment size relative to HSC	**Best fit value**	**95% c.i.**
$n_{\text{ST-HSC}}$	2.9	[2.3, 3.5]
n_{MPP}	9	[7, 11]
n_{CMP}	39	[25, 50]
n_{CLP}	13	[8, 16]

Supplementary Discussion

Main ideas of the mathematical analysis (related to Fig. 3)

Here we discuss the ideas and assumptions that allow us to infer the parameters of adult hematopoiesis from the experimental data. The corresponding mathematical treatment is detailed in Supplementary Methods.

Residence times of the hematopoietic stem and progenitor cell compartments. The measurement of the labelling frequency in a hematopoietic compartment over time after label induction in HSC provides direct information on how long incoming cells and their progeny reside in this compartment. The underlying ideas are as follows:

1. The residence time of a compartment measures how long incoming cells and their progeny reside in the compartment before being lost by differentiation and cell death. Formally, we define the residence time as the time period over which the compartment size would be reduced to 37% of its initial value after the input is switched off.

2. When hematopoiesis is in steady state, the total number of cells in each compartment remains constant and, hence, the frequency of labelled cells increases as they replace unlabelled cells. Therefore, the residence time of a compartment determines how long it takes for labelled cells to achieve equilibrium with the labelling frequency in the preceding compartment.

3. With respect to label equilibration between HSC and downstream compartments, the residence times are additive. The label equilibration between HSC and ST-HSC is determined by the residence time for ST-HSC; the label equilibration between HSC and MPP is determined by the sum of the residence times for ST-HSC and MPP, etc.

4. These ideas define how the residence times can be inferred ('read off') from the label progression data. To this end, we fit a series of exponential functions, defined by differential equations (Supplementary Methods, Eq. 8), to the label progression data. The first measured compartment (ST-HSC) is described by one exponential function, the second (MPP) by two exponential functions (one accounting for the kinetics of the ST-HSC input and the second for the MPP kinetics) and so on.

5. These exponential functions can then be used to predict the labelling frequencies in the compartments of interest at later time points. The predictions can be compared with further experimental data that have not been used for model fitting (Fig. 3c, g, green points).

Degree of self-renewal and rates of cell differentiation and net proliferation. Perfect self-renewal implies that cell loss by differentiation and cell death is exactly balanced by proliferation; this is a key property of HSC. Downstream compartments require HSC input for their maintenance. However, cell proliferation might almost balance cell loss and such a downstream compartment then operates close to self-renewal (as we find for the ST-HSC

in steady state). In the following, we explain that the degree of self-renewal in downstream compartments is directly linked to its residence time established from the kinetics of label progression. Moreover, we show that the label progression data together with a set of additional measurements, namely the ratios of compartment sizes, are sufficient to infer the rates of cell differentiation and net proliferation for the hematopoietic stem and progenitor compartments.

6. The following definitions are needed: (i) Differentiation rate = inverse of the mean waiting time before an incoming cell differentiates; (ii) Proliferation rate = inverse of the mean cell cycle time; (iii) Death rate = inverse of the mean waiting time from the birth of a cell to cell death; (iv) Net proliferation rate = proliferation rate minus death rate.

7. The net proliferation rate is an important quantity because it measures how many new cells effectively appear in a given time interval, after first taking away the cells that have died during this time interval. Because the (unlabelled) cells, whose numbers we also measure, have already played this give-and-take game, we can directly infer the net proliferation rate from such data without knowing the real proliferation and death rates.

8. Without cell proliferation and death in a compartment, the residence time is equal to the mean time an incoming cell waits before differentiating (i.e., leaving the compartment again). It is practical to define the inverse of this waiting time as the rate of differentiation (see 6). Hence in this special case: residence time = 1/(rate of differentiation).

9. With cell proliferation and cell death, it turns out that: residence time = 1/ (rate of differentiation – rate of net proliferation). The result is obtained using mathematics (Supplementary Methods) but is also intuitive. Suppose the differentiation rate in a compartment equals the net proliferation rate. Such a compartment exactly self-renews: all cells that are lost by differentiation are replaced by proliferation. We then have an infinite residence time. Generally, the closer the proliferation rate comes to the differentiation rate, the closer the compartment is to self-renewal (large residence time).

10. The knowledge of the residence time for a compartment can be interpreted as an equation for two unknowns: residence time (measured by label progression) = 1/ (unknown rate of differentiation – unknown rate of net proliferation).

11. To get a second equation that involves the same two unknowns, we realise that at steady state, the flux of cells that come into a compartment equals the flux of cells that leave the compartment per unit time. This gives a second equation of the type: measured ratio of compartment sizes = function of unknown rate of differentiation and unknown rate of net proliferation.

12. These two types of equations are defined by two independent measurements (label progression and ratio of compartment sizes) and contain two unknowns: rate of differentiation and rate of net proliferation for the compartment of interest. This is sufficient to calculate the two unknowns for each compartments in a given topology. In particular, we find

that the experimental data can be adequately interpreted in terms of the classical model of hematopoiesis.

