university of groningen

University Medical Center Groningen

University of Groningen

Clonal analysis of young and aged hematopoietic stem cells using cellular barcoding

Verovskaya, Evgenia

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2014

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):
Verovskaya, E. (2014). Clonal analysis of young and aged hematopoietic stem cells using cellular barcoding [S.l.]: s.n.

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Download date: 11-02-2018

Clonal analysis of young and aged hematopoietic stem cells using cellular barcoding

Evgenia Verovskaya

Clonal analysis of young and aged hematopoietic stem cells using cellular barcoding

University Medical Center Groningen, University of Groningen
Groningen, The Netherlands

Copyright © 2014 by Evgenia Verovskaya

Cover design: Ed Rochford

Printed by Off Page, Book design & production, Amsterdam, the Netherlands

ISBN (printed version): 978-90-367-7064-4

ISBN (electronic version): 978-90-367-7063-7

rijksuniversiteit groningen

Clonal analysis of young and aged hematopoietic stem cells using cellular barcoding

Proefschrift

ter verkrijging van de graad van doctor aan de
Rijksuniversiteit Groningen
op gezag van de
rector magnificus prof. dr. E. Sterken
en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op

woensdag 2 juli 2014 om 09.00 uur

door

Evgenia Verovskaya

geboren op 19 maart 1984
te Wolgograd, Rusland

Promotor
Prof dr. G. de Haan

Copromotor
Dr. L. Bystrykh

Beoordelingscommissie
Prof. dr. C. Eaves
Prof. dr. I. Roeder
Prof. dr. T.N.M. Schumacher

To my mother
Моей маме

Paranymphen

Mathilde Broekhuis
Ekaterina Ovchinnikova

CONTENTS

CHAPTER **1**

GENERAL INTRODUCTION & OUTLINE OF THE THESIS

Partially based on:

Verovskaya, E. and de Haan, G.

Noncanonical Wnt comes of age
in hematopoietic stem cells.
Cell Stem Cell, 3(6), 642-3 (2013).

GENERAL INTRODUCTION AND OUTLINE OF THE THESIS

Trillions of blood and immune cells are produced daily to sustain the functioning of the organism. This process is maintained by activity of hematopoietic stem cells (HSCs) that reside in the bone marrow (Figure 1). HSCs are unique in their ability to self-renew (produce more HSCs) and differentiate into all mature blood lineages. The equilibrium between these processes is crucial for tissue maintenance during the lifetime of the organism. For instance, functional decline in the pool of HSCs leads to development of hematopoietic diseases, such as anemia, cancers of blood and immune system or bone marrow failure (1).

Understanding the functioning and regulation of HSCs is important for multiple reasons. First,

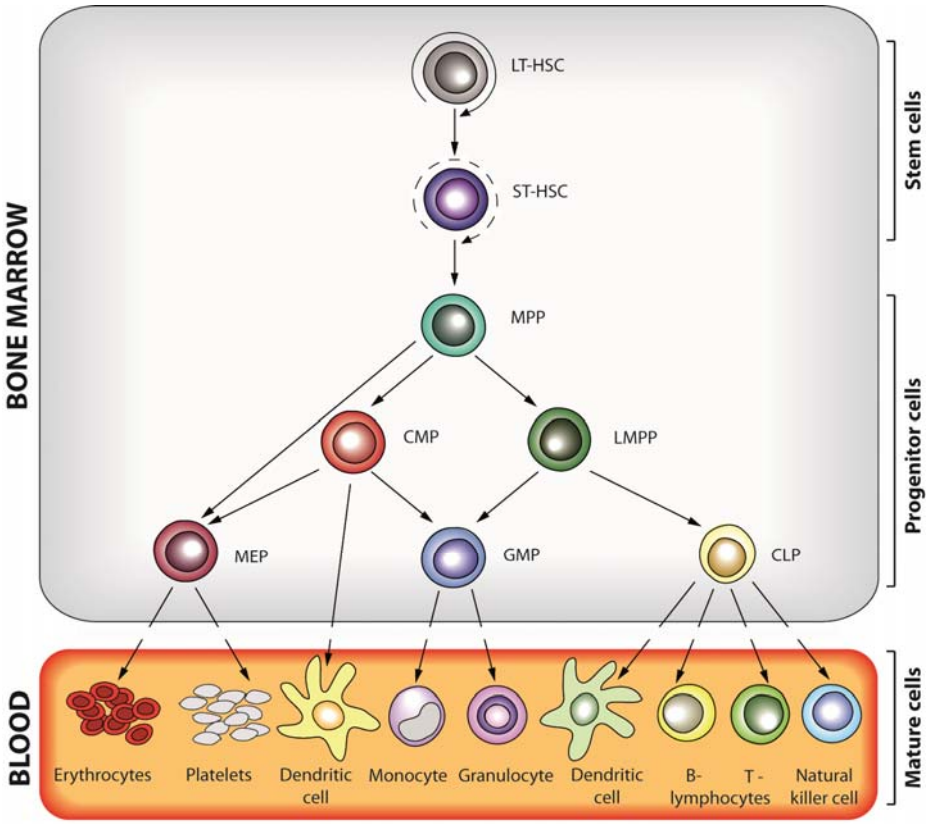

Figure 1. Traditional view on hematopoietic hierarchy. Long-term HSCs (LT-HSCs) reside on the top of hematopoietic hierarchy. They are capable of self-renewal (indicated with round arrows) and differentiation into all immune and blood cells. These properties are gradually lost upon progression down the hematopoietic tree. During this process the cells become progressively more committed towards either myeloid (left side) or lymphoid lineage (right side). Some of the progenitor stages are indicated and are as follows: ST-HSC - short-term HSCs; MPP – multipotent progenitor; CMP – common myeloid progenitor; LMPP – lymphoid primed multipotent progenitor; MEP – megakaryocyte- erythroid progenitor; GMP – granulocyte-macrophage progenitor; CLP – common lymphoid progenitor.

HSCs often serve as an example for other adult stem cells (2). Currently, it is commonly appreciated that the maintenance and repair of most adult organs is supported by presence of a small population of tissue-specific stem cells. Already in the beginning of 1960s, it has been shown that a population of cells within the bone marrow possesses abilities for self-renewal and differentiation into multiple lineages (3). This discovery paved the way to the regenerative medicine field and shaped basic concepts of stem cell biology. Second, changes in HSC pool underlie development of certain hematopoietic diseases and alteration of the hematopoietic function in aging, for instance, decline of immune system and increased rate of myeloid malignancies. Third, HSCs are the first adult stem cells used in clinics. The first bone marrow transplantation was performed in 1958 for treatment of leukemia (4). Currently, bone marrow transplantations are widely used for treatment of hematopoietic and non-hematopoietic disorders. HSCs have been the most extensively studied in mice, allowing elucidation of multiple mechanisms that has been prospectively confirmed in humans.

Activity of HSCs is regulated both by intrinsic (cell specific) and extrinsic factors from the niche - specialized bone marrow environment where the HSCs reside. Several parameters are often used to characterize the functionality of HSCs (Figure 2):

1) Repopulating ability – how much of mature cells can a stem cell produce upon transplantation into recipient animal. The host's own hematopoietic system is usually ablated by irradiation.
2) Turnover rate – division rate of an HSCs.
3) Developmental potential – ability of an HSC to produce mature cells of different hematopoietic lineages.
4) Homing ability – describes whether a cell is capable to engraft the bone marrow niche upon transplantation. This feature is utilized in the clinics and allows to repopulate bone marrow by injecting HSCs into systemic circulation.
5) Mobilization capacity – property complementary (opposite) to homing/engraftment, characterizes an ability of a cell to leave the bone marrow niche and to enter blood stream. Cytokine-induced mobilization is used for harvesting donor HSCs for clinical transplantations.

Some authors presumed that the HSC pool was homogeneous and *bona fide* stem cell can be isolated based on a set of surface marker expression (2). With development of HSC purification methods over the last decade, a number of research laboratories employed single-cell transplantation to demonstrate a profound degree of heterogeneity of various features of individual HSCs originating from a single animal (Figure 2) (5,6). For instance, contribution of a single HSC towards production of myeloid and lymphoid cells can be very different (5,7,8). Discovery of these features require understanding of how different HSC types can interact, and

how changing of the balance between different subtypes, rather than changes in a population of similar cells, can contribute to pathogenesis and aging.

In the hematopoietic system, aging is associated with the development of immunodeficiencies, anemias and an elevated risk of malignancies (1). A growing body of data suggests that these phenotypes result from changes in the pool of HSCs. The hallmarks of aging of HSCs include: 1) increased HSC pool size (7,16-18); 2) skewing of hematopoiesis towards the myeloid lineage and an increase of myeloid-dominant HSCs in the HSC pool (7,17-20); and 3) decreased repopulation potential per HSC (7,16,17,21). These features are schematically represented in Figure 3 (7,12).

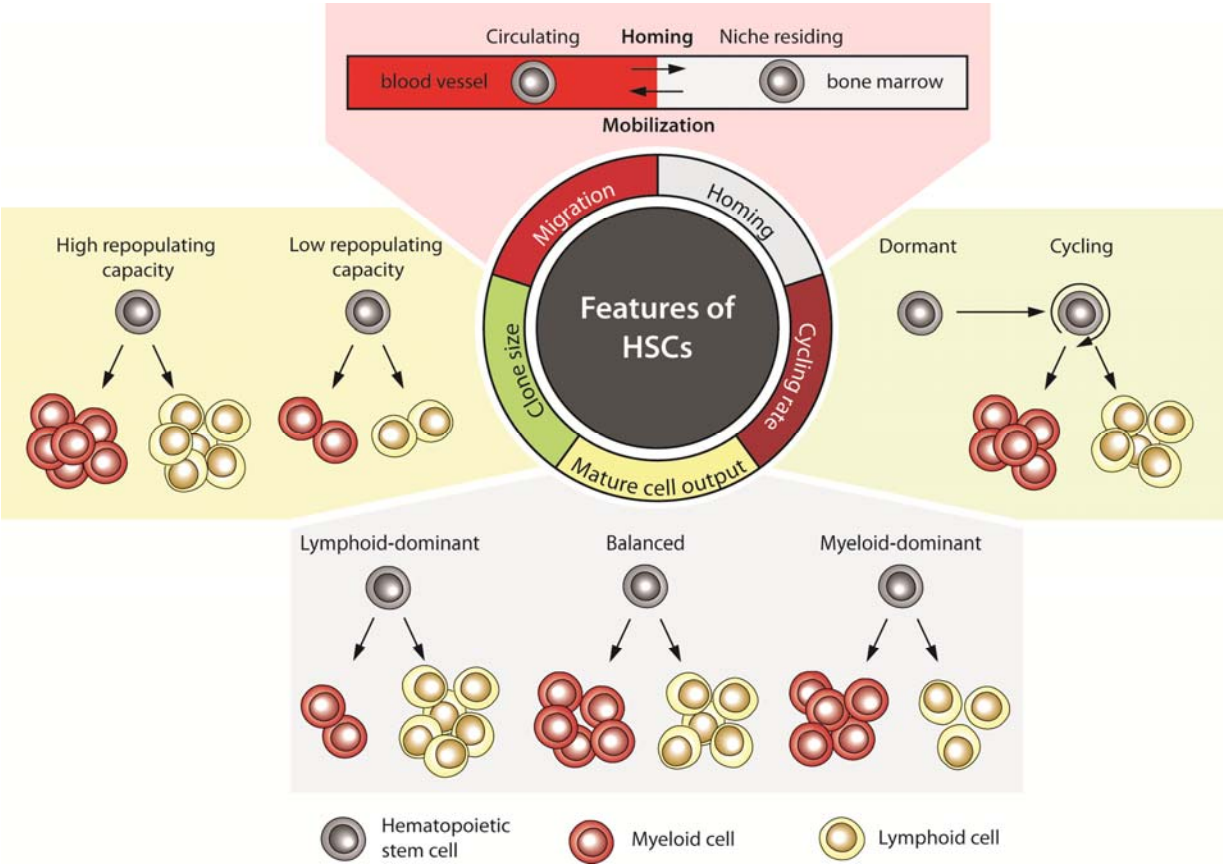

Figure 2. Properties of HSCs and heterogeneity of their manifestation. Mature cell output of HSCs describes how many of blood cells of a certain lineage is produced. Studies over the last decade demonstrate that ability of HSCs to produce cells of lymphoid and myeloid lineage can be remarkably biased (5,9). HSCs also have different repopulation capacity, which results in different sizes of clone that is formed (6). Based on their cycling behavior, HSCs can be classified as quiescent (dormant) and homeostatic (actively cycling). Dormant HSCs are dividing once in several months, while homeostatic HSCs – once in several weeks (10,11). Upon stimulation (e.g. with mobilizing cytokines), HSCs are capable of leaving the bone marrow niche and entering the circulation. The reverse process of entering bone marrow niche from blood is called homing.

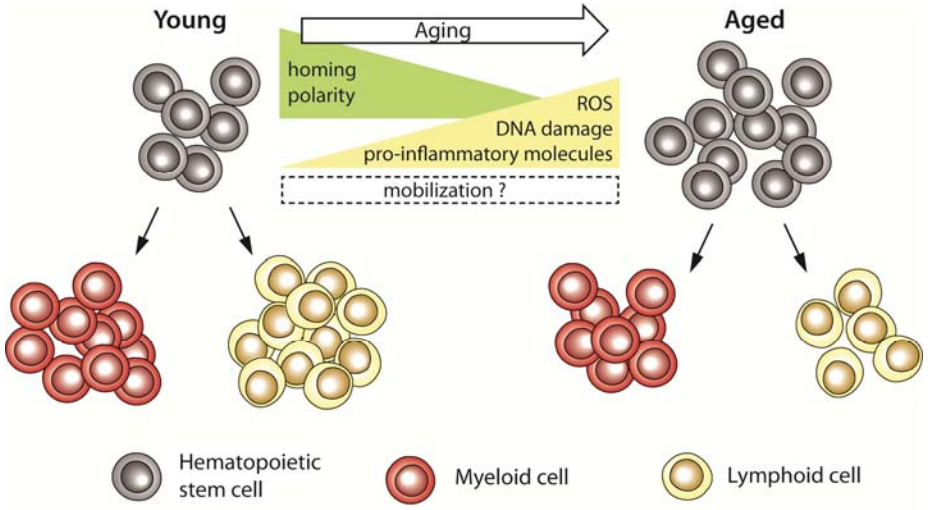

Figure 3. Changes in HSC pool with aging. During aging, the number of phenotypically-defined HSCs increases dramatically, whereas both the frequency of functional HSCs within the putative pool and the repopulation capacity per HSC decreases (12). Production of mature cells becomes skewed towards the myeloid lineage. Ability of HSCs to home bone marrow niches declines. Studies on changes in mobilization upon aging are scarce and come to opposite conclusions (13-15). Increased levels of DNA damage, ROS, pro-inflammatory molecules and loss of polarization of distribution of certain proteins in the cell have also been described (12).

Additionally, HSCs from old mice demonstrate a reduced ability to home and engraft bone marrow niches, and show alternated mobilization in response to cytokines (7,14,22). Although the phenotype of aged HSCs has now been well described, our view on the underlying molecular mechanisms is not yet well developed. Transcriptome profiling data show that HSC aging has been associated with an upregulation of myeloid and a concomitant downregulation of lymphoid genes, increased production of reactive oxygen species (23), elevated levels of DNA damage and an activation of pro-inflammatory molecules (24-26). Recently, loss of cell polarity, modulated by the small Rho GTPase Cdc42, in old HSCs has been described as a novel aging mechanism (27).

At the same time, despite of over 5 decades of research on HSCs, many old fundamental questions of HSC biology are still a matter of discussion. For instance, while thousands of HSCs reside in the bone marrow, there is no general consensus on the number of HSCs that are simultaneously participating in blood production at a moment of time (28). Similarly, it is unknown how HSCs regulate their cycling rate (28). While some authors suggest that contribution of HSCs to blood is stable over time (29,30), others propose that HSCs can exist in two stages – quiescent (dormant) and homeostatic (actively cycling) (10,11). To address these questions reliably, it is necessary to employ a method that will allow quantitative high-resolution analysis of contributions of high numbers of HSCs simultaneously (28).

To study behavior of an individual HSC, one has to be able to distinguish this cell and all of its progeny (HSC clone) from all other cells (31). Several groups of methods are used to study clonal behavior of HSCs and other stem cell types (reviewed in (28,32-34)). For clonal discrimination, these techniques exploit naturally occurring or introduced genetic heterogeneity or are based on single-cell analysis.

A number of clonal assays take advantage of naturally occurring mutations, chromosomal abnormalities, copy number variations, and rearrangements and are mostly used for analysis of hematopoietic malignancies (32,34). While these methods don't require external manipulation, they are limited to retrospective analysis and contributions of different clones are hard to quantify (32).

The majority of clonal tracking approaches to study the normal HSC behavior in mice are based on transplantation of purified single HSCs or individually marked HSCs. Single-cell transplants do not require discrimination of clones, but demand large number of mice to achieve sufficient numbers and preclude analysis in polyclonal situation (5,6). On contrary, viral marking of hematopoietic cells allows transplanting and simultaneously following large number of clones. Retroviral transduction results in semi-random insertion of viral DNA into mouse genome, which can be consequently retrieved by integration site analysis. The major drawback of these techniques is their dependency on genome restriction (insertions outside of proximity to restriction sites are missed) and quantitative bias in fragment amplification (larger fragments are amplified less efficiently) (28).

Alternatively, multi-color marking can be used to discriminate (populations of) HSCs in a single animal. Such mice can be initiated either by transduction of HSCs with different colored vectors (35-37), co-transplantation of HSCs from genetically modified mice (Dykstra et al., unpublished data), or by recombination (38,39). These methods have been useful for *in situ* analysis, however, the number of (distinguishable) colors is limited, and therefore truly clonal analysis of blood is challenging.

To summarize, performing quantitative high-resolution analysis of in polyclonal situation is challenging due to limitations of technology (28). Recently, to circumvent the drawbacks of available approaches, cellular barcoding method has been introduced for analysis of polyclonal populations in hematopoietic system (40,41). This technique is based on insertion of a random sequence fragment into the viral vector (Figure 4). Upon transduction, barcoded vector introduces a unique tag into the genome of target cells that gets inherited by all its progeny. Currently, the barcodes are detected by sequencing (41-46).

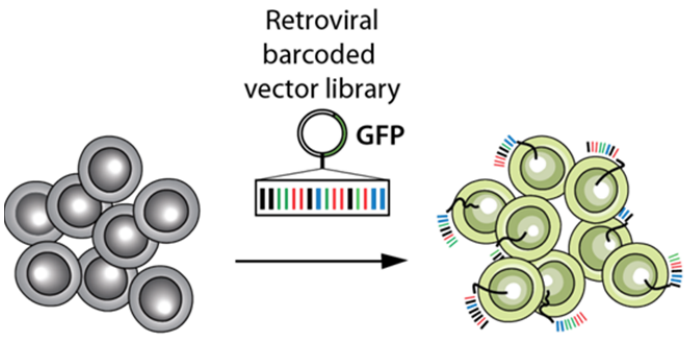

Figure 4. Cellular barcoding of HSCs. In a naïve mouse, all HSCs are genetically identical (left). Transducing HSCs with barcoded vector library introduced unique genetic label into genome of each HSC (right). Barcoded vectors used in our lab carry green fluorescent protein (GFP) tag that allows to distinguish transduced cells.

The work described in this thesis employs cellular barcoding to characterize clonal behavior of young and aged HSCs.

Chapter 2 reviews the literature focused on heterogeneity in the HSC pool and describes the clinical relevance of changes in representation of different HSC subsets.

Chapter 3 discusses the long standing controversy regarding the number of HSCs participating in blood production over the lifespan. It further illustrates how technology used for assessment can affect the conclusions of clonal analysis in hematopoietic system. It also introduces cellular barcoding as a method for unbiased clonal analysis and specifies criteria for successful application of the barcoding technique.

Chapter 4 provides experimental validation of barcoding technique. Further, barcoding method is used to study short-term and long-term clonal fluctuations in polyclonal murine hematopoiesis. At last, it demonstrates HSC pool size changes occurring during aging.

Chapter 5 demonstrates that both young and old HSC clones are asymmetrically distributed among different skeletal niches and can be readily distributed upon administration of mobilizing cytokines. It also describes differences between HSC pool composition and blood in transplant recipient.

Chapter 6 summarizes the main findings of this thesis, puts them in general perspective and discusses their potential impact.

REFERENCES

(1) Balducci L, Ershler W, Gaetano G editors. Blood Disorders in the Elderly. New York: Cambridge University Press; 2008.

(2) Bryder D, Rossi DJ, Weissman IL. Hematopoietic stem cells: the paradigmatic tissue-specific stem cell. Am J Pathol 2006 Aug;169(2):338-346.

(3) Till JE, McCulloch EA. A direct measurement of the radiation sensitivity of normal mouse bone marrow cells. 1961. Radiat Res 2011 Feb;175(2):145-149.

(4) Thomas ED, Lochte HL,Jr, Cannon JH, Sahler OD, Ferrebee JW. Supralethal whole body irradiation and isologous marrow transplantation in man. J Clin Invest 1959 Oct;38:1709-1716.

(5) Dykstra B, Kent D, Bowie M, McCaffrey L, Hamilton M, Lyons K, et al. Long-term propagation of distinct hematopoietic differentiation programs in vivo. Cell Stem Cell 2007 Aug 16;1(2):218-229.

(6) Ema H, Sudo K, Seita J, Matsubara A, Morita Y, Osawa M, et al. Quantification of self-renewal capacity in single hematopoietic stem cells from normal and Lnk-deficient mice. Dev Cell 2005 Jun;8(6):907-914.

(7) Dykstra B, Olthof S, Schreuder J, Ritsema M, de Haan G. Clonal analysis reveals multiple functional defects of aged murine hematopoietic stem cells. J Exp Med 2011 Dec 19;208(13):2691-2703.

(8) Benz C, Copley MR, Kent DG, Wohrer S, Cortes A, Aghaeepour N, et al. Hematopoietic stem cell subtypes expand differentially during development and display distinct lymphopoietic programs. Cell Stem Cell 2012 Mar 2;10(3):273-283.

(9) Muller-Sieburg CE, Sieburg HB. Clonal diversity of the stem cell compartment. Curr Opin Hematol 2006 Jul;13(4):243-248.

(10) Wilson A, Laurenti E, Oser G, van der Wath RC, Blanco-Bose W, Jaworski M, et al. Hematopoietic stem cells reversibly switch from dormancy to self-renewal during homeostasis and repair. Cell 2008 Dec 12;135(6):1118-1129.

(11) Foudi A, Hochedlinger K, Van Buren D, Schindler JW, Jaenisch R, Carey V, et al. Analysis of histone 2B-GFP retention reveals slowly cycling hematopoietic stem cells. Nat Biotechnol 2009 Jan;27(1):84-90.

(12) Geiger H, de Haan G, Florian MC. The ageing haematopoietic stem cell compartment. Nat Rev mmunol 2013 May;13(5):376-389.

(13) Pozotrigo M, Adel N, Landau H, Lesokhin A, Lendvai N, Chung DJ, et al. Factors impacting stem cell mobilization failure rate and efficiency in multiple myeloma in the era of novel therapies: experience at Memorial Sloan Kettering Cancer Center. Bone Marrow Transplant 2013 Aug;48(8):1033-1039.

(14) Xing Z, Ryan MA, Daria D, Nattamai KJ, Van Zant G, Wang L, et al. Increased hematopoietic stem cell mobilization in aged mice. Blood 2006 Oct 1;108(7):2190-2197.

(15) Morris CL, Siegel E, Barlogie B, Cottler-Fox M, Lin P, Fassas A, et al. Mobilization of CD34+ cells in elderly patients (>/= 70 years) with multiple myeloma: influence of age, prior therapy, platelet count and mobilization regimen. Br J Haematol 2003 Feb;120(3):413-423.

(16) Rossi DJ, Bryder D, Zahn JM, Ahlenius H, Sonu R, Wagers AJ, et al. Cell intrinsic alterations underlie hematopoietic stem cell aging. Proc Natl Acad Sci U S A 2005 Jun 28;102(26):9194-9199.

(17) Morrison SJ, Wandycz AM, Akashi K, Globerson A, Weissman IL. The aging of hematopoietic stem cells. Nat Med 1996 Sep;2(9):1011-1016.

(18) Sudo K, Ema H, Morita Y, Nakauchi H. Age-associated characteristics of murine hematopoietic stem cells. J Exp Med 2000 Nov 6;192(9):1273-1280.

(19) Challen GA, Boles NC, Chambers SM, Goodell MA. Distinct hematopoietic stem cell subtypes are differentially regulated by TGF-beta1. Cell Stem Cell 2010 Mar 5;6(3):265-278.

(20) Cho RH, Sieburg HB, Muller-Sieburg CE. A new mechanism for the aging of hematopoietic stem cells: aging changes the clonal composition of the stem cell compartment but not individual stem cells. Blood 2008 Jun 15;111(12):5553-5561.

(21) Yilmaz OH, Kiel MJ, Morrison SJ. SLAM family markers are conserved among hematopoietic stem cells from old and reconstituted mice and markedly increase their purity. Blood 2006 Feb 1;107(3):924-930.

(22) Liang Y, Van Zant G, Szilvassy SJ. Effects of aging on the homing and engraftment of murine hematopoietic stem and progenitor cells. Blood 2005 Aug 15;106(4):1479-1487.

(23) Jang YY, Sharkis SJ. A low level of reactive oxygen species selects for primitive hematopoietic stem cells that may reside in the low-oxygenic niche. Blood 2007 Oct 15;110(8):3056-3063.

(24) Chambers SM, Shaw CA, Gatza C, Fisk CJ, Donehower LA, Goodell MA. Aging hematopoietic stem cells decline in function and exhibit epigenetic dysregulation. PLoS Biol 2007 Aug;5(8):e201.
(25) Ershler WB, Keller ET. Age-associated increased interleukin-6 gene expression, late-life diseases, and frailty. Annu Rev Med 2000;51:245-270.
(26) Bruunsgaard H, Pedersen AN, Schroll M, Skinhoj P, Pedersen BK. Impaired production of proinflammatory cytokines in response to lipopolysaccharide (LPS) stimulation in elderly humans. Clin Exp Immunol 1999 Nov;118(2):235-241.
(27) Florian MC, Dorr K, Niebel A, Daria D, Schrezenmeier H, Rojewski M, et al. Cdc42 activity regulates hematopoietic stem cell aging and rejuvenation. Cell Stem Cell 2012 May 4;10(5):520-530.
(28) Bystrykh LV, Verovskaya E, Zwart E, Broekhuis M, de Haan G. Counting stem cells: methodological constraints. Nat Methods 2012 May 30;9(6):567-574.
(29) Harrison DE, Astle CM, Lerner C. Number and continuous proliferative pattern of transplanted primitive immunohematopoietic stem cells. Proc Natl Acad Sci U S A 1988 Feb;85(3):822-826.
(30) Jordan CT, Lemischka IR. Clonal and systemic analysis of long-term hematopoiesis in the mouse. Genes Dev 1990 Feb;4(2):220-232.
(31) Glauche I, Bystrykh L, Eaves C, Roeder I, other participants. Stem cell clonality -- theoretical concepts, experimental techniques, and clinical challenges. Blood Cells Mol Dis 2013 Apr;50(4):232-240.
(32) Hope K, Bhatia M. Clonal interrogation of stem cells. Nat Methods 2011 Apr;8(4 Suppl):S36-40.
(33) Buckingham ME, Meilhac SM. Tracing cells for tracking cell lineage and clonal behavior. Dev Cell 2011 Sep 13;21(3):394-409.
(34) Gilliland DG, Blanchard KL, Bunn HF. Clonality in acquired hematologic disorders. Annu Rev Med 1991;42:491-506.
(35) Cornils K, Thielecke L, Huser S, Forgber M, Thomaschewski M, Kleist N, et al. Multiplexing clonality: combining RGB marking and genetic barcoding. Nucleic Acids Res 2014 Jan 29.
(36) Weber K, Thomaschewski M, Warlich M, Volz T, Cornils K, Niebuhr B, et al. RGB marking facilitates multicolor clonal cell tracking. Nat Med 2011 Apr;17(4):504-509.
(37) Malide D, Metais JY, Dunbar CE. Dynamic clonal analysis of murine hematopoietic stem and progenitor cells marked by 5 fluorescent proteins using confocal and multiphoton microscopy. Blood 2012 Dec 20;120(26):e105-16.
(38) Wang L, Benedito R, Bixel MG, Zeuschner D, Stehling M, Savendahl L, et al. Identification of a clonally expanding haematopoietic compartment in bone marrow. EMBO J 2013 Jan 23;32(2):219-230.
(39) Livet J, Weissman TA, Kang H, Draft RW, Lu J, Bennis RA, et al. Transgenic strategies for combinatorial expression of fluorescent proteins in the nervous system. Nature 2007 Nov 1;450(7166):56-62.
(40) Schepers K, Swart E, van Heijst JW, Gerlach C, Castrucci M, Sie D, et al. Dissecting T cell lineage relationships by cellular barcoding. J Exp Med 2008 Sep 29;205(10):2309-2318.
(41) Gerrits A, Dykstra B, Kalmykowa OJ, Klauke K, Verovskaya E, Broekhuis MJ, et al. Cellular barcoding tool for clonal analysis in the hematopoietic system. Blood 2010 Apr 1;115(13):2610-2618.
(42) Naik SH, Perie L, Swart E, Gerlach C, van Rooij N, de Boer RJ, et al. Diverse and heritable lineage imprinting of early haematopoietic progenitors. Nature 2013 Apr 11;496(7444):229-232.
(43) Lu R, Neff NF, Quake SR, Weissman IL. Tracking single hematopoietic stem cells in vivo using high-throughput sequencing in conjunction with viral genetic barcoding. Nat Biotechnol 2011 Oct 2;29(10):928-933.
(44) Verovskaya E, Broekhuis MJ, Zwart E, Ritsema M, van Os R, de Haan G, et al. Heterogeneity of young and aged murine hematopoietic stem cells revealed by quantitative clonal analysis using cellular barcoding. Blood 2013 May 29.
(45) Grosselin J, Sii-Felice K, Payen E, Chretien S, Roux DT, Leboulch P. Arrayed lentiviral barcoding for quantification analysis of hematopoietic dynamics. Stem Cells 2013 Apr 4.
(46) Cheung AM, Nguyen LV, Carles A, Beer P, Miller PH, Knapp DJ, et al. Analysis of the clonal growth and differentiation dynamics of primitive barcoded human cord blood cells in NSG mice. Blood 2013 Sep 12.

CHAPTER **2**

THE POWER OF DIVERSITY: HEMATOPOIETIC STEM CELL HETEROGENEITY AND ITS CLINICAL RELEVANCE

Verovskaya, E. and de Haan, G.

Hematology Education, 5(1) (2011).

ABSTRACT

Multiple subsets of functionally distinct hematopoietic stem cells (HSCs) have recently been identified in murine studies. Individual HSCs can differ in their potential to generate certain mature blood cell lineages, repopulating capacity, and turnover rate. Changes in the clonal representation of distinct HSC subsets in the bone marrow are associated with aging and certain types of hematopoietic disorders. Therefore, understanding and addressing HSC heterogeneity is of major clinical importance and may lead to the development of superior treatment protocols. For instance, HSC transplantations should ensure re-establishment of a heterogeneous HSC pool in recipient patients. Although heterogeneity in the stem cell pool has been studied less extensively in humans than in mice, clonal tracking in recent gene therapy trials serves as an important source of information on stem cell behavior. Development of novel barcoding and deep sequencing technologies can provide more detailed and quantitative data on stem cell clonality.

INTRODUCTION

More than 10^{11} blood cells are produced daily in the process of hematopoiesis (1). The turnover of these cells is assured by the enormous proliferative capacity of hematopoietic stem cells (HSCs) residing in the bone marrow. HSCs are capable of generating and regenerating at least nine distinct types of mature blood cells in the complex process of differentiation (Figure 1). The current prevailing view implies that during hematopoiesis, HSCs undergo several unidirectional division steps in which they generate distinct intermediate cell populations (2) and gradually lose their ability to self-renew. The ability for self-renewal and multilineage differentiation is the unique feature of HSCs.

Traditionally, it has been assumed that all stem cells possess similar developmental potential and are equally capable of producing cells of all lineages (Figure 1). Multiple strategies were developed that were aimed to purify a homogeneous stem cell population, where each and every single stem cell possessed equal potential. However, the notion of stem cell homogeneity was challenged by data indicating a high level of heterogeneity of murine HSC in terms of differentiation program (3,4), repopulating capacity (5,6), and turnover rate (7,8). Although much effort is put in uncovering improved ways to characterize and (prospectively) isolate the putative hematopoietic stem cell, a growing body of data points out that THE hematopoietic stem cell does not exist. In contrast, intrinsic heterogeneity may be a key characteristic of HSCs.

Existence of stem cells with different behavior provides a new perspective on current views of normal and pathological hematopoiesis. Recent data on HSC heterogeneity suggests that the equilibrium of different stem cell subsets can be crucial in maintaining blood homeostasis, while the clonal disbalance of HSC underlies abated hematopoietic function in aging and disease. Although evidence for intrinsic heterogeneity of the HSC compartment originates from studies in mice, there is no reason to assume that a similar extent of heterogeneity is absent in human HSCs. Understanding the facets of HSC heterogeneity in humans is important to ensure both the current and future success of stem cell therapies and for the apprehension of blood disorders, such as various types of leukemias and other proliferative and dysplastic bone marrow diseases. The unique opportunity to track individually marked human HSCs provided by gene therapy trials, combined with rapid development of high-throughput sequencing analysis, holds great promise to obtain insight in such cellular heterogeneity. In the current review, we will address recent advances in our understanding of HSC heterogeneity and will highlight putative clinical consequences.

HSC HETEROGENEITY

Stem cells comprise less than 0.01% of all bone marrow cells and are morphologically very difficult to distinguish from progenitor cells. HSCs can be partly characterized by

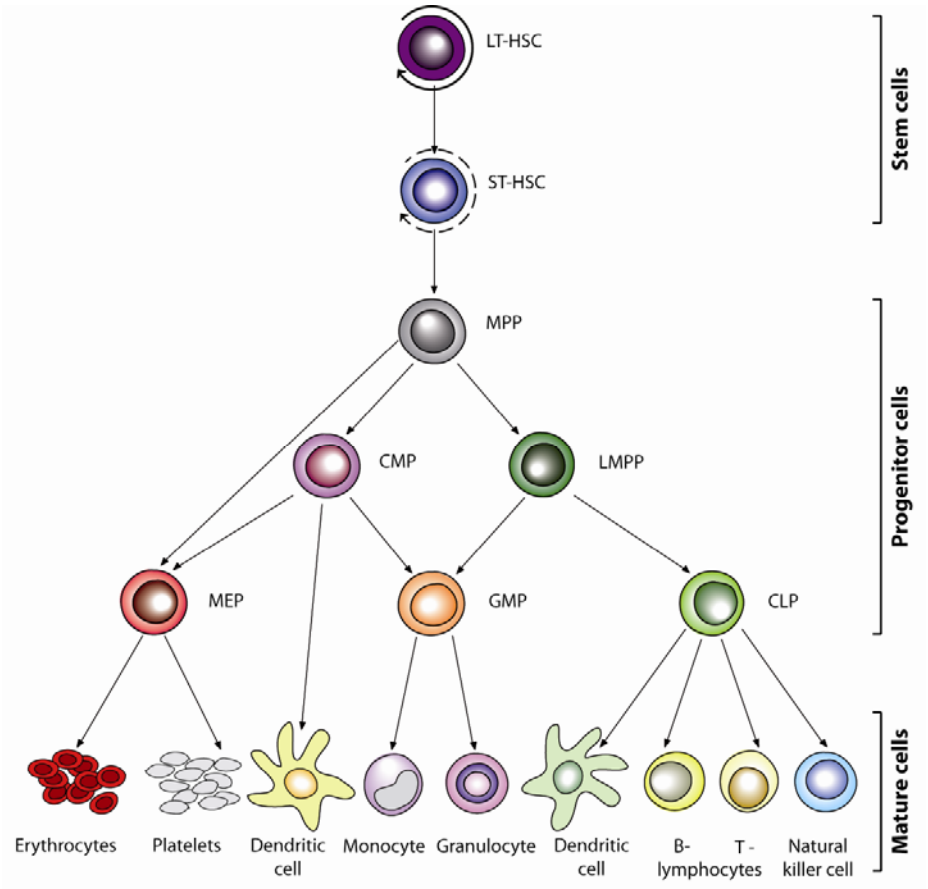

Figure 1. Traditional model of differentiation and commitment of hematopoietic stem cells. Long-term (LT-HSC) and short-term (ST-HSC) HSCs reside at the top of hematopoietic hierarchy. Both these populations have multilineage developmental potential, but ST-HSCs have lower self-renewal capacity (75). Upon differentiation, HSCs are equally capable of producing all mature blood lineages via a chain of progenitor populations with restricted lineage commitment. MPP - multipotent progenitor, CMP - common myeloid progenitor, LMPP - lymphoid-primed multipotent progenitor, MEP - megakaryocyte-erythroid progenitor, GMP - granulocyte-macrophage progenitor, CLP - common lymphoid progenitor. Self-renewal is indicated in the round arrows.

immunophenotype, but unlike, for instance, the situation in the intestine, where Lgr5 can be used as a single stem cell marker (9); such mono-markers have neither been found in the hematopoietic system (10), nor in most other tissues. In a young mouse, differential expression of several surface molecules allows isolation of populations containing up to 50% of long-term repopulating HSCs (11–13). In the human system, the knowledge of HSC surface proteins is much more limited, and typically only a few markers are used in purification protocols. These include lineage negative, CD34, CD90/Thy1 (14), CD38 (15) and CD45RA (16).

