Hoxb5 marks long-term haematopoietic stem cells and reveals a homogenous perivascular niche

James Y. Chen1,2*, Masanori Miyawaki1,2*, Sean K. Wang1,2, Satoshi Yamazaki1, Rahul Sinha1,2, Kevin S. Kao1,2, Jun Seita1,2, Debashis Sahoo1,2, Hiromitsu Nakauchi1,3 & Irving L. Weissman1,2

Haematopoietic stem cells (HSCs) are arguably the most extensively characterized tissue stem cells. Since the identification of HSCs by prospective isolation1, complex multi-parameter flow cytometric isolation of phenotypic subsets has facilitated studies on many aspects of HSC biology, including self-renewal2–4, differentiation, ageing, niche5, and diversity6–8. Here we demonstrate by unbiased multi-step screening, identification of a single gene, homeobox B5 (Hoxb5, also known as Hox-2.1), with expression in the bone marrow that is limited to long-term (LT)-HSCs in mice. Using a mouse single-colour tri-mCherry reporter driven by endogenous Hoxb5 regulation, we show that only the Hoxb5+ HSCs exhibit long-term reconstitution capacity after transplantation in primary transplant recipients and, notably, in secondary recipients. Only ~7–35% of various previously defined immunophenotypic HSCs are LT-HSCs. Finally, by in situ imaging of mouse bone marrow, we show that >94% of LT-HSCs (Hoxb5+) are directly attached to VE-cadherin cells, implicating the perivascular space as a near-homogenous location of LT-HSCs.

Prospective isolation of HSCs requires that the isolated cells are capable of long-term production of all blood cell types in primary irradiated hosts, as well as self-renewal, such that the cells can be transplanted to secondary hosts to give rise to long-term multilineage repopulation. From the first enrichment and isolation of candidate HSCs9,10, this activity has been entirely contained in cell-surface-marker-defined cell populations, and more recently in fluorescent reporters11–13. However, the precise fraction of cells in those populations that are true LT-HSCs remains unknown.

To enable further purification of LT-HSCs, we sought to identify genes expressed exclusively in HSCs within cell populations resident in mouse bone marrow, detectable by flow cytometry and in situ fluorescence, and thus performed the following four-step screening (Fig. 1d).

First, we compared microarray gene expression assays among 28 distinct populations of the haematopoietic system (Extended Data Fig. 1a and Supplementary Table 1). Using the Gene Expression Commons platform14, we identified 118 candidate HSC-specific genes (Fig. 1a and Supplementary Table 2). Surprisingly, this list did not include all previously reported HSC-specific markers1–13 (Extended Data Fig. 1b and Supplementary Table 2). Second, to identify HSCs in situ, we excluded candidates that also label non-haematopoietic cell populations in the bone marrow such as stromal and endothelial cells15,16. Consequently, we excluded genes expressed in eight distinct non-haematopoietic bone marrow populations, thereby narrowing the list to 45 candidate genes (Fig. 1a).

Next, to ensure that the expression of any candidates could be detected by both flow cytometry and in situ fluorescence, we used RNA-sequencing (RNA-seq) combined with a threshold gene standard to estimate the fragments per kilobase of transcript per million mapped reads (FPKM) value that could serve as a detection threshold. From the bone marrow of 12-week-old mice, we sorted and RNA-sequenced immunophenotypically defined (Lin−c-Kit+Sca-1+CD150+CD34−/−Flk2−) HSCs (hereafter referred to as pHSCs), multipotent progenitors subset A (MPPs; Lin−c-Kit+Sca-1+CD150+CD34−/−Flk2−), and multipotent progenitors subset B (MPPb; Lin−c-Kit+Sca-1−CD150+CD34−/−Flk2−) (Fig. 1b) to determine the FPKM value of candidate genes. On the basis of Bmi-1–eGFP knock-in reporter expression17, we found that a single copy of eGFP is detectable at an estimated FPKM value of ~20. However, this high threshold would have excluded all candidate genes. Therefore, we designed a targeting construct (Fig. 1e) with three copies of mCherry, bringing the theoretical detection limit to ~7 FPKM. Lastly, to minimize aberrant detection, we set threshold FPKM values for both the MPPa and MPPb fractions to 2.5. Only three genes, Hoxb5, Rnf208, and Smtnl1, met these criteria (Fig. 1b).