Limiting dilution analysis consistent with kinetic model of hematopoiesis (related to Fig. 1b and Fig. 3)

In vivo limiting dilution analysis (LDA) inferred the probability that an HSC has given rise to the observed progeny within the time period of observation. To quantify the LDA data, we assumed that the HSC are independent. The proportionality ('single-hit kinetics'), between YFP^+ HSC number and the logarithm of the fraction of mice without YFP^+ pro B cells and DN thymocytes (Fig. 1b) indicates that this assumption is correct. It is important to note that the probability to observe labeled progeny from labeled HSC depends on the parameters of all compartments from HSC to the final product.

Here we developed a data-driven model for the cell-population dynamics of the hematopoietic stem and progenitor compartments during the maintenance of the system in adult mice. We can compare the LDA estimate of at least 30% active HSC during an observation period of 6 – 34 weeks with the rate-based mathematical model based on arguments from elementary probability theory, as follows: Given the estimated differentiation rate of HSC $\alpha_{\text{HSC}\to\text{ST-HSC}}$, the probability that one HSC has not produced differentiated progeny within time T is:

$$P_0(T) = e^{-\alpha_{\text{HSC}\to\text{ST-HSC}}T}$$

Accordingly, the probability to find no labelled HSC progeny in a mouse with N labelled HSC, at time T after labelling, is

$$P_N(T) = P_0(T)^N = e^{-N\alpha_{\text{HSC}\to\text{ST-HSC}}T} \tag{14}$$

This estimate can be compared to the LDA data for overall $CD45^+YFP^+$ progeny from labelled HSC (Extended Data Fig. 1h; a total of 60 mice were analysed). We focus on the subgroup of mice for which it was most likely not to recover labelled progeny, because they had the lowest number of labelled HSC and were analysed at early time points. From the 15 mice with the lowest HSC labelling frequencies, we consider the subset analysed in the early time window of 6 – 10 weeks after label induction. These 12 mice had on average $N = 8$ labelled HSC (not shown). Using the estimated HSC differentiation rate $\alpha_{\text{HSC}\to\text{ST-HSC}} = 1/110$ per day, the probabilities to find no $CD45^+YFP^+$ progeny are 4.9% after 6 weeks and 0.6% after 10 weeks (Equation 14). Hence the kinetic model implies that fewer than 1/20 mice (at 6 weeks) and 1/150 mice (at 10 weeks) will not have generated labelled progeny. Consistent with these figures, we found that all analysed mice (12) from the respective group had $CD45^+YFP^+$ progeny. For the further mice included in the LDA (Extended Data Fig.1g, h), all of which contained $CD45^+YFP^+$ progeny, the kinetic model would predict an even smaller probability to not find such progeny (because they had more

labelled HSC and/or were analysed at later time points). Hence the finding of the LDA that all 60 mice analysed between 6 and 34 weeks after label induction in HSC also contained labelled progeny is consistent with the differentiation rate of 1/110 HSC per day.

Formally, we cannot exclude that only a defined subset (fraction $0 < \rho \leq 1$) of the HSC is active. Hence the true number of HSC producing progeny would only by ρN in Equation 14. However, the probability to find no labelled progeny $P_N(T)$ will remain unchanged compared to the case of a uniform participation of the HSC: As the estimated differentiation rate of HSC is an average over all HSC, the differentiation rate of the active HSC will then be proportionally higher, $\alpha_{\text{HSC}\rightarrow\text{ST-HSC}}/\rho$. This will exactly offset the effect of decreasing the effective number of producers, so that the probability to find no labelled progeny $P_N(T)$ will still be given by Equation 14.

However, if only a defined subset of HSC were active, it is no longer guaranteed that active stem cells will be included among a small number of labelled HSC. If we require that the probability of labelling an active HSC exceeds 95% (99%) when we label in total 8 HSC (as we find for the subset of mice discussed above), then the proportion of active HSC, given through the binomial distribution, must be at least 31% (44%). Thus our finding that all of the mice with the smallest number of labelled HSC produced labelled progeny lends further support to the LDA result of at least 30% active HSC over a period of 6 – 34 weeks.

Simulation of response to 5-FU-induced leukopenia (related to Fig. 4)

The label equilibration between HSC and downstream stem and progenitor cell compartments is accelerated after transient leukopenia in the blood induced by a single application of 5-FU (Fig. 4). In compartments downstream of HSC, labelling ratios can only be increased by accelerated influx (an activation of cells within a compartment affects labelled and unlabelled cells alike and does not change their ratio in this compartment). We used the mathematical model for label propagation in steady state to examine the cellular dynamics that underlie the accelerated label equilibration. To this end, we first simulated the labelling dynamics for the control conditions (injection of PBS) with the steady-state rates of differentiation and net proliferation (Supplementary Table 1). Model and control data (PBS injection) were consistent; the labelling frequencies shown by the grey bars in Fig. 4c obtained at 37–43 days post labelling (12–18 days post PBS) were fit by the labelling frequencies in the model simulations at these time points (see Fig. 3c).