The most stringent way to demonstrate the “stemness” of a cell is to test its function by transplantation into a recipient in which endogenous blood cell formation is destroyed or

otherwise impaired. Several criteria, based on either long-term repopulating ability, self-renewal capacity in serial transplantation, or multilineage differentiation potential, were proposed to define a stem cell. In the mouse system, a very common definition of a HSC requires a cell to be able to contribute at least 1% of chimerism in both myeloid and lymphoid lineages for 16 weeks post transplantation into a recipient animal (6,17). Yet, several substantially different populations of stem cells fit this definition. In a series of *in vivo* transplantation studies, a spectrum of different subcategories in the HSC pool were uncovered by scrutinizing three key properties of HSCs – multilineage developmental capacity, proliferative potential, and self-renewal ability.

The first insights into differences of clonal behavior of distinct HSC were provided by experiments on retrovirally transduced stem cells that demonstrated a high degree of diversity in the HSC pool. Contributions of individual HSC clones to myeloid and lymphoid lineages and their engraftment kinetics were strikingly different (5,18). Since then, the development of murine HSC enrichment methods, allowing isolation of stem cells with sufficient purity (19), has propelled single-cell HSC studies. It was demonstrated that individual HSCs possess markedly diverse self-renewal capacity and repopulating activity upon transplantation into irradiated host mice (6,20). For instance, in competitive repopulation assays, single purified cells could contribute from as low as 2.5% to as high as 63% of white blood cell chimerism in recipient mice (6). Interestingly, the size of the HSC clone was correlated with the ability of an HSC for self-renewal. The cells with the highest repopulating activity were also more likely to produce transplantable progeny in secondary transplantations (6). However, the ability to produce cells of myeloid, T- or B-lymphoid lineages varied significantly, even between cells with similar repopulating capacity.

Notably, this fundamental feature allowed classification of each HSC: data on both limiting dilution (3) and single-cell transplantations (4) have shown existence of at least three stem cell developmental subtypes (Figure 2A). These cells differ in their ability to produce mature cells of the myeloid or lymphoid lineage. The HSCs are myeloid-biased (or alpha), balanced (or beta), and lymphoid-biased (or gamma) stem cells (3,4). This “differentiation program” is inherited by the progeny of the original cells and exhibited upon serial transplantation (3,4) and therefore, strongly suggests epigenetic differences to be the cause of observed variation.

Another level of HSC heterogeneity originates from differences in stem cell turnover rate, which underlies the proliferative capacity of the HSCs (7,8). It has been proposed that a high repopulation potential of HSCs is associated with slow turnover, and that proliferative quiescence protects these cells against mutagenic stress and prevents their exhaustion (20–22). Slow cycling was confirmed by analysis of phenotypically characterized primitive cells (7,23,24).

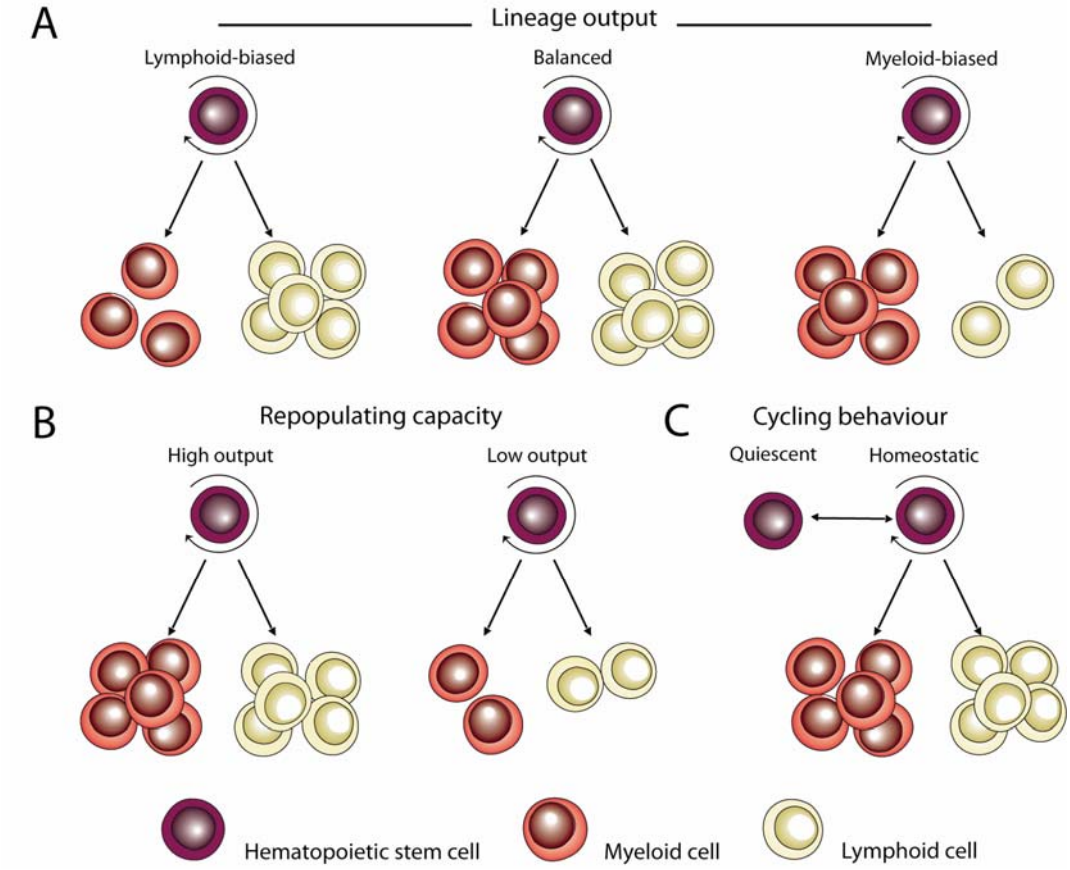

Figure 2. Different levels of heterogeneity in the HSC compartment. Although the behavior of individual cells taken from any of these categories can be very different, certain stem cell qualities are related. For instance, quiescent HSCs are more likely to provide high output in repopulating assay than their counterparts. Myeloid bias, high repopulating capacity, and quiescence have all been associated with high self-renewal (6–8,31).

Seventy percent of most primitive HSCs were found to be in the G_0 stage of the cell cycle, in contrast to less than 10% in more differentiated progenitor cells (7).The concept of dormancy has been discussed for several decades. For example, it has been shown that the most primitive HSCs survive a very high dose of the S-phase specific cytotoxic drug 5-fluorouracil (5-FU) (25) and that cytokines can be used to increase sensitivity of HSCs towards 5-FU (26). However, recently this concept was proven more directly, when two laboratories have combined label-retaining assays, flow cytometry, and mathematical modeling to demonstrate that less than 30% of HSCs are contained in a dormant state (7,8). In these studies, cells were labeled *in vivo* by traceable markers that upon cellular division were diluted and became undetectable. Cells that retained the label after a prolonged period of time represented slowly cycling cells. Measuring percentages of these label-retaining cells (LRCs) within the stem cell population at different time points allowed the authors to deduce the cell cycling history. Mathematical

modeling of these data demonstrated that the stem cell pool was constituted by two cell populations with different cycling times: “dormant” HSCs that divided just once in approximately 170 days and “homeostatic” HSCs, which divided every approximate 30 days (7,27) (Figure 2C).

To test whether “dormant” label-retaining HSCs had a higher proliferative potential than their actively cycling “homeostatic” counterparts, the authors have isolated these HSCs populations by flow cytometry and transplanted them into irradiated recipients (7,8). While all LRCs exhibited ability to repopulate both primary and secondary transplant recipients, only a minor part of non-LRCs were able to do so (7,8). Remarkable differences in multilineage reconstitution capacity confirmed the long-standing hypotheses linking dormancy and proliferative potency.

Although HSC heterogeneity is multidimensional, different aspects of it are correlated (Figure 2). What drives distinct behaviors in individual HSC clones is a topic of frequent discussion.

One of the fundamental questions related to HSC heterogeneity is the role of deterministic, “programmed” decisions and stochastic factors, or biological noise, in stem cell decisions. It is tempting to divide HSC subtypes into binary categories: quiescent/cycling, myeloid-/lymphoid-biased, and so on. However, these states can be temporal and interchangeable for individual cells. Moreover, additional levels of heterogeneity within each of the stem cell subtypes indicate that the HSC pool could be a continuum of cells with a gradient of features. In concordance with this idea, a recent paper from Takizawa et al. (28) provides experimental evidence on the inter-changeability of quiescent and cycling states of HSCs. When fast-cycling cells (more than five divisions over 14-week period in the primary recipients) were re-transplanted into non-irradiated secondary recipients, some of them remained quiescent over the period of 6 weeks, demonstrating that stem cells can regulate their division rate (28). This notion is further supported by mathematical models and simulations. For instance, Glauche et al. have re-analyzed the original data on stem cell quiescence/activation by Wilson et al. (7), taking into consideration individual decisions that single stem cells make (29). The authors have demonstrated that rather than being separated into categories with distinct cell cycle, individual HSCs could adapt their turnover rate in response to demands in blood production (29). Although general changes occurring during aging or in disease are likely to shift HSC pool in a certain direction, some authors propose that it is impossible to predict the behavior of an individual stem cell (30).

HSC heterogeneity may be an important attribute of balanced hematopoiesis. Both data of clonal skewing of lymphomyeloid potential during aging in mice (31–33), and models of clonal expansions preceding leukemia development (34) support this notion. Consequently, HSC transplantations should ensure re-establishment of a heterogeneous HSC pool in the recipient patient.

CLINICAL RELEVANCE OF HSC HETEROGENEITY

Understanding the biology of HSCs is important for the potential development of more efficient treatments to fight human blood diseases and to compensate for aging-related loss of hematopoietic function. Several hematopoietic disorders have been already directly linked to defects in HSCs, and more evidence arises implicating their role in other diseases. Since different subtypes of HSCs exist, expansion or loss of clones of stem cell with distinct properties during the lifespan of the organism can predispose to the development of hematopoietic malignancies (34) and contribute to impaired hematopoietic function in the elderly. Data demonstrating changes of HSC compartment in aging will be summarized below.

Aging of the hematopoietic system

In humans, aging is associated with an elevated occurrence of anemias, a decline in immune response, and an increased frequency of hematopoietic malignancies and certain autoimmune diseases (35). Molecular studies in mice provide cues about how changes in the HSC pool can contribute to impaired hematopoiesis in aging. The hallmarks of this process include skewing of hematopoiesis to produce more myeloid cells at the expense of lymphoid cell production (36,37), elevated HSC numbers (36–39), impaired function per stem cell (36,37,40), and decreased homing efficiencies after transplant (41) (reviewed in (42)).

Hematopoietic aging is orchestrated by the interplay between intrinsic and extrinsic factors, which are likely to be related both to the aging environment (hematopoietic niche) and functional changes in the HSCs. Strikingly, the number of phenotypically characterized HSCs increases through the lifespan of a mouse (36–39). Several studies report a 7- to 16-fold expansion of the HSC pool (42). However, the repopulation potential of individual aged stem cells declines, possibly due to accumulation of DNA damage (43,44). The ability of HSCs to generate cells of different lineages also changes and becomes “skewed” towards myeloid cells, while lymphoid lineage output declines (37,38,41). Recently, it has been demonstrated that this effect is accompanied by clonal expansion of myeloid-biased HSCs in the aging bone marrow (31–33,45). It was proposed that both higher self-renewal potential (31) and unequal cytokine response (32) allow myeloid-primed HSCs to outperform their lymphoid-biased counterparts. Further, a significantly higher proportion of the aged HSCs are cycling (28,37,38,40), in contrast to predominantly dormant young HSCs.

Changes of the HSC pool during aging in humans have been much less studied. In general, human hematopoiesis is associated with decreased bone marrow cellularity during aging (46). As in mice, the frequency of primitive cells identified by immunophenotype ($CD34^+CD38^-$ or $lineage^-CD34^+CD38^-CD90^+$) increases over the lifespan (47–49). Although it is not certain whether the repopulating ability of the individual human HSCs decreases over time, the quality

of the whole bone marrow decays with age. A study of 6978 bone marrow transplantation patients demonstrated that older donor age is significantly associated with lower disease-free survival in transplant recipients (50).

Interestingly, lineage changes in the murine HSC compartment mirror the human aging phenotype, as reflected in immunological decline and increased frequency of myeloid cancers. It is likely that age-dependent changes in the human HSC pool are caused by these clinically important features. Moreover, if the composition of the HSC pool predisposes to hematological disease, one may expect that the same oncogenic mutation will cause a distinct type of disease in old and young patients. This argument is fueled by data indicating that transformation of cells with BCR-ABL, the oncogene often associated with chronic myeloid leukemia (CML), causes myeloproliferative disease (MPD) in old mice, while young mice develop both B-cell leukemia and MPD (51).

Complementary, heterogeneity of leukemias in the elderly population could be a reflection of heterogeneity in underlying HSC pool. The current models of leukemic development in aging presume that over the lifespan, HSCs accumulate mutations and combination of several events by chance could lead to malignant transformation (52). Clonal selection of myeloid-biased HSCs could explain the elevated proportion of AML during aging (52).

Besides a general aging-related clonal shift, hematopoietic injuries by infections or cytotoxic agents could affect the balance of dormant and active cells in the organism.

Loss of HSC quiescence – the evil and the good

Since dormant HSCs are a reservoir preserving most of hematopoietic repopulating activity, clinical implications could arise if this compartment is disrupted. Stem cell proliferative potential is not unlimited, and self-renewal is lost upon long-term divisions. For instance, serial transplantations of HSCs are only possible for four to six rounds (53). It seems plausible that repeated hematopoietic stress will lead to recurrent HSC proliferation, which has been shown to decrease HSC quality (23). As a consequence, certain treatments, such as repeated chemotherapy, are likely to result in activation of dormant HSCs. It remains to be investigated what the long-term effects of this therapy will be. Several recent studies indicated that activation of quiescent stem cells by cytokines associated with viral and bacterial infection rendered them vulnerable to competition with HSCs resistant to stimulation (54,55).

Conversely, in some cases, it could be clinically beneficial to activate quiescent stem cells. Certain types of stress, such as radiation-induced damage, are more detrimental to dormant than to cycling cells, making them prone to apoptosis (56). Pre-stimulation of quiescent HSCs before radiotherapeutic treatment could possibly help to decrease the damage (57). Moreover, use of agents activating HSCs have been proposed to break the resistance of quiescent leukemic

stem cells to chemotherapeutic drugs, and consequently to increase response to chemotherapeutics (reviewed in (10)). The first clinical study employing a combination of G-CSF and imatinib mesolate is underway (58).

Factors mediating the exit of HSCs from the quiescent state are mostly unidentified but are likely to be cytokines and growth factors. Current known signaling molecules include granulocyte colony-stimulating factor (G-CSF) (59), interferon-α (55), and interferon-γ (54). It is also likely that molecules that were used many years ago to explore their potential as radioprotective agent, such as stem cell factor (SCF) (26), interleukin-1 (IL-1) (60), and FLT3-ligand (61) will activate dormant HSCs.

HSC transplantation

Transplantation of HSCs is used for treatment of a range of congenital and acquired diseases, including hematopoietic cancers (leukemias, lymphomas, and myelomas), myelodysplastic and myeloproliferative disorders, anemias, and immunodeficiencies. The studies referred to above that document heterogeneity of bone marrow-derived HSCs allow speculation about the quality of cells derived from different sources (bone marrow, peripheral blood, or cord blood).

It can be expected that cord blood will bear the cells with both the higher proliferative capacity (62) and lymphoid developmental potential than HSCs from mobilized blood or adult bone marrow, and with increased donor age, the HSC pool will lose the ability to generate lymphoid cells and will have an overall reduced repopulating ability. Longer neutrophil and platelet recovery times in recipients of UCB graft compared with peripheral blood HSC or bone marrow HSCs could also reflect more quiescent, slower cycling behavior of UCB stem cells (63). Interestingly, one of HSC activating molecules, G-CSF (59) is routinely used to mobilize HSCs for bone marrow transplantations. It still remains to be investigated whether heterogeneity within the mobilized HSC pool in the blood reflects the bone marrow situation and how this could affect recipients of the HSC transplants.

CLONAL BEHAVIOR OF HUMAN HEMATOPOIETIC CELLS – LESSONS FROM GENE THERAPY

Correction of genetic defects by introducing a therapeutic gene in a patient's own stem cells – gene therapy – is one of the most promising applications of stem cells. HSC gene therapy has already been proven successful for correction of both hematopoietic and non-hematopoietic diseases, including X-linked severe-combined immunodeficiency (SCID-X) (64), β-thalassemia (65), chronic granulomatous disease (CGD) (66), adenosine deaminase deficiency (67), and adrenoleukodystrophy (68). The essence of the treatment is the isolation of patients' own HSCs, which are corrected by (retroviral) gene transfer ex vivo and infused in the blood stream.

However, the wide-spread usage of this technique has been hampered by the observation of a high incidence of leukemia in a severe combined immunodeficiency gene therapy trial – 4 of 10 patients in a French SCID-X study developed T-cell acute lymphoid leukemia (T-ALL) (64). One additional case of leukemia was registered in a similar trial of 10 patients in the United Kingdom (69,70). Leukemia development was associated with activation of pro-oncogenes (including LMO2 and CCND2) after integration of the virus that was used for gene delivery in close proximity to these genes, resulting in expansion of clones marked by these integration sites. Similarly, clonal dominance preceded the development of myelodysplasia and leukemia in two patients with X-linked CDG treated with gene therapy (66). Clonal restriction was triggered by retroviral integration in proximity to EVI1 and MDS-EVI1 loci (66).

The development of safer therapeutic vectors and the establishment of methods to control dynamics of retrovirally marked clones was a prerequisite for the restart of gene therapy trials. Several methods for tracking individual HSC clones by analysis of integration sites have been reported. The most advanced methods combine detection of insertion sites with high-throughput sequencing. Integration site analysis is based on the cutting of genomic DNA isolated from transduced cells by restriction enzymes. Subsequently, the fragments containing parts of viral sequences are amplified by polymerase chain reaction (PCR) and analyzed by sequencing, allowing identification of viral insertion site. If the relative contribution of an integration site over time increases, such clonal expansion can mark development of a myeloproliferative disease. Coupling of integration site analysis with high-throughput sequencing allows simultaneous detection of multiple clones in a single sample. For instance, a longitudinal study of clonal fluctuations in the French SCID-X study reported variation in 9767 integration sites in the blood of eight patients (71). However, these data need careful interpretation. Aspects, such as efficiency of integration site recovery, sensitivity of the method, robustness against sequencing noise, and statistical analysis can influence the scope of reported integration sites.

Importantly, clonal tracking in gene therapy studies would provide a unique opportunity to assess the heterogeneity in the human HSC pool. By analyzing the presence of different HSC clones in subtypes of mature blood cells (granulocytes, B and T cells) over time, it would be possible to distinguish different human HSC types. Remarkably, a study of a patient who underwent HSC gene therapy for treatment of β-thalassemia, demonstrated long-term (33 months) labeling of erythroid and myeloid, but not lymphoid cells with vector insertion in the HMGA2 gene, indicating possible gene transfer into myeloid-biased stem cell (65).

Data on clonal changes in the blood of gene therapy patients can be an important source of information on HSC heterogeneity in humans. Multiclonality observed in two patients in X-linked adrenoleukodystrophy trial (68) provides an exciting starting point for observations of

clonal changes. Development of unbiased quantitative methods for clonal tracking is an important factor to ensure reliability of such information. Unfortunately, currently used methods of integration site analysis are non- or semi-quantitative, and integration site discovery is biased to the shorter size of DNA fragments (less than 500 DNA base pairs (72)). Furthermore, even extensive analysis fails to detect 30–40% of the clones (73).

Recently, our laboratory developed and validated a novel cellular barcoding technique, which can be used for tracking clonality in transduced cell populations (74). We have shown that in contrast to previously available techniques, the barcoding method provides unbiased, quantitative results. The barcoding technique and its possible application in gene therapy trials will be discussed in the next paragraph.

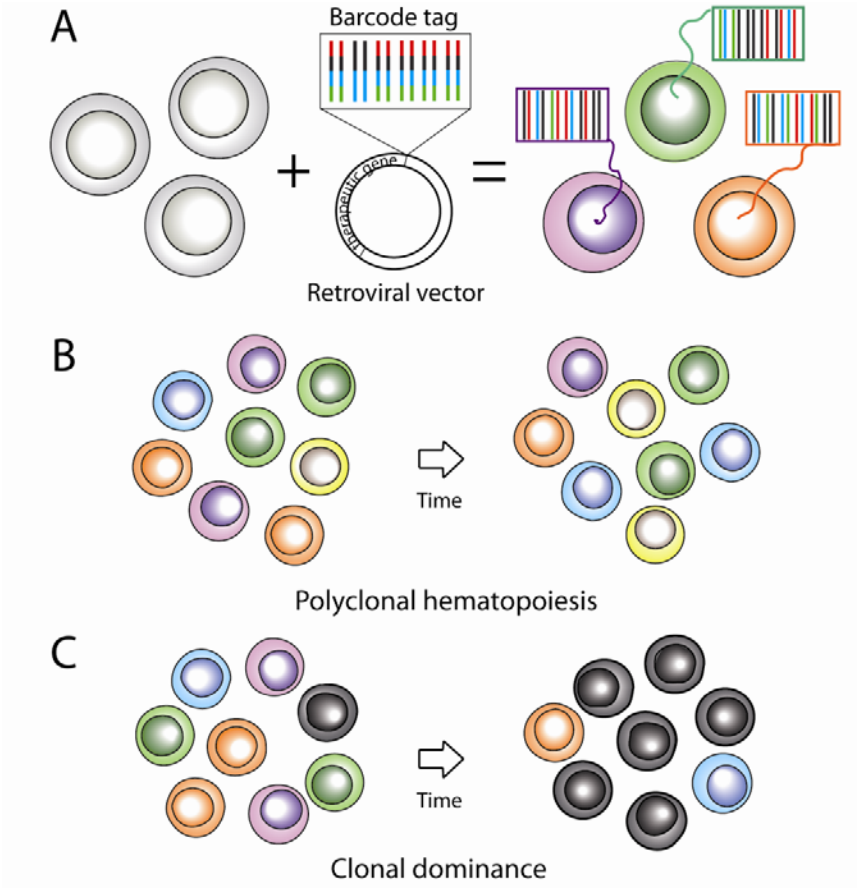

Figure 3. Use of barcoded vectors for tracking clonal changes in hematopoietic system. Under normal circumstances, HSCs are genetically identical (A, left), therefore, analysis of the output of HSC clones is impossible. Cellular barcoding uniquely labels individual HSCs enabling clonal tracking of their progeni (A). Barcode analysis can be used in gene therapy trials both to study the clonal fluctuations in the polyclonal system (B) and to detect clonal expansion events (C).

CELLULAR BARCODING AS A POWERFUL TOOL TO STUDY STEM CELL CLONALITY

Our cellular barcoding method is based on introduction of a short 27 nucleotide-long random sequence tag (barcode) in the retroviral vector (Figure 3A). Upon viral integration, a barcode provides a heritable traceable label into the genome of transduced cells. Optimal design of the barcode tag allows construction of barcoded vector libraries with complexity up to several million combinations. Although the barcode tag itself is highly variable, the structure of the vector backbone is constant, allowing amplification of equal size barcode fragments with primers against internal virus sequences. After PCR amplification of genomic DNA, barcode tags can be detected by sequencing. In contrast with integration site analysis methods, our barcoding technique permits unbiased analysis of clonal composition in the sample. Simplicity of barcode tag design allows introducing such a tag in virtually all vector systems.

Previously, we have described barcode analysis based on Sanger sequencing. Recently, we coupled cellular barcoding to the Illumina Solexa high-throughput sequencing method, which allowed us to observe fluctuations of more than 100 different clones in transduced bone marrow cultures *in vitro* (Bystrykh, et al., unpublished data, 2010). Tracking clonality of hematopoietic cells by barcoding provides a setup, which can be easily applied in clinical gene therapy studies and would allow extracting data on human HSC behavior with more precision. Both crucial targets of clonal tracking in gene therapy could be addressed: first, resolution and sensitivity of the method allow detailed analysis of the polyclonal hematopoiesis (Figure 3B); second, quantitativeness of barcode detection can be utilized to monitor expansion of (potentially malignant) clones (Figure 3C).

CONCLUSIONS

1. Heterogeneity in the HSC compartment involves all the crucial features of stem cells – Their repopulating ability, self-renewal capacity, cycling time, and multilineage differentiation potential. Maintenance of heterogeneity might be required to ensure tissue homeostasis.
2. Clonal selection is associated with development of hematopoietic disorders and function loss during aging.
   - Hematopoietic aging is accompanied by clonal expansion of myeloid-biased cells.
   - Viral and bacterial infections activate dormant cells.
3. Gene therapy provides a context in which HSC clonal changes in humans can be investigated.

ACKNOWLEDGEMENTS

The authors would like to thank Leonid V. Bystrykh, Brad J. Dykstra, Ronald van Os and Hein Schepers, Department of Cell Biology, Section Stem Cell Biology, University Medical Centre Groningen for critical reading of the manuscript and valuable comments.

FUNDING

This work was supported by a grant of the Netherlands Organization for Scientific Research (TopTalent, 021-002-084 (E.V.), and VICI**,** 918-76-601 (G.d.H.)), and by the European Community (EuroSystem, 200720).

REFERENCES

(1) Lansdorp PM. Stem cell biology for the transfusionist. Vox Sang. 1998;74(Suppl 2):91-4.
(2) Bryder D, Rossi DJ, Weissman IL. Hematopoietic stem cells: the paradigmatic tissue-specific stem cell. Am J Pathol. 2006 Aug;169(2):338-46.
(3) Muller-Sieburg CE, Cho RH, Thoman M, Adkins B, Sieburg HB. Deterministic regulation of hematopoietic stem cell self-renewal and differentiation. Blood. 2002 Aug 15;100(4):1302-9.
(4) Dykstra B, Kent D, Bowie M, McCaffrey L, Hamilton M, Lyons K, et al. Long-term propagation of distinct hematopoietic differentiation programs in vivo. Cell Stem Cell 2007 Aug 16;1(2):218-29.
(5) Jordan CT, Lemischka IR. Clonal and systemic analysis of long-term hematopoiesis in the mouse. Genes Dev. 1990 Feb;4(2):220-32.
(6) Ema H, Sudo K, Seita J, Matsubara A, Morita Y, Osawa M, et al. Quantification of self-renewal capacity in single hematopoietic stem cells from normal and Lnk-deficient mice. Dev Cell. 2005 Jun;8(6):907-14.
(7) Wilson A, Laurenti E, Oser G, van der Wath RC, Blanco-Bose W, Jaworski M, et al. Hematopoietic stem cells reversibly switch from dormancy to self-renewal during homeostasis and repair. Cell. 2008 Dec 12;135(6):1118-29.
(8) Foudi A, Hochedlinger K, Van BD, Schindler JW, Jaenisch R, Carey V, et al. Analysis of histone 2B-GFP retention reveals slowly cycling hematopoietic stem cells. Nat Biotechnol. 2009 Jan;27(1):84-90.
(9) Barker N, van Es JH, Kuipers J, Kujala P, van den BM, Cozijnsen M, et al. Identification of stem cells in small intestine and colon by marker gene Lgr5. Nature. 2007 Oct 25;449(7165):1003-7.
(10) Trumpp A, Essers M, Wilson A. Awakening dormant haematopoietic stem cells. Nat Rev Immunol. 2010 Mar;10(3):201-9.
(11) Kiel MJ, Yilmaz OH, Iwashita T, Yilmaz OH, Terhorst C, Morrison SJ. SLAM family receptors distinguish hematopoietic stem and progenitor cells and reveal endothelial niches for stem cells. Cell. 2005 Jul 1;121(7):1109-21.
(12) Kent DG, Copley MR, Benz C, Wohrer S, Dykstra BJ, Ma E, et al. Prospective isolation and molecular characterization of hematopoietic stem cells with durable self-renewal potential. Blood. 2009 Jun 18;113(25):6342-50.
(13) Challen GA, Boles N, Lin KK, Goodell MA. Mouse hematopoietic stem cell identification and analysis. Cytometry A. 2009 Jan;75(1):14-24.
(14) Baum CM, Weissman IL, Tsukamoto AS, Buckle AM, Peault B. Isolation of a candidate human hematopoietic stem-cell population. Proc Natl Acad Sci U S A. 1992 Apr 1;89(7):2804-8.
(15) Terstappen LW, Huang S, Safford M, Lansdorp PM, Loken MR. Sequential generations of hematopoietic colonies derived from single nonlineage-committed CD34+. Blood. 1991 Mar 15;77(6):1218-27.
(16) Majeti R, Park CY, Weissman IL. Identification of a hierarchy of multipotent hematopoietic progenitors in human cord blood. Cell Stem Cell. 2007 Dec 13;1(6):635-45.
(17) Miller CL, Eaves CJ. Expansion in vitro of adult murine hematopoietic stem cells with transplantable lympho-myeloid reconstituting ability. Proc Natl Acad Sci U S A. 1997 Dec 9;94(25):13648-53.
(18) Lemischka IR, Raulet DH, Mulligan RC. Developmental potential and dynamic behavior of

hematopoietic stem cells. Cell. 1986 Jun 20;45(6):917-27.
(19) Osawa M, Hanada K, Hamada H, Nakauchi H. Long-term lymphohematopoietic reconstitution by a single CD34-low/negative hematopoietic stem cell. Science. 1996 Jul 12;273(5272):242-5.
(20) Dykstra B, Ramunas J, Kent D, McCaffrey L, Szumsky E, Kelly L, et al. High-resolution video monitoring of hematopoietic stem cells cultured in single-cell arrays identifies new features of self-renewal. Proc Natl Acad Sci U S A. 2006 May 23;103(21):8185-90.
(21) Orford KW, Scadden DT. Deconstructing stem cell self-renewal: genetic insights into cell-cycle regulation. Nat Rev Genet. 2008 Feb;9(2):115-28.
(22) Arai F, Suda T. Maintenance of quiescent hematopoietic stem cells in the osteoblastic niche. Ann N Y Acad Sci. 2007 Jun;1106:41-53.
(23) Passegue E, Wagers AJ, Giuriato S, Anderson WC, Weissman IL. Global analysis of proliferation and cell cycle gene expression in the regulation of hematopoietic stem and progenitor cell fates. J Exp Med. 2005 Dec 5;202(11):1599-611.
(24) Cheshier SH, Morrison SJ, Liao X, Weissman IL. In vivo proliferation and cell cycle kinetics of long-term self-renewing hematopoietic stem cells. Proc Natl Acad Sci U S A. 1999 Mar 16;96(6):3120-5.
(25) van Zant G. Studies of hematopoietic stem cells spared by 5-fluorouracil. J Exp Med. 1984 Mar 1;159(3):679-90.
(26) Down JD, de Haan G, Dillingh JH, Dontje B, Nijhof W. Stem cell factor has contrasting effects in combination with 5-fluorouracil or total-body irradiation on frequencies of different hemopoietic cell subsets and engraftment of transplanted bone marrow. Radiat Res. 1997 Jun;147(6):680-5.
(27) van der Wath RC, Wilson A, Laurenti E, Trumpp A, Lio P. Estimating dormant and active hematopoietic stem cell kinetics through extensive modeling of bromodeoxyuridine label-retaining cell dynamics. PLoS One. 2009;4(9):e6972.
(28) Takizawa H, Regoes RR, Boddupalli CS, Bonhoeffer S, Manz MG. Dynamic variation in cycling of hematopoietic stem cells in steady state and inflammation. J Exp Med. 2011 Feb 14;208(2):273-84.
(29) Glauche I, Moore K, Thielecke L, Horn K, Loeffler M, Roeder I. Stem cell proliferation and quiescence--two sides of the same coin. PLoS Comput Biol. 2009 Jul;5(7):e1000447.
(30) Potten CS, Loeffler M. Stem cells: attributes, cycles, spirals, pitfalls and uncertainties. Lessons for and from the crypt. Development. 1990 Dec;110(4):1001-20.
(31) Beerman I, Bhattacharya D, Zandi S, Sigvardsson M, Weissman IL, Bryder D, et al. Functionally distinct hematopoietic stem cells modulate hematopoietic lineage potential during aging by a mechanism of clonal expansion. Proc Natl Acad Sci U S A. 2010 Mar 23;107(12):5465-70.
(32) Challen GA, Boles NC, Chambers SM, Goodell MA. Distinct hematopoietic stem cell subtypes are differentially regulated by TGF-beta1. Cell Stem Cell. 2010 Mar 5;6(3):265-78.
(33) Cho RH, Sieburg HB, Muller-Sieburg CE. A new mechanism for the aging of hematopoietic stem cells: aging changes the clonal composition of the stem cell compartment but not individual stem cells. Blood. 2008 Jun 15;111(12):5553-61.
(34) Bagby GC, Meyers G. Bone marrow failure as a risk factor for clonal evolution: prospects for leukemia prevention. Hematology Am Soc Hematol Educ Program. 2007;40-6.
(35) Balducci L, Ershler W, de Gaetano G, editors. Blood disorders in the elderly. New York: Cambridge University Press; 2008.
(36) Rossi DJ, Bryder D, Zahn JM, Ahlenius H, Sonu R, Wagers AJ, et al. Cell intrinsic alterations underlie hematopoietic stem cell aging. Proc Natl Acad Sci U S A. 2005 Jun 28;102(26):9194-9.
(37) Morrison SJ, Wandycz AM, Akashi K, Globerson A, Weissman IL. The aging of hematopoietic stem cells. Nat Med. 1996 Sep;2(9):1011-6.
(38) Sudo K, Ema H, Morita Y, Nakauchi H. Age-associated characteristics of murine hematopoietic stem cells. J Exp Med. 2000 Nov 6;192(9):1273-80.
(39) de Haan G, Nijhof W, Van ZG. Mouse strain-dependent changes in frequency and proliferation of hematopoietic stem cells during aging: correlation between lifespan and cycling activity. Blood. 1997 Mar 1;89(5):1543-50.
(40) Yilmaz OH, Kiel MJ, Morrison SJ. SLAM family markers are conserved among hematopoietic stem cells from old and reconstituted mice and markedly increase their purity. Blood. 2006 Feb 1;107(3):924-30.
(41) Liang Y, van Zant G, Szilvassy SJ. Effects of aging on the homing and engraftment of murine hematopoietic stem and progenitor cells. Blood. 2005 Aug 15;106(4):1479-87.