Given previous reports of heterogeneity within pHSCs18–20, we analysed single cells to determine whether the remaining candidate genes were heterogeneously expressed among pHSCs. We reasoned that an ideal pan-HSC candidate gene would label the majority of pHSCs, with quantitative differences potentially reflecting HSC heterogeneity/diversity. We thus performed single-cell quantitative PCR (qPCR) analysis of pHSCs, and evaluated expression of Hoxb5, Rnf208, and Smtnl1. Only Hoxb5 exhibited bimodal expression, in comparison to the unimodality of Rnf208 and Smtnl1 (Fig. 1c). Therefore, from the entire HSC transcriptome, only Hoxb5 satisfied the criteria of our extensive unbiased screening (Fig. 1d).

We next sought to generate a Hoxb5 reporter with minimal disruption of endogenous Hoxb5 function. Thus we designed our targeting construct and CRISPR guide sequences to facilitate an in-frame knock-in to the endogenous Hoxb5 gene locus immediately 5′ of the only endogenous stop codon. We used three tandem mCherry cassettes separated by porcine teschovirus-1 2A (P2A) sequences, with the terminal mCherry carrying a CAAX membrane localization sequence (Hoxb5–tri-mCherry) (Fig. 1e).

To evaluate the specificity of this reporter, we isolated whole bone marrow cells from 12-week-old reporter mice and measured mCherry+ cells in the following immunophenotypic populations: pHSC, MPPa, MPPb, Flk2+ multipotent progenitor, megakaryocyte erythrocyte progenitor, granulocyte monocyte progenitor, common myeloid progenitor, common lymphoid progenitor fractions; differentiated cell populations (B cell, T cell, natural killer (NK) cell, neutrophil, eosinophil, monocyte, macrophage, dendritic cell, red blood cell, and megakaryocyte); and in CD45+ stromal fractions (Fig. 1f, Extended Data Fig. 2a, b, Extended Data Fig. 3, and unpublished data). Consistent with our initial screen (Fig. 1a–d), and using wild-type mice as a fluorescence minus one (FMO) threshold15, mCherry-labelled cells were highly enriched in the pHSC fraction (21.8% ± 0.90%), had a low
frequency in the MPPa fraction (1.88% ± 0.64%), and background frequencies in the remaining fractions (Fig. 1f, Extended Data Figs 2b and 3 and unpublished data). Interestingly, as only a minority of pHSCs were mCherry+ , this suggested that either our reporter labelled only a subfraction of pHSCs or only a subfraction of pHSCs were indeed HSCs.

To distinguish between these two possibilities and to determine whether Hoxb5 is a reporter of LT-HSCs, we characterized Hoxb5-expressing cells by transplantation. In order to be inclusive of all events in the pHSC gate, we used wild-type FMO to define Hoxb5 negativity (Hoxb5neg or sometimes referred to as Hox5neg) and divided the positive fraction (sometimes referred to as Hox5pos) into Hoxb5hi (top 5th percentile) and Hoxb5lo (14.1 ± 7.7%) (Extended Data Fig. 4). Ten-cell and three-cell grafts of Hoxb5hi, Hoxb5lo, or Hoxb5neg pHSCs were transplanted with supporting bone marrow cells into irradiated mouse recipients. We used CD45.2 expression to assess donor HSC contribution to haematopoietic lineages at 4-week intervals (Fig. 2a).

Analysis of peripheral blood 16 weeks after transplantation of ten cells demonstrated that multilineage reconstitution was present in 78% of Hoxb5hi, 70% of Hoxb5lo, and 44% Hoxb5neg pHSC recipients (Fig. 2b and Extended Data Fig. 5a). Three-cell transplants exhibited similar kinetics (Extended Data Fig. 5c). Notably, the chimera of the Hoxb5neg pHSCs decreased over time, in particular between 4 to 8 weeks and shifted towards a predominantly lymphoid chimera (Fig. 2b, d and Extended Data Fig. 5a), suggesting that either the Hoxb5neg fraction comprised of lymphoid-biased HSCs or, more likely, transiently self-renewing short-term (ST)-HSCs and/or MPPs that had given rise to long-lived lymphocytes.