Next, we increased in the model the activity of various hematopoietic compartments after 5-FU application to simulate the stem and progenitor cell response to leukopenia (see Fig. 4a for the experimental setup on which the simulations were based). We expect that both differentiation rates and proliferation rates will be affected. Accelerated differentiation will increase hematopoietic flux, and enhanced proliferation will be needed to counter cell loss by 5-FU or/and by increased differentiation. For simplicity, we accelerated the rates of differentiation and net proliferation in a given compartment by the same numerical factor.

We found that a 7-fold acceleration of HSC differentiation and net proliferation increased the labelling frequency in the ST-HSC to the measured value. However, when such increase was limited to HSC, the label failed to propagate to downstream progenitor compartments (Fig. 4d, green dots). The reason for this behaviour is the nearly self-renewing reservoir property of the ST-HSC; the ST-HSC 'absorb' the increased input from the HSC, and only very slowly (over a time scale of 300 days) relay the label to the next downstream compartment. In view of the rapid recovery from the leukopenia (Fig. 4b), and the observed increase of label in the progenitor compartments (Fig. 4c), we reasoned that ST-HSC activity must also be stimulated after hematopoietic injury. Indeed, a 7-fold increase in HSC differentiation and net proliferation combined with a 16-fold increase of ST-HSC differentiation and net proliferation matched the experimentally observed acceleration of label propagation in all stem and progenitor cell compartments along the myeloid pathway (Fig. 4d, red dots). Despite the fact that CLP also appeared to respond (Fig. 4c), high variability prevented statistical significance. Finally, we tested whether the stimulation of ST-HSC alone would be sufficient to explain the data. This was clearly not the case, as the stimulation of the ST-HSC compartment does not change the labelling ratio in this compartment but does so only in downstream compartments (Fig. 4d, blue dots).

In summary, these model simulations show that the accelerated label equilibration after 5-FU application can be explained by activation of both HSC and ST-HSC compartments. It could be possible that, at least transiently, progenitor compartments are also stimulated. Such a stimulation of HSC and ST-HSC activities could be achieved by feedback inhibition from downstream compartments (including the blood). In turn, this would imply that the very low rate of HSC activity in steady-state hematopoiesis is, at least in part, controlled by feedback inhibition.

Quantitation of HSC engraftment and post-transplantation hematopoiesis (related to Fig. 5)

In the transplantation experiments, the YFP^+ HSC constitute a small fraction of the total transferred HSC (input: 0.2 – 2.7%; see Fig. 5c), allowing for a limiting dilution analysis. We find that 1/33 donor HSC produce engrafted HSC that we recovered in the bone marrow 4 months after transplantation (Extended Data Fig. 8a–c). This engraftment probability of HSC is consistent with a previous estimate[1]; hence, engraftment presents a strong bottleneck for transferred HSC.

After successful engraftment, the fraction of YFP^+ HSC in the bone marrow varied widely (Fig. 4d). Given the small absolute numbers of transferred YFP^+ HSC (4 – 30, on average 9 cells), and the fact that the mice with the two largest YFP^+ HSC contributions after 4 months had received only few YFP^+ HSC (4 – 6 cells), the unequal expansion of engrafted HSC in the bone marrow cannot be explained by different numbers of engrafted YFP^+ HSC; rather, engrafted HSC undergo skewed expansion. Note that the probability that

in both mice with the largest expansion of YFP^+ HSC after 4 months more than 1 founding YFP^+ HSC has contributed is below 0.5%. Given an engraftment probability of 1/33 and the numbers of YFP^+ HSC transferred per mouse, we estimate that on average 1.16 YFP^+ HSC per recipient have engrafted, expanded and thus produced the recovered YFP^+ HSC counts after 4 months shown in Fig. 5d (data points labelled 'within sampling error' and 'overrepresented').

As the bone marrow is fully functional again at this time point, we assume that in the order of 10,000 HSC have been rebuilt from the transplant. Based on the estimated number of engrafted YFP^+ HSC and the measured fractions of HSC that are YFP^+ HSC at 4 months, we calculated for each mouse the number of generations of YFP^+ HSC. These numbers provide a lower estimate of HSC proliferation because cell death and differentiation into ST-HSC have been neglected for lack of information. This estimate of self-renewing proliferation of the engrafted YFP^+ HSC shows a broad distribution (Extended Data Fig. 8d), with an average of 7.1 ± 2.0(SD) generations. Averaged over four months, this corresponds to a mean proliferation rate of 1/16 HSC per day, which is nearly seven-fold larger than the net proliferation rate in steady state (1/110 HSC per day). Importantly, the large variation between mice indicates that the participation of individual engrafted HSC in the repopulation of the bone marrow is highly heterogeneous. In summary, our label tracing data after transplantation show that engrafted HSC are on average much more active than HSC in the bone marrow at steady state. Moreover, the data indicate that both initial engraftment and subsequent expansion after transplantation display strong heterogeneity.

References

35. Efron, B. & Tibshirani, R.J. An Introduction to the Bootstrap. Chapman & Hall, New York, 1993.

36. Venzon, D.J. & Moolgavkar, S.H. A Method for computing profile-likelihood-based confidence intervals. *Appl. Statistics*, **37**, 87-94 (1988).