(42) Dykstra B, de Haan G. Hematopoietic stem cell aging and self-renewal. Cell Tissue Res. 2008 Jan;331(1):91-101.
(43) Rossi DJ, Bryder D, Seita J, Nussenzweig A, Hoeijmakers J, Weissman IL. Deficiencies in DNA damage repair limit the function of haematopoietic stem cells with age. Nature. 2007 Jun 7;447(7145):725-9.
(44) Nijnik A, Woodbine L, Marchetti C, Dawson S, Lambe T, Liu C, et al. DNA repair is limiting for haematopoietic stem cells during ageing. Nature. 2007 Jun 7;447(7145):686-90.
(45) Morita Y, Ema H, Nakauchi H. Heterogeneity and hierarchy within the most primitive hematopoietic stem cell compartment. J Exp Med. 2010 Jun 7;207(6):1173-82.
(46) Ogawa T, Kitagawa M, Hirokawa K. Age-related changes of human bone marrow: a histometric estimation of proliferative cells, apoptotic cells, T cells, B cells and macrophages. Mech Ageing Dev. 2000 Aug 15;117(1-3):57-68.
(47) Taraldsrud E, Grogaard HK, Solheim S, Lunde K, Floisand Y, Arnesen H, et al. Age and stress related phenotypical changes in bone marrow CD34+ cells. Scand J Clin Lab Invest. 2009;69(1):79-84.
(48) Beerman I, Maloney WJ, Weissmann IL, Rossi DJ. Stem cells and the aging hematopoietic system. Curr Opin Immunol. 2010 Aug;22(4):500-6.
(49) Kuranda K, Vargaftig J, de la Rochere P, Dosquet C, Charron D, Bardin F, et al. Age-related changes in human hematopoietic stem/progenitor cells. Aging Cell. In press 2011.
(50) Kollman C, Howe CW, Anasetti C, Antin JH, Davies SM, Filipovich AH, et al. Donor characteristics as risk factors in recipients after transplantation of bone marrow from unrelated donors: the effect of donor age. Blood. 2001 Oct 1;98(7):2043-51.
(51) Signer RA, Montecino-Rodriguez E, Witte ON, McLaughlin J, Dorshkind K. Age-related defects in B lymphopoiesis underlie the myeloid dominance of adult leukemia. Blood. 2007 Sep 15;110(6):1831-9.
(52) Rossi DJ, Jamieson CH, Weissman IL. Stems cells and the pathways to aging and cancer. Cell. 2008 Feb 22;132(4):681-96.
(53) Harrison DE, Astle CM. Loss of stem cell repopulating ability upon transplantation. Effects of donor age, cell number, and transplantation procedure. J Exp Med. 1982 Dec 1;156(6):1767-79.
(54) Baldridge MT, King KY, Boles NC, Weksberg DC, Goodell MA. Quiescent haematopoietic stem cells are activated by IFN-gamma in response to chronic infection. Nature. 2010 Jun 10;465(7299):793-7.
(55) Essers MA, Offner S, Blanco-Bose WE, Waibler Z, Kalinke U, Duchosal MA, et al. IFNalpha activates dormant haematopoietic stem cells in vivo. Nature. 2009 Apr 16;458(7240):904-8.
(56) Milyavsky M, Gan OI, Trottier M, Komosa M, Tabach O, Notta F, et al. A Distinctive DNA Damage Response in Human Hematopoietic Stem Cells Reveals an Apoptosis-Independent Role for p53 in Self-Renewal. Cell Stem Cell. 2010 Aug 6;7(2):186-97.
(57) Mohrin M, Bourke E, Alexander D, Warr MR, Barry-Holson K, Le Beau MM, et al. Hematopoietic stem cell quiescence promotes error-prone DNA repair and mutagenesis. Cell Stem Cell. 2010 Aug 6;7(2):174-85.
(58) Drummond MW, Heaney N, Kaeda J, Nicolini FE, Clark RE, Wilson G, et al. A pilot study of continuous imatinib vs pulsed imatinib with or without G-CSF in CML patients who have achieved a complete cytogenetic response. Leukemia. 2009 Jun;23(6):1199-201.
(59) Morrison SJ, Wright DE, Weissman IL. Cyclophosphamide/granulocyte colony-stimulating factor induces hematopoietic stem cells to proliferate prior to mobilization. Proc Natl Acad Sci U S A.. 1997 Mar 4;94(5):1908-13.
(60) Oppenheim JJ, Neta R, Tiberghien P, Gress R, Kenny JJ, Longo DL. Interleukin-1 enhances survival of lethally irradiated mice treated with allogeneic bone marrow cells. Blood. 1989 Nov 1;74(6):2257-63.
(61) Gratwohl A, John L, Baldomero H, Roth J, Tichelli A, Nissen C, et al. FLT-3 ligand provides hematopoietic protection from total body irradiation in rabbits. Blood. 1998 Aug 1;92(3):765-9.
(62) Lansdorp PM, Dragowska W, Mayani H. Ontogeny-related changes in proliferative potential of human hematopoietic cells. J Exp Med. 1993 Sep 1;178(3):787-91.
(63) Eapen M, Rocha V, Sanz G, Scaradavou A, Zhang MJ, Arcese W, et al. Effect of graft source on unrelated donor haemopoietic stem-cell transplantation in adults with acute leukaemia: a retrospective analysis. Lancet Oncol. 2010 Jul;11(7):653-60.
(64) Hacein-Bey-Abina S, Hauer J, Lim A, Picard C, Wang GP, Berry CC, et al. Efficacy of gene therapy for X-linked severe combined immunodeficiency. N Engl J Med. 2010 Jul 22;363(4):355-64.

(65) Cavazzana-Calvo M, Payen E, Negre O, Wang G, Hehir K, Fusil F, et al. Transfusion independence and HMGA2 activation after gene therapy of human beta-thalassaemia. Nature. 2010 Sep 16;467(7313):318-22.
(66) Stein S, Ott MG, Schultze-Strasser S, Jauch A, Burwinkel B, Kinner A, et al. Genomic instability and myelodysplasia with monosomy 7 consequent to EVI1 activation after gene therapy for chronic granulomatous disease. Nat Med. 2010 Feb;16(2):198-204.
(67) Aiuti A, Cattaneo F, Galimberti S, Benninghoff U, Cassani B, Callegaro L, et al. Gene therapy for immunodeficiency due to adenosine deaminase deficiency. N Engl J Med. 2009 Jan 29;360(5):447-58.
(68) Cartier N, Hacein-Bey-Abina S, Bartholomae CC, Veres G, Schmidt M, Kutschera I, et al. Hematopoietic stem cell gene therapy with a lentiviral vector in X-linked adrenoleukodystrophy. Science. 2009 Nov 6;326(5954):818-23.
(69) Gaspar HB, Parsley KL, Howe S, King D, Gilmour KC, Sinclair J, et al. Gene therapy of X-linked severe combined immunodeficiency by use of a pseudotyped gammaretroviral vector. Lancet. 2004 Dec 18;364(9452):2181-7.
(70) Howe SJ, Mansour MR, Schwarzwaelder K, Bartholomae C, Hubank M, Kempski H, et al. Insertional mutagenesis combined with acquired somatic mutations causes leukemogenesis following gene therapy of SCID-X1 patients. J Clin Invest. 2008 Sep;118(9):3143-50.
(71) Wang GP, Berry CC, Malani N, Leboulch P, Fischer A, Hacein-Bey-Abina S, et al. Dynamics of gene-modified progenitor cells analyzed by tracking retroviral integration sites in a human SCID-X1 gene therapy trial. Blood. 2010 Jun 3;115(22):4356-66.
(72) Kim S, Kim N, Presson AP, An DS, Mao SH, Bonifacino AC, et al. High-throughput, sensitive quantification of repopulating hematopoietic stem cell clones. J Virol. 2010 Nov;84(22):11771-80.
(73) Harkey MA, Kaul R, Jacobs MA, Kurre P, Bovee D, Levy R, et al. Multiarm high-throughput integration site detection: limitations of LAM-PCR technology and optimization for clonal analysis. Stem Cells Dev. 2007 Jun;16(3):381-92.
(74) Gerrits A, Dykstra B, Kalmykowa OJ, Klauke K, Verovskaya E, Broekhuis MJ, et al. Cellular barcoding tool for clonal analysis in the hematopoietic system. Blood. 2010 Apr 1;115(13):2610-8.
(75) Morrison SJ, Wandycz AM, Hemmati HD, Wright DE, Weissman IL. Identification of a lineage of multipotent hematopoietic progenitors. Development. 1997 May;124(10):1929-39.

CHAPTER **3**

COUNTING HEMATOPOIETIC STEM CELLS: METHODOLOGICAL CONSTRAINS

Bystrykh, L.V., Verovskaya, E., Broekhuis, M.J.C., Zwart, E., de Haan, G.

Nature Methods, 9(6), 567-74 (2012).

ABSTRACT

The number of stem cells contributing to hematopoiesis has been a matter of debate. Many studies use retroviral tagging of stem cells to measure clonal contribution. Here we argue that methodological factors can impact such clonal analyses. Whereas early studies had low resolution, leading to underestimation, recent methods may result in noisy data and overestimation of stem cell counts. We discuss how restriction enzyme choice, PCR bias, high-throughput sequencing depth and tagging method could affect the conclusions of clonal studies.

Proliferating tissues in the adult body contain a small population of stem cells that establish the system during ontogeny and maintain homeostasis during life. Assessing the number of these stem cells and the extent to which their clonal offspring contribute to tissue regeneration has been a matter of interest and controversy for many decades. Pioneering work on clonal stem cell tracking has been performed in the hematopoietic system and serves as a paradigm for other adult organs. Not only do these analyses provide insight into the function and heterogeneity of hematopoietic stem cells during ontogeny, normal life and aging, they are also important for our understanding of hematopoietic failures, which often have a clonal origin, for instance, in the case of leukemia (1,2). Here we discuss the use of genetic marking of hematopoietic stem cells with retro- or lentiviral vectors, followed by transplantation in human or animal recipients, for the purpose of counting stem cells and estimating clonal contribution in the hematopoietic system.

Models of clonal kinetics in hematopoiesis

Discussions of clonality in the hematopoietic system revolve around two main theories: clonal succession and clonal stability (Figure 1). Although useful, these models simplify other stem cell properties, such as stochastic or predetermined functional and dynamic heterogeneity. Nevertheless, initial studies aimed at resolving the clonal succession versus stability question, most intensively performed in mice, arrived at variable conclusions: some were in favor of clonal succession(3-6), and other findings were at odds with this theory (7-10). A validation of either model relies on counting clones at different time points, but reports of the number of

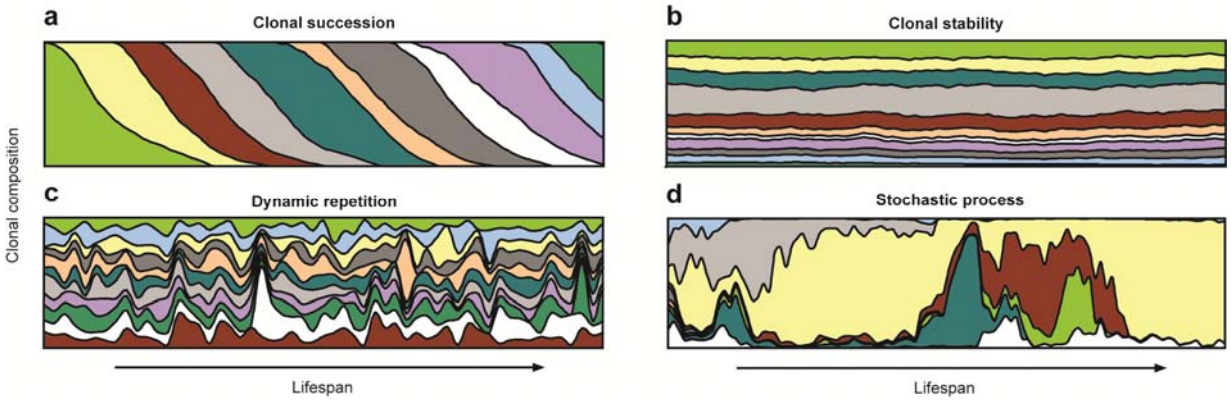

Figure 1. Models of clonal kinetics in hematopoiesis. (a) Clonal succession model (colors depict individual clones), which implies that stem cells have intrinsic limits to their repopulating ability and that only few stem cells would contribute to the blood at any given time (15). Exhausted clones would be replaced by new, previously dormant, clones. (b) Clonal maintenance model, which implies that many stem cells contribute simultaneously and indefinitely to the blood, that is therefore expected to be highly polyclonal with stable clonal contribution over time (7). (c) Dynamic repetition model, in which frequency of stem cell division and their clonal contributions oscillate, thus introducing cycling heterogeneity in clones. (d) A representation of a case of extreme stochasticity, such that behavior of individual stem cells in a population fluctuates according to stochastic intrinsic and extrinsic signals.

active clones supporting hematopoiesis have been variable. Early publications described small numbers of active clones in analyzed blood samples (3,5), whereas several recent studies report the simultaneous presence of hundreds of viral insertion sites (11-14), generally considered to reflect the number of active clones (Figure 2). Whereas the few clones found in early publications are often seen as an evidence of clonal succession, most of the recent data incline toward clonal maintenance (although there are some exceptions) (14).

Counting stem cells by retroviral DNA tagging

Retroviral transduction strategies were initially developed to insert a transgene into the genome of deficient cells. More recently, the detection of viral integration sites in genomic DNA also has been used to count stem cell clones. Various methods have been used to identify the unique genetic mark left by the integrating viral vector, allowing the identification of single stem cell–derived clones in the transplant recipient (Figure 3). In general terms, the protocols include fragmentation of genomic DNA isolated from transplanted cells followed by PCR amplification of the sequence adjacent to the vector integration site. In early studies, variation in PCR-amplified products was assessed by gel electrophoresis. These studies typically reported a small number of insertions (Figure 2). Currently, most advanced methods require digestion of genomic DNA with multiple restriction enzymes followed by PCR-based amplification of integration site and high-throughput sequencing (12,13,16). In some recent studies many viral insertion sites have been detected, for instance, as reported in studies of human gene therapy patients (1380–9659 integration sites) (12,13) as well as in studies of dogs (11) and mice (14). These numbers of viral insertion sites may be interpreted as if the hematopoietic system is highly polyclonal.

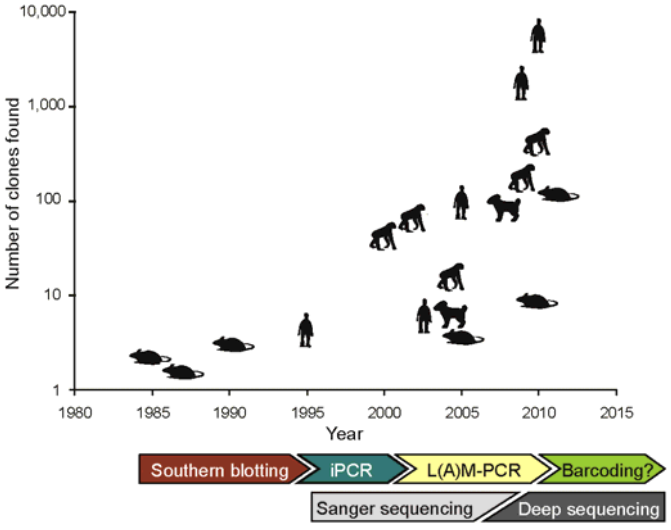

Figure 2. Reports of hematopoietic clonality in multiple studies. We depict schematically the approximate numbers of detected viral integration sites or predicted transduced cell clones in the blood and/or bone marrow, per animal, reported in multiple studies in different species. Readout techniques predominantly used in certain time periods are indicated. iPCR – inverse PCR, L(A)M-PCR reflects the use of either LAM-PCR - linear-amplification mediated PCR, or LM-PCR – ligation-mediated PCR. The values shown here are not intended to reflect exact numbers, but are rather typical or average values per recipient; data used to generate this figure are in Table 1. For papers where predictions are made for numbers of clones, we plot predicted values.

Table 1 Estimates of clonality in the hematopoietic system in representative studies

Year	Species				Techniques	References
	Mouse	Dog	Non-human primate	Human		
1985	1–2 ISs per mouse; 1–5 ISs per spleen colony				Southern blotting	17
1986	2–11 ISs per mouse; 1–2 stem cells predicted				Southern blotting	3
1990	1–6 ISs per mouse				Southern blotting	9
1995				3, 4 or 5 bands in each of three patients	iPCR	18
2000			35–45 bands in each of two macaques		iPCR	19
2002			>20–50 ISs per time point per baboon or macaque; 50–100 clones per animal predicted		LAM-PCR	20
2003				4 to 7 IS in two patients	LAM-PCR	21
2004		1–14 IS per dog	2–26 ISs per macaque, 8–28 ISs per baboon		LAM-PCR	22
2005	19–22 ISs in 6 mice			Approx. 100 predicted transduced clones[62]	LM-PCR; LAM-PCR	23,24
2008		535 ISs in 6 dogs			LAM-PCR, LM-PCR, 454	11
2009			519 ISs in three macaques	1,380 or 2,217 IS per patient; 900 or 1,719 ISs per patient	LAM-PCR, 454 sequencing	12,25
2010	4–13 barcodes per mouse		401 ISs in a primate	9,767 ISs in eight patients; 1784-9,659 transduced clones predicted per patient	Barcoding, bi-directional PCR, LAM-PCR, 454 sequencing	13,26,27
2011	>7,000 ISs in 56 mice				LM-PCR, 454 sequencing	14

Number of integration sites (ISs) reported in representative retroviral marking studies; papers do not include all retroviral marking studies published during this period but are chosen to be representative of the literature. Where possible (a subset of manuscripts explicitly predict clone numbers from the data), the number of predicted clones or stem cells is indicated.

However, techniques used to count clones have limitations, as do all experimental methods: they have detection limits, and sequencing or mapping errors can hamper identification and discrimination of genetic marks. We argue that technical aspects of the methods used to assess clonal contribution could affect the number of estimated active hematopoietic stem cell clones. In particular we address: (i) dependency of most methods on restriction endonucleases; (ii) bias in PCR-amplification of DNA fragments of variable length; (iii) insufficient depth of high-throughput sequencing analysis and (iv) inappropriate filtering of false positive high-throughput sequencing results. In addition, biological questions such as functional heterogeneity and formal definitions of stem cells also influence the conclusions of clonal studies.

Optimal choice of restriction enzymes and PCR bias

With the exception of non-restrictive linear amplification-mediated PCR (nrLAM-PCR)(28), PCR with arbitrary primers (29) and barcoding (26), methods for detecting viral integration sites, and hence for identifying virally tagged clones, require digestion of genomic DNA with restriction enzymes. As has recently been discussed in several papers (28,30-32), insufficient attention to

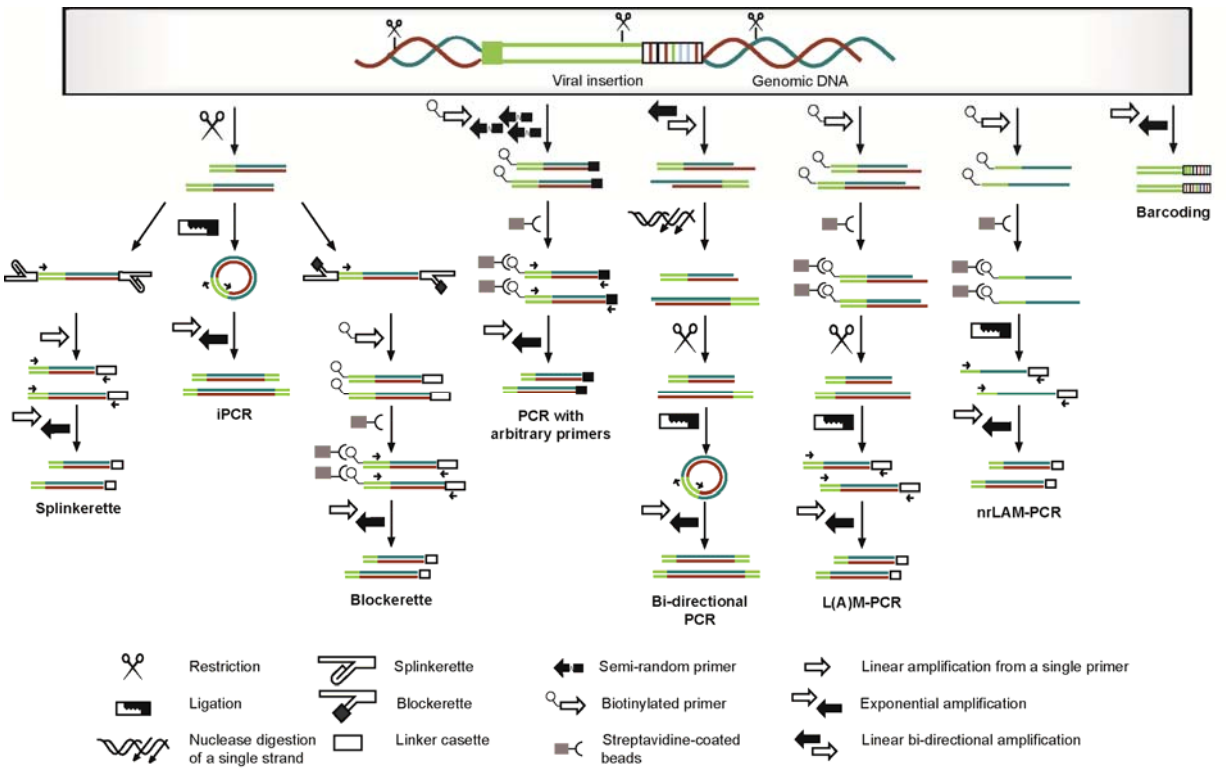

Figure 3. Strategies for identifying retroviral tags for clonal analysis. Shown is a schematic of the experimental steps in clonal analysis of retrovirally tagged cells: splinkerette PCR (33), inverse PCR (34), blockerette-ligated capture PCR (35), PCR with arbitrary primers (29), bi-directional PCR (one of two described experimental setups in shown) (27), LM-PCR and LAM-PCR (24,36), nrLAM-PCR (28) and cellular barcoding (26).

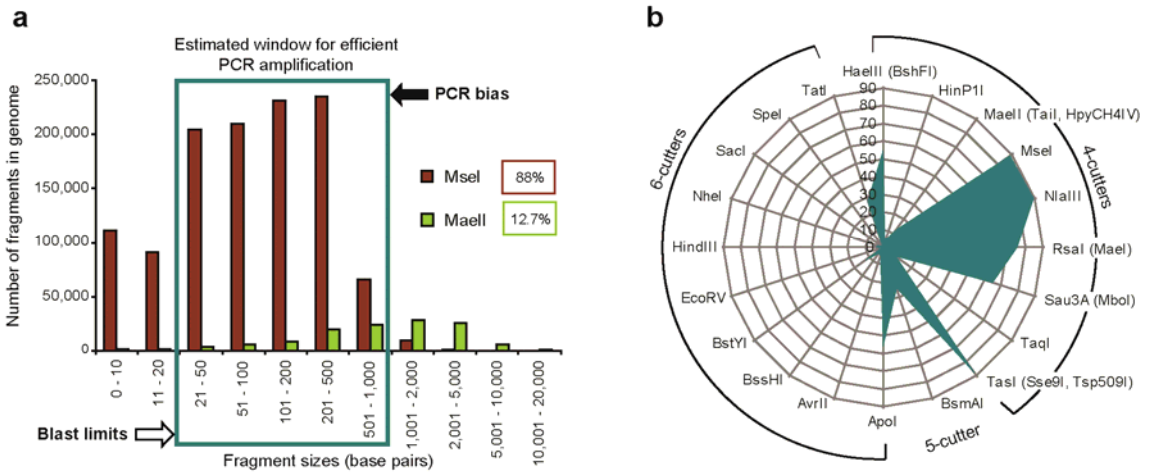

Figure 4. Predicted PCR amplification efficiency of genomic fragments digested with different enzymes. (a) Results of an *in silico* DNA restriction digest of mouse genome with MseI and MaeII. The boxed region indicates an effective PCR-amplification window, using upper (determined by PCR bias) and lower (determined by BLAST limitations) thresholds. The fraction of the genome found within the effective window defines genome accessibility. MseI will ensure 88% of genome accessibility. With MaeII only 12.7% will be accessible. (b) Genome accessibility percentage for every enzyme used for viral integration-site analysis.

the selected restriction enzymes (and the frequency of corresponding restriction sites) will mean that genomic fragments in which restriction sites are rare will not be properly digested, and viral integration in these regions will remain undetected. Only ~100–500-base-pair fragments will be subsequently amplified by PCR, whereas longer fragments will be underrepresented (27,28,30-32) or completely absent in the resulting data.

In an attempt to systematically characterize whether such problems are widespread in the literature, we analyzed papers reporting clonal analysis in the hematopoietic system over the past 10 years and retrospectively determined the expected odds of detection of a clone based on the restriction enzymes used. We performed an *in silico* restriction digestion of the mouse and human genomes and identified the effective window of fragment sizes in which PCR amplification would enable detection of a particular viral integration site (Figure 4). We assumed 1,000 base pairs as the maximum size of an amplifiable restricted fragment. In case of bi-directional PCR, the most recently described method, each integration site will then be situated 500 base pairs or closer to one of restriction sites in adjacent genomic DNA (27). We assumed 20 base pairs to be the minimal fragment size that can be mapped by a Basic Local Alignment Search tool (BLAST) search. Using this size window, we determined for each enzyme the fraction of genomic DNA that can be cut for further amplification. We used a custom written script in R to perform virtual restrictions with a range of selected enzymes (available upon request). We sorted the results in bins by fragment length and counted frequencies per bin. It is evident that, for many restriction enzymes, a large percentage of the genome lacks appropriate

restriction sites and therefore remains inaccessible to interrogation. For instance, only 2–13% of all DNA fragments are accessible for amplification and sequence analysis upon digestion with MaeII (37), TaiI (38) or TaqI (19).

As others have also discussed (28,31), our *in silico* analysis revealed that 5- or 6-base cutters have too low a frequency of genome-wide restriction to be useful for integration-site analysis, and therefore their use should be avoided. In contrast, restriction with 4-base cutters such as TasI (39,40), Sse9I (36), Tsp509I (24,40,41) is predicted to result in more than 85% of all fragments being accessible for PCR and sequence analysis (Figure 4b). However, it should be noted that among 4-base cutters, differences in restriction frequencies are quite large (28). C+G-containing palindromes occur with the lowest frequency, and A+T-rich palindromes occurred most frequently among all possible 4-base restriction sites, which reflects the nonrandom structure of genomic DNA (28,42-44).

Enzymes for multi-arm DNA restriction

Several authors have attempted to empirically improve restriction-based methods by using combinations of restriction enzymes (31). For instance, Biasco *et al.* (45) showed that restriction with one enzyme (MseI) retrieved only 2–6% of integration sites compared to all sites retrieved with a mixture of four restriction enzymes (also including MseI). Our simulations confirmed that multi-arm approaches suggested by Harkey *et al.* (31) and Wang *et al.*(13) improved the accessibility of the genomic fragments from approximately 80% to 95%.

Strategies to select combinations of restriction enzymes have been proposed; for example, Gabriel *et al.* (28) have suggested starting with the most frequent cutter and adding the next best cutter and so on, to increase coverage of the accessible genome. However, with any particular choice of restriction enzymes attention should be paid to the similarities between sequences in each selected restriction site. A balanced combination of subsequences presented in the sites of selected restriction enzymes is recommended. A good illustration as to why this is important is the case of (G+C)-rich regions: although (A+T)-rich restriction enzymes are more efficient than (G+C)-rich restriction enzymes genome-wide, they are inefficient in (G+C)-rich regions, which frequently occur in promoters and are potential loci for retroviral integrations.

In Figure 5 we present one example of such a combination: every four-nucleotide restriction site has its first two nucleotides reverse-complemented by the last two nucleotides such that 16 possible palindromes can be created in total. Three sets of dinucleotides are included in each palindrome; for instance, AATT contains AA, AT and TT. If we do not want to avoid DNA loci that are either enriched for or missing any particular dinucleotide combination, the rational approach would be to ensure that the selected combination of restriction enzymes recognizes sites that include as many dinucleotides as possible. Note that 5- or 6-cutter enzymes will target a subset of sequences that are recognized by a 4-cutter with a similar recognition specificity,

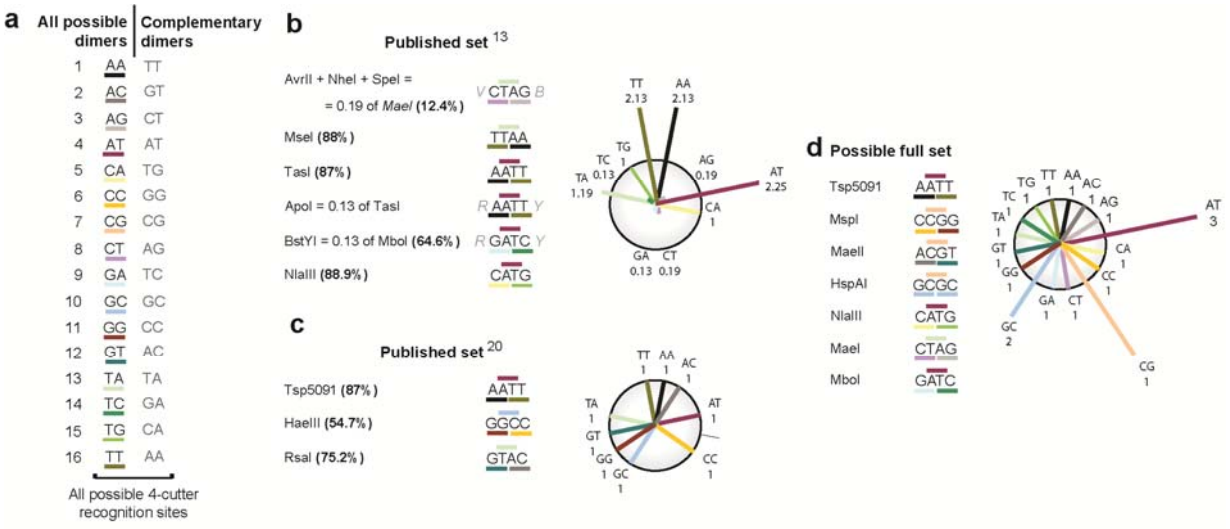

Figure 5. Combinatorial strategy to select optimal combination of restriction enzymes. (a) The sixteen possible recognition sites for 4-cutter restriction enzymes. (b,c) Analysis of combinations of dinucleotides present in enzyme mixtures in ref. 13 (b) and ref. 20 (c). The two circular diagrams depict the estimated efficiency with which each of the 16 dinucleotides will be recognized by the used enzyme mixture. Dinucleotides that are not shown in the circular diagram are not recognized. (d) Set of restriction enzymes that will cover all possible dimer combinations and cut 98% of the genome.

limited by additional outer nucleotide(s) in the recognition site. For instance, the enzyme ApoI (recognition site RAATTY) can cut 4 × 4 / 2 = 8 times fewer DNA sequences than restriction enzyme TasI (recognition site AATT) (Figure 5b; restriction set of Wang *et al.* (13)). Of course, optimal combinations of enzymes must be selected for each individual experiment (the restriction site must be present in the integrated viral sequence, and at sufficient distance from the end of the long terminal repeat to allow vector recognition).

Sequencing depth, data validation and clonal counts

A critical part of integration-site analysis is identification of the location of the insert, now typically done by deep sequencing of PCR-amplified fragments. This type of analysis is invaluable in understanding genomic integration preferences of viral vectors as well as their possible impact on the dysregulation of adjacent genes. Quantitative implementation of this method to count clones, however, is fraught with potential technical problems.

If sequencing depth is low, many clones may escape detection, and read frequency is no longer reliable for discriminating true and false integration sites. Moreover, detection of a single sequence read is insufficient evidence for stem-cell presence: a single read is by definition present only in a single cell type and therefore cannot demonstrate robust repopulating potential in different blood lineages. Restriction and ligation–based approaches also carry the risk of accidental re-ligation of irrelevant DNA fragments, which may be misinterpreted as

unique integration sites, although methods have been designed to prevent such events (46). Contamination of samples is another source of inconsistent reads; several papers have reported so-called 'collision reads'—identical insertion sites detected in samples from different individuals—that resulted from cross-contamination between samples (12,32). Barcoded primers can identify cross-contamination in the sequencing facility except if it occurs before PCR amplification.

To compensate for low sequencing depth and limited overlap of sequence data from repeat samples, a recapture concept (first suggested by Kim *et al.* (19)) has been used (13). This approach presumes that partial overlap of data acquired from the same donor tested multiple times is due to the high numbers of clones in the samples, such that only a subset of clones is detected in each case. Based on the frequency of sites detected in one or multiple samples, the total number of clones is estimated. However, this type of prediction must take potential technical errors into account. Noisy and shallow sequencing of repeatedly analyzed samples could boost the predicted clonality to several thousands, leading to an overestimation of stem cell numbers.

In our opinion, insufficient attention is sometimes paid to the reporting and validation of sequence reads for clonal studies. Approaches used to prefilter raw sequence data, to identify integration sites and to determine mapping accuracy are quite variable. Most reports do not contain primary sequencing results, making re-analysis impossible. To properly interpret stem-cell clonality data based on sequence analysis, detailed information on sequence statistics, analysis algorithms and a summary of sequence data are required, as provided for instance in ref. 13. It is important to report the number and frequencies of unique reads as well as their measure of uniqueness (sequence differences between all reads with respect to the number of single polymorphisms). Instead, most authors compare sample read sequence with a reference genome and discard integration sites that cannot be uniquely mapped; however, unmappable reads may be true and valid. Integration-site analysis could be improved by increasing sequencing depth and introducing robust criteria for noise discrimination, similar to those applied in analyses of data from chromatin immunoprecipitation followed by high-throughput sequencing (ChIP-Seq).

Clonal analysis with DNA barcodes

We and others have used cellular barcoding techniques to estimate clonal complexity in blood and other tissues (26,47-50). The basic idea of barcoding is to create a cell-specific unique DNA mark that can be used to identify clones. Barcodes should be of equal size in all cells, be identified by uniform primers, resist PCR bias, be sufficiently unique to allow unequivocal identification and also resist cross contamination owing to naturally occurring sequence read errors. Recently we had described hematopoietic stem cell barcoding based on the introduction

of a semi-random tag in a retroviral vector (26) followed by direct sequencing of barcoded fragments. This approach can avoid several of the technical problems described above; in particular, barcode analysis avoids problems with restriction-enzyme choice and random ligand attachment.

A critical aspect of successful cellular barcoding is the uniqueness of each tag: this depends on the total number of possible barcode variations, the real number of barcode variations in the vector libraries, and the number of clonogenic cells to be barcoded and transplanted. If the barcode sequence is based on random nucleotide synthesis, the theoretical number of possible combinations should be several orders of magnitude above the real barcode library size to ensure good randomization and uniqueness (which must also be confirmed experimentally). For proper interpretation of the data, it is imperative to validate the initial barcoded vector library size and avoid excessive biases in the barcode frequency distribution. Effective library size may otherwise be dramatically reduced to a few dominant barcodes. As a simple rule, the size of the library should exceed the number of clonogenic cells subjected to analysis, with a 10:1 ratio of barcodes to target cells resulting in >90% of uniquely labeled cells. Finally, to correct for multiple vector insertions, distribution of the number of barcode per individual stem cell must be assessed as part of the analysis.