To evaluate these two possibilities, we carried out a secondary transplantation of whole bone marrow from primary Hoxb5hi, Hoxb5lo, or Hoxb5neg pHSC recipients into lethally irradiated secondary recipients (Fig. 2a). Sixteen weeks after secondary transplantation, peripheral blood analyses revealed robust multilineage chimaerism from all Hoxb5hi and Hoxb5lo transplant recipients, with minimal chimaerism from the Hoxb5neg fraction (Fig. 2c, e and Extended Data Fig. 5b). Furthermore, bone marrow analysis of primary recipients for donor pHSCs revealed that the Hoxb5hi pHSC recipients contained Hoxb5hi, Hoxb5lo, and Hoxb5neg cells (100%, n = 10 mice), whereas the recipients of Hoxb5neg cells were only Hoxb5neg (36%, n = 4 mice) or devoid of donor cells (64%, n = 7 mice). These results further suggested that the Hoxb5neg pHSCs were in fact transiently self-renewing ST-HSCs/MPPs. To determine if Hoxb5 expression could also distinguish LT-HSCs in a primary transplant, we normalized the number of donor cells used for secondary transplant by sorting 100 Lin c-Kit Sca-1+ (LSK) donor cells from the primary Hoxb5hi recipients and transplanted them into irradiated secondary recipients (n = 24 mice) (Fig. 2a). As with the primary bone marrow transplants, the chimera was minimal in Hoxb5hi compared to Hoxb5lo recipients (Extended Data Fig. 5d). Limiting dilution analysis revealed that the frequency of long-term plus short-term HSCs in primary...
hosts at 16 weeks was 1 in 2.1 for Hoxb5hi, 1 in 2.4 for Hoxb5lo, and 1 in 16.1 for Hoxb5neg recipients (Extended Data Fig. 6). Taken together, these results demonstrate that Hoxb5 labels functional LT-HSCs.

Given that Hoxb5- cells are non-LT-HSCs, we re-examined the specificity of past definitions (Fig. 3a, f), including previously reported refinements to the LT-HSC immunophenotype18-20 and the most widely used in situ definition over the past decade22. We found that 78.5% ± 2.6% of CD11a+ HSCs19, 63.9% ± 3% of the HSC-1 (Lin-c-Kit+Sca-1+CD48+CD150+CD229+CD244-)20, and 82.5% ± 0.4% of fraction I HSCs (Lin-c-Kit+Sca-1+CD34-CD150+CD41-)21 were Hoxb5+ (Fig. 3b, c, and d, respectively). Surprisingly, 91.3% ± 0.4% of Lin-c-Kit+Sca-1+CD48+CD150+ (ref. 22) cells were Hoxb5hi (Fig. 3e). As this subset was initially used for localization of HSCs in situ22, we re-examined the in situ location of HSCs using Hoxb5 expression.

Visualizing LT-HSCs in bone marrow and identifying the cellular constituents and structures of the HSC niche remains challenging. Despite this, multiple constituent cell types have been proposed, including mesenchymal stromata, endostial osteoblasts, glia, endothelia, and pericytes. In situ studies are made difficult by several technical limitations, including the number of fluorescent colours, difficulty in identifying HSCs surrounded by non-HSCs, and difficulty in translating the same fluorescently defined positive and negative thresholds from flow cytometry to tissue sections.

To address these issues, we used the Hoxb5-tri-mCherry reporter. Using flow cytometry, we determined the utility of Hoxb5 alone in identifying LT-HSCs. After logical exclusion of autofluorescence by comparison to wild-type control mice, we found that 80% of Hoxb5hi events are within the c-Kit+ fraction and 62% ± 12.8% of all Hoxb5+ events are located in the pHSIC gate (Extended Data Fig. 8), representing an eight- to nine-fold enrichment compared to previous in situ labelling of HSCs (Lin-c-Kit+Sca-1+CD48+CD150-)22 (Fig. 3e).