If these conditions are met, viral barcoding can be used to measure clonal contributions in a quantitative manner. Owing to the compact size of typical barcodes, the approach is amenable

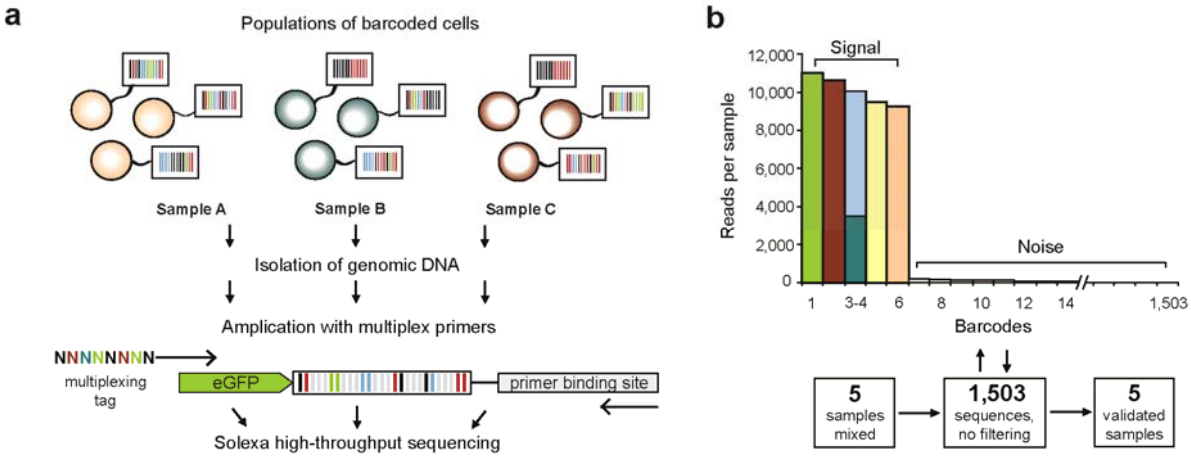

Figure 6. Clonal analysis in barcoded cells. (a) Schematic of the setup for a multiplex barcoding experiment. To read the barcodes in transduced hematopoietic cells, genomic DNA is extracted and amplified with tagged primers against an internal vector sequence. A library of tagged multiplex primers allows several hundred blood or bone marrow samples to be individually labeled. These samples are then pooled and subjected to Solexa high-throughput sequencing. (b) Signal/noise discrimination based on barcode frequency. A mixture of five uniquely barcoded clones (four clones singly barcoded, one clone doubly barcoded) was sequenced. Different colors in the frequency histogram indicate different barcodes.

to readout using deep sequencing platforms. As for the analysis of viral integration sites, however, several important technical aspects must be considered. True and false (noise-borne) barcodes in the analyzed blood sample can be most easily distinguished on the basis of cross-comparison with the barcode vector library. A full report of the library in publications is therefore desirable; knowledge of its size and the identity of all barcodes are crucial for the success of the barcode analysis. Second, sequence analysis must address the origin of single-nucleotide mutants and the frequency of mutations. False barcodes that result from errors during PCR and sequencing will most frequently differ from true barcodes by one substitution or insertion-deletion and will have considerably lower frequencies. These mutant reads can be detected by cross-alignment with the library of barcodes sequenced before stem-cell transduction and removed if necessary. For example, we sequenced barcodes extracted from a mixture of 5 clonally expanded cells (Figure 6). Of the 1,503 different barcodes we recovered in the sample, six barcodes represented the five original clones. Noise could be discriminated from real barcodes based on the frequencies of the barcode sequences. We remove single reads from our analyses because they frequently correspond to false barcodes. Although the occurrence of mutant ('false') barcode sequences is usually within the lowest 5% of all reads, calculations of stem-cell clone numbers can differ substantially (between 10- and 1000-fold) before and after removing such false tags. It is of interest to note that when Biffi *et al.* (51), in their clinical study using vector-integration-site analysis, removed 2% of the sequences with the lowest frequencies, their total clonal count dropped substantially. Similarly, Cavazzana-Calvo *et al.* (16) in a comparable study showed that kinetics of the most frequent 16 integration sites could be followed in time with 100% overlap.

In our experience, clonal kinetics are quite slow in mice, barcode sequences are highly consistent in samples taken at consecutive time points, and only a few stem cells actively contribute to hematopoiesis. Early time points after transplantation are typically characterized by a wide spectrum of barcodes in both granulocytes and T cells but the signature becomes smaller and more stable at later time points (Figure 7).

Hematopoietic stem cell heterogeneity and counts

Counting stem cells is affected not only by technical constraints; there are biological considerations that are critical for the interpretation of such studies. Although these have not been the focus of our discussion, a few points are worth noting because analysis of clonal contribution can be interpreted differently depending on the prevailing stem cell definition. The very criteria used to define hematopoietic stem cells are subject to continuous discussion. True 'stemness' can only be confirmed retrospectively by the ability of the cell to achieve robust, long-term multilineage repopulation of an irradiated recipient. Usually, to confirm the presence of a stem cell, several conditions must be met. First, the progeny of a cell must be detected in blood for an extended time (the threshold for separating stem cells from progenitors is

currently set at 4 months after transplantation). Second, a stem cell clone must be seen to contribute to several blood lineages; myeloid, T and B cells are usually analyzed. Of those, only short-lived cells, for example, granulocytes, reflect recent proliferative activity in the bone marrow. Mature lymphoid cells, in contrast, can persist in blood for an extended time after exhaustion of corresponding precursor cells. Therefore, using only total blood or lymphoid chimerism as a criterion may lead to an overestimation of stem-cell counts. A third parameter is the robustness of hematopoietic reconstitution. In single-cell transplantation studies, bona fide stem cells should reconstitute more than 1% of the blood of recipients. The presence of an occasional donor-derived blood cell is not sufficient to claim stem-cell activity. Finally, especially because the volume of blood aliquots used for clonal analysis is small relative to the total volume of blood, it is critical that genomic DNA is efficiently extracted and analyzed to ensure representative and reproducible results.

Several recent studies have demonstrated heterogeneity of hematopoietic stem cells (HSC) behavior in terms of their repopulation potential, lineage output and turnover rate (reviewed elsewhere (52-54)) and this can constrain the discovery of all clones. Strict cutoff values for chimerism levels in a limited time interval, as we described above, could preclude detection of particular HSC types irrespective of the readout technique used, because not all progeny of a clone may be present at a single time point in a polyclonally reconstituted animal. Limiting the readout time to 4 months after transplant could also result in exclusion of quiescent HSCs from analysis(55,56). The general implications of HSC heterogeneity for experimental design are that samples must be collected over a considerable time interval, that different cell lineages must be assessed and that data must be cross-checked for the presence of common clones (a good example of such an analysis is in (13)).

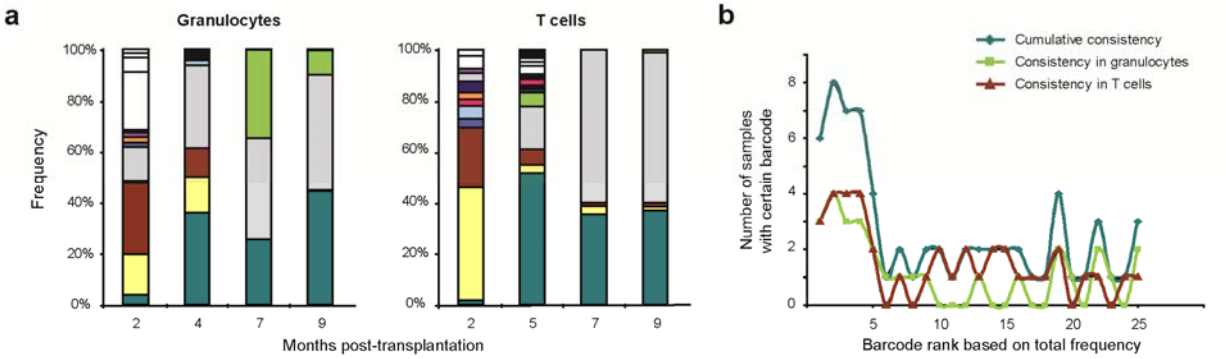

Figure 7. Clonal kinetics in mice transplanted with barcoded cells. (a) An example of clonal fluctuations in blood samples taken from one mouse transplanted with retrovirally barcoded bone marrow cells (as described previously in ref. (26)). We analyzed 8 samples (4 time points × 2 cell types). Colors indicate individual barcodes. (b) Consistency score - number of samples in which the same barcode is found.

How many stem cell clones do we actually expect?

In general, it is ill-advised to rely on a single approach to count stem cells; limiting dilution of bone marrow followed by transplant is an alternative approach. In the mouse, this approach has shown that there are ~2–8 stem cells per 10^5 freshly isolated bone marrow cells (7,8,57,58). In 5-fluoruracil–treated bone marrow this frequency increases to 1 in 2,000 (refs. 47,48). However, *ex vivo* expansion, transduction and transplantation procedures lead to loss of stem cells and need to be taken into consideration in interpreting these experiments. Specifically, we observe a 1 to 2-log loss of stem cell frequency due to cell sorting, *ex vivo* stem cell manipulation and culturing in limiting dilution experiments. With these considerations, by transplanting 2×10^6 to 5×10^6 5-fluoruracil–treated, retrovirally transduced bone marrow cells into recipient mice after a three- to tenfold expansion in culture, the expected clone numbers would be 1–10 per recipient, quite close to what we observe in our experiments with barcoded cells(26).

In humans, the frequency of stem cells has been estimated to be 1–4 stem cells per 10^7 mononuclear cells or 5–100 stem cells per 10^7 $CD34^+$ cells (38,59,60) based on limiting dilution of bone marrow in non-obese diabetic severe combined immunodeficient (NOD/SCID) mice, and capture and release calculations. Pediatric patients in gene therapy trials are transplanted with $5–10 \times 10^6$ $CD34^+$ cord blood cells per kilogram body weight (61). If stem cell loss resulting from experimental manipulation is close to what we observe in mouse studies, the final number of stem cells is likely to be less than 1 stem cell per 2 million $CD34^+$ cells, or 3–5 stem cells per kilogram of body weight. These estimates are several orders of magnitude lower than the numbers of integration sites detected in human recipients receiving a transplant. An interpretation of these numbers as reflective of polyclonality should therefore be taken with caution, and additional experiments will be required to resolve this discrepancy.

Concluding remarks

We discussed the use of retroviral labeling to count the number of active hematopoietic stem cells and argued that technical constraints can undermine the reliability of these measurements and yield variable stem cell counts. Any method used for a quantitative application in molecular biology should meet two fundamental requirements, namely calibration and noise discrimination, which will define the dynamic range and detection limits of the method. In our view, an advantage of cellular barcoding for stem-cell counting is that it allows calibration of dynamic range, can test detection limits with respect to the physical size of the sample and the detectable number of barcodes it contains, and permits identification of false positives by sequence comparison of samples with a reference library. It should be noted, however, that retroviral cell labeling and cell transplantation are experimental perturbations. Although these approaches capitalize on the natural abilities of HSCs to home, self-renew and differentiate,

caution is warranted as to how observations made using these techniques relate to normal hematopoiesis.

Hematopoiesis may include stochastic components for which multiple dynamic scenarios are possible, such that clonal succession or clonal stability are two extreme outcomes of a range of possibilities. Future studies are likely to focus on describing the full spectrum of scenarios and estimating the likelihood at which they occur. With the rapid development of sequencing technologies, it is likely that whole-genome sequencing for clonality studies will soon offer a realistic alternative for some of the approaches we have discussed here. Future experiments are likely to exploit induced or naturally occurring barcode-like sequence variations to improve our understanding of clonal dynamics of multicellular systems.

ACKNOWLEDGMENTS

We thank B. Dykstra for critical reading of the manuscript, and M. Ritsema, H. Moes, G. Mesander and R.-J. van der Lei for technical assistance. This study was supported by the Netherlands Organization for Scientific Research (VICI grant to G.d.H. and TopTalent grant to E.V.), and the Netherlands Genomics Initiative (Horizon Grant 050-71-055 to G.d.H. and the EU FP7 grant EuroSystem).

AUTHOR CONTRIBUTIONS

L.V.B. conceptualized the paper; L.V.B., M.B. and E.V. performed experiments; L.V.B., E.V. and E.Z. analyzed the data; L.V.B., E.V. and G.d.H. wrote the paper; and G.d.H. provided funding.

REFERENCES

(1) Anderson K, Lutz C, van Delft FW, Bateman CM, Guo Y, Colman SM, et al. Genetic variegation of clonal architecture and propagating cells in leukaemia. Nature 2011 Jan 20;469(7330):356-361.

(2) Notta F, Mullighan CG, Wang JC, Poeppl A, Doulatov S, Phillips LA, et al. Evolution of human BCR-ABL1 lymphoblastic leukaemia-initiating cells. Nature 2011 Jan 20;469(7330):362-367.

(3) Lemischka IR, Raulet DH, Mulligan RC. Developmental potential and dynamic behavior of hematopoietic stem cells. Cell 1986 Jun 20;45(6):917-927.

(4) Snodgrass R, Keller G. Clonal fluctuation within the haematopoietic system of mice reconstituted with retrovirus-infected stem cells. EMBO J 1987 Dec 20;6(13):3955-3960.

(5) Capel B, Hawley R, Covarrubias L, Hawley T, Mintz B. Clonal contributions of small numbers of retrovirally marked hematopoietic stem cells engrafted in unirradiated neonatal W/Wv mice. Proc Natl Acad Sci U S A 1989 Jun;86(12):4564-4568.

(6) Abkowitz JL, Linenberger ML, Newton MA, Shelton GH, Ott RL, Guttorp P. Evidence for the maintenance of hematopoiesis in a large animal by the sequential activation of stem-cell clones. Proc Natl Acad Sci U S A 1990 Nov;87(22):9062-9066.

(7) Harrison DE, Astle CM, Lerner C. Number and continuous proliferative pattern of transplanted

primitive immunohematopoietic stem cells. Proc Natl Acad Sci U S A 1988 Feb;85(3):822-826.
(8) Harrison DE, Astle CM, Stone M. Numbers and functions of transplantable primitive immunohematopoietic stem cells. Effects of age. J Immunol 1989 Jun 1;142(11):3833-3840.
(9) Jordan CT, Lemischka IR. Clonal and systemic analysis of long-term hematopoiesis in the mouse. Genes Dev 1990 Feb;4(2):220-232.
(10) Takizawa H, Regoes RR, Boddupalli CS, Bonhoeffer S, Manz MG. Dynamic variation in cycling of hematopoietic stem cells in steady state and inflammation. J Exp Med 2011 Feb 14;208(2):273-284.
(11) Hai M, Adler RL, Bauer TR,Jr, Tuschong LM, Gu YC, Wu X, et al. Potential genotoxicity from integration sites in CLAD dogs treated successfully with gammaretroviral vector-mediated gene therapy. Gene Ther 2008 Jul;15(14):1067-1071.
(12) Cartier N, Hacein-Bey-Abina S, Bartholomae CC, Veres G, Schmidt M, Kutschera I, et al. Hematopoietic stem cell gene therapy with a lentiviral vector in X-linked adrenoleukodystrophy. Science 2009 Nov 6;326(5954):818-823.
(13) Wang GP, Berry CC, Malani N, Leboulch P, Fischer A, Hacein-Bey-Abina S, et al. Dynamics of gene-modified progenitor cells analyzed by tracking retroviral integration sites in a human SCID-X1 gene therapy trial. Blood 2010 Jun 3;115(22):4356-4366.
(14) Maetzig T, Brugman MH, Bartels S, Heinz N, Kustikova OS, Modlich U, et al. Polyclonal fluctuation of lentiviral vector-transduced and expanded murine hematopoietic stem cells. Blood 2011 Mar 17;117(11):3053-3064.
(15) KAY HE. How Many Cell-Generations? Lancet 1965 Aug 28;2(7409):418-419.
(16) Cavazzana-Calvo M, Payen E, Negre O, Wang G, Hehir K, Fusil F, et al. Transfusion independence and HMGA2 activation after gene therapy of human beta-thalassaemia. Nature 2010 Sep 16;467(7313):318-322.
(17) Dick JE, Magli MC, Huszar D, Phillips RA, Bernstein A. Introduction of a selectable gene into primitive stem cells capable of long-term reconstitution of the hemopoietic system of W/Wv mice. Cell 1985 Aug;42(1):71-79.
(18) Kohn DB, Weinberg KI, Nolta JA, Heiss LN, Lenarsky C, Crooks GM, et al. Engraftment of gene-modified umbilical cord blood cells in neonates with adenosine deaminase deficiency. Nat Med 1995 Oct;1(10):1017-1023.
(19) Kim HJ, Tisdale JF, Wu T, Takatoku M, Sellers SE, Zickler P, et al. Many multipotential gene-marked progenitor or stem cell clones contribute to hematopoiesis in nonhuman primates. Blood 2000 Jul 1;96(1):1-8.
(20) Schmidt M, Zickler P, Hoffmann G, Haas S, Wissler M, Muessig A, et al. Polyclonal long-term repopulating stem cell clones in a primate model. Blood 2002 Oct 15;100(8):2737-2743.
(21) Schmidt M, Carbonaro DA, Speckmann C, Wissler M, Bohnsack J, Elder M, et al. Clonality analysis after retroviral-mediated gene transfer to CD34+ cells from the cord blood of ADA-deficient SCID neonates. Nat Med 2003 Apr;9(4):463-468.
(22) Kiem HP, Sellers S, Thomasson B, Morris JC, Tisdale JF, Horn PA, et al. Long-term clinical and molecular follow-up of large animals receiving retrovirally transduced stem and progenitor cells: no progression to clonal hematopoiesis or leukemia. Mol Ther 2004 Mar;9(3):389-395.
(23) Schmidt M, Hacein-Bey-Abina S, Wissler M, Carlier F, Lim A, Prinz C, et al. Clonal evidence for the transduction of CD34+ cells with lymphomyeloid differentiation potential and self-renewal capacity in the SCID-X1 gene therapy trial. Blood 2005 Apr 1;105(7):2699-2706.
(24) Kustikova O, Fehse B, Modlich U, Yang M, Dullmann J, Kamino K, et al. Clonal dominance of hematopoietic stem cells triggered by retroviral gene marking. Science 2005 May 20;308(5725):1171-1174.
(25) Kim YJ, Kim YS, Larochelle A, Renaud G, Wolfsberg TG, Adler R, et al. Sustained high-level polyclonal hematopoietic marking and transgene expression 4 years after autologous transplantation of rhesus macaques with SIV lentiviral vector-transduced CD34+ cells. Blood 2009 May 28;113(22):5434-5443.
(26) Gerrits A, Dykstra B, Kalmykowa OJ, Klauke K, Verovskaya E, Broekhuis MJ, et al. Cellular barcoding tool for clonal analysis in the hematopoietic system. Blood 2010 Apr 1;115(13):2610-2618.
(27) Kim S, Kim N, Presson AP, An DS, Mao SH, Bonifacino AC, et al. High-throughput, sensitive quantification of repopulating hematopoietic stem cell clones. J Virol 2010 Nov;84(22):11771-11780.
(28) Gabriel R, Eckenberg R, Paruzynski A, Bartholomae CC, Nowrouzi A, Arens A, et al. Comprehensive genomic access to vector integration in clinical gene therapy. Nat Med 2009 Dec;15(12):1431-1436.

(29) Gentner B, Laufs S, Nagy KZ, Zeller WJ, Fruehauf S. Rapid detection of retroviral vector integration sites in colony-forming human peripheral blood progenitor cells using PCR with arbitrary primers. Gene Ther 2003 May;10(9):789-794.
(30) Mantovani J, Holic N, Martinez K, Danos O, Perea J. A high throughput method for genome-wide analysis of retroviral integration. Nucleic Acids Res 2006;34(19):e134.
(31) Harkey MA, Kaul R, Jacobs MA, Kurre P, Bovee D, Levy R, et al. Multiarm high-throughput integration site detection: limitations of LAM-PCR technology and optimization for clonal analysis. Stem Cells Dev 2007 Jun;16(3):381-392.
(32) Wang GP, Garrigue A, Ciuffi A, Ronen K, Leipzig J, Berry C, et al. DNA bar coding and pyrosequencing to analyze adverse events in therapeutic gene transfer. Nucleic Acids Res 2008 May;36(9):e49.
(33) Mikkers H, Allen J, Knipscheer P, Romeijn L, Hart A, Vink E, et al. High-throughput retroviral tagging to identify components of specific signaling pathways in cancer. Nat Genet 2002 Sep;32(1):153-159.
(34) Silver J, Keerikatte V. Novel use of polymerase chain reaction to amplify cellular DNA adjacent to an integrated provirus. J Virol 1989 May;63(5):1924-1928.
(35) Lenvik T, Lund TC, Verfaillie CM. Blockerette-ligated capture T7-amplified RT-PCR, a new method for determining flanking sequences. Mol Ther 2002 Jul;6(1):113-118.
(36) Schmidt M, Hoffmann G, Wissler M, Lemke N, Mussig A, Glimm H, et al. Detection and direct genomic sequencing of multiple rare unknown flanking DNA in highly complex samples. Hum Gene Ther 2001 May 1;12(7):743-749.
(37) Mitsuhashi J, Tsukahara S, Suzuki R, Oh-hara Y, Nishi S, Hosoyama H, et al. Retroviral integration site analysis and the fate of transduced clones in an MDR1 gene therapy protocol targeting metastatic breast cancer. Hum Gene Ther 2007 Oct;18(10):895-906.
(38) Hayakawa J, Washington K, Uchida N, Phang O, Kang EM, Hsieh MM, et al. Long-term vector integration site analysis following retroviral mediated gene transfer to hematopoietic stem cells for the treatment of HIV infection. PLoS One 2009;4(1):e4211.
(39) Bauer TR,Jr, Hai M, Tuschong LM, Burkholder TH, Gu YC, Sokolic RA, et al. Correction of the disease phenotype in canine leukocyte adhesion deficiency using ex vivo hematopoietic stem cell gene therapy. Blood 2006 Nov 15;108(10):3313-3320.
(40) Schwarzwaelder K, Howe SJ, Schmidt M, Brugman MH, Deichmann A, Glimm H, et al. Gammaretrovirus-mediated correction of SCID-X1 is associated with skewed vector integration site distribution in vivo. J Clin Invest 2007 Aug;117(8):2241-2249.
(41) Deichmann A, Hacein-Bey-Abina S, Schmidt M, Garrigue A, Brugman MH, Hu J, et al. Vector integration is nonrandom and clustered and influences the fate of lymphopoiesis in SCID-X1 gene therapy. J Clin Invest 2007 Aug;117(8):2225-2232.
(42) Cuticchia AJ, Ivarie R, Arnold J. The application of Markov chain analysis to oligonucleotide frequency prediction and physical mapping of Drosophila melanogaster. Nucleic Acids Res 1992 Jul 25;20(14):3651-3657.
(43) Karlin S, Mrazek J. Compositional differences within and between eukaryotic genomes. Proc Natl Acad Sci U S A 1997 Sep 16;94(19):10227-10232.
(44) Chor B, Horn D, Goldman N, Levy Y, Massingham T. Genomic DNA k-mer spectra: models and modalities. Genome Biol 2009;10(10):R108-2009-10-10-r108. Epub 2009 Oct 8.
(45) Biasco L, Ambrosi A, Pellin D, Bartholomae C, Brigida I, Roncarolo MG, et al. Integration profile of retroviral vector in gene therapy treated patients is cell-specific according to gene expression and chromatin conformation of target cell. EMBO Mol Med 2011 Feb;3(2):89-101.
(46) Ciuffi A, Ronen K, Brady T, Malani N, Wang G, Berry CC, et al. Methods for integration site distribution analyses in animal cell genomes. Methods 2009 Apr;47(4):261-268.
(47) Lu R, Neff NF, Quake SR, Weissman IL. Tracking single hematopoietic stem cells in vivo using high-throughput sequencing in conjunction with viral genetic barcoding. Nat Biotechnol 2011 Oct 2;29(10):928-933.
(48) Golden JA, Fields-Berry SC, Cepko CL. Construction and characterization of a highly complex retroviral library for lineage analysis. Proc Natl Acad Sci U S A 1995 Jun 6;92(12):5704-5708.
(49) Cepko CL, Ryder E, Austin C, Golden J, Fields-Berry S, Lin J. Lineage analysis using retroviral vectors. Methods 1998 Apr;14(4):393-406.
(50) Schepers K, Swart E, van Heijst JW, Gerlach C, Castrucci M, Sie D, et al. Dissecting T cell lineage relationships by cellular barcoding. J Exp Med 2008 Sep 29;205(10):2309-2318.

(51) Biffi A, Bartolomae CC, Cesana D, Cartier N, Aubourg P, Ranzani M, et al. Lentiviral vector common integration sites in preclinical models and a clinical trial reflect a benign integration bias and not oncogenic selection. Blood 2011 May 19;117(20):5332-5339.
(52) Trumpp A, Essers M, Wilson A. Awakening dormant haematopoietic stem cells. Nat Rev Immunol 2010 Mar;10(3):201-209.
(53) Schroeder T. Hematopoietic stem cell heterogeneity: subtypes, not unpredictable behavior. Cell Stem Cell 2010 Mar 5;6(3):203-207.
(54) Cavazzana-Calvo M, Fischer A, Bushman FD, Payen E, Hacein-Bey-Abina S, Leboulch P. Is normal hematopoiesis maintained solely by long-term multipotent stem cells? Blood 2011 Apr 28;117(17):4420-4424.
(55) Hock H. Some hematopoietic stem cells are more equal than others. J Exp Med 2010 Jun 7;207(6):1127-1130.
(56) Morita Y, Ema H, Nakauchi H. Heterogeneity and hierarchy within the most primitive hematopoietic stem cell compartment. J Exp Med 2010 Jun 7;207(6):1173-1182.
(57) Abkowitz JL, Golinelli D, Harrison DE, Guttorp P. In vivo kinetics of murine hemopoietic stem cells. Blood 2000 Nov 15;96(10):3399-3405.
(58) Szilvassy SJ, Humphries RK, Lansdorp PM, Eaves AC, Eaves CJ. Quantitative assay for totipotent reconstituting hematopoietic stem cells by a competitive repopulation strategy. Proc Natl Acad Sci U S A 1990 Nov;87(22):8736-8740.
(59) Wang JC, Doedens M, Dick JE. Primitive human hematopoietic cells are enriched in cord blood compared with adult bone marrow or mobilized peripheral blood as measured by the quantitative in vivo SCID-repopulating cell assay. Blood 1997 Jun 1;89(11):3919-3924.
(60) van der Loo JC, Hanenberg H, Cooper RJ, Luo FY, Lazaridis EN, Williams DA. Nonobese diabetic/severe combined immunodeficiency (NOD/SCID) mouse as a model system to study the engraftment and mobilization of human peripheral blood stem cells. Blood 1998 Oct 1;92(7):2556-2570.
(61) Aiuti A, Cattaneo F, Galimberti S, Benninghoff U, Cassani B, Callegaro L, et al. Gene therapy for immunodeficiency due to adenosine deaminase deficiency. N Engl J Med 2009 Jan 29;360(5):447-458.

CHAPTER **4**

HETEROGENEITY OF YOUNG AND AGED MURINE HEMATOPOIETIC STEM CELLS REVEALED BY QUANTITATIVE CLONAL ANALYSIS USING CELLULAR BARCODING

Verovskaya, E., Broekhuis, M.J., Zwart, E., Ritsema, M., van Os, R., de Haan, G., Bystrykh, L.V.

Blood, 122(4):523-32 (2013).

ABSTRACT

The number of hematopoietic stem cells (HSCs) that contributes to blood formation and the dynamics of their clonal contribution is a matter of on-going discussion. Here we use cellular barcoding combined with multiplex high-throughput sequencing to provide a quantitative and sensitive analysis of clonal behavior of hundreds of young and old HSCs. The majority of transplanted clones steadily contributes to hematopoiesis in the long-term, although clonal output in granulocytes, T cells and B cells is substantially different. Contributions of individual clones to blood are dynamically changing; most of the clones either expand or decline with time. Finally, we demonstrate that the pool of old HSCs is composed of multiple small clones while the young HSC pool is dominated by fewer, but larger, clones.

INTRODUCTION

HSCs provide a paradigm of how stem cells maintain the function of an organ over the lifespan of an organism. Multiple views have been formulated on how HSCs contribute to hematopoiesis throughout life. The main hypotheses include clonal succession, which argues that distinct HSCs contribute to hematopoiesis in a sequential order (1-3), clonal stability, which implies that HSC clones steadily contribute to hematopoiesis (4), and more recently, dynamic repetition (5), where stem cells undergo a reversible switch between dormant and active self-renewal state also affecting their contribution to differentiation.

These models need to be reconciled with recent data that demonstrate profound heterogeneity within the HSC pool. Single-cell transplantation studies argue that individual HSCs are very different in terms of their lineage differentiation potential (6,7), repopulation capacity (7,8) and self-renewal ability (8). Additional levels of HSC heterogeneity were shown to exist with respect to their cycling behavior (2,3) and homing and migration ability (9,10). Further, proportions of HSCs with different characteristics change during aging (11-13). Aging results in an increase of the numbers of HSCs with a lymphoid-deficient differentiation program (6,11-13), impaired repopulation potential per stem cell (14,15) and decreased homing efficiencies (10,13). Given such a high level of heterogeneity, it is unclear how different HSCs co-exist in a polyclonal environment, and how many HSC clones are simultaneously active (16).

Reliable clonal analysis depends on the method of clonal labeling and detection. Cellular barcoding was recently introduced for the analysis of clonal fluctuations in T-cells (17) and hematopoietic cell populations (18,19). These techniques are based on inserting a random DNA sequence tag into viral vectors. Upon transduction, the tag is stably integrated into the genome of a target cell and is inherited by its progeny.

Here, we employed a cellular barcoding method of highly purified young and old HSCs in combination with multiplex parallel sequencing to perform detailed clonal tracking in the hematopoietic system. We characterized clone size, developmental potential and homing ability in several hundreds of young and old HSC clones. Our data provide detailed insight into functional differences, and homing ability of young and old HSCs. We demonstrate that most of transplanted clones consistently contribute to hematopoiesis, arguing in favor of clonal stability, while there are directional changes (growth/decline) and age-dependencies in the clonal output of individual HSCs.

MATERIALS AND METHODS

Mice

C57BL/6 (B6) mice were purchased from Harlan. C57BL/6.SJL and C57BL/6.SJLxC57BL/6 mice were bred in Central Animal Facility of University Medical Centre Groningen. C57BL/6J-kitW-41J/kitW-41J (W41) mice were obtained from Prof. E. Dzierzak (Rotterdam, Netherlands), and were crossed with B6.SJL to obtain CD45.1 W41 as described before (13). All experiments were approved by the local Animal Experimentation Committee.

Barcoded vector libraries

Design and validation of the MIEV-based barcode library (totally 800 barcodes) has previously been described (18). The pGIPZ vector was purchased from Open Biosystems and modified in the following way: an IRES-puro-miR cloning site was cut with BsrgI-MluI restriction enzymes and replaced with a short linker carrying ClaI, BamHI, SmaI restriction sites. A barcode of the following structure GTACAAGTAAGGNNNACNNNGTNNNCGNNNTANNNCANNNTGNNNGACGGCCAGTGAC was cloned via BsrgI-BamHI sites as previously described (18). While a modification was introduced in design of barcode backbone of two libraries, preparation of libraries and retrieval of barcodes was exactly the same. The pGIPZ-based library (450 barcodes) was used in the calibration experiment (Figure 1) for establishing the technical limitations of the method, whereas the MIEV library was used in all other experiments.

Purification and transduction of hematopoietic stem and progenitor cells

Bone marrow cells were isolated from bones of donor mice and stained with a cocktail of antibodies against Sca1, c-Kit, CD150, CD48 and lineage markers (Ter119, CD11b, CD3ε, B220 and Gr1) (BioLegend). Sorted lineage$^-$ Sca1$^+$ c-Kit$^+$ CD150$^+$ CD48$^-$ (LSK48$^-$150$^+$) cells were pre-stimulated in Stemspan medium (Stem Cell Technologies) supplemented with 300 ng/mL of stem cell factor (SCF), 1 ng/mL Flt3 ligand (both Amgen) and 20 ng/mL interleukin-11 (IL-11) (R&D Systems) for 24 hours. Production of MIEV-supernatant was described earlier (18). For production of pGIPZ supernatant 293T human embryonic kidney cells were transfected with pCMV, VSV-G and pGIPZ plasmids as previously described (20). After 24 hours medium was changed to Stemspan (Stemcell Technologies). Supernatant containing lentiviral particles was harvested after 12 hours and stored at -80ºC. Retro- and lentiviral transduction was performed overnight in viral supernatant supplemented with SCF, FLT3-ligand, IL-11 in retronectin-coated plates (Takara) in the presence of 2 μg/mL of polybrene.

Monoclonal liquid cultures of barcoded cells

20-22 hours post-transduction, GFP^+ cells were single-cell sorted in 96-well round-bottom plates in Stemspan medium supplemented with 10% fetal calf serum (FCS), 300 ng/mL of SCF, 1 ng/mL Flt3 ligand and 20 ng/mL IL-11. For vector copy number analysis, cells were harvested once the colony reached the size of about 30000 cells. For generation of calibration samples, monoclonal colonies were passaged to 12-well plate and expanded until a culture reached several million cells.

Hematopoietic cell transplantation and blood analysis

For limiting dilution transplantation of GFP^+ cells, cells were sorted 20-22 hours post-transduction. Doses of 10, 50, 70, 700, 800 and 1600 GFP^+ cells were transplanted into 6 to 10 W41 recipient animals irradiated with 3,5 Gy. For transplantations of unfractionated transduced cells donor cells were transplanted into lethally (9,5 Gy) irradiated B6 mice, simultaneously with 2×10^6 radioprotective cells of either W41 origin or from previously transplanted B6 mice. For limiting dilution studies of unfractionated cells, the doses were 100, 500 and 17360 cells per mouse and the transduction efficiency was 20%. Six to eight mice per dose were transplanted. As not all transduced cells express GFP at the moment of transplantation, gene transfer efficiencies were determined 3-5 days post-transduction in a small aliquot of cells left in culture. Blood samples were stained with antibodies against Gr1, B220 and CD3ε. Antibodies against CD45.1 and CD45.2 were used to discriminate between donor and recipient cells. Granulocytes ($Gr1^+$ side $scatter^{high}$ cells), B cells ($B220^+$) and T cells ($CD3^+$) were sorted. For estimation of HSC frequencies, we used extreme limiting dilution analysis software (ELDA) (21).

Barcode analysis in $LSK48^-150^+$ cells after transplantation

$LSK48^-150^+$ cell purification from transplant recipients was performed as described above with addition of antibodies against CD45.2 and CD45.1 for young and old HSC discrimination. Viable GFP^+ $lineage^-$ $Sca1^+$ $c\text{-}Kit^+$ $CD48^-$ $CD150^+$ donor-derived cells were individually sorted to generate monoclonal cultures as described above. In some cases, the origin (old or young donor) of a clone was established after monoclonal expansion and detection of CD45 polymorphisms using PCR and restriction analysis (22).

Barcode detection in monoclonal cultures

Genomic DNA was extracted from $LSK48^-150^+$ cell-derived colonies by the use of the REDExtract-N-Amp Tissue PCR Kit (Sigma-Aldrich). Barcode sequences were amplified with primers against MIEV vector sequences and sequenced as previously described (18). Sanger sequencing was used here since the majority of monoclonal colonies contained only one or two barcodes.

Design of multiplexing tags for deep sequencing

For the present study we generated a list of more than 300 individually tagged primers against the barcode vector sequence. The structure of multiplex primers ensured at least three nucleotide differences between different tags to allow unambiguous sample identification post-sequencing. In addition, such differences allow correction of single nucleotide substitution errors, if necessary (23).

Preparation of batches for multiplex sequencing

Extraction of genomic DNA was performed as described before (18). Individual samples for each deep sequencing run were amplified with assigned multiplexing primers. 35-cycle amplification was performed using DreamTaq Green mastermix (Fermentas). Each sample was amplified in duplicate. PCR products were purified with the use of the High Pure PCR Cleanup Micro Kit (Roche), and pooled. Before sequencing, pools of products were phosphorylated with polynucleotide kinase (Fermentas) (30' at 37^{o}C, inactivation for 10' at 75^{o}C) and an additional round of purification was performed using High Pure PCR Cleanup Micro Kit. Sequencing was performed using an Illumina HiSeq2000 sequencer in the sequencing facility of University Medical Centre Groningen or at BaseClear Group (Leiden, the Netherlands).

Data processing and noise removal

Custom-written scripts were applied to FASTQ sequence data to remove low quality reads and sequences occurring only once. Samples were retrieved on the basis of exact match to sample tag and adjacent primer sequence. Each sequencing batch comprised up to 211 samples; 10^3 to 10^6 reads per sample were retrieved. Barcode sequences were retrieved and cross-compared for linear similarity. In barcode pairs differing by a single nucleotide, the lower frequency barcode was removed.

RESULTS

Cellular barcoding combined with high-throughput sequencing permits sensitive quantitative clonal analysis

Before clonal analysis of biological samples can be reliably carried out *in vivo*, it was important to test the sensitivity and accuracy of our barcoding method. Therefore, defined mixtures of clonally expanded barcoded cells were analyzed in several dilutions (Figure 1). Purified lineage-negative LSK48^{-}150^{+} cells were transduced with a barcoded vector library and transgene-positive cells were single-cell sorted and cultured to generate large uniquely barcoded colonies (Figure 1A). Expanded cells were mixed in different ratios and samples ranging from 1000 to

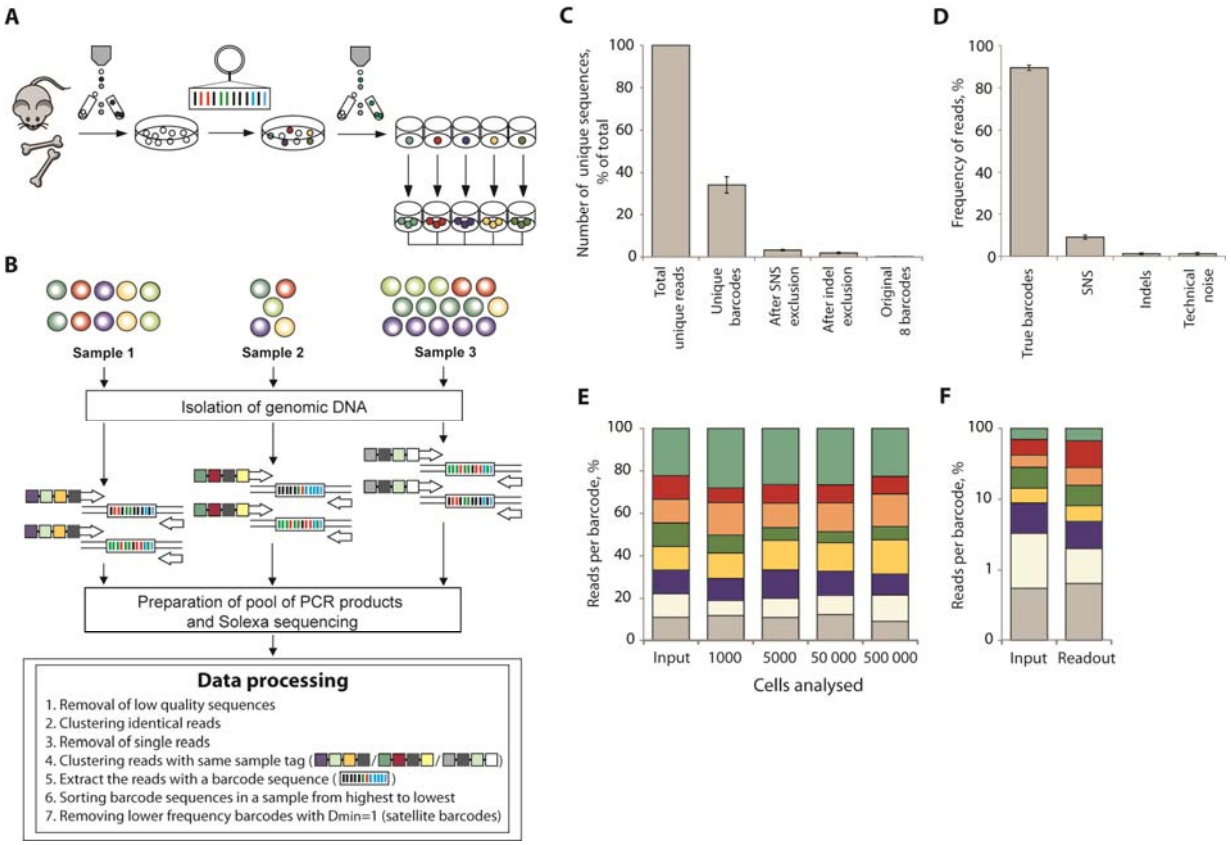

Figure 1. Combining cellular barcoding with multiplex deep sequencing – setup and method validation. (A) Individually barcoded LSK48^{-}150^{+} cells were monoclonally expanded in liquid culture. (B) Different numbers of cells from expanded barcode cultures were mixed to generate samples with different ratios of barcodes and different total cell numbers. After isolation of genomic DNA, individual samples were amplified with primers bearing multiplex tags. Pooled PCR products were analyzed on Illumina HiSeq 2000. Steps of data processing and noise filtering are described. D_{min} refers to minimal distance, or nucleotide difference between two barcodes. (C) Number of unique sequencing reads (sequences different at any sequence position from all other reads) that remain after removal of noise calculated based on calibration samples. (D) Calculated reads frequencies (proportions of total number of reads in a multiplexed sample) related to true barcodes and various sources of noise. (E) Distribution of barcode frequencies in calibration samples with different cell content (1000 to 500000). The original ratio of mixed barcodes was 1:1:1:1:1:1:1:2 (top barcode, green). Each color represents a distinct barcode. (F) Barcode analysis in a sample with highly unequal barcode composition (1:5:10:10:25:25:50:55). Barcodes comprising 0.55% of the total mix (gray) could be quantitatively detected.

500000 cells were compared (Figure 1B). Individual DNA samples were amplified with primers containing a unique multiplex tag and pooled multiplexed PCR products were sequenced (Figure 1B). After removal of low quality and single reads from each sample, we retrieved >100000 reads per sample. The initial mixture of clones contained eight barcodes, which were identified in ~88.5% of all sequencing reads per sample (Figure 1D). The remaining 11,5% of reads were constituted by 3578 to 5359 unique barcode sequences per sample and must be explained by

sequencing/PCR errors and technical noise. We observed that 96,5% of these unique barcode sequences occurred due to single nucleotide substitutions (SNS), derived from the true barcodes, and a further 1,7% due to single nucleotide insertion/deletion mutations of the true barcodes. Remaining sequences unrelated to true barcodes (technical noise) had low frequencies and cumulatively constituted ~1% of all reads (Figure 1D). The proportions of reads and unique barcode sequences related to different sources of noise are shown in Figure 1 (panels C and D). We developed a filtering protocol for noise removal and applied it for all other data sets described below.

Barcodes could be reliably quantified in as few as 1000 cells (Figure 1E) and clones that represent only 0,55% of the starting population could successfully be detected (Figure 1F). We estimated that the relative measurement error of barcode detection was on average 22% (Figure 1E). Thus, if the real contribution of a barcode is 10%, values for clonal contribution between 8% and 12% can be expected.

Concordance between estimated numbers of transplanted HSCs and barcode count in granulocytes, but not lymphocytes

To test whether this method is accurate and sensitive enough to follow clonal contributions *in vivo*, barcode readout was combined with limiting dilution analysis, currently the gold standard for HSCs quantification. This allowed us to assess whether the estimated numbers of barcoded clones agreed with numbers of HSCs predicted from limiting dilution. Further, it allowed us to establish an upper limit of expected clones in recipients transplanted with a high HSCs dose. To this end we transplanted different numbers of non-transduced freshly isolated or transduced barcoded LSK48^{-}150^{+} cells into irradiated recipients and measured donor and GFP^{+} chimerism (Figure 2A). Reconstituted mice were defined by at least 1% donor chimerism (for naive cells) or 1% GFP chimerism (for transduced cells) in granulocytes 16 weeks or longer post-transplantation. The HSC frequency equaled 1 in 9 (95% confidence interval (CI) 1 in 5-16,5) for freshly isolated LSK48^{-}150^{+} cells and decreased to 1 in 134 (95% CI 1 in 53,9-334) after transduction. When GFP^{+} cells were re-sorted prior to transplantation, the frequency of functional long-term repopulating stem cells further decreased to 1 in 243 (95% CI 1 in 132-448) (Figure 2B). It should be noted that 20-22 hours post-transduction when cells were resorted not all of the transduced cells already expresses GFP, and the final gene transfer efficiency was established in cell aliquot 3-5 days after transduction. The frequency of functionally-defined HSCs is higher among cells with delayed GFP expression (cells that express GFP later than 20-22 hours after gene transfer) (data not shown). This explains the lower HSCs prevalence in the sorted cells (Figure 2B).

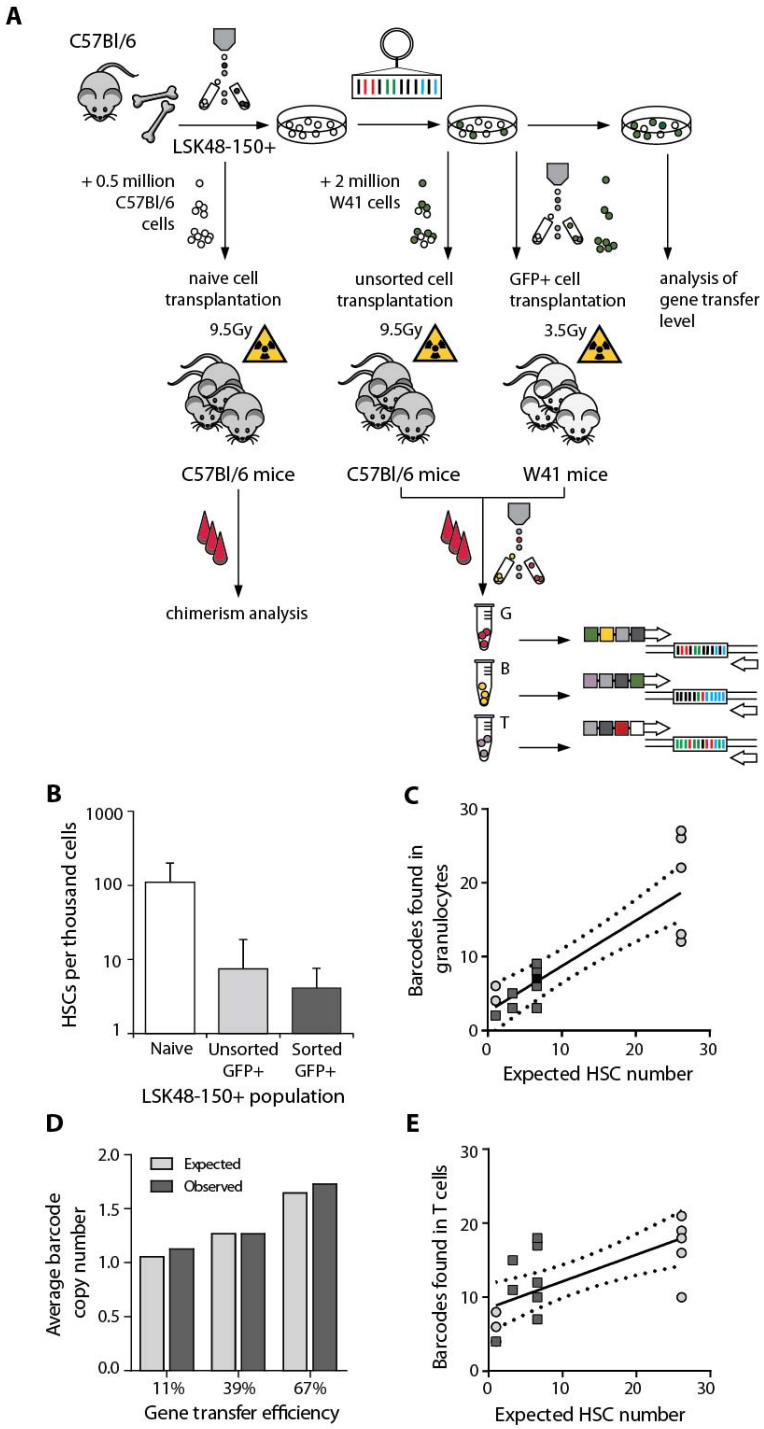

Figure 2. Relationship between the number of transplanted HSCs and detected barcodes. (A) Overview of limiting dilution experiments. LSK48^{-}150^{+} cells were purified by cell sorting. For establishing the frequency of functional HSCs in this population, naïve cells were transplanted into irradiated recipients. Alternatively, LSK48^{-}150^{+} cells were transduced with barcoded viruses and different doses of transduced cells were transplanted into irradiated hosts, either without (middle arrow) or with (right arrow) selection for GFP^{+}. Mouse strains, competitors and irradiation regimen used in every experiment are indicated. Congenic donor and recipient B6 animals were used to allow donor and recipient cell discrimination. Granulocytes (G), B- and T-lymphocytes were isolated for further barcode analysis at regular time points after transplant. (B) Assessment of HSCs frequencies by limiting dilution analysis in the naive LSK48^{-}150^{+} population, in transduced non-sorted and in sorted GFP^{+} cells. (C) Relationship between the expected number of transplanted HSCs and the number of barcodes, as detected in granulocytes. Each dot represents an individual mouse. Light grey circles indicate data generated using non-sorted cells, dark grey squares reflect experiments with sorted cells. In one mouse, data on granulocytes were not available for week 20, so week 28 data is shown instead with a black square. The best-fit line (line equation Y=0.61X+2.58) and 95% confidence intervals are plotted. Note that the 95% confidence interval includes X/Y intercept. (D) Observed and expected vector copy number (VCN) per transduced cell at different transduction efficiencies. The average number of barcodes in 15 to 22 colonies is depicted as a function of gene transduction efficiency. (E) Same as Panel C, but now data are shown for T cell clones. Equation for best-fit line was Y=0.36X+8.5.

To determine how many hematopoietic clones are active, we analyzed the barcode composition in short-lived granulocytes, and long-lived B and T lymphocytes, 18-20 weeks post-transplantation (Figure 2A) in 17 mice from cohorts of mice transplanted with sorted and unfractionated barcoded cells. A conservative threshold for barcode identification of 0,5% of total barcode reads was used. In each mouse the range of barcodes varied from 2 to 27 in granulocytes, 4 to 21 in T cells and 6 to 33 in B cells. We compared the number of barcodes detected in mice transplanted with different doses of transduced cells with HSC estimates from limiting dilution analysis (Figure 2 C, E). In granulocytes, the number of detected barcodes linearly increased with transplanted HSC dose (Figure 2C) indicating consistency between the two methods. This linearity, however, was less evident in the case of T cells (Figure 2E) and B cells (data not shown). While 95% confidence intervals for data of granulocytes intercepts the X/Y coordinate, this is not the case for T cells. These 17 reconstituted mice were used for time-course clonal tracking experiments described further.

Correction of clonal counts for multiple vector integrations

Integration of multiple vectors in a single cell can lead to an overestimation of stem cell counts. In order to correct for this we compared the occurrence of multiple barcodes integrating in single HSCs at different transduction efficiencies (11%, 39% and 67% GFP^+ cells). Transduced $LSK48^-150^+$ cells were monoclonally expanded to large colonies. Sanger sequencing was used to analyze the number of barcodes in 15-22 of such colonies for each group. Between 1 and 6 barcodes per clone were detected. The distributions of vector copy numbers per HSC followed a Poisson distribution (data not shown) and therefore could be considered a random event (24). The average number of vector insertions per cell varied from 1.1 at 11% gene transfer to 1.7 at 67% (Figure 2D). To calculate the number of active hematopoietic clones in blood, we divided the numbers of barcodes detected in blood by average barcode copy number at the respective transduction efficiency. For clonal analysis, these observations implied that the proportion of transduced HSCs with multiple integrated vectors may vary from 13% (at 11% GFP^+) to 40% (at 67% GFP^+).

Clonal composition of granulocytes, T cells, and B cells is stable over time, but these lineages are maintained by different subsets of primitive cells

Next, we studied whether the clonal composition of blood cells of the mice described above shows any signs of variation with time in a cohort of 9 mice transplanted with unfractionated barcoded cells. As an illustrative example, the barcode composition of granulocytes, T and B cells of one of these mice is shown in Figure 3A. Barcode contributions were compared among different time points and among the three blood lineages. To do so we used Pearson correlation as a measure of the similarity for all data points. Results of analysis in 9 mice are shown at

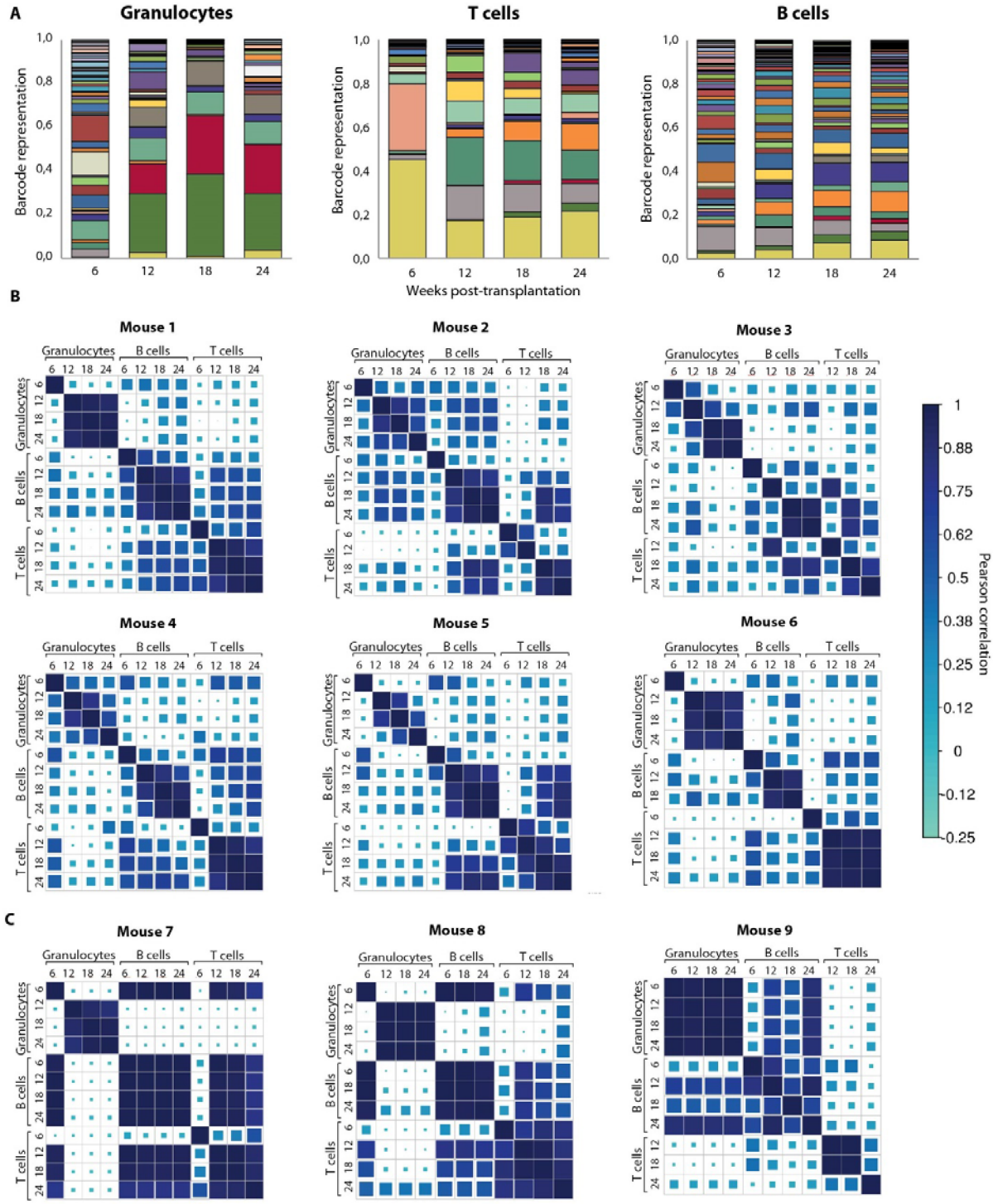

Figure 3. Clonal dynamics in mice transplanted with barcoded cells. (A) Barcode composition of granulocytes, T- and B lymphocytes in one of the mice transplanted with ~ 26 barcoded HSCs. Different colors reflect different barcodes. This and other mice shown in this figure originate from a cohort of mice transplanted with non-sorted transduced LSK48^{-}150^{+} cells (these mice are identified with light grey circles in Figure 2). (B) Pearson correlations between barcode compositions of three cell types at 4 time points from 6 mice transplanted with ~26 HSCs. Note good correlations between samples collected from 12 to 24 weeks within each of the cell types. Mouse 1

corresponds to the data shown in panel A. (C) Same, but for 3 mice transplanted with 100 barcoded cells. At this dose, Poisson distribution predicts that more than two-thirds of the positive animals are transplanted with a single HSC.

Figure 3. As expected, we observed that initial time points after transplantation are substantially different from later time points for the same lineage. At this time point short-lived progenitors will contribute to blood cell development. Starting from week 12, clonal composition within each lineage showed very good correlation, demonstrating that already at 12 weeks after transplant peripheral blood cell counts reflect HSC engraftment. Interestingly, when we analyzed mice transplanted with HSCs at limiting dilution, barcodes found in granulocytes, B cells and T cells were consistently different (Figure 3C). This provides a cautionary tale when interpreting limiting dilution experiments as truly clonal.

The barcode compositions between the lineages were only moderately correlated, indicating differences in clonal composition in each of the lineages. We followed the behavior of 350 clones found in 6 polyclonally-repopulated mice. Contributions of all clones to at least one of these three lineages is shown in Figure 4. Importantly, although we transduced highly purified cells that have robust multilineage potential, B and T lymphocyte populations were frequently supported by different clones.

Additionally, to establish the correlation between barcode composition of HSCs and mature blood cells after transplantation, we analyzed four hematopoietic cells types – blood granulocytes, T and B cells and bone marrow LSK48^{-}150^{+} cells - in a cohort of three mice (data not shown). LSK48^{-}150^{+} cells were cultured in cytokine-supplemented medium to generate colonies of about 30 000 cells. Monoclonal expansion served two purposes. First, it allowed us to confirm functionality of phenotypically defined HSCs, as it assesses the high proliferative potential of these cells. Second, the large number of cells allowed us to perform robust barcode analysis of each colony. Correlation analysis revealed that clonal spectrum of bone marrow and blood compartments were highly variable, ranging from high overlap between LSK48^{-}150^{+} and blood to medium and low correlation (data not shown).

Majority of clones in granulocytes and T cells are either expanding or declining in time

As long as 6 months after transplantation the same clones were persistently detected in the same cell types. Although this may be considered as “clonal stability”, this was only qualitative, meaning that no major clonal changes occurred between two consecutive time points among all mice and cell lineages. Quantitatively, however, many clones showed systematic changes in their contribution. A cohort of seven mice, representing 228 barcoded clones, was followed for a period of up to 1 year after transplant. We used this cohort to analyze time trends. Clonal

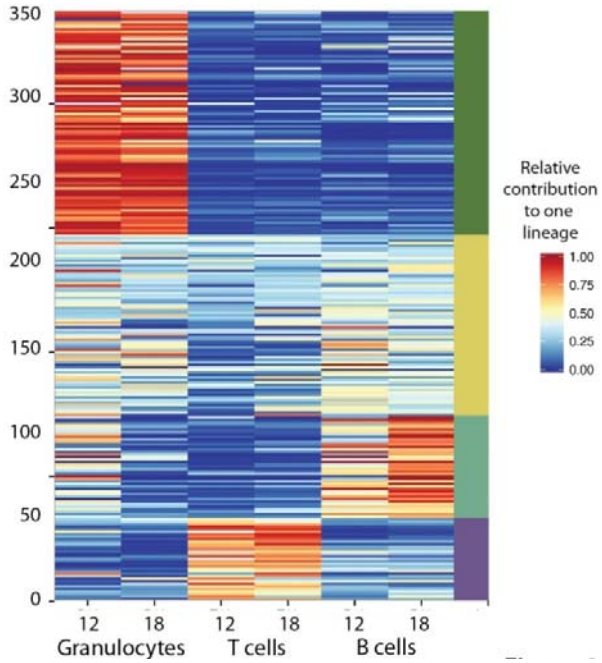

Figure 4. Relative lineage contributions of 350 barcoded clones found in 6 polyclonally-repopulated mice. Only clones that contributed at least 0,5% to one of the lineages 12-24 weeks post-transplant are shown. To assess relative contributions to granulocytes (and other lineages), the barcode representation in granulocytes was divided by total barcode representation (granulocytes + T cells + B cells). Clonal fluctuations within these mice is shown in Figure 3B.

kinetics in one of these mice is shown in Figure 5 (panels A and B). Since the number of data points was limited (5-6 time points) we restricted our dynamic analysis only to linear trends in time (positive or negative) using Pearson correlation. In this analysis, a positive correlation indicates consistent clonal growth, a negative correlation detects gradual decline, whereas the absence of correlation implies fluctuation of clone size around a certain value without a defined direction. We summarized the values of Pearson correlation for 107 individual clones in granulocytes and 187 clones in T cells (Figure 5C). The data were compared with a random model, thus allowing establishing thresholds for significant non-random dynamic behavior. As shown in Figure 5C, in randomly simulated data, most clones do not follow a time trend. Experimental data, on the other hand, show a clear bimodal distribution of correlation values, indicating that most clones undergo systematic change in their clone size. In granulocytes, most of the clones were declining in time, while in T cells both growing and declining populations were seen. Although it is possible that clonal fluctuations will differ from one experiment to the other, may depend on the number of clones that contributes to hematopoiesis, and may also reflect the presence of progenitor cells in the transplanted population, it is apparent that a large fraction of clones will systematically change in time. Therefore, barcode tracking over multiple time points is necessary for establishing the direction of clonal dynamics.

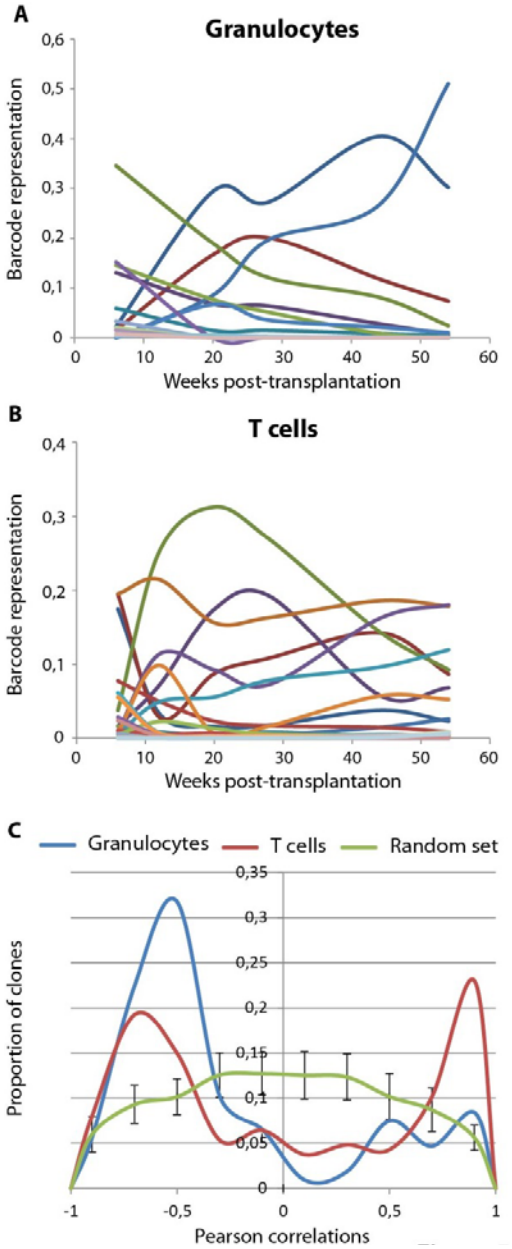

Figure 5. Clonal dynamics in long-term hematopoiesis. (A,B) Barcode fluctuations from 6 to 54 weeks after transplant are shown for one mouse in granulocytes (panel A) and T lymphocytes (panel B). Different colors represent different barcodes. (C) Pearson correlations of clonal sizes with time trend for clones detected at 0,5% or higher frequency at any of the time points in the respective lineage. Positive correlation indicates that the clone is consistently growing, and negative correlation reflects decline. Proportions of T cells (red) and granulocytes (blue) are plotted. The green line shows randomly expected correlations for 200 clones simulated 20 times (average values and standard deviations for 20 simulations are shown). Fluctuations in granulocytes reflect behavior of 107 barcodes, and in T cells – of 187 barcodes. Mice used for this analysis were transplanted with sorted barcoded cells (identified in Figure 2 with dark squares).

Clonal tracking of young and old HSCs

Finally, we asked how the composition of HSC pool changes with aging. To this end, we co-transplanted barcoded LSK48^{-}150^{+} cells isolated from young (4 months) and old (24 months) donors into 8 recipients. Studies of unmanipulated old and young HSCs have indicated a ~2-fold reduced functional activity of old cells (13). To compensate for the expected decrease of functional activity in old HSCs we initiated the experiment with LSK48^{-}150^{+} cells from young and

old mice co-transplanted in a 1:2 ratio (Figure 6A). 24-38 weeks post-transplantation, we analyzed blood chimerism and barcode composition of engrafted LSK48^{-}150^{+} bone marrow cells. Although two-fold more old LSK48^{-}150^{+} cells were transplanted, chimerism in the peripheral blood was predominantly derived from young HSCs (Figure 6B). Old HSCs contributed the least to lymphoid populations (Figure 6B). Interestingly, whereas old HSC were functionally inferior in producing peripheral cells, the frequency of old LSK48^{-}150^{+} cells in the bone marrow was substantially higher than predicted from blood values, indicating hampered differentiating activity of aged HSCs (Figure 6B). The number and size distributions of old HSC clones was subsequently determined. We sorted single transduced LSK48^{-}150^{+} HSCs of old and young origin

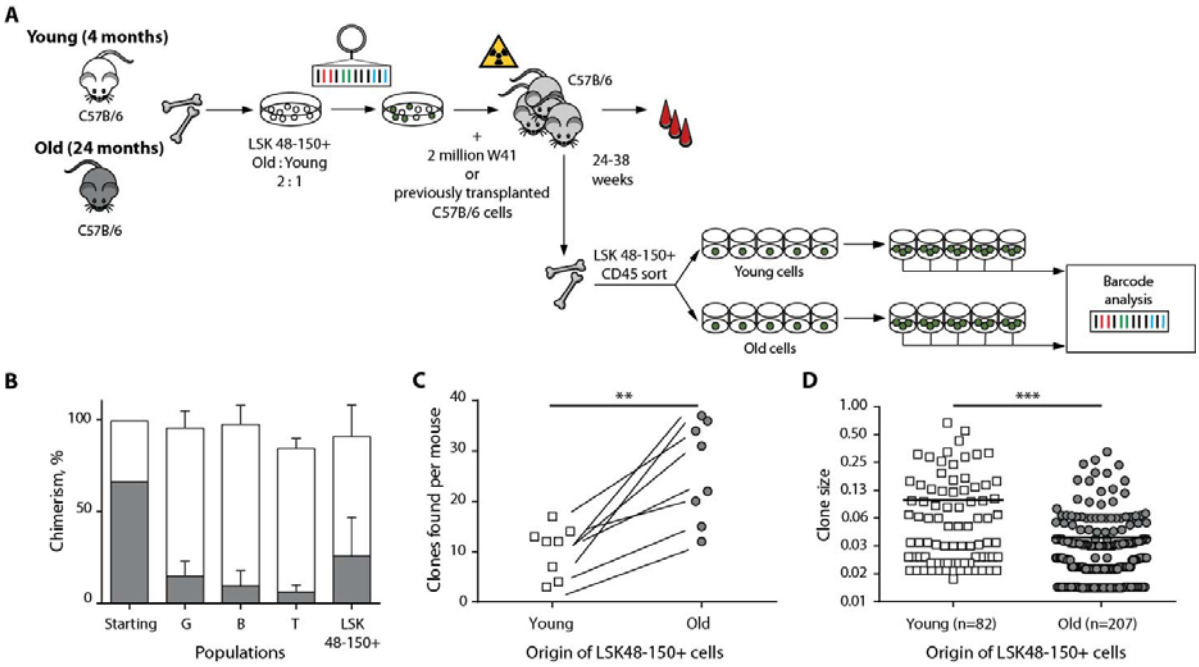

Figure 6. Analysis of HSC pool in mice transplanted with old and young cells. (A) Experimental setup. LSK48^{-}150^{+} cells were purified from CD45 congenic young (4 months) and old (24 months) donor mice, mixed in 1:2 ratio and transduced with the barcoded vector library. 20500 transduced cells were transplanted simultaneously with W41 or previously transplanted B6 cells into 2 cohorts of 4 lethally irradiated B6 mice. 6 mice were sacrificed 6 months post-transplantation, and 2 at 8 months after transplantation. GFP^{+} LSK48^{-}150^{+} cells of young and old origin were single-cell sorted in 96-well plates and expanded in liquid culture in presence of cytokines for subsequent barcode analysis. (B) Contribution from young (white bars) and old (grey bars) to different cell populations before (starting) and after transplantation. (C) Number of uniquely barcoded clones detected in expanded colonies of young and old LSK48^{-}150^{+} cells. Lines connect data points derived from the same mice. The number of clones detected within the old compartment was significantly higher than the number of clones within the young population (p=0.0011, paired two-sided T test). (D) Contribution of individual young and old LSK48^{-}150^{+} HSCs to the stem cell compartment. Horizontal lines indicate mean values. Young LSK48-150+ cells produced larger clones than old LSK48^{-}150^{+} cells (p<0.0001, two-tailed Mann Whitney non-parametric test).

from the 8 transplanted recipients into cytokine-supplemented medium and grew colonies. Between 38 and 146 single cell colonies per mouse were successfully expanded and barcode sequences were retrieved and analyzed. Each animal contained between 15 and 53 uniquely barcoded HSC clones. Contrary to our expectations (13), in each mouse, the number of unique old LSK48^{-}150^{+} clones exceeded the number of young clones (Figure 6C) in the 2:1 ratio of originally mixed HSCs. These results argue against the previously suggested decreased homing potential of old HSCs. However, old clones were significantly smaller than their young counterparts, resulting in lower overall contributions to the LSK48^{-}150^{+} pool and to blood cell production (Figure 6D). We did not observe enrichment for clones carrying multiple inserts nor larger clone size, both in young and old HSCs (data not shown).

DISCUSSION

Detailed investigations of HSC clonality in the hematopoietic system require reliable heritable marking of HSCs to trace their progeny in time (reviewed in (16)). In this paper, a barcoding method has been used to provide quantitative and dynamic tracking of individual HSCs regarding their lineage contribution and pool composition upon aging.