To reveal the three-dimensional (3D) HSC niche architecture, we applied the CUBIC technique (clear, unobstructed brain imaging cocktails and computational analysis23) to tibial bone marrow, facilitating depletion of autofluorescent cellular components (Fig. 4a). Given
knowledge, for identification of LT-HSCs, and could facilitate model of Fig. 1b and Supplementary Table 3). These results suggest that the implicated5, including less homogenous association of candidate HSCs homogenously perivascular. Although other compartments have been bone marrow HSCs30), irradiation, and their influence on engraftment estimates the functional potential of candidate HSCs. Transplantation, in adult mouse bone marrow. Limiting dilution assay shows that at LT-HSCs from ST-HSCs has never been fully accomplished. Our study HSCs18–20,22,24,25. Identification of CD150, CD34, and CD48 enabled LT-HSC niche (Fig. 4b–d). cells. This implied a near-homogenous perivascular location for the proximal epiphysis, 39.0
longitudinal axis of the tibia (Fig. 4e and Extended Data Fig. 10a–c). In situ
images of direct (top panel) and non-direct (bottom panel) association of
of VE-cadherin in 3D-reconstructed images. Scale bar, 30

Figure 4 | LT-HSCs exhibit near-homogenous attachment to VE-cadherin+ cells. a, Tissue preparation and representative images of Tibial bone marrow plug after paraformaldehyde fixation (day 0) and treatment with reagent-1 (see ref. 23) (day 7, day 14). b, Localization of Hoxb5+ cells (red and arrows) and VE-cadherin+ cells (green) in 3D-reconstructed images. Scale bar, 30 μm. c, Representative 2D images of direct (top panel) and non-direct (bottom panel) association of
Hoxb5+ cells (red) with VE-cadherin+ cells (green). Scale bar, 10 μm. d, Frequency of Hoxb5+ cells (n = 287 cells, from n = 3 mice) and random spots (n = 600 spots, from n = 3 mice) plotted against proximity to VE-cadherin+ cells. **P < 0.0034. e, Average number of Hoxb5+ cells in proximal, medial, and distal regions of tibia (n = 3 mice). NS, not significant. Unpaired Student’s t-test (d, e).

Online Content Methods, along with any additional Extended Data display items, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 16 June; accepted 21 December 2015.

1. Spangrude, G. J., Heimfeld, S. & Weissman, I. L. Purification and characterization of mouse hematopoietic stem cells. Science 241, 58–62 (1988).

8. Lu, R., Neff, N. F., Quake, S. R. & Weissman, I. L. Tracking single hematopoietic stem cells in vivo using high-throughput sequencing in conjunction with viral genetic barcoding. Nature Biotechnol. 29, 928–933 (2011).

12. Gazit, R. et al. Fgd5 identifies hematopoietic stem cells in the murine bone marrow. J. Exp. Med. 211, 1315–1311 (2014).

15. Beerman, I. et al. Functionally distinct hematopoietic stem cells modulate hematopoietic lineage potential during aging by a mechanism of clonal expansion. Proc. Natl Acad. Sci. USA 107, 5465–5470 (2010).

16. Chan, C. K. F. et al. Identification and specification of the mouse skeletal stem cell. Cell 160, 285–298 (2015).

17. Hosen, N. et al. Bmi-1-green fluorescent protein-knock-in mice reveal the dynamic regulation of bmi-1 expression in normal and leukemic hematopoietic cells. Stem Cells 25, 1635–1644 (2007).

285–298 (2015).

34. Guo, W. et al. Multi-genetic events collaboratively contribute to Pten-null leukemia stem-cell formation. Nature 453, 529–533 (2008).

35. Yilmaz, O. H. et al. Pten dependence distinguishes hematopoietic stem cells from leukemia-initiating cells. Nature 441, 475–482 (2006).

36. Ito, K. et al. Regulation of oxidative stress by ATM is required for self-renewal of hematopoietic stem cells. Nature 431, 997–1002 (2004).

37. Morrison, S. J. & Scadden, D. T. The bone marrow niche for hematopoietic stem cells. Nature 505, 327–334 (2014).

38. Dykstra, B. et al. Long-term propagation of distinct hematopoietic differentiation programs in vivo. Cell Stem Cell 1, 218–229 (2007).

39. Beerman, I. et al. Functionally distinct hematopoietic stem cells modulate hematopoietic lineage potential during aging by a mechanism of clonal expansion. Proc. Natl Acad. Sci. USA 107, 5465–5470 (2010).

40. Lu, R., Neff, N. F., Quake, S. R. & Weissman, I. L. Tracking single hematopoietic stem cells in vivo using high-throughput sequencing in conjunction with viral genetic barcoding. Nature Biotechnol. 29, 928–933 (2011).

41. Uchida, N. & Weissman, I. L. Searching for hematopoietic stem cells: evidence that Thy-1.1−Lin− Sca-1− cells are the only stem cells in C57BL/Ka-Thy-1.1 bone marrow. J. Exp. Med. 175, 175–184 (1992).

42. Morrison, S. J. & Weissman, I. L. The long-term repopulating subset of hematopoietic stem cells is deterministic and isolatable by phenotype. Immunity 1, 661–673 (1994).

43. Acar, M. et al. Deep imaging of bone marrow shows non-dividing stem cells are mainly perisinusoidal. Nature 526, 126–130 (2015).

44. Gazit, R. et al. Fgd5 identifies hematopoietic stem cells in the murine bone marrow. J. Exp. Med. 211, 1315–1311 (2014).

45. Hills, D. et al. Hoxb4-YPF reporter mouse model: a novel tool for tracking HSC development and studying the role of Hoxb4 in hematopoiesis. Blood 117, 3521–3528 (2011).

46. Seita, J. et al. Gene expression commons: an open platform for absolute gene expression profiling. PLoS ONE 7, e40321 (2012).

47. Chan, C. K. F. et al. Identification and specification of the mouse skeletal stem cell. Cell 160, 285–298 (2015).

48. Chan, C. K. F. et al. Clonal precursor of bone, cartilage, and hematopoietic niche stromal cells. Proc. Natl Acad. Sci. USA 110, 12643–12648 (2013).

49. Hosen, N. et al. Bmi-1-green fluorescent protein-knock-in mice reveal the dynamic regulation of bmi-1 expression in normal and leukemic hematopoietic cells. Stem Cells 25, 1635–1644 (2007).
18. Yamamoto, R. et al. Clonal analysis unveils self-renewing lineage-restricted progenitors generated directly from hematopoietic stem cells. Cell 154, 1112–1126 (2013).

19. Fathman, J. W. et al. Upregulation of CD11A on hematopoietic stem cells denotes the loss of long-term reconstitution potential. Stem Cell Reports 3, 707–715 (2014).

20. Oguro, H., Ding, L. & Morrison, S. J. SLAM family markers resolve functionally distinct subpopulations of hematopoietic stem cells and multipotent progenitors. Cell Stem Cell 13, 102–116 (2013).

21. Herzenberg, L. A., Tung, J., Moore, W. A., Herzenberg, L. A. & Parks, D. R. Interpreting flow cytometry data: a guide for the perplexed. Nature Immunol. 7, 681–685 (2006).

22. Kiel, M. J. et al. SLAM family receptors distinguish hematopoietic stem and progenitor cells and reveal endothelial niches for stem cells. Cell 121, 1109–1121 (2005).

23. Susaki, E. A. et al. Whole-brain imaging with single-cell resolution using chemical cocktails and computational analysis. Cell 157, 726–739 (2014).

24. Christensen, J. L. & Weissman, I. L. Fli-2 is a marker in hematopoietic stem cell differentiation: a simple method to isolate long-term stem cells. Proc. Natl Acad. Sci. USA 98, 14541–14546 (2001).

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

26. Wright, D. E., Wagers, A. J., Gulati, A. P., Johnson, F. L. & Weissman, I. L. Long-term leukemia-initiating hematopoietic stem cells. PLoS ONE 5, e8785 (2010).

27. Forsberg, E. C. et al. Molecular signatures of quiescent, mobilized and leukemia-initiating hematopoietic stem cells. Proc. Natl Acad. Sci. USA 98, 14541–14546 (2001).

28. Fleming, W. H. et al. Functional heterogeneity is associated with the cell cycle status of murine hematopoietic stem cells. J. Cell Biol. 122, 897–902 (1993).

29. Passegué, E., Wagers, A. J., Giuriato, S., Anderson, W. C. & Weissman, I. L. Global analysis of proliferation and cell cycle gene expression in the regulation of hematopoietic stem and progenitor cell fates. J. Exp. Med. 202, 1599–1611 (2005).

30. Jaiswal, S. et al. CD47 is upregulated on circulating hematopoietic stem cells and leukemia cells to avoid phagocytosis. Cell 138, 271–285 (2009).

Supplementary Information is available in the online version of the paper.