We show here that cellular barcoding reliably measures the number of clones contributing to hematopoiesis. The observed concordance between barcoding and limiting dilution in granulocytes indicates that most of HSC clones contribute to hematopoiesis. However, neither method can exclude the presence of quiescent stem cells. The notion of dormant HSCs is supported by the detection of clones that were prominently contributing to LSK48^{-}150^{+}, but not to mature blood cells (data not shown). Additional hematopoietic stress such as secondary transplantation or longer observation time might be necessary to activate these cells.

This study confirmed the view that blood sampling at early time points post-transplantation is not representative for long term repopulating cells (25). However, the clonal make-up of blood starting 3 months post-transplantation was consistent for all time points for all individual lineages. At the same time, tracking clones for longer time periods was essential for understanding the dynamics of clonal fluctuations. Our data agree best with the theory of clonal stability of hematopoiesis. However, the relevance of other models cannot be excluded, since the definition of clones varies and clonal behavior is heterogeneous. For instance, clonal fluctuation between active and quiescent states will be unnoticed if only a fraction of one barcoded clone will undergo those changes. We also cannot exclude that progeny of an originally barcoded HSC does not participate in hematopoiesis sequentially on the basis of their division history as proposed by the theory of clonal succession (1).

Although most differentiated cells originated from a common precursor, the number of uniquely barcoded clones in myeloid and lymphoid cells varied substantially. Interestingly, also

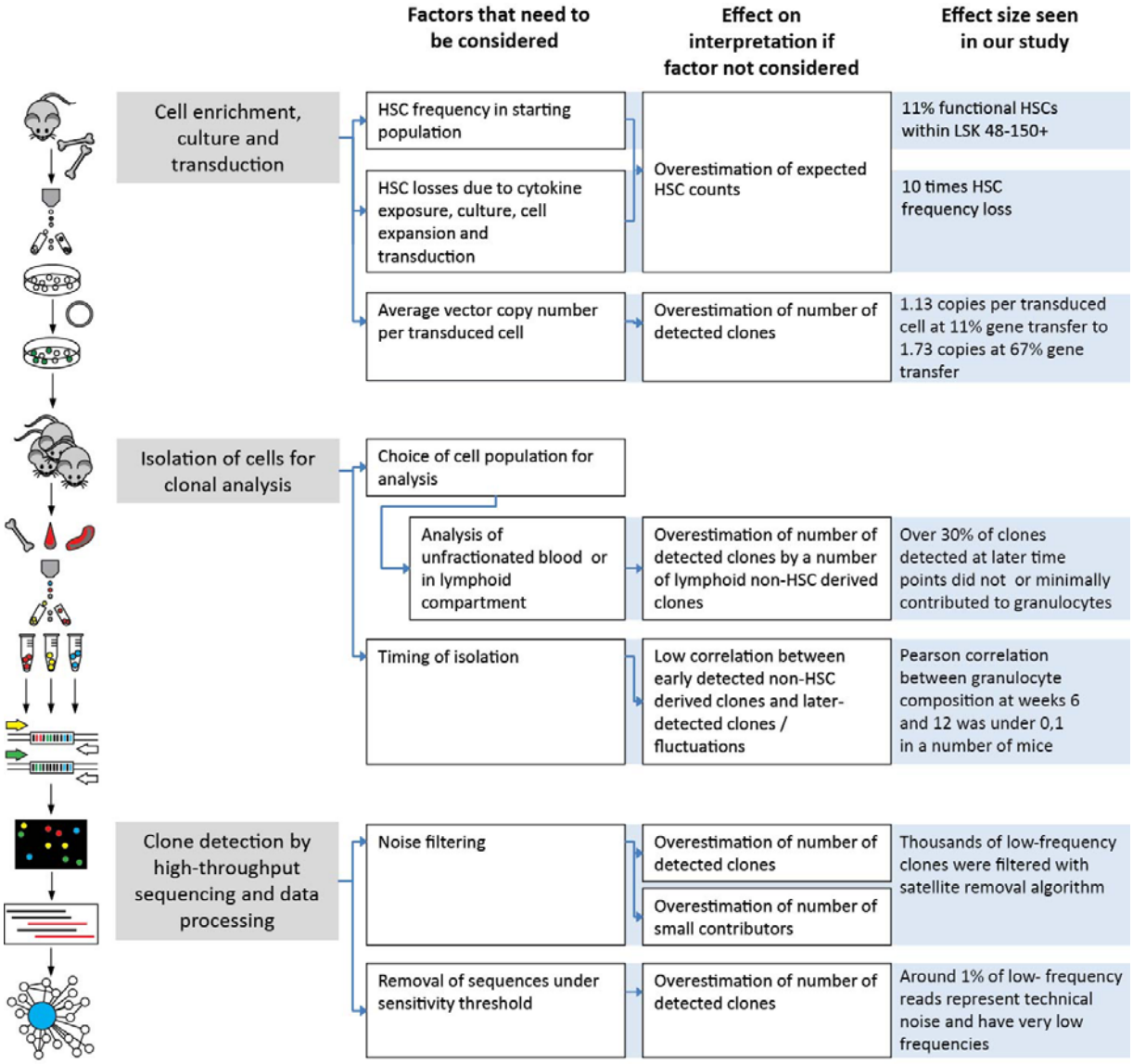

Figure 7. Factors that influence HSC clonal counts at different stages of analysis. Methodological steps of planning, implementation of experiments, and barcode data analysis are shown. For each step, we indicate how experimental setup and approach may influence conclusions of analysis. The diagram aims to point out the factors that we found critical in the currently described experiments but is not exhaustive. For instance, the setup of transplantation experiments, number of transplanted cells, irradiation regimens can also influence clonal counts.

the clonal repertoire of B and T cells was only moderately correlated, demonstrating that HSC clones do not only differentially contribute to the myeloid and lymphoid lineages, but similar bias was observed within the lymphoid lineage. In the future, barcode analysis of progenitor populations can help to identify the stages of differentiation where such commitment occurs.

Unexpectedly, in mice transplanted with a cell dose close to limiting dilution, lymphoid and myeloid populations were supported by different groups of barcodes, suggesting more hematopoietic clones than predicted (Figure 3C).

Previous studies by us and others (13,14) suggested a ~2-fold decrease in functionality of old purified stem cells. This was mostly attributed to a defect in homing of aged HSCs (10,13). In concordance with these data, we observed that transplantation of twice as many old than young LSK48^{-}150^{+} cells resulted in lower chimerism of old LSK48^{-}150^{+} cells in recipient animals (Figure 6). However, barcode analysis of individual LSK48^{-}150^{+} cells showed that the initial 2:1 relation between the number of old and young clones was preserved (Figure 6), although old clones were smaller compared to clones from young HSCs. This argues against the previously described homing defect of old HSCs (10,13). In addition, the output of old HSCs to mature blood lineages was drastically lower than of young cells. In classical competitive repopulation assays this would result in a lower frequency of old HSC, because the contribution of many old HSCs would remain below the detection threshold.

It has previously been suggested that transduction of hematopoietic cells leads to clonal dominance(26) and preferential survival of clones with multiple insertions (27). Observations made in our study did not confirm these data. Firstly, while we did detect abundant clones contributing more than 20% of a respective cell population, the number of barcoded clones found > 3 months post-transplantation correlated well with the expected HSC frequency, arguing against clonal selection (Figure 2). The presence of such large clones is likely reflecting intrinsic differences in repopulating capacity of HSCs. Secondly, we observed neither preferential survival nor larger clone size in HSC with multiple insertions among ~150 young and old HSCs 6 months post-transplantation (data not shown).

Recently, we discussed methodological constraints that can lead to misinterpretation of clonal tracking studies (16). Limited resolution, insufficient quantification, and failure of noise detection can severely impact conclusions of the analysis. Here we define principles of establishing cut-off values for minimizing noise. We demonstrate that this approach allows robust quantitative clonal tracking in clones present at frequencies of less than 1% of the transduced population. We believe that our current experimental design compares favorably to several related barcoding approaches, including our own, that had lower sensitivity (17), did not use high-throughput sequencing (18) or did not allow quantitative comparison of clones within the same animal (19). We briefly summarized several methodological steps that can influence clonal count analysis in Figure 7.

While the setup of our experiments, using a conservative 0,5% detection threshold, would preclude detection of more than 200 (equally represented) barcodes, we would argue that an

exact counting of HSCs in complex polyclonal situations is neither possible nor relevant. First, there is the problem of discrimination between signal and noise at the very small clone sizes. We expect that some false positive and false negative clones will persist through any statistical filtering of noise. Second, the question of biologically relevant clone size has to be addressed. Large clones are most important whereas small clones are biologically less relevant. We expect that some small contributors identified in our study would not satisfy the "stem cell definition" accepted in single-cell transplantation studies.

To conclude, our data document heterogeneity in multiple aspects of HSC functioning and demonstrates how HSC clones co-manifest themselves in polyclonally reconstituted recipient animals. Further studies will demonstrate how such HSC behavior can be influenced by hematopoietic stress or can contribute to the development of blood malignancies.

ACKNOWLEDGEMENTS

The authors thank H. Moes, G. Mesander and R.-J. van der Lei for expert cell sorting assistance, E. Weersing, E. Wojtowicz and L.Bosman for technical assistance, B. Dykstra and H. Schepers for valuable discussions and suggestions and assistance in the lab and P. van der Vlies and J. Bergsma for assistance with high-throughput sequencing. This study was supported by the Netherlands Organization for Scientific Research (VICI grant to G.d.H. and TopTalent grant to E.V.), the National Roadmap for Large Scale Infrastructure (Mouse Clinic for Cancer and Aging), and the Netherlands Institute for Regenerative Medicine (NIRM).

AUTHORSHIP CONTRIBUTIONS AND DISCLOSURE OF CONFLICTS OF INTEREST

E.V., L.V.B. and G.d.H. designed research; E.V., M.J.C.B., M.R. and R.v.O. performed research, E.V., L.V.B., and E.Z. analyzed and interpreted data, E.V., L.V.B. and G.d.H. wrote the manuscript with contributions from R.v.O. The authors declare no competing financial interests.

REFERENCES

(1) Kay HE. How Many Cell-Generations? Lancet 1965 Aug 28;2(7409):418-419.

(2) Wilson A, Laurenti E, Oser G, van der Wath RC, Blanco-Bose W, Jaworski M, et al. Hematopoietic stem cells reversibly switch from dormancy to self-renewal during homeostasis and repair. Cell 2008 Dec 12;135(6):1118-1129.

(3) Foudi A, Hochedlinger K, Van Buren D, Schindler JW, Jaenisch R, Carey V, et al. Analysis of histone 2B-GFP retention reveals slowly cycling hematopoietic stem cells. Nat Biotechnol 2009 Jan;27(1):84-90.

(4) Harrison DE, Astle CM, Lerner C. Number and continuous proliferative pattern of transplanted primitive immunohematopoietic stem cells. Proc Natl Acad Sci U S A 1988 Feb;85(3):822-826.

(5) Takizawa H, Regoes RR, Boddupalli CS, Bonhoeffer S, Manz MG. Dynamic variation in cycling of hematopoietic stem cells in steady state and inflammation. J Exp Med 2011 Feb 14;208(2):273-284.

(6) Benz C, Copley MR, Kent DG, Wohrer S, Cortes A, Aghaeepour N, et al. Hematopoietic stem cell subtypes expand differentially during

development and display distinct lymphopoietic programs. Cell Stem Cell 2012 Mar 2;10(3):273-283.
(7) Dykstra B, Kent D, Bowie M, McCaffrey L, Hamilton M, Lyons K, et al. Long-term propagation of distinct hematopoietic differentiation programs in vivo. Cell Stem Cell 2007 Aug 16;1(2):218-229.
(8) Ema H, Sudo K, Seita J, Matsubara A, Morita Y, Osawa M, et al. Quantification of self-renewal capacity in single hematopoietic stem cells from normal and Lnk-deficient mice. Dev Cell 2005 Jun;8(6):907-914.
(9) Mazurier F, Doedens M, Gan OI, Dick JE. Rapid myeloerythroid repopulation after intrafemoral transplantation of NOD-SCID mice reveals a new class of human stem cells. Nat Med 2003 Jul;9(7):959-963.
(10) Liang Y, Van Zant G, Szilvassy SJ. Effects of aging on the homing and engraftment of murine hematopoietic stem and progenitor cells. Blood 2005 Aug 15;106(4):1479-1487.
(11) Beerman I, Bhattacharya D, Zandi S, Sigvardsson M, Weissman IL, Bryder D, et al. Functionally distinct hematopoietic stem cells modulate hematopoietic lineage potential during aging by a mechanism of clonal expansion. Proc Natl Acad Sci U S A 2010 Mar 23;107(12):5465-5470.
(12) Cho RH, Sieburg HB, Muller-Sieburg CE. A new mechanism for the aging of hematopoietic stem cells: aging changes the clonal composition of the stem cell compartment but not individual stem cells. Blood 2008 Jun 15;111(12):5553-5561.
(13) Dykstra B, Olthof S, Schreuder J, Ritsema M, de Haan G. Clonal analysis reveals multiple functional defects of aged murine hematopoietic stem cells. J Exp Med 2011 Dec 19;208(13):2691-2703.
(14) Rossi DJ, Bryder D, Zahn JM, Ahlenius H, Sonu R, Wagers AJ, et al. Cell intrinsic alterations underlie hematopoietic stem cell aging. Proc Natl Acad Sci U S A 2005 Jun 28;102(26):9194-9199.
(15) Morrison SJ, Wandycz AM, Akashi K, Globerson A, Weissman IL. The aging of hematopoietic stem cells. Nat Med 1996 Sep;2(9):1011-1016.
(16) Bystrykh LV, Verovskaya E, Zwart E, Broekhuis M, de Haan G. Counting stem cells: methodological constraints. Nat Methods 2012 May 30;9(6):567-574.
(17) Schepers K, Swart E, van Heijst JW, Gerlach C, Castrucci M, Sie D, et al. Dissecting T cell lineage relationships by cellular barcoding. J Exp Med 2008 Sep 29;205(10):2309-2318.
(18) Gerrits A, Dykstra B, Kalmykowa OJ, Klauke K, Verovskaya E, Broekhuis MJ, et al. Cellular barcoding tool for clonal analysis in the hematopoietic system. Blood 2010 Apr 1;115(13):2610-2618.
(19) Lu R, Neff NF, Quake SR, Weissman IL. Tracking single hematopoietic stem cells in vivo using high-throughput sequencing in conjunction with viral genetic barcoding. Nat Biotechnol 2011 Oct 2;29(10):928-933.
(20) Kranc KR, Schepers H, Rodrigues NP, Bamforth S, Villadsen E, Ferry H, et al. Cited2 is an essential regulator of adult hematopoietic stem cells. Cell Stem Cell 2009 Dec 4;5(6):659-665.
(21) Hu Y, Smyth GK. ELDA: extreme limiting dilution analysis for comparing depleted and enriched populations in stem cell and other assays. J Immunol Methods 2009 Aug 15;347(1-2):70-78.
(22) Ramos CA, Zheng Y, Colombowala I, Goodell MA. Tracing the origin of non-hematopoietic cells using CD45 PCR restriction fragment length polymorphisms. BioTechniques 2003 Jan;34(1):160-162.
(23) Bystrykh LV. Generalized DNA barcode design based on Hamming codes. PLoS One 2012;7(5):e36852.
(24) Fehse B, Kustikova OS, Bubenheim M, Baum C. Pois(s)on--it's a question of dose.. Gene Ther 2004 Jun;11(11):879-881.
(25) Jordan CT, Lemischka IR. Clonal and systemic analysis of long-term hematopoiesis in the mouse. Genes Dev 1990 Feb;4(2):220-232.
(26) Kustikova O, Fehse B, Modlich U, Yang M, Dullmann J, Kamino K, et al. Clonal dominance of hematopoietic stem cells triggered by retroviral gene marking. Science 2005 May 20;308(5725):1171-1174.
(27) Modlich U, Kustikova OS, Schmidt M, Rudolph C, Meyer J, Li Z, et al. Leukemias following retroviral transfer of multidrug resistance 1 (MDR1) are driven by combinatorial insertional mutagenesis. Blood 2005 Jun 1;105(11):4235-4246.

CHAPTER **5**

ASYMMETRY IN SKELETAL DISTRIBUTION OF MURINE HEMATOPOIETIC STEM CELL CLONES AND THEIR EQUILIBRATION BY MOBILIZING CYTOKINES

Verovskaya, E., Broekhuis, M.J., Zwart, E., Weersing, E., Ritsema, M., Bosman, L.J., van Poele, T., de Haan, G., Bystrykh L.V.

Journal of Experimental Medicine, 211:487-497 (2014)

ABSTRACT

Hematopoietic stem cells (HSCs) are able to migrate through the blood stream and engraft bone marrow niches. These features are key factors for successful stem cell transplantations that are used in cancer patients and in gene therapy protocols. It is unknown to what extent transplanted HSCs distribute throughout different anatomical niches in the bone marrow, and whether this changes with age. Here we determine the degree of hematopoietic migration at a clonal level by transplanting individual young and aged murine HSCs labeled with barcoded viral vector, followed by assessing the skeletal distribution of hundreds of HSC clones. We detected highly skewed representation of individual clones in different bones at least 11 months post-transplantation. Importantly, a single challenge with the clinically relevant mobilizing agent G-CSF caused rapid redistribution of HSCs across the skeletal compartments. Old and young HSC clones showed a similar level of migratory behavior. Clonal make-up of blood of secondary recipients recapitulates the barcode composition of HSCs in the bone of origin. These data demonstrate a previously unanticipated high skeletal disequilibrium of the clonal composition of HSC pool long-term after transplantation. Our findings have important implications for experimental and clinical and stem cell transplantation protocols.

INTRODUCTION

Continuous generation and regeneration of all blood and immune cells over the lifespan of an organism is ensured by a limited number of HSCs. The vast majority of HSCs reside in the bone marrow, while a small fraction of functional HSCs can be found in the blood circulation, both in mice and humans (1-4). In early development, the ability of HSCs to migrate and engraft niches is important at the stage when HSCs exit the fetal liver and populate the bone marrow (5). In adults, HSCs have been shown to move towards the site of injury or inflammation and participate in tissue repair (6). The migrating ability of HSCs is routinely used in clinical transplantation and gene therapy protocols, which are applied in the treatment of an increasing number of hematopoietic and non-hematopoietic diseases. Thus far, it is unknown how individual HSC clones migrate and distribute among skeletal niches after transplantation and how this is affected by mobilization-inducing cytokines.

Our limited knowledge of HSCs migration is primarily based on results from parabiotic rodents sharing a common circulation (2,8-10). These studies suggest that egress of HSCs into blood is continuous. Migrating cells are capable of re-engrafting the bone marrow and further contributing to hematopoiesis (10). Based on approximate calculations it was claimed that 1-5% of all HSCs are circulating daily (11). If this claim were correct, HSC distribution within the same mouse or across parabiotic mice would approach an equilibrium within a few months. Direct measurements of chimerism in parabiotic mice, however, demonstrated relatively slow rates of equilibration (10). Although this rate was dramatically increased upon administration of granulocyte-colony stimulating factor (G-CSF), it did not result in full equilibration of HSCs between parabiotic mice (8).

G-CSF-induced mobilization is routinely used in clinical bone marrow transplantation and gene therapy protocols, allowing the harvest of HSC-enriched fraction from the donors' blood (12). Stem cell mobilization in patients has been claimed to decline with age (13,14), however experimental data underlying this phenomenon are limited and contradictory. While a number of studies found a homing defect of old murine HSCs (15,16), other studies suggest that G-CSF-induced mobilization in aged mice was more efficient than in young (17).

In this study, we analyzed post-transplantation skeletal localization of hundreds of young and old hematopoietic clones. To track individual stem cell clones, we labeled highly-purified HSCs with a viral barcode label prior to transplantation (7,18). We questioned whether old and young HSCs would respond differently to mobilizing stimuli. Our data demonstrate that migration of clones under steady-state conditions is very limited, such that clonal distribution does not reach an equilibrium up to 11 month post-transplantation. However, migration was strongly activated and led to complete clonal equilibration upon a single mobilizing challenge. Clonal differences in

HSC composition of specific skeletal sites were inherited upon secondary transplantations from those particular bones, and also resulted in different functional activity in secondary recipients.

MATERIALS AND METHODS

Mice

C57BL/6 mice were purchased from Harlan and C57BL/6.SJL and C57BL/6.SJLxC57BL/6 (all referred to as B6) mice were bred in the Central Animal Facility of University Medical Centre Groningen. C57BL/6J-kitW-41J/kitW-41J (W41) mice were originally obtained from Prof. E. Dzierzak (Rotterdam, Netherlands), and were crossed with B6.SJL to obtain CD45.1 W41 as described before (16). All experiments were approved by the University of Groningen Animal Experimentation Committee.

Purification and transduction of bone marrow cells

LSK48^{-}150^{+} cell sorting and transduction was performed as described before (18). Briefly, bone marrow cells were isolated from bones of hind legs, spines, and sterna of naive B6 mice and stained with antibodies against Sca1, c-Kit, CD48, CD150, CD3ε, Gr1, CD11b, Ter119 and B220. LSK48^{-}150^{+} cells were sorted using MoFlo XDP and MoFlo Astrios (Beckman Coulter) cell sorters, and pre-stimulated in StemSpan medium (StemCell Technologies) supplemented with 300 ng/mL of stem cell factor (SCF), 1 ng/mL Flt3 ligand (both Amgen) and 20 ng/mL interleukin-11 (IL-11) (R&D Systems) for 24 hours. Cells were transduced with viral supernatant containing barcoded MIEV vectors in retronectine-covered plates (Takara) in the presence of 2 μg/mL of polybrene (Sigma).

Bone marrow cells transplantation

20-22 hours post-transduction, transduced cells were transplanted into lethally irradiated recipients (9-9.5 Gy) in the presence of 1-2 million radioprotective cells from W41 or B6 origin. Efficiency of gene transfer, measured by flow cytometry (LSR-II, BD Biosciences), was similar in all experiments and ranged from 19 to 34% GFP^{+} cells. CD45 congenic B6 and W41 strains were employed to discriminate between donor and recipient cells and to distinguish young and aged cells in co-transplant setting. In the first young/aged cohort, aged cells were CD45.2^{+}, young cells – CD45.1^{+}/CD45.2^{+} and recipients and donors of radioprotector cells – CD45.1^{+}. In the second cohort, aged cells were CD45.1^{+}, young cells – CD45.2^{+}, recipients/radioprotector cells – CD45.1^{+}/CD45.2^{+}.: Young donors were 4 months old, aged donors were, 24 months old and recipient animals were 2-6 months old. Transplanted transduced cells supported robust multilineage blood production. Donor and GFP chimerism determined in blood 6 months post transplantation as described below. In old/young co-transplants (Figure 1 and Figure 2J), old cells contributed 8.0±7.3% of B cells (17.2±13.2% GFP), 6.3±5.1% of T cells (17.1±14.1% GFP) and

15.7±6.8% granulocytes (22.7±16.0% GFP), while young HSCs generated 81.2±8.5% granulocytes (14.3±10.9% GFP), 75.4±12.6% T cells (9.7±6.6% GFP) and 89.5±8.8% B cells (9.7±4.2% GFP). In recipients of young cells (Figure 2, panels B-F, Figure 3), donor HSCs produced 81.0±8.2% granulocytes (15.5±8.2% GFP), 65.1±11.8% B cells (9.9±4.8% GFP) and 61.2±8.0% T cells (10.4±4.7% GFP). In recipients of old HSCs (Figure 2, panels G-I), these values were as follows: 87.8±4.3% granulocytes (23.1±5.5% GFP), 75.9±10.4% B cells (16.8±10.7% GFP), 60.9±16.8% T cells (14.1±15.6% GFP). For secondary transplantations, 5-8 million whole bone marrow cells were transplanted into lethally-irradiated recipients.

G-CSF administration

To induce mobilization, we performed 2 injections of pegylated G-CSF intraperitonially (25 µg/mouse) with a three-day interval between treatments. In two mice sacrificed one week after G-CSF administration, the efficiency of mobilization was confirmed by the presence of colony forming units (CFUs) in the blood samples.

Isolation of LSK48$^-$150$^+$ from distinct skeletal locations after transplantation

Transplant recipients were sacrificed by cervical dislocation under isoflurane anaesthesia. Bones were isolated separately into four groups: 1) bones of fore limbs and sternum (front), 2) left femur, tibia, and left half of pelvic bone (left hind bones), 3) right femur, tibia, and right half of pelvic bone (right hind bones), 4) bones of spine. After removal of muscle, cells were harvested by crushing bones in lysis solution (NH_4Cl) and filtered through a 100-µm filter to remove debris. Nucleated cells were washed and stained with an antibody cocktail. Cells were washed and resuspended in 1 µg/mL solution of propidium iodide (PI). Antibodies were directed against Sca1, c-Kit, a panel of lineage markers (Ter119, CD11b, CD3ε, B220 and Gr1), CD48 and CD150 to discriminate different cell populations. CD45.1 and/or CD45.2 staining was used further to distinguish donor and recipient cells. LSK CD48$^-$ cells from different locations were isolated in "enrich mode", donor-derived (young and/or old) LSK48$^-$150$^+$ GFP$^+$ cells were further single-cell sorted in round-bottom 96-well plates in cytokine-supplemented medium to generate monoclonal cultures. Only 60 inner wells were used to minimize medium evaporation. ~120 cells were sorted per bone group per mouse.

Monoclonal expansion of LSK48$^-$150$^+$ cells

Colony growth was assessed by light microscopy at low (5x) magnification at several time points 10-35 days post-sorting. Most large colonies formed within two weeks of culture. Colonies reaching ~30000 cells or more (colony diameter reaching ~60% of the diameter of the viewing field) were used for barcode analysis. ~45% of the sorted LSK48$^-$150$^+$ cells formed large colonies and were harvested (this efficiency varied from 38 to 49% over four cohorts of mice used for

analysis). In some cases, young/old origin of the clones was established after expansion by PCR and restriction analysis for CD45 polymorphisms (19).

Clonal analysis of LSK48$^-$150$^+$ colonies

Genomic DNA was extracted and barcoded DNA was amplified with primers against internal vector sequence. Sanger sequencing was used to detect barcodes present in the samples. When less than 5 colonies could be retrieved from a bone marrow sample, data were excluded from analysis.

Clonal analysis of spleen samples

Spleens were excised, weighed and pressed through a 100-µm filter to generate single cell suspensions. Red blood cells were lysed in NH_4Cl solution and cell suspensions were stained with an antibody cocktail as described above. Splenic LSK48$^-$150$^+$ GFP$^+$ cells were individually sorted to generate monoclonal cultures as described above.

Clonal analysis of blood samples

Clonal analysis in blood was performed by next generation deep sequencing, as previously described (18). Briefly, blood samples were collected every 6 weeks for at least 6 months and at the moment of sacrifice. Erythrocytes were lysed in NH_4Cl buffer and cells were stained with fluorophore-conjugated antibodies against Gr1, B220, CD3ε, CD45.1 and CD45.2, washed and resuspended in propidium iodide solution. Mature blood populations were purified by FACS. Viable (propidium iodide$^-$) granulocytes (Gr1$^+$ side scatter high), T cells (CD3$^+$) and B cells (B220$^+$) of donor origin (CD45.1$^+$) were sorted. Barcode sequence was amplified with tagged primers to allow multiplex analysis, pooled and analyzed on Solexa Illumina platform. Data extraction was performed using in-house developed scripts in Python, Perl, Bioconductor, R and VBA, as previously described (18). In short, high quality reads were extracted and compressed into sets of unique sequences. Samples were retrieved on the basis of exactly matching the primer tag. Barcode sequences were extracted using Motif Occurrence Detection Suite (MOODS) (20). Barcodes were sorted by read frequencies in descending order, barcode pairs varying by a single nucleotide were compared and lower frequency barcodes removed (removing barcodes with single-nucleotide substitution). Unique barcodes with frequencies under 0.5% of total within the sample were removed (technical threshold). Samples with less than 1000 reads were excluded from further analysis.

Statistical analysis

For testing the statistical significance of variations of clone size among different bone groups, experimental data series were compared with expected random fluctuation around the average.

As the number of barcoded clones retrieved from different bones varied, we first normalized all clone frequencies to the bone location with the smallest number of successfully retrieved clones. A custom script in Python was used to test the probability of random deviation from average (10000 trials) for a given set of a data points (number of bone compartments), as well as total sequence reads (number of trials). The script allows to simulate distribution of a certain number of reads (normalized number of reads related to a single clone) over several bone groups. Based on 10000 trials, the script calculates probability of observing a deviation from the average in a single bone. When the probability to detect the observed deviation was less than $p < 0.05$, this deviation was considered significant. This protocol allows to identify all significantly deviating barcodes individually.

For correlation and linear regression analyses, coefficients of correlation (in the text *r* for Pearson, and ρ for Spearman) and p-values were derived using GraphPad Prism (GraphPad Software, La Jolla, USA). For linear regression analysis, when the slope deviated from zero, this used as an indicator of linear relationship between the populations. For comparing RSD values in G-CSF treated and non-treated animals we used non-parametric Mann Whitney test which allows comparison of not-normally distributed data. P-values were calculated using GraphPad Prism.

Throughout this paper we mostly employed correlation coefficients (Pearson and Spearman) as a measure of similarity between clonal composition of different samples. These functions are quantitative and therefore have particular limitations, especially in case of poorly overlapping clone sets and sets with small numbers of barcodes. Nevertheless, they are well known and widely used, and when used properly provide sufficient statistical information.

RESULTS

Quantification of HSC migration by clonal analysis of distinct skeletal locations

Experiments with parabiotic animals demonstrated quite a slow exchange of HSC pools during an observation time up to 22 weeks (8,10). Other authors, however, claimed a continuously high level of hematopoietic stem cell migration (11). Based on these observations one would predict the clonal composition of the HSC compartment in all skeletal sites to equilibrate within a few months post-transplantation. To address the extent of skeletal distribution of old and young HSC clones, we compared the barcode composition in four skeletal sites in recipients of barcoded bone marrow cells (Figure 1A and B). Purified young (4 months) and aged (24 months) lineage$^-$ Sca1$^+$ c-Kit$^+$ CD48$^-$ CD150$^+$ (LSK48$^-$150$^+$) cells were mixed and transduced with a barcoded vector library and transplanted into irradiated hosts. To achieve a statistically sufficient number of young and old clones, several spatially closely related bones were pooled (Figure 1B). Young and old stem cells were isolated from four distinct skeletal regions: 1) spine;

2) sternum and bones of fore limbs; 3) left femur, left tibia and left half of pelvic bone; 4) right counterparts of these bones (Figure 1B). Single GFP^+ $LSK48^-150^+$ cells from young and old origin were sorted from primary recipients and clonally expanded in the presence of cytokines to confirm the high proliferative capacity of these cells and to generate enough cells for robust barcode analysis (Figure 1A).

Both old and young HSC clones demonstrate a highly skewed skeletal distribution for at least 11 months post-transplantation

Mice were sacrificed 6 to 11 months post-transplantation to assess the dynamics of the migration process. Six mice were analyzed 6-month post-transplant, while additional groups of two mice each were analyzed 10 and 20 weeks later (at 8.5 and 11 months, respectively). Approximately 100 colonies per mouse (average 98 ± 35; ranging from 38 to 160) were obtained and barcodes successfully analyzed. Detailed results of such an analysis for one of the mice are shown in Figure 1C. The distribution of different barcoded clones was highly diverse: while some clones were detected in all four analyzed locations (e.g. bottom (pink, #5) clone in panel on old cells), others were found only in one site. Such heterogeneity of spatial distribution was observed both for old and young HSC clones, and it was not related to the total number of clones found in an animal. The observed differences in clonal distribution were of two kinds: some clones were present in some bones and absent in others. Other clones were present in all observed bones, yet their sizes were very different. As differences in the occurrence of barcode frequencies might be purely random, or related to differences in the total number of clones analyzed (i.e. minor error of measurement due to a relatively small colony number), we tested whether the observed skewed distributions were statistically significant or rather happening by chance. To this end the data set was tested against the null hypotheses of equal skeletal distribution of HSCs. Only relatively abundant clones represented by at least 4 colonies were taken for analysis (smaller clones per definition cannot be found in 4 locations). The data for all mice sacrificed 6 months post-transplantation are summarized in panels D (clones derived from young HSCs) and E (clones derived from old HSCs). For individual clones, the contribution to the local stem cell compartment size at different locations could vary as much as from 8 to 83% or 0 to 79%, (Figure 1D and E). It appeared that 8 out of 42 clones were significantly deviating from the expected average ($p<0.05$, marked with asterisks (*)). To test whether skeletal distribution was equalized in time, 4 additional mice were analyzed 10 and 20 weeks later (Figure 1F and G). Similarly to clones analyzed 6 months post-transplantation, we detected skewed clonal representation among the 4 skeletal sites. For 3 of 17 clones, this skewing was statistically significant, confirming that local dominance of HSC clones in certain skeletal regions is retained for at least 11 months after transplantation.

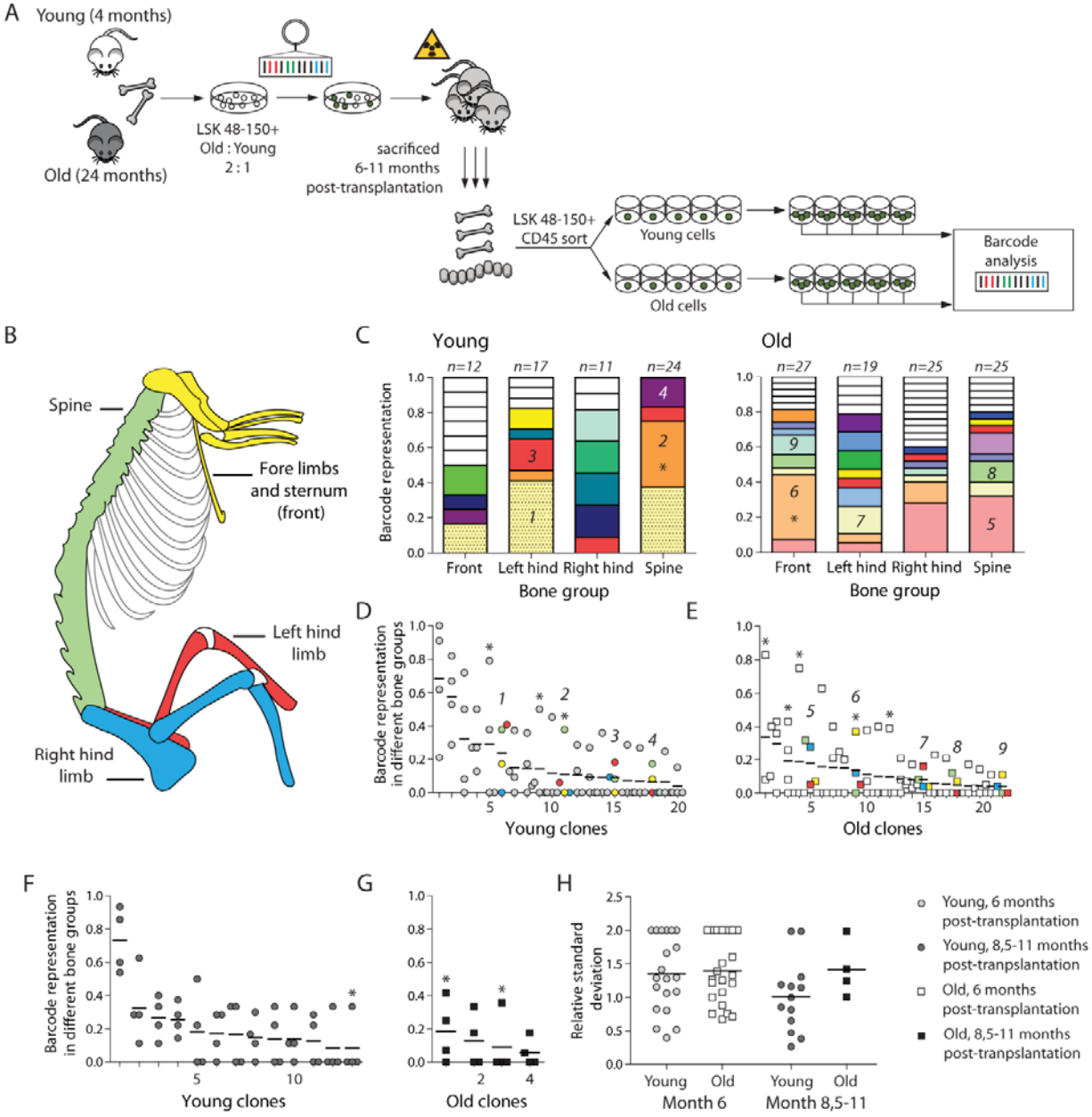

Figure 1. Old and young HSCs are asymmetrically distributed among various skeletal niches and migrate at a similar rate. (A) Experimental setup. Bone marrow cells were isolated from young and aged donors and LSK48^{-}150^{+} cells were separately purified by cell sorting, and subsequently pooled. After transduction with a barcoded vector library, cells were transplanted into lethally irradiated animals. Recipient mice were sacrificed 6-11 months post-transplant. Bone marrow cells were isolated from four skeletal locations and LSK48^{-}150^{+} cells of young and old origin were single-cell sorted into cytokine-rich medium. Colonies expanded to ~30 000 cells were harvested for barcode analysis. (B) Schematic representation of the four bone groups that were analyzed. Bones analyzed in the same group are shown in the same color. (C) Skeletal distribution of barcoded clones in young (left) and old (right) LSK48^{-}150^{+} cells in a single mouse, analyzed 6 months post-transplantation. Clones that are colored white are unique, i.e. were found only once. The asterisks identify clones with distributions that are significantly different from random ($p < 0.05$ based on Monte Carlo simulations). The number of barcoded colonies analyzed for every

skeletal location is indicated above the column. Clones found in 4 or more colonies and further used for distribution analysis are numbered. (D) Summary of all clones of young origin, found in 4 or more colonies 6 months post-transplantation (n=6, data were derived from two cohorts of mice). X-axis shows the rank of clone size, based on average value over all 4 locations, ranked from most frequent to the least frequent. Y-axis shows the contribution of a certain clone to the total pool in one location. Dots on the same vertical line reflect the size of a particular clone in each of the 4 different locations. The average contribution for a given clone is shown with a horizontal line. As in panel C, clones significantly deviating from equal distribution ($p < 0.05$, based on Monte Carlo simulations), are marked with an asterisk (*). For clarity, four numbered clones from panel C are colored according to the bone group colors as in panel B. (E) Same as panel D, but shown for old clones (summary of two independent cohorts of mice, n=6). (F) Same as panel D, but for young clones from mice sacrificed 8.5-11 months post-transplantation (one cohort, n=4). (G) Same as panel F, but for old clones (n=4). (H) Summary of the relative standard deviations for all clones shown in panels D-G. Horizontal lines indicate mean value.