Acknowledgements We would like to acknowledge N. Neff and G. Mantalas for advice regarding RNA sequencing; B. Yu and A. Beel for providing critical input on imaging data; H. Nishikii for advice regarding imaging data; S. Karten for help in editing the manuscript; L. Jerabek and T. Storm for laboratory management; A. McCarty and C. Wang for animal care; P. Lovelace and J. Ho for FACs facility management; H. Zeng, Y. Li, and C. Wang for collaboration in generating the mouse model; and Y. Sato for technical assistance in Imaris software analysis. The authors would like to acknowledge ongoing support for this work: NCI and NHLBI of the NIH under award numbers R01 CA086065, U01 HL099999, and R01 HL058770, and by the Virginia and D. K. Ludwig Fund for Cancer Research (I.L.W.); Stanford University Medical Scientist Training Program (T32 GM007365) and NHLBI Ruth L. Kirschstein National Research Service Award (F30-HL122096) (J.Y.C.); and Human Frontier Science Program Long-Term Fellowships, the Uehara Memorial Foundation Research Fellowship, Toyobo Biotechnology Foundation Research Fellowship, and Kanzawa Medical Research Foundation Overseas study grants (M.M.). The content is solely the responsibility of the authors and does not necessarily represent the official views of NIH.

Author Contributions J.Y.C. and M.M. contributed equally to this work, and either has the right to list himself first in bibliographic documents. M.M. and J.Y.C. conceived, performed, analysed, and oversaw the experiments, with suggestions from I.L.W. M.M. and J.Y.C. identified Hoxb5 as a LT-HSC marker, and made and characterized the Hoxb5–tri-mCherry mouse. S.K.W. and K.S.K. performed experiments and prepared figures under the supervision of M.M. and J.Y.C. S.Y. generated CUBIC data and evaluated the association with VE-cadherin vasculature. R.S. designed and performed RNA-seq and associated data analysis. J.S. and D.S. designed the gene expression commons for microarray analysis. D.S. provided critical advice regarding combined analysis of microarray and RNA-seq data. M.M., J.Y.C., S.K.W., K.S.K., and I.L.W. wrote the manuscript. H.N. and R.S. provided comments on the manuscript.

Author Information Microarray data was deposited at GEO under accession number GSE77078. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.M. (supamasa@stanford.edu) or I.L.W. (irv@stanford.edu).
Transplantations and peripheral blood analyses. B6.SJL-Peprc Pepc+/BoyJ (Jackson Laboratory) recipient mice were lethally irradiated at a single dose of 9.1 Gy. For reconstitution assays, donor cells were first combined with 2 × 10^6 whole bone marrow supporting cells (B6.SJL-Peprc Pepc+/BoyJ × C57BL/6J F1; mice CD45.1^+CD45.2^−) in 200 μl of PBS with 2% FBS, then injected into the retro-orbital venous plexus. Peripheral blood analyses were performed at 4, 8, 12, and 16 weeks after primary and secondary transplantations. At each time point, 50 μl of blood was collected from the tail vein and added to 100 μl of PBS with 2% EDTA. Flow cytometry reagents were then stained by using BD Pharm Lyse Buffer (BD Pharmingen), as per the manufacturer’s protocol, for 5 min on ice, followed by blocking with 5 μg/ml rat IgG. Leukocytes were stained with antibodies against (refer to Supplementary Table 4 for specific clones and colours): CD45.1 (FITC), CD45.2 (PE), CD11b (BV235; Gr-1 (Alexa-Flour700)), B220 (BV786), CD3 (BV421), TCRβ (BV421), and NK-1.1 (PerCP-cy5.5). For each mouse, the percentage of donor chimaerism in the peripheral blood was defined as the percentage of CD45.1^+CD45.2^− cells among total CD45.1^+CD45.2^+ and CD45.1^−CD45.2^− cells. To control for variability in host response to lethal irradiation, mice with host chimaerism of 50% or higher 16 weeks after transplantation were excluded from our analyses. The frequency of chimaerism in peripheral blood was analysed as follows. For evaluation of donor (CD45.1^−CD45.2^+) chimaerism kinetics, after exclusion of recipient (CD45.1^−CD45.2^−) fraction, the frequency of the donor fraction was calculated. Within the whole donor fraction, the frequency of each lineage (NK cell, B cell, T cell, granulocyte, and monocyte) was determined. For evaluation of lineage contribution kinetics, after gating each lineage, the frequency of the donor fraction (CD45.1^−CD45.2^−) was calculated. Any recipients that exhibited lower than 1% of chimaerism were treated as negative to exclude ambiguous cases.

Limiting dilution analysis. The frequency of long-term and short-term HSCs was calculated using the transplantation data of ten- and three-cell Hoxb5h, Hoxb5a, and Hoxb3a HSC transplants. Analysis were showing long-term (>16 weeks) multi-lineage reconstitution (>1% in each lineage) were counted as positive recipients. A nonlinear regression semi-log best fit line was used to calculate the frequency of LT/ST-HSCs at F_0 = 0.368 (GraphPad Prism 6).

CUBIC bone marrow imaging. Bone clearing protocol was modified from the original CUBIC protocol27. Specifically, tibias were collected and fixed in 4% PFA for two days, after which bone marrow plugs were extracted from the distal end by flushing method with a 25-gauge syringe. For nuclear staining, bone marrow plugs were immersed in DAPI/PBS solution at 37°C for three days with gentle shaking. For clearing, marrow plugs were immersed in ScaleCUBIC-1 (Reagent-1) at 37°C for two weeks with gentle shaking. The solution was changed every 48 h. To visualize vasculature, 20 μg of Alexa488-conjugated anti-mouse VE-cadherin antibody (BV13) was administered intravenously (retro-orbital) with tibias collected 30 min later. Processed plugs were embedded in 4-mm diameter glass capillaries with 2% agarose for imaging. Images were acquired using a Zeiss Z1 Lissht microscope (Zeiss) and reconstituted into 3D images using Zen software (Zeiss). Acquired 3D images were analysed using Imaris software (Bitplane). After exclusion of outliers including events of extraordinary size (>30 μm) or intensity, all other mCherry^− cells were analysed in the tibia (n = 287 cells in total from four mice). The threshold for determining the presence of the intensity level of the wild-type control tibial plugs. These mCherry^+ events with intensities ranging from 0.002–75.978 (75.978 representing the upper bound of mCherry intensity) were transformed into an integer list of 75,000 values (75,000 intensities ranging from 0.002–75.998) as a linear regression of intensity values.

Statistics. All comparative analyses were performed using unpaired Student’s t-tests. Pearson’s chi-square test was performed using online software (http://vassarstats.net/).

31. Moraga, L. et al. Tuning cytokine receptor signaling by re-orienting dimer geometry with surrogate ligands. *Cell* **160**, 1196–1208 (2015).
32. Wu, J., Anczkiewicz, O., Kainer, A. R., Zhang, M. Q. & Zhang, C. Olego: fast and sensitive mapping of spliced mRNA-Seq reads using small seeds. *Nucleic Acids Res.* **41**, 5149–5163 (2013).
33. Sanjana, N. E. et al. Transcript activator-like effector toolbox for genome engineering. *Nature Protocols* **7**, 171–192 (2012).
34. Cong, L. et al. Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).
Extended Data Figure 1 | GEXC expression of previously reported HSC markers in mouse bone marrow. a, Ideal expression pattern of HSC-specific genes (pink represents increased expression, blue represents decreased expression). b, Relative expression of Hoxb5 (top left), α-catulin/Ctnmal1 (top middle), Fgd5 (top right), CD150/Slamf1 (bottom left), Hoxb4 (bottom middle), Gfi-1 (bottom right) in haematopoietic and stromal populations as determined by microarray analysis.
Extended Data Figure 2 | Gating scheme for HSC and progenitors.

**a**, Representative flow cytometry gating to isolate pHSCs, MPPs, and oligopotent progenitors from mouse bone marrow. Panels gated as shown after exclusion of doublets and dead cells.