Since previous studies demonstrated a decreased homing or binding to stroma of old HSCs (15-17), we questioned whether the extent of migration/spatial distribution of old HSCs across the skeleton was different from their young counterparts. To address this issue, we calculated the relative standard deviation (RSD) for each of the clones found 4 or more times. The RSD allows to compare variations in clones with different mean contributions. An RSD<1 demonstrates low-level variance of distribution, while an RSD>1 indicates high-level variance. Most young and old clones have RSD values >1 but the average RSD for both age groups is very similar (Figure 1H), demonstrating once more that both populations have highly unequal skeletal distributions.

G-CSF-induced mobilization leads to redistribution of HSC clones

G-CSF administration mobilizes HSCs and progenitors into the circulation. G-CSF has also been shown to increase partner chimerism when administered to a parabiotic animal (8). To assess whether G-CSF affects clonal skeletal distribution, we analyzed a cohort of 7 animals transplanted with young (4 months old) barcoded LSK48^{-}150^{+} cells. As only one cell type was transplanted, it was possible to generate more *in vitro* colonies per clone and increase accuracy of the measurements. Three mice were mobilized 7 months post-transplantation while 4 mice were left untreated. 10-13 weeks after G-CSF treatment mice were sacrificed and the barcode composition of the LSK48^{-}150^{+} pool was analyzed in different locations (Figure 2A). Additionally, LSK48^{-}150^{+} cells were sorted from the spleen. Similar to our results described above, skeletal distribution of HSC clones in all 4 untreated animals was highly asymmetrical (Figure 2B,D), fifteen of 31 clones were significantly skewed across bone sites. Strikingly, a single G-CSF challenge led to complete randomization of HSC clones across the 4 studied locations (Figure 2C,E). Skewing completely disappeared in this group (Figure 2E). This indicates that HSC clones are readily mobilized and reenter the niches after mobilization. We have also been able to retrieve up to three barcoded LSK48^{-}150^{+} colonies in spleen per mouse. The clones present in spleen are marked with a hashtag (#) in Figure 2.

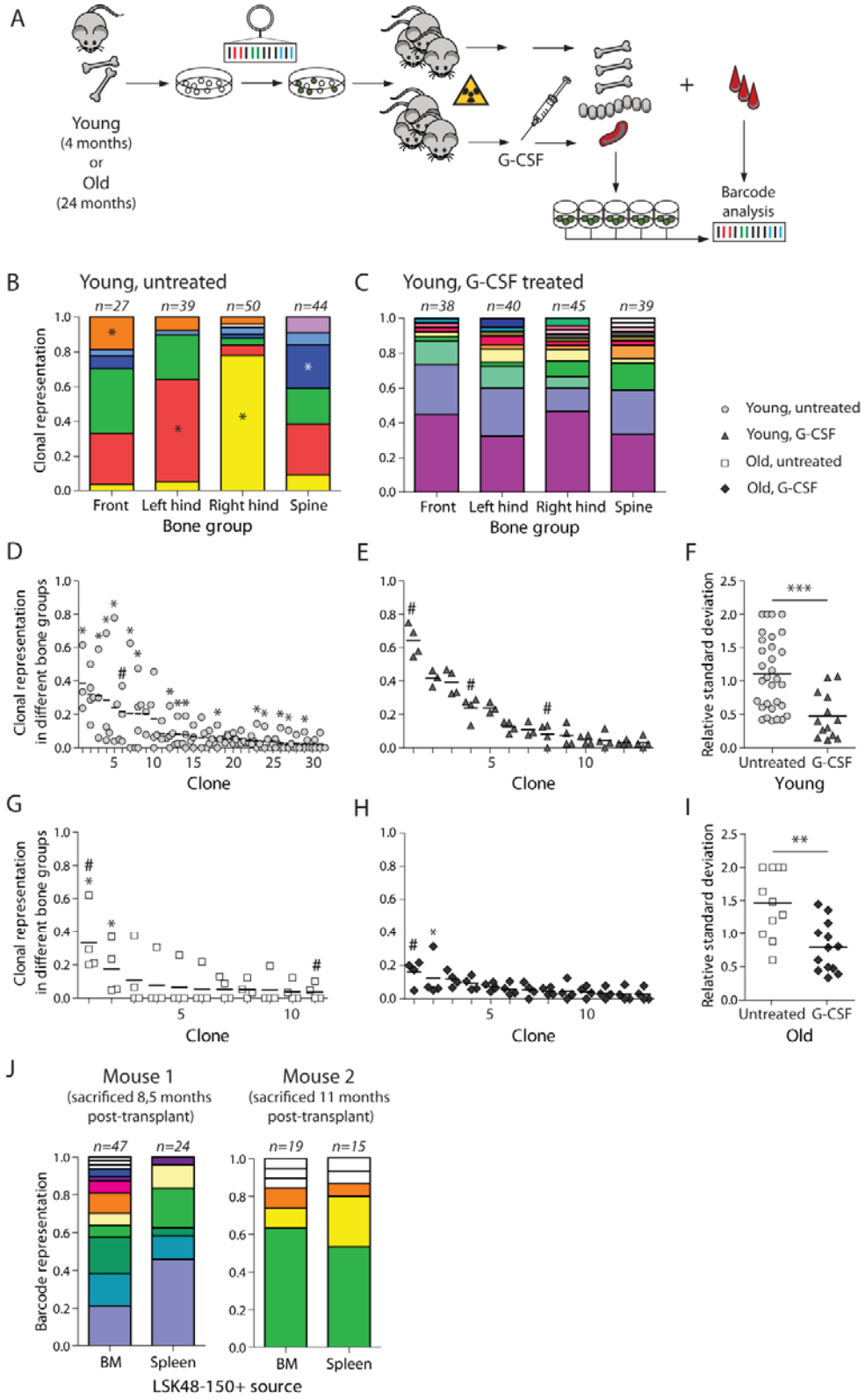
A
Young
(4 months)
or
Old
(24 months)
G-CSF
Barcode
analysis
B
Young, untreated
n=27
n=39
n=50
n=44
Clonal representation
Front
Left hind
Right hind
Spine
Bone group
C
Young, G-CSF treated
n=38
n=40
n=45
n=39
Young, untreated
Young, G-CSF
Old, untreated
Old, G-CSF
D
Clonal representation
in different bone groups
Clone
E
Clone
F
Relative standard deviation
***
Untreated
G-CSF
Young
G
Clonal representation
in different bone groups
Clone
H
Clone
I
Relative standard deviation
**
Untreated
G-CSF
Old
J
Mouse 1
(sacrificed 8,5 months
post-transplant)
Mouse 2
(sacrificed 11 months
post-transplant)
n=47
n=24
n=19
n=15
Barcode representation
BM
Spleen
BM
Spleen
LSK48-150+ source

Figure 2. G-CSF-induced mobilization leads to homogenization of skeletal HSC distribution of both young and old clones. (A) Experimental setup. Recipient mice were transplanted with young or old transduced LSK48$^-$150$^+$ cells. The groups of mice were challenged by G-CSF seven months (young mice) or eight months (old mice) post-transplantation. 10-18 weeks after treatment, all mice (both treated and untreated groups) were sacrificed. Clonal composition of different skeletal locations, spleen and blood was investigated by barcode analysis. Recipients of young cells belong to one cohort of seven animals. Recipients of old LSK48$^-$150$^+$ cells relate to two cohorts (n=4, one of untreated animals originated from a separate cohort). (B) Example of clonal barcode representation in an untreated mouse transplanted with young cells. As in Figure 1, identical colors relate to the same clone, clones found only once are marked white. (C) Same, but for a G-CSF-treated mouse. (D) Summary of skeletal clonal distributions in 4 recipient mice transplanted with young stem cells. The asterisks identify clones with distributions that are significantly different from random ($p < 0.05$ based on Monte Carlo simulations). (E) Same, but for G-CSF treated recipients (n=3). (F) Summary of relative standard deviations for all clones found in 4 or more colonies for the groups shown in panels D and E. The difference between the groups was tested by two-tailed Mann Whitney test ($p = 0.0004$). (G) Analysis of clonal skeletal distribution in 2 recipient animals transplanted with old LSK48$^-$150$^+$ cells. Statistical analysis performed as in panels D and E. (H) Same for 2 G-CSF treated recipients transplanted with old cells. (I) Summary of relative standard deviations for all clones found in 4 or more colonies for the groups shown in panels G and H. The difference between the groups was assessed by two-tailed Mann Whitney test ($p = 0.0037$). (J) Comparison of clonal composition of LSK48$^-$150$^+$ cells in bone marrow and spleen/blood, 7 days after G-CSF treatment (n=2). BM (bone marrow) refers to pooled data from all four skeletal sites. Correlation between clone sizes in spleen and bone marrow was significant for young cells in both mice (Pearson correlation, $r=0.62$, $p=0.032$ and $r=0.91$, $p<0.001$, respectively). Data for old cells is not shown due to low old cell chimerism in one of the analyzed mice.

It has been reported that mobilization potential of HSCs might be affected by the age of the patient. For instance, older myeloma patients have been shown to be poorer mobilizers compared to young counterparts (13,14). Mobilization in old mice, however, was shown to be more efficient than in young individuals (17). We addressed this controversy at the clonal level: we studied the extent of skeletal allocation of old barcoded HSC clones (24 months at time of transduction and transplantation) (Figure 2G and H). Transplantations were performed as for recipients of young cells and G-CSF was administered to half of the animals 8 months post-transplant. One of G-CSF-treated mice suffered from a spleen rupture and died one week after injection. The two remaining mice were sacrificed 10 weeks after mobilization. The response was similar to young clones (Figure 2F and I), with the exception of one out of 13 clones (clone 2, Figure 2H). While most of the clones equilibrated after G-CSF challenge, clone 2 was dominant in spine (11 of 35 colonies carried this barcode). While all young clones seem to be equally prone to mobilization, within the old cell population we cannot exclude the presence of both good and bad mobilizers.

To test this in a more direct manner, we studied clonal representation of the LSK48$^-$150$^+$ cells in bone marrow and spleen at the peak of mobilization (one week after G-CSF administration). Recipients of co-transplanted young and old cells (as shown at Figure 1A) were used to study the response within the same animal. The animals were mobilized 8.5 or 11 months post-transplantation and LSK48$^-$150$^+$ clones were harvested from bone marrow and spleen (Figure

2J). The clonal composition of "young" LSK48^{-}150^{+} cells in bone marrow and spleen was very similar (r = 0.62, p = 0.032; and r = 0.91, p < 0.001 for mouse 1 and 2 respectively, Figure 2J). Data for aged cells are not shown due to low cell chimerism.

LSK48^{-}150^{+} cells represent the functional HSC population with highly variable developmental potential

In naive mice, the LSK48^{-}150^{+} phenotype allows to purify a highly enriched HSC fraction (21). However, *ex vivo* culture, transduction and transplantation could in theory affect the phenotype of primitive cells. To confirm that LSK48^{-}150^{+} colonies originate from functional HSCs, we compared barcode spectra in mature blood populations with those from single cell expanded clones. Barcode fluctuations in blood samples were followed for a period of 10-11 months in primary recipients, and for an additional 6 months in secondary recipients (Figure 3A). Granulocytes, T and B lymphocytes were purified from blood of primary and secondary transplant recipients by FACS, and barcodes were analyzed by high-throughput Solexa sequencing. Previously, we showed that clonal representation of the hematopoietic system becomes stable 3 months post-transplantation (18), therefore here we focused on maximum representation of a particular barcode in blood 12 weeks and later post-transplant. Results of such analysis for four primary recipients of young LSK48^{-}150^{+} cells and 13 secondary recipients are summarized in Figure 3B. We considered a barcode present if it was detected at a level of at least 0.5% in a relevant population (18). Of 70 barcodes detected in LSK48^{-}150^{+} cells, 43 were found in primary recipients, 17 of them were found in both primary and secondary recipients, and the 10 of remaining 27 barcodes were detected only in secondary animals. Surprisingly, the probability to detect a particular barcode in blood of primary recipients was not proportional to the corresponding clone size in LSK48^{-}150^{+} cells (Figure 3B). There seems to be no apparent relation between clone size of a particular stem cell and its repopulating activity measured in blood samples.

We decided to study this issue of clone size representation in more detail. For the barcodes found in both blood and HSCs (reflecting "active" stem cells), we next asked whether the frequency of a barcode in LSK48^{-}150^{+} pool was correlated to frequencies in granulocytes, B and T lymphocytes. Figure 3C summarizes such analysis in primary recipients. It is apparent that the barcode representation data are quite disparate, and no obvious trend can be seen. To test whether the barcode representation in blood and LSK48^{-}150^{+} colonies from bone marrow showed a linear relationship, we calculated the relation between these two parameters by testing whether the slope of the regression line was significantly different from zero (Figure 3D). For granulocytes, the slope was significantly different from null hypothesis (p=0.0125). For T cells and B cells, the difference was insignificant, indicating no linear relationship between two

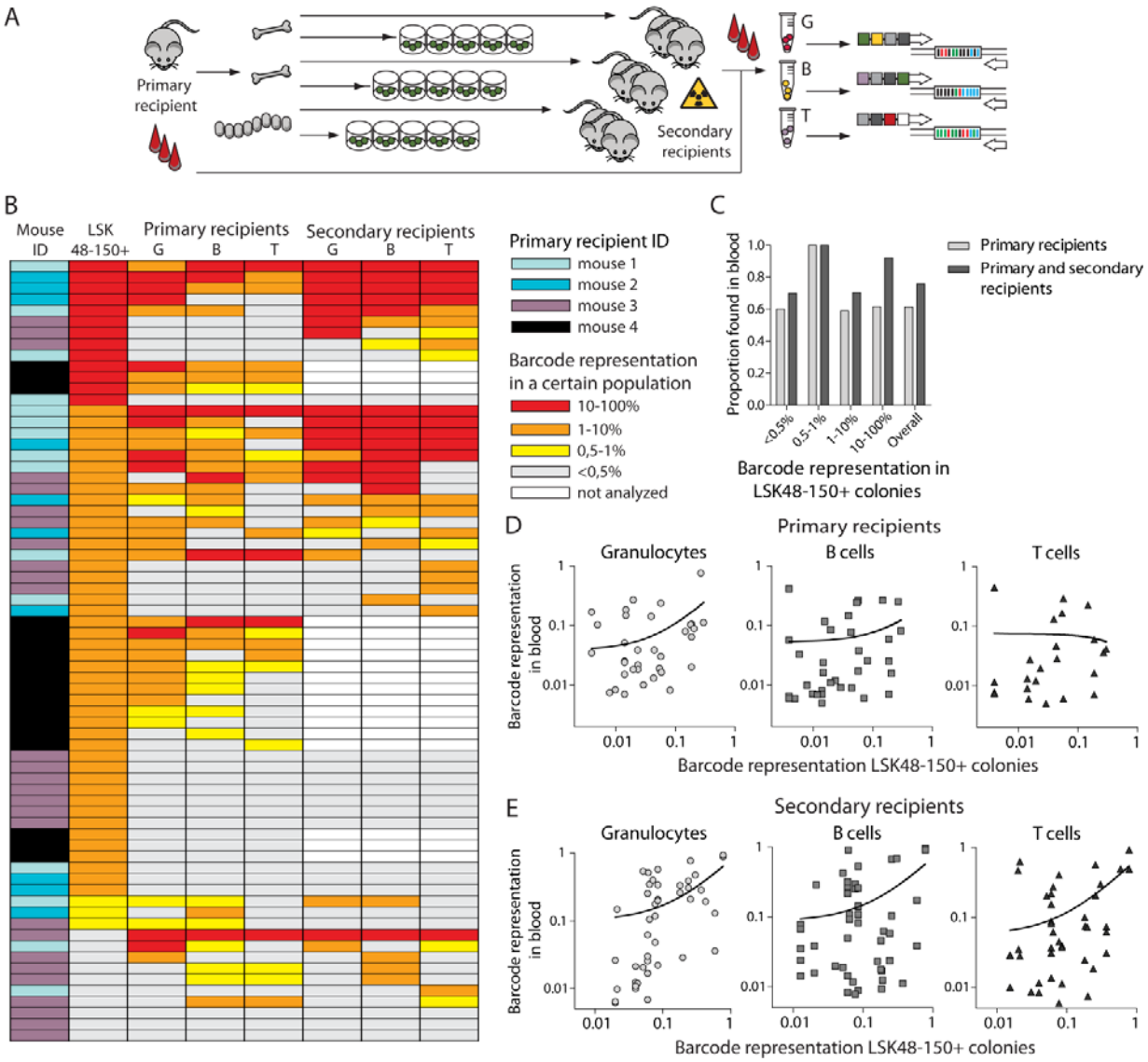

Figure 3. Developmental potential of migrating HSC clones in primary and secondary recipients. (A) Experimental setup. Mature cells were purified from blood of primary and secondary transplant recipients by FACS and barcodes were analyzed by high-throughput sequencing. Clonal composition in blood was compared with barcodes from expanded single stem cell-initiated colonies. (B) Barcode representation in primary LSK48^{-}150^{+} colonies and in three blood populations in primary or secondary recipients is plotted. The graph summarizes data from 70 clones found in four primary recipients of young cells and 13 secondary recipients transplanted with cells from three of these primary recipients. For one of the primary recipients (corresponding to 18 barcodes), no secondary transplantations were performed. The colored legend identifies each of the four primary recipients from which the barcode data originate. Each row reflects a single clone/barcode, and shows the maximum frequency with which that particular clone contributes to a particular lineage, measured over a period of observation starting at 12 weeks post-transplantation. (C) Proportion of clones contributing to blood formation in primary and secondary recipients as a function of the clone size in the LSK48^{-}150^{+} pool. Note that overall ~40% and ~20% of the clones detected in bone marrow LSK48^{-}150^{+} cells remains undetected in blood in primary or secondary recipients, respectively. (D) Correlation between barcode representation in bone marrow LSK48^{-}150^{+} colonies and three mature blood populations in primary recipients (in one cohort of 4 mice). The best-fit line of linear regression analysis is shown

(for granulocytes, slope ($\pm CI_{95\%}$) = 0.70 ± 0.26, p = 0.0125; for B cells, slope = 0.24 ± 0.21, p = 0.2776; for T cells, slope = - 0.07, p = 0.8). (E) Same as in D, but now in secondary recipients (13 mice). The slopes of the best-fit lines and p-values were as follows: for granulocytes, slope = 0.70 ± 0.16, p < 0.0001; for B cells, slope = 0.63 ± 0.18, p = 0.001; for T cells, slope = 0.63 ± 0.15, p = 0.0001).

variables (p=0.2776 and p=0.8, respectively). We also tested Spearman correlation between different cell subsets: only LSK48$^-$150$^+$ and B cells significantly correlated (ρ=0.3725, p=0.0275).

Several factors could contribute to the lack of linearity between barcode size in lymphocytes and LSK48$^-$150$^+$. First, lymphocytes are long-lived blood cells that persist in the circulation for months, in contrast to granulocytes that are replenished every two days. Thus, detected lymphocytes can originate from stem cells that have exhausted long before analysis took place. Second, as only a fraction of bones was taken for analysis, a number of clones (contributing to B and T cell production) could be missed in the analysis. As these factors are irrelevant in secondary transplantations, we compared barcodes in blood of 13 secondary recipients to the LSK48$^-$150$^+$ pool composition in primary animals (Figure 3E). For all three populations (granulocytes, B and T cells), clonal distribution in blood was proportional to corresponding LSK48$^-$150$^+$ pool of the donor, and the slope was significantly different from zero (p<0.0001, p=0.001, and p=0.0001, respectively). While in the primary animals, only B cells and bone marrow composition were significantly correlating, in secondary recipients, clonality of both granulocytes and T cells significantly correlated with that of primitive cells originating from primary recipient (for granulocytes, ρ=0.6704, p<0.0001; for T cells, ρ=0.3303, p=0.0349). Note, however, that this is an average regression for all clones. Individual data are scattered, and the degree of variation is substantial. It strongly argues against a simple relation between clone size in bone marrow and blood. Obviously, there are other factors than clone size in the bone marrow which affect clonal representation in the blood.

Clonal composition of secondary transplant recipients depends on donor bone of origin and resembles the clonal composition of donor LSK48$^-$150$^+$

As the clonal composition of the HSC pool differs in different skeletal locations, this could have ramifications on the interpretation of bone marrow biopsy composition, or on the clonality of secondary recipients, resulting in an underestimated clonal representation compared to the primary recipient. We tested whether the clonal composition of blood in secondary recipients, transplanted with bone marrow from a single skeletal site, was similar to the HSC (LSK48$^-$150$^+$) pool of the corresponding donor bone. As a “negative control” we compared the data with those obtained from another skeletal site from the very same donor mouse. As some clones found in the LSK48$^-$150$^+$ cell population were not contributing to blood (Figure 3), these could not be taken for similarity analysis using correlation estimates. Therefore, we focused only on

the barcodes found in blood at levels of >1% in either primary or secondary animals that correspond to the active stem cells, and excluded barcodes that occurred less frequently. We further analyzed correlations and tested linear relationships between barcode make-up of the LSK48^{-}150^{+} pool in a certain skeletal site and matched (marked in green) or unmatched (marked in white) secondary recipients (Figure 4). Despite large variation in the amount of output of mature blood cells per LSK48^{-}150^{+} clone (Figure 3), both linearity and correlation were higher in the matched samples (Figure 4A, green). In 8 out of 12 cases of blood vs. bone marrow comparisons, barcode makeup of blood in secondary recipients was significantly similar to the barcode composition of LSK48^{-}150^{+} cells in the donor bone of origin ($p<0.05$). In contrast, correlations were insignificant in all unmatched samples. Similarly, we observed good linearity between clone sizes in blood and HSCs in 6 groups ($p<0.05$). Among unmatched groups, linear dependence was observed in one case (where the granulocytic population showed good fit to the bones of right hind leg, while originating from spine), further underscoring restricted interdependence between primitive and mature cells. This example demonstrates that, while Pearson correlation and linear regression both are testing linear relationship between two sets of data, the results might be different, and conclusions should be drawn with care.

B

Secondary recipients blood analysis

LSK48^{-}150^{+}	*Primary recipient 1*		G	B	T	G	B	T
			Left hind			Right hind		
Left hind	Correlation		-	-	*	-	-	-
	Linearity		-	-	-	-	-	-
Right hind	Correlation		-	-	-	*	*	**
	Linearity		-	-	-	***	***	***

Secondary recipients blood analysis

LSK48^{-}150^{+}	*Primary recipient 2*		G	B	T	G	B	T
			Spine			Right hind		
Spine	Correlation		*	*	-	-	-	-
	Linearity		**	*	-	-	-	-
Right hind	Correlation		-	-	-	*	-	*
	Linearity		***	-	-	***	-	-

Figure 4. Clonal blood composition in secondary recipients reflects the clonal HSC makeup of bone of origin. (A) Experimental setup. Bone marrow cells were isolated from two primary recipients (left and right hind legs from recipient 1, and spine and right hind leg from recipient 2) and were transplanted in 4 secondary recipients each. In parallel with secondary transplant barcode frequencies were assessed in donor LSK48^{-}150^{+} cells. Granulocytes (G), B and T cells were isolated from blood of secondary recipients 12-24 weeks post-transplant and clonal analysis performed by high-throughput sequencing. The relation between barcode representation in LSK48^{-}150^{+} colonies and blood of secondary transplant recipients was evaluated in matched (green) and unmatched (non-marked) samples. (B) The clonal composition of HSCs in the LSK48^{-}150^{+} pool in a certain skeletal site was compared with blood composition of related and unrelated secondary recipients using Spearman correlation and linear regression analysis. The tables summarize conditions where correlation was significant and where the slope was significantly different from zero. Significance levels were as follows: * - $p<0.05$; ** - $p < 0.01$; *** - $p <0.001$.

DISCUSSION

Understanding the kinetics and molecular events underlying trafficking of HSC is important both in basic hematology research, and for implementation and interpretation of experimental and clinical bone marrow transplantation protocols. Here, we studied post-transplantation skeletal distribution of hundreds of HSCs clones to address the extent of migration in steady-state conditions and upon G-CSF-induced mobilization. Genetic barcoding of highly purified hematopoietic cells was used to quantitatively analyze HSC clone sizes in different bones and contributions of these clones to blood.

Our findings demonstrate that, at very extended time intervals (at least 11 months) after transplantation, the distribution of both old and young HSC clones across multiple skeletal sites is highly skewed. This is in line with data showing relatively slow rates of HSC equilibration in parabiotic mice (8). However, our findings are in contrast with studies that suggested that 1-5% of HSC pool circulates daily (11). Simulations indicate that at such rates the clonal make-up of the HSC pool would equalize within 4-6 months (data not shown). This prediction strongly contrasts our experimental observations. Notably, previous studies were never performed on the clonal HSC level, and therefore inequality in the level of egress of distinct HSC clones within one mouse was never addressed.

Strikingly, a single G-CSF mobilization regimen resulted in uniform spreading of HSC clones, which was preserved for many weeks after treatment. Here our data are quite distinct from results obtained in parabiotic models, in which only partial equilibration of the HSC pool occurs (8). Our data confirm that mobilized HSCs retain the ability to re-engraft bone marrow niches. They also strongly suggest that all “young” HSCs are equally prone to mobilization; we did not find any evidence of mobilization of only a selected subset of stem cell clones.

Heterogeneity of the spatial clonal distribution of stem cells, as demonstrated here, is likely the combined result of multiple factors, such as the number, size and migratory potential of distinct HSCs . It has recently been shown that multiple (biologically and physically) separate niches may exist (22-24). It is interesting to speculate that certain HSCs might home and expand in one but not another niche. While we have not addressed whether specific clones were found in specific niches, differential niche preference could potentially contribute to heterogeneity in the observed localization patterns as well. This issue could potentially be resolved by further subfractionation of HSCs from increasingly more defined physical regions of the bone marrow. It is recognized that G-CSF induces mobilization through affecting specific ligand-receptor interactions, and by inducing proteolytic cleavage of receptors (6). G-CSF affects not only HSCs, but also mature hematopoietic cells, as well as other cell types collectively referred to as the niche (osteoblasts, osteoclasts, mesenchymal cells) (6). G-CSF treatment will inevitably cause

modification of the niche. How this modification is contributing to the randomization of the clonal distribution is a matter of a separate study.

The fact that G-CSF efficiently equilibrates practically all clones, indicates that each individually labeled HSC achieved sufficient clonal outgrowth to guarantee redistribution of its offspring across all bones tested. This observation also predicts that naturally occurring infections in mice, when endogenous G-CSF levels are elevated, will induce spatial equilibration of HSC clones. The physiological relevance of HSC redistribution remains unexplored.

Previous reports suggest differences in mobilization abilities of young and aged HSCs (13,14,17), but extrinsic and intrinsic effects could not be discriminated Here, we observed similar migratory abilities of young and old populations in co-transplantation settings, in transplants of only young or old cells, and upon mobilization, demonstrating that putative changes in mobilization with age must be primarily stem cell extrinsic, highlighting the possible role of niche ageing in modulating HSC mobilization.

There are several putative caveats of our approach. First, HSC activation during gene transfer procedures could affect their function. While we (inevitably) did use *ex vivo* transduction and transplantation to follow HSCs, we demonstrate that these cells retain the ability for multilineage differentiation in irradiated hosts. The extent of heterogeneity of repopulating ability and developmental capacity of individual HSCs observed is in a good agreement with experiments with transplantation of unmanipulated single cells (25,26). Therefore, we are confident that qualitatively our observations are not affected by the method, but rather seem to reflect the natural clonal behavior of repopulating HSCs. Second, we determined barcode identities only in single LSK48^{-}150^{+} cells that were able to form a colony *in vitro*. While this approach allowed us to verify the functional activity of purified cells, *bona fide* stem cells that fail to produce a colony in our culture conditions would remain unnoticed.

Our findings have important implications for the design and interpretation of experimental and clinical transplantation protocols. First, we show that the current routine of using a single femur and tibia as a source for bone marrow cells to initiate secondary transplants in mice results in an underestimation of the clonal repertoire, and will exclude (potentially dominant) clones located elsewhere. This can contribute to a lack of consistency between the clonal make-up of bone marrow and blood, and the use of cells obtained from multiple skeletal bones is advisable. Second, the same limitations apply for using bone marrow biopsies to monitor clonal fluctuations in mice, cats, dogs, monkeys, or indeed in human gene therapy patients. Previous research on xenotransplantation of human cells into immunodeficient mice indicated that the clonal composition of bone marrow in individual bones varied (27). However, the authors interpreted local clonal dominance as a sign of functional differences between subsets of human HSCs, rather than reflecting general patterns of HSC distribution (27).

In conclusion, the data presented here demonstrate that upon transplant functional HSC clones preferentially expand in certain skeletal locations, exhibiting only limited migration towards other niches. High cytokine levels that are characteristic for infections and are clinically widely used to induce stem cell mobilization, result in rapid and permanent homogenization of clonal stem cell distribution.

ACKNOWLEDGEMENTS

The authors thank H. Moes, G. Mesander and R.-J. van der Lei for expert cell sorting assistance, E. Wojtowicz for laboratory assistance, B. Dykstra, H. Schepers and R.van Os for valuable discussions and suggestions and P. van der Vlies and J. Bergsma for assistance with high-throughput sequencing. This study was supported by the Netherlands Organization for Scientific Research (VICI grant to G.d.H. and TopTalent grant to E.V.), the National Roadmap for Large Scale Infrastructure (Mouse Clinic for Cancer and Aging), and the Netherlands Institute for Regenerative Medicine (NIRM).

REFERENCES

(1) Korbling M, Burke P, Braine H, Elfenbein G, Santos G, Kaizer H. Successful engraftment of blood derived normal hemopoietic stem cells in chronic myelogenous leukemia. Exp Hematol 1981 Jul;9(6):684-690.

(2) Dorie MJ, Maloney MA, Patt HM. Turnover of circulating hematopoietic stem cells. Exp Hematol 1979 Oct;7(9):483-489.

(3) Goodman JW, Hodgson GS. Evidence for stem cells in the peripheral blood of mice. Blood 1962 Jun;19:702-714.

(4) Richman CM, Weiner RS, Yankee RA. Increase in circulating stem cells following chemotherapy in man. Blood 1976 Jun;47(6):1031-1039.

(5) Orkin SH, Zon LI. Hematopoiesis: an evolving paradigm for stem cell biology. Cell 2008 Feb 22;132(4):631-644.

(6) Lapid K, Glait-Santar C, Gur-Cohen S, Canaani J, Kollet O, Lapidot T. Egress and Mobilization of Hematopoietic Stem and Progenitor Cells: A Dynamic Multi-facet Process. StemBook Cambridge (MA): Kfir Lapid, Chen Glait-Santar, Shiri Gur-Cohen, Jonathan Canaani, Orit Kollet and Tsvee Lapidot; 2008.

(7) Abkowitz JL, Robinson AE, Kale S, Long MW, Chen J. Mobilization of hematopoietic stem cells during homeostasis and after cytokine exposure. Blood 2003 Aug 15;102(4):1249-1253.

(8) WARREN S, CHUTE RN, FARRINGTON EM. Protection of the hematopoietic system by parabiosis. Lab Invest 1960 Jan-Feb;9:191-198.

(9) Wright DE, Wagers AJ, Gulati AP, Johnson FL, Weissman IL. Physiological migration of hematopoietic stem and progenitor cells. Science 2001 Nov 30;294(5548):1933-1936.

(10) Bhattacharya D, Czechowicz A, Ooi AG, Rossi DJ, Bryder D, Weissman IL. Niche recycling through division-independent egress of hematopoietic stem cells. J Exp Med 2009 Nov 23;206(12):2837-2850.

(11) To LB, Haylock DN, Simmons PJ, Juttner CA. The biology and clinical uses of blood stem cells. Blood 1997 Apr 1;89(7):2233-2258.

(12) Pozotrigo M, Adel N, Landau H, Lesokhin A, Lendvai N, Chung DJ, et al. Factors impacting stem cell mobilization failure rate and efficiency in multiple myeloma in the era of novel therapies: experience at Memorial Sloan Kettering Cancer Center. Bone Marrow Transplant 2013 Aug;48(8):1033-1039.

(13) Tricot G, Jagannath S, Vesole D, Nelson J, Tindle S, Miller L, et al. Peripheral blood stem cell transplants for multiple myeloma: identification of favorable variables for rapid engraftment in 225 patients. Blood 1995 Jan 15;85(2):588-596.