**b**, Hoxb5 reporter expression (red) in Flk2⁺ MPPs, megakaryocyte erythrocyte progenitor (MEP), granulocyte monocyte progenitor (GMP), common myeloid progenitor (CMP), and common lymphoid progenitor (CLP) populations compared to wild-type controls (blue). Values indicate the percentage of mCherry⁺ cells ± s.d. in each fraction for \( n = 3 \) mice.
Extended Data Figure 3 | Hoxb5 is not expressed in CD45− bone marrow. Hoxb5 reporter expression in the CD45− compartment within bone marrow of wild-type (red) and three Hoxb5−tri-mCherry mice (blue, orange, and green, n = 3 mice).
Extended Data Figure 4 | FMO gating for Hoxb5+ signal. Representative flow cytometry gating to separate mCherry (Hoxb5)-high, -low, and -negative populations in both wild-type and Hoxb5-tri-mCherry mice.
Extended Data Figure 5 | Hoxb5 distinguishes between LT-HSCs and non-LT-HSCs. a, Reconstitution kinetics in primary recipients 4, 8, and 12 weeks after receiving ten Hoxb5\(^{neg}\) (n = 9 mice), Hoxb5\(^{lo}\) (n = 13 mice), or Hoxb5\(^{hi}\) (n = 18 mice) pHSCs. Each column represents an individual mouse. b, Reconstitution kinetics 4, 8, and 12 weeks after whole bone marrow secondary transplant. c, Reconstitution kinetics in primary recipients receiving three Hoxb5\(^{neg}\) (n = 11 mice), Hoxb5\(^{lo}\) (n = 12 mice), or Hoxb5\(^{hi}\) (n = 14 mice) pHSCs. Each column represents an individual mouse. d, Reconstitution kinetics following secondary transplant of 100 sorted LSK Hoxb5\(^{−}\) (n = 14 mice) or Hoxb5\(^{+}\) (n = 9 mice) cells and 2 \times 10^5 supporting cells.
Extended Data Figure 6 | Limiting dilution analysis of Hoxb5^+ and Hoxb5^- pHSCs. Limiting dilution results of ten- and three-cell transplants of Hoxb5^hi (red, \(n = 18\) mice for ten-cell and \(n = 14\) mice for three-cell), Hoxb5^lo (green, \(n = 13\) mice for ten-cell and \(n = 12\) mice for three-cell), and Hoxb5^- (blue, \(n = 9\) mice for ten-cell and \(n = 11\) mice for three-cell). Frequency of LT/ST-HSCs by limiting dilution analysis is 1 in 2.1 for Hoxb5^hi, 1 in 2.4 for Hoxb5^lo, and 1 in 16.1 for Hoxb5^- cells.
Extended Data Figure 7 | Previously defined HSC immunophenotypes contain Hoxb5<sup>−</sup> cells. Representative HSC gating strategy for various HSC definitions after exclusion of doublets and dead cells. a, CD11a<sup>−</sup> (LSK CD150<sup>+</sup>CD34<sup>−</sup>/loCD11a<sup>−</sup>)<sup>21</sup>. b, HSC-1 (LSK CD150<sup>+</sup>CD48<sup>−</sup>/loCD229<sup>−</sup>/loCD244<sup>−</sup>)<sup>20</sup>. c, Fraction I (LSK CD150<sup>+</sup>CD34<sup>−</sup>/loCD41<sup>−</sup>)<sup>18</sup>. d, CD150<sup>+</sup>CD48<sup>−</sup>CD41<sup>−</sup> cells<sup>22</sup> (n = 5 mice).
Extended Data Figure 8 | Specificity of Hoxb5 as a single marker for LT-HSCs. a. Flow cytometry plots of wild type (top row) and Hoxb5–tri-mCherry (bottom row) after excluding doublets, dead cells, autofluorescence, and gating on Hoxb5⁺ events. Frequencies shown are percentage in gate ± s.d. in each fraction (n = 3 mice).
Extended Data Figure 9 | Comparison of processing methods on pHSC and Hoxb5<sup>+</sup> LT-HSC yield. a, b, Relative frequency of pHSCs (a) and Hoxb5<sup>+</sup> LT-HSCs (b) in tibial plugs (flushed) (n = 6 mice) compared to tibial plugs plus bones (crushed) (n = 6 mice).
Extended Data Figure 10 | Hoxb5+ HSCs are evenly distributed in the tibia. a, Distribution of Hoxb5+ cells (red and arrows) in bone marrow in 3D-reconstructed images. Nuclei are counterstained with DAPI (blue) wild-type (top panel) Hoxb5-tri-mCherry (middle and bottom panel). Scale bar, 100 μm. b, Cartoon representing the location of the proximal, medial, and distal sampling. c, Representative 3D-reconstructed images of Hoxb5+ cells (red) in proximal (left column), medial (middle column), and distal (right column) regions of the tibia. Scale bar, 150 μm. Nuclei are counterstained with DAPI (blue); n = 3 mice.