(14) Liang Y, Van Zant G, Szilvassy SJ. Effects of aging on the homing and engraftment of murine

hematopoietic stem and progenitor cells. Blood 2005 Aug 15;106(4):1479-1487.
(15) Dykstra B, Olthof S, Schreuder J, Ritsema M, de Haan G. Clonal analysis reveals multiple functional defects of aged murine hematopoietic stem cells. J Exp Med 2011 Dec 19;208(13):2691-2703.
(16) Xing Z, Ryan MA, Daria D, Nattamai KJ, Van Zant G, Wang L, et al. Increased hematopoietic stem cell mobilization in aged mice. Blood 2006 Oct 1;108(7):2190-2197.
(17) Gerrits A, Dykstra B, Kalmykowa OJ, Klauke K, Verovskaya E, Broekhuis MJ, et al. Cellular barcoding tool for clonal analysis in the hematopoietic system. Blood 2010 Apr 1;115(13):2610-2618.
(18) Verovskaya E, Broekhuis MJ, Zwart E, Ritsema M, van Os R, de Haan G, et al. Heterogeneity of young and aged murine hematopoietic stem cells revealed by quantitative clonal analysis using cellular barcoding. Blood 2013 May 29.
(19) Ramos CA, Zheng Y, Colombowala I, Goodell MA. Tracing the origin of non-hematopoietic cells using CD45 PCR restriction fragment length polymorphisms. BioTechniques 2003 Jan;34(1):160-162.
(20) Korhonen J, Martinmaki P, Pizzi C, Rastas P, Ukkonen E. MOODS: fast search for position weight matrix matches in DNA sequences. Bioinformatics 2009 Dec 1;25(23):3181-3182.
(21) Kiel MJ, Yilmaz OH, Iwashita T, Yilmaz OH, Terhorst C, Morrison SJ. SLAM family receptors distinguish hematopoietic stem and progenitor cells and reveal endothelial niches for stem cells. Cell 2005 Jul 1;121(7):1109-1121.
(22) Lo Celso C, Scadden DT. The haematopoietic stem cell niche at a glance. J Cell Sci 2011 Nov 1;124(Pt 21):3529-3535.
(23) Ding L, Morrison SJ. Haematopoietic stem cells and early lymphoid progenitors occupy distinct bone marrow niches. Nature 2013 Mar 14;495(7440):231-235.
(24) Greenbaum A, Hsu YM, Day RB, Schuettpelz LG, Christopher MJ, Borgerding JN, et al. CXCL12 in early mesenchymal progenitors is required for haematopoietic stem-cell maintenance. Nature 2013 Mar 14;495(7440):227-230.
(25) Dykstra B, Kent D, Bowie M, McCaffrey L, Hamilton M, Lyons K, et al. Long-term propagation of distinct hematopoietic differentiation programs in vivo. Cell Stem Cell 2007 Aug 16;1(2):218-229.
(26) Ema H, Sudo K, Seita J, Matsubara A, Morita Y, Osawa M, et al. Quantification of self-renewal capacity in single hematopoietic stem cells from normal and Lnk-deficient mice. Dev Cell 2005 Jun;8(6):907-914.
(27) Mazurier F, Doedens M, Gan OI, Dick JE. Rapid myeloerythroid repopulation after intrafemoral transplantation of NOD-SCID mice reveals a new class of human stem cells. Nat Med 2003 Jul;9(7):959-963.

CHAPTER **6**

SUMMARY & DISCUSSION

SUMMARY

Studying the composition of hematopoietic stem cell (HSC) pool is essential for understanding the fundamental HSC biology and for gaining insight into hematopoietic malignancies. As bone marrow contains thousands of HSC clones, it is crucial to follow their behavior in the polyclonal setting. Current understanding of HSC functions is primarily based on studies of cell populations, single-cell transplantations or retroviral marking. Due to the methodological limitations, none of these approaches allows quantitative high-resolution analysis of clonality.

The work presented in this thesis is focused on uncovering heterogeneity of clonal behavior of HSCs in young mice and upon aging. To investigate this question, we developed, validated and applied cellular barcoding technique. We then addressed several basic questions regarding HSC biology, such as kinetics following bone marrow transplantation, HSC migration and effect of aging on clonal behavior of HSCs. This Chapter summarizes the main findings of the thesis and puts these studies into the perspective.

In **Chapter 2**, we provide an overview of the available data on heterogeneity in HSC compartment and discuss potential clinical relevance of this phenomenon. We introduce different aspects of HSC behavior that can vary dramatically between distinct HSC clones, such as cycling rate, mature blood cell output and repopulation potential. Next, we discuss how representation of particular HSC subsets in the HSC pool can change with the aging. Afterwards, we speculate how knowledge of HSC heterogeneity can contribute to our understanding of human hematopoietic disorders. Finally, we provide examples of clonal studies in human gene therapy trials and discuss how they support the notion of HSC heterogeneity in men.

In **Chapter 3**, we describe methodological caveats of studying clonality in hematopoietic system. This Chapter illustrates how suboptimal technology can lead to discrepancies in interpretation of clonal studies providing an incentive for development of quantitative sensitive methods of analysis. We compare available methods of clonal tracking based on retroviral marking of hematopoietic cells and discuss how drawbacks of different experimental procedures can affect the clonal readout. Such experimental steps include genome restriction that produces fragments of variable size what could result in inefficient PCR amplification, or in missing integration sites during analysis. We analyze the theoretical genome coverage by restriction enzymes frequently used in the literature, and propose combinatorial approach for optimal genome restriction. Next, we discuss how high-throughput sequencing detection of insertions, on contrary, can exaggerate the clone numbers if appropriate noise filtering is not applied. We argue that recently developed cellular barcoding technique represents a method of choice for quantitative analysis. This approach is based on insertion of a random sequence tag, or a barcode, into the genome of viral vector. Upon transduction, barcode tag individually labels the HSCs and is passed on to the their progeny. The tags are detected by sequencing following

amplification of barcode fragment in genomic DNA. As barcodes are uniform in structure, their amplification and readout are also uniform. A number of requirements, including validation of barcoded vector library size, consideration of number of barcodes needed per target cell, and signal-to-noise discrimination, need to be fulfilled to ensure successful implementation of the method.

In **Chapter 4**, we validate the clonal analysis method based on combining cellular barcoding with high-throughput sequencing detection. As Illumina (Solexa) sequencing can generate millions of reads per run, we perform the analysis in multiplexed format by amplifying barcodes with indexed primers. This allows us to combine up to 300 samples in a single pool. To determine the sensitivity and resolution of this protocol, we analyzed defined mixtures of barcoded cells containing different cell numbers and varying ratios of barcodes. Barcodes with frequency of 0,5% could be quantitatively detected, and barcodes were reliably analyzed in samples with as low as 1000 cells. Further, we used this approach to follow the clonal behavior of hundreds of young and old HSCs in the polyclonal transplantation setting. HSCs were purified by flow cytometry based on their immunophenotype, transduced with barcoded vector library and transplanted into irradiated hosts. Barcode composition of short-lived granulocytes, and long-lived B and T lymphocytes was analyzed in sequentially drawn blood samples. We demonstrate that the most of the transplanted HSCs stably contribute to blood production from 3 to 6 months post-transplantation. However, in the long term (up to a year after transplant), clonal representation of both myeloid and lymphoid clones changed uni-directionally: barcoded clones either declined or expanded. Last, we analyzed composition of pool of young and old HSCs in the transplant recipients. The young HSC pool was composed of a few potent clones, while the aged HSC pool – of numerous smaller clones.

Chapter 5 reports the skeletal distribution of young and aged HSC clones following bone marrow transplantation. Barcode labeling is used to individually mark HSCs prior to injection into irradiated recipients. We isolate HSC fraction from 4 bone groups 6-11 months post-transplatant and compare the barcode composition of different skeletal sites. We demonstrate that HSC clones are predominantly localized at one bone group. Such preferential localization persists for at least 11 months post-transplantation arguing for limited migration of HSC clones. Notably, a single challenge with granulocyte colony stimulating factor (G-CSF), which is routinely used to harvest HSCs for human bone marrow transplantations, leads to a complete equilibration of skeletal spreading. Further, composition of HSC pool in bone marrow is compared to clonal make-up of blood in secondary recipients. We show that barcode representation in blood is defined by HSCs present in the bones used to initiate the transplantation however clone sizes differ substantially between marrow and blood.

DISCUSSION AND FUTURE PERSPECTIVES

The work described in this thesis had two main focuses. First, we aimed to develop and validate the cellular barcoding as a tool for clonal analysis, to explore limitations of the clonal tracking methods and formulate recommendations for design of barcoding experiments. Second, we used this novel technology for clonal analysis of young and old hematopoiesis. We will further discuss the progress achieved in these directions, and speculate how the field will develop in the future.

Analysis of heterogeneity in the HSC pool in mice

In contrast to the traditional view on hematopoiesis that implies that in a homogeneous pool of HSCs the decisions on lineage commitment are made in a binary fashion on all levels of differentiation, recent data indicate that intrinsic heterogeneity in developmental program exists on several levels of hematopoietic hierarchy (1). Such heterogeneity is now considered one of the fundamental characteristics of the HSC pool (1). Lineage imprinting of the more committed progenitors has also been recently demonstrated in multipotent progenitors (2).

Data obtained with cellular barcoding method in our and other labs confirm the claim of heterogeneity in the polyclonal hematopoiesis (2-4). In Chapter 4 we demonstrate that the lineage skewing of transplanted barcoded HSCs is stable and is preserved for at least 6 months post-transplant. Interestingly, changes in the lineage bias were detected in a proportion of clones upon serial transplantation both in single-cell transplant recipients (5-7) and in barcoding experiments (3). The recent data in our lab demonstrates that serial transplantation is not requisite for such 'switching': we have observed transitions of developmental program of certain clones upon a long period of observation (>1 year post-transplantation).

Data on barcoded HSCs generated in this thesis and in other current work also allows to demonstrate on the clonal level presence of HSCs with different turnover rates in the bone marrow. As shown in Chapter 5 and demonstrated by Grosselin *et al.* (3), a number of clones present in bone marrow HSC population is only detected in blood upon secondary transplantation arguing that they were initially quiescent.

The mechanisms responsible for heterogeneity of HSC features are starting to be elucidated and studies focused on this question are likely to be continued in the coming years. Only a handful of regulatory processes have been documented thus far. Both intrinsic and extrinsic factors are shown to influence the lineage commitment, repopulating potential and cycling rate of an HSC (1). Recently, existence of separate niches harboring cells of different developmental stages (8,9) and different cycling rates (10) has been reported. As our data on post-transplantation HSC distribution shows preferential expansion of clones within one location, it remains a question whether certain HSCs only home to certain niches.

Changes in the representation of different HSC subsets are characteristic for HSC aging. While the majority of previous studies compared young and aged populations in either single-cell setup or on population basis, in this thesis we observed their behavior in polyclonal situation.

Insights into HSC aging

The quest for understanding the mechanisms of aging becomes increasingly urgent as the population of Earth ages. The aging phenotype of HSCs has been relatively well documented (overviewed in Chapter 1, reviewed in detail in (11,12). In short, it is characterized by decreased lymphoid cell production and increased fraction of myeloid-biased HSCs in the pool; the number of phenotypically and functionally defined HSCs in the bone marrow is also increased with aging, however decreased repopulating potential per stem cell is reduced.

In this thesis we aimed to extend the current understanding of clonal behavior of aged HSC by analyzing homing, migration and mobilization behavior of hundreds of old clones co-transplanted with young HSCs (Chapters 4 and 5). Puzzlingly, while previous (not clonal) data on homing suggests aging-induced decline (6,13), detailed analysis of HSC pool demonstrated preservation of original relationship between number of aged to young clones for up to 8 months post-transplantation, however the output of old HSC pool to both primitive and mature cells is dramatically decreased. This finding underscores the necessity to study HSC behavior using clonal methods.

Further, post-transplantation skeletal distribution of HSCs and their equilibration by G-CSF show similar migratory behavior of young and aged HSCs. As both migration and mobilization in these experiments are performed in the young recipients of young/old co-transplant, the data indicates similar intrinsic migratory ability of both populations, and effect of niche aging on both aspects needs to be studied.

Development of technology for clonal analysis in hematopoietic system

Clonal tracking of HSCs focuses on establishing relationship between HSCs and downstream effector cells and characterizing how this relationship changes in time or upon development of pathology. Ideally, a tool for clonal analysis would allow to follow every HSC and its output quantitatively, and to ensure simple and reliable discrimination between distinct clones. Moreover, the methodology used to mark the cells (if any) should be neutral and not change their properties.

Cellular barcoding presently coupled with high-throughput detection system became a method of choice for analysis in hematopoietic system. As discussed in Chapter 3, barcoding surpasses the previously available assays based on analysis of integration sites as it allows unbiased quantitation of clonal contributions in polyclonal setting. Not surprisingly, it is currently applied

in a growing number of tissues and model organisms: first two studies reporting clonal analysis in human hematopoietic and mammary systems have been recently published (14,15), and studied on non-human primates are underway. While current reports document normal hematopoiesis, barcoding can be applied to follow clonal expansion following oncogene-driven leukemia, and to analyze changes during treatment and secondary transplantation. The first study that tracks development of leukemia following overexpression of barcoded ΔTkr-expressing vector has recently been published (16).

An obvious limitation of barcoding experiments is the necessity for *ex vivo* cell manipulation and transplantation. In line with that, in Chapter 4, we confirm that even relatively short 2 day transduction protocol decreases frequencies of functional stem cells among purified primitive cells. Similarly, transplantation of HSCs into irradiated animals or mice with compromised hematopoietic system creates conditions favoring expansion of donor HSCs rather than steady maintenance mode of hematopoietic system at normal conditions. Irradiation of recipients by itself could also influence engrafting cells, however recent data generated in our lab do not support this possibility, demonstrating similar lineage skewing, engraftment and migration between skeletal niches in irradiated and unconditioned mice.

While barcoding experimental setup - bone marrow cell isolation followed by gene transfer and transplantation – closely models human gene therapy protocols, it might not reflect behavior of HSCs in homeostatic conditions. Recent developments in high-throughput sequencing and single-cell techonologies might enable clonal analysis in unmanipulated animals. As the process of DNA replication is accompanied by some, usually low rate of errors, every cell division leads to accumulation of somatic mutations in the newly synthesized DNA strands. Therefore, every cell in a multicellular organism is genetically distinct. With sufficient number of cells analyzed with high precision, one can study genomic signatures of individual cells and can reconstruct the lineage trees (17). The proof of principle experiments based on analysis the polymorphisms in microsatellite regions allowed to study clonal architecture of human leukemia (18). Given the speed of development of sequencing technology, it is highly likely that the whole-genome characterization of thousands of cells from individual animal will soon be an everyday reality. It would also require sufficient computational power to be able to process such data.

REFERENCES

(1) Copley MR, Beer PA, Eaves CJ. Hematopoietic stem cell heterogeneity takes center stage. Cell Stem Cell 2012 Jun 14;10(6):690-697.

(2) Naik SH, Perie L, Swart E, Gerlach C, van Rooij N, de Boer RJ, et al. Diverse and heritable lineage imprinting of early haematopoietic progenitors. Nature 2013 Apr 11;496(7444):229-232.

(3) Grosselin J, Sii-Felice K, Payen E, Chretien S, Roux DT, Leboulch P. Arrayed lentiviral barcoding for quantification analysis of hematopoietic dynamics. Stem Cells 2013 Apr 4.

(4) Lu R, Neff NF, Quake SR, Weissman IL. Tracking single hematopoietic stem cells in vivo using high-throughput sequencing in conjunction with viral

genetic barcoding. Nat Biotechnol 2011 Oct 2;29(10):928-933.
(5) Benz C, Copley MR, Kent DG, Wohrer S, Cortes A, Aghaeepour N, et al. Hematopoietic stem cell subtypes expand differentially during development and display distinct lymphopoietic programs. Cell Stem Cell 2012 Mar 2;10(3):273-283.
(6) Dykstra B, Olthof S, Schreuder J, Ritsema M, de Haan G. Clonal analysis reveals multiple functional defects of aged murine hematopoietic stem cells. J Exp Med 2011 Dec 19;208(13):2691-2703.
(7) Dykstra B, Kent D, Bowie M, McCaffrey L, Hamilton M, Lyons K, et al. Long-term propagation of distinct hematopoietic differentiation programs in vivo. Cell Stem Cell 2007 Aug 16;1(2):218-229.
(8) Ding L, Morrison SJ. Haematopoietic stem cells and early lymphoid progenitors occupy distinct bone marrow niches. Nature 2013 Mar 14;495(7440):231-235.
(9) Greenbaum A, Hsu YM, Day RB, Schuettpelz LG, Christopher MJ, Borgerding JN, et al. CXCL12 in early mesenchymal progenitors is required for haematopoietic stem-cell maintenance. Nature 2013 Mar 14;495(7440):227-230.
(10) Kunisaki Y, Bruns I, Scheiermann C, Ahmed J, Pinho S, Zhang D, et al. Arteriolar niches maintain haematopoietic stem cell quiescence. Nature 2013 Oct 31;502(7473):637-643.
(11) Geiger H, de Haan G, Florian MC. The ageing haematopoietic stem cell compartment. Nat Rev Immunol 2013 May;13(5):376-389.
(12) Dykstra B, de Haan G. Hematopoietic stem cell aging and self-renewal. Cell Tissue Res 2008 Jan;331(1):91-101.
(13) Liang Y, Van Zant G, Szilvassy SJ. Effects of aging on the homing and engraftment of murine hematopoietic stem and progenitor cells. Blood 2005 Aug 15;106(4):1479-1487.
(14) Cheung AM, Nguyen LV, Carles A, Beer P, Miller PH, Knapp DJ, et al. Analysis of the clonal growth and differentiation dynamics of primitive barcoded human cord blood cells in NSG mice. Blood 2013 Sep 12.
(15) Nguyen LV, Makarem M, Carles A, Moksa M, Kannan N, Pandoh P, et al. Clonal Analysis via Barcoding Reveals Diverse Growth and Differentiation of Transplanted Mouse and Human Mammary Stem Cells. Cell Stem Cell 2014 Feb 6;14(2):253-263.
(16) Cornils K, Thielecke L, Huser S, Forgber M, Thomaschewski M, Kleist N, et al. Multiplexing clonality: combining RGB marking and genetic barcoding. Nucleic Acids Res 2014 Jan 29.
(17) Shapiro E, Biezuner T, Linnarsson S. Single-cell sequencing-based technologies will revolutionize whole-organism science. Nat Rev Genet 2013 Sep;14(9):618-630.
(18) Shlush LI, Chapal-Ilani N, Adar R, Pery N, Maruvka Y, Spiro A, et al. Cell lineage analysis of acute leukemia relapse uncovers the role of replication-rate heterogeneity and microsatellite instability. Blood 2012 Jul 19;120(3):603-612.

APPENDICES

SUMMARY IN DUTCH
ACKNOWLEDGEMENTS
CURRICULUM VITAE
LIST OF PUBLICATIONS

DUTCH SUMMARY / NEDERLANDSE SAMENVATTING

Dagelijks worden miljarden bloedcellen geproduceerd om een organisme goed te laten functioneren. Deze bloedcellen zijn afkomstig van hematopoëtische stamcellen (HSC) in het beenmerg. Door het onderzoeken van deze HSC's kunnen we meer te weten komen over de fundamentele biologie van deze cellen en inzicht krijgen in het ontstaan van hematologische maligniteiten. HSC's worden vaak gebruikt als model voor andere stamcellen. Tegenwoordig wordt ervan uitgegaan dat het onderhouden en herstellen van organen wordt gereguleerd door een kleine populatie stamcellen die specifiek zijn voor dat weefsel. Al in het begin van de jaren 60 heeft men laten zien dat een populatie cellen in het beenmerg de potentie bezitten om zich zelf te vernieuwen en te differentiëren tot verschillende type bloedcellen. Deze ontdekking was de basis voor de regeneratieve geneeskunde en stamcelbiologie.

Ook is bekend dat veranderingen in de HSC populatie kunnen leiden tot hematologische ziektebeelden en veranderingen van hematopoëtische cellen tijdens veroudering, zoals verzwakking van het immuunsysteem en een verhoging van de kans op myeloïde maligniteiten. HSC's zijn de eerste stamcellen die zijn toegepast in de kliniek, de eerste beenmerg transplantatie is uitgevoerd in 1958 voor de behandeling van leukemie. Tegenwoordig worden beenmerg transplantaties veelvuldig gebruikt voor de behandeling van hematologische en niet hematologische afwijkingen. HSC's zijn het best onderzocht in muismodellen waardoor een beter inzicht ontstond in verschillende mechanismen die later zijn bevestigd in mensen.

Doordat het beenmerg vele duizenden HSC klonen bevat is het belangrijk ze te volgen in een polyklonale omgeving. Het huidige inzicht van het functioneren van HSC functie is gebaseerd op verschillende studies van celpopulaties, transplantaties met één enkele cel en retrovirale markering. Deze technieken hebben technische beperkingen en kunnen moeilijk een goed beeld geven van kwantitatieve, hoge resolutie analyse van klonaliteit.

Het werk dat wordt beschreven in dit proefschrift is gericht op het onderzoeken van de heterogeniteit van het gedrag van HSC's in jonge muizen en gedurende veroudering. Hiervoor hebben we een nieuwe techniek opgezet, gevalideerd en toegepast voor het barcoderen van cellen. We hebben enkele basale vraagstellingen over de biologie van HSC's onderzocht, waaronder de kinetiek van hematopoietisch herstel na beenmergtransplantaties, HSC migratie en het effect van veroudering op de klonaliteit van HSC's.

Hoofdstuk 1 bevat een algemene introductie van dit proefschrift. Hierin wordt het hematopoëtische systeem beschreven en de biologische kenmerken van HSC's. Ook beschrijven we de technieken die zijn gebruikt voor het in kaart brengen van het gedrag van HSC's, waaronder het barcoderen van cellen die voor de rest van dit onderzoek is gebruikt.

Hoofdstuk 2 bevat een overzicht van de beschikbare data omtrent de heterogeniteit van HSC's en beschrijft de mogelijke klinische relevantie hiervan. We bespreken verschillende aspecten van het gedrag van HSC's die grote variatie tussen verschillende klonen kennen, zoals de snelheid waarmee ze delen, de productie van uitgerijpte bloedcellen en de mogelijkheid tot repopulatie. Vervolgens beschrijven we hoe een bepaalde subpopulatie HSC's kunnen veranderen gedurende veroudering. Daarna speculeren we hoe de kennis omtrent heterogeniteit van HSC kan bijdragen tot het doorgronden van humane hematologische ziektes. Tenslotte geven we voorbeelden van studies van klonaliteit in gentherapie trials in de mens en bespreken we hoe ze bij kunnen dragen aan het begrip van HSC heterogeniteit in de mens.

In **hoofdstuk 3** beschrijven we de technische details en beperkingen van het onderzoeken van klonaliteit in het hematopoëtische systeem. Dit hoofdstuk beschrijft hoe suboptimale technologie kan leiden tot verschillen in interpretatie van studies over klonaliteit en moedigt aan tot het ontwikkelen van kwantitatieve en gevoelige methodes voor analyse. We vergelijken beschikbare methodes voor het traceren van klonen gebaseerd op het retroviraal markeren van hematopoëtische cellen en beschrijven hoe de nadelen van experimentele handelingen het begrip van klonaliteit beïnvloeden. Deze experimentele stappen omvatten o.a. het fragmenteren van het genoom op basis van restrictie enzymen, waardoor fragmenten ontstaan van variabele lengtes, hetgeen kan resulteren in inefficiënte PCR (polymerase kettingreactie) amplificatie, of het niet vinden van integratiesites gedurende analyse. We maken een theoretische analyse van de dekking van het genoom met restrictie enzymen die vaak worden gebruikt in de literatuur, en doen een voorstel voor een combinatie van enzymen voor een optimale dekking van het genoom. Vervolgens beschrijven we hoe door een gebrek aan ruis filtering virale genoominserties die gedetecteerd worden door middel van "high-throughput sequencing", een overschatting kunnen geven van het aantal klonen We beargumenteren dat de recent ontwikkelde techniek van het barcoderen van cellen een goede methode is voor kwantitatieve analyse. Deze methode is gebaseerd op de integratie van een willekeurig DNA sequentie (ook wel barcode genoemd) in het genoom van een virale vector. Door transductie worden HSC's individueel gelabeld met een barcode en geven dit door aan hun nakomelingen. De labels worden gedetecteerd door amplificatie van het barcodefragment in het genomisch DNA en gesequenced. Doordat de structuur van de barcodes uniform is, is hun amplificatie en analyse ook uniform. Er zijn echter wel een aantal voorwaarden voor het succesvol toepassen van deze techniek, waaronder validatie van de grootte van de gebarcodeerde vector library, de beschouwing van het aantal barcodes dat nodig is per cel en het onderscheiden van data en ruis.

In **hoofdstuk 4** valideren we bovenbeschreven techniek voor klonale analyse door het combineren van het barcoderen van cellen en high-throughput sequentie detectie. Omdat Illumina (Solexa) sequencing miljoenen sequenties kan produceren per run, gebruiken we een

multiplexed format voor het amplificeren van barcodes met geïndexeerde primers. Dit maakt het voor ons mogelijk om tot 300 samples te combineren in een enkele mix. Om de gevoeligheid en resolutie van deze methode te bepalen hebben we analyses uitgevoerd op mengsels van gebarcodeerde cellen met vooraf bekende hoeveelheid cellen en variërende ratio's van barcodes. Barcodes met een frequentie van 0,5% konden kwantitatief worden gedetecteerd en barcodes konden betrouwbaar worden geanalyseerd in samples van minimaal 1000 cellen . Vervolgens hebben we deze benadering gebruikt voor het volgen van het klonale gedrag van honderden jonge en oude HSC's in polyklonale transplantaties. HSC's werden op basis van hun immunofenotype gezuiverd met flowcytometrie, getransduceerd met een gebarcodeerde vector library en getransplanteerd in bestraalde ontvangers. De compositie van barcodes van kortlevende granulocyten en langlevende B- en T-lymfocyten werd geanalyseerd in sequentieel afgenomen bloed monsters. Tussen 3 en 6 maanden na transplantatie levert het grootste deel van de getransplanteerde HSC's een bijdrage aan de bloedproductie. Echter, op lange termijn (tot 1 jaar na transplantatie) verandert de klonale representatie van zowel myeloïde als lymfoïde klonen in één richting, of ze verminderen, of ze expanderen. Tenslotte hebben we de compositie van een populatie van jonge en oude HSC's geanalyseerd in getransplanteerde ontvangers. De populatie jonge HSC bevatte slechts enkele potente klonen, terwijl de oude HSC pool bestond uit meerdere, maar kleine klonen.

Hoofdstuk 5 beschrijft de distributie van jonge en oude HSC klonen over het skelet na beenmergtransplantatie. Een barcode label is gebruikt om HSC's individueel te markeren voordat deze getransplanteerd werden in bestraalde ontvangers. We isoleerden HSC uit 4 verschillende soorten botten, 6 tot 11 maanden na transplantatie, en vergeleken de barcode compositie op verschillende locaties in het skelet. We laten zien dat de meeste HSC klonen gelokaliseerd zijn op één enkele locatie. Deze lokalisatie houdt aan tot tenminste 11 maanden na transplantatie, hetgeen suggereert dat migratie van HSC zeer beperkt is. Echter, één enkele dosis van Granulocyte Colony Stimulating Factor (G-CSF), dat veelvuldig wordt gebruikt om HSC's te oogsten voor beenmergtransplantaties, leidt tot een evenwichtige verspreiding over het skelet. Verder is de compositie van de HSC pool in het beenmerg vergeleken met de klonale samenstelling van het bloed in secundaire ontvangers. We laten zien dat de barcodes zoals ze voorkomen in het bloed na transplantatie overeenkomen met die zoals ze voorkomen in de botten die gebruikt zijn voor de transplantatie, echter de grootte van de klonen wijkt substantieel af tussen beenmerg en bloed.

Hoofdstuk 6 is een samenvatting van de belangrijkste bevindingen van dit proefschrift, geeft deze weer in een algemeen perspectief en bediscussieerd de potentiele impact.

ACKNOWLEDGEMENTS

After so many years, my stay in Groningen is almost over. While I hoped my PhD project will be smooth as a Dutch highway, I couldn't escape the "Russian road" with a lot of bumps and gaps. Before making the next step, I would like to thank all the people who made this thesis happen – and it is a long list!

First of all, my deepest gratitude goes to my promotor **Gerald de Haan** and my co-promotor **Leonid Bystrykh**. I was truly lucky to get supervision from both of you, as the things I learnt were very complementary. You made a huge contribution to what I am as a scientist.

Gerald, looking back I understand that taking into the lab a student with no biology background has been quite a risk. Thank you for your trust in me! Thank you for introducing me into the exciting world of stem cells and letting me pursue my PhD in an excellent scientific environment. I am always inspired by your enthusiasm about science, your positive attitude, and your ability to make others excited about the research. I also appreciate the freedom you gave me during the project. You have always been extremely supportive in helping me make the career choices, and you always took time to discuss my concerns. I also really value all the exposure you gave to me and my results by letting me participate in so many conferences and introducing me to the leading scientists in the field.

Leonid, **Леонид Витальевич,** I am very grateful for being supervised by you over the last 6 years. You are an exceptional scientist – very critical, with very original ideas, and you always have the courage to ask difficult to-the-point questions. I was always impressed by your ability to look directly into the root of a problem. I really appreciate all the time and effort you invested in me to teach me how to analyze experiments in systematic way, how to look at math and statistics as friends and not enemies, and how to view every new result as an extension of our knowledge. I also really value your support and encouragement through all these years!

Next, I would like to thank the members of reading committee – **Prof. Connie Eaves, Prof. Ingo Röder** and **Prof. Ton Schumacher** - for taking your time to assess and comment on my thesis. I highly respect your work and it has been an honor to have you approve my dissertation.

Being a PhD means spending a lot of time in the lab, and I would like to thank all the colleagues who are met and interacted with through the years. I have been lucky to meet a lot of extremely talented and inspiring people, and many of them have become my good friends.

Mathilde and **Katya**, thank you for agreeing to be my paranymphs and for helping me organize this day! **Mathildushka**, I cannot imagine how I would go through this PhD without you in the lab! Your optimism, your jokes always helped to go through a hard day. Also I think it is safe to say at least half of the data presented in this book has been generated with your golden hands

(how many thousand monoclones was it?). You are extremely responsible, hard-working and creative and it is nice to see how important you became not only for functioning of our lab, but of the whole Eriba. **Katya**, for the last couple years, not a day came by without me sharing all of my joys and sorrows with you. You are a great friend, neighbor and, (for the last year) a colleague. You are a very talented, determined and dedicated scientist, and I believe there is a great career in front of you whichever trajectory you will choose.

For a beginning scientist, it is often hard to choose the best path to follow at the crossroad of all options. **Brad, Ronald** and **Hein**, your experience, expertise and advice were extremely important to guide me through the start of my PhD. **Olya**, it was worth starting a PhD just to meet you! You made my torturous first attempts in using molecular biology techniques feel better. I am still hopeful "Oncoengine" will find its way into a publication and our lost Christmas vacation will be useful. I am really happy you are my friend! **Brad**, you have been a great example of answering a scientific question in a best possible way, and I learnt a lot from you – both methodologically and conceptually. Thank you for always being willing to help and share your knowledge, and for your excitement about science. I am sure you will become a wonderful group leader! **Ronald,** you always contributed to good atmosphere in the lab and at conferences by your jokes and dancing. I personally also profited from your extensive experience with *in vivo* experiments, and I am looking forward to getting the irradiation story done! (All these old papers to read still…) **Erik**, what you do is not only science but also ART! Will we ever get these Delftse blauw pictures on silk? Thank you for all the help with bioinformatics and for your patience! (Zucht). **Martha**, how would I manage all the mouse experiments without you? Thank you for all your work and for always being so kind! Also the best of luck with the mouse aging clinic! **Ellen**, you have a huge amount of lab experience, and no wonder we all want support from you. Sometimes our requests are weird – who would expect to plan a "non-sexy experiment"? I really appreciate all your help! **Bertien**, I don't know how we would get through all the levels of bureaucracy without you. Thank you for your help with animals experiments and for sharing your expertise on cell culture! **Jaring**, what a coincidence you moved from one barcode lab to another one! It was great to know you as a fellow student, but I also am very happy you established the "Jaring assay" with Brad, as it is in the core of publications in this thesis. I hope you are enjoying Australia! **Sandra O**, your speed in performing experiments was legendary in the lab and you surely were missed! Thank you for your help! **Martusia**, you are our escapee from the academic world, and I am very glad to see your happy photos. I tell about my talented friend to every Stemcelltechnologies employee I ever meet ☺ It was great to share office with you, to have a fellow PhD sufferer to share the experience, but also to spend so much time outside of the lab (Andrzejki were so much fun!). It was also a lot of fun to be in the Gopher board with you and Thomas, and I am proud we were the founding members. Soon we will share the same coast, so I hope to see you in the near future! **Edyta**, you used to be my student, but now I can learn a lot of things from you! You have profound knowledge in many

scientific areas, and I believe you will be very successful. Don't forget to relax and have some fun in your PhD! I will miss "Ptasie mleczko" when I am gone! **Visnja**, it is also almost your time to graduate! It was nice to share an office with you. The best of luck for your postdoc! **Karin**, you became a very successful PhD student, and now decided to stay in our lab for a postdoc which I am sure will be very successful as well. I will never forget our Australian trip with **Alice** and **Brad** (speedies, spiders and flying foxes?!). **Alice**, you were the first offspring of Topmaster project, and you remain an example of true perfectionist in the lab. The new additions to de Haan group – **Sara, Seka, Mirjam, Taco** and **Johannes** – good luck with your projects! Starting up is always hard, but I am sure you will produce a lot of nice results at the end. Thank you for gezelligheid at work and outside work! **Sophia**, it must be very exciting to start your own lab, and I wish you a lot of success! I would also like to thank **Vincent, Sandra R, Esmee, Evert-Jan** and **Niek** for the time shared in the lab and at work discussions.

I spent an uncountable number of hours at the FACS facility, and I would like to thank **Henk, Geert** and **Roelof-Jan** for their help, nice conversations and improving my Dutch skills. You really had to have patience with my samples (who else could spend the whole day analyzing a single mouse?) and I appreciate your professionalism and expertise. **Henk**, bedankt voor alle gesprekken over Nederland en Rusland, voor jouw steun en vriendlijkheid! Ik ga je heel erg missen! **Geert**, with your impressive knowledge of the sorters, we knew every problem would be solved very quickly. **Roelof-Jan**, thank you for being willing for stay that extra (half-)hour needed to finish the sort!

I had a pleasure of supervising a number of great students over the course of my PhD: **Edyta, Arend, Lisette, Taco** and **Theo**. All of you generated a lot of nice data, some of which was (or will be) used in the publications and I had a lot of fun interacting with you. I am happy to see two of you – **Edyta** and **Taco** – as my colleagues now. I wish you all the best of luck in your further careers.

Dear **Ed**, thank you for designing the cover of my thesis. You are a great illustrator, you always have a backup in case you want to quit science.

I would like to thank **Pieter van der Vlies** and **Jelkje** from Department of Genetics for the help with analyzing high-thoughput sequencing data. I would also like to thank the the stuff of **CDP** for their support in performing the animal experiments.

For the last year, I have been working on collaborative project with the lab of **Prof. Eugene Berezikov**. I would like to thank **Zhenya** for letting me join the lab, and all lab members – **Daniel, Dan, Frank, Jacub, Katya, Lisa, Magda, Philipp** and **Stijn** – for all the discussions, interactions and introducing me to a completely new world of *Macrostomum lignano*.

I would also like to acknowledge the group of **Dr. Sjef Copray** for teaching me to work with induced pluripotent stem cells. Dear **Koen**, thank you for your clear instructions and your time!

I would also like to thank the members of groups of **Prof. Harrie Kampinga, Prof. Rob Coppes** and **Prof. Ody Sibon** for all the nice interactions at the time our lab was still located on the 5th floor. Our Christmas dinners, Sinterklaas parties and lab days were always a lot of fun, and it was great to be surrounded by so many friendly faces. **Sarah, Martti** and **Yamini**, it was exciting to follow your progress in salivary stem cell research during our work discussions.

Working in a large organization always involves a lot of bureaucracy, but I almost didn't feel it thanks to wonderful support from our secretaries, lab managers and receptionists. For the large part of my PhD, while we were located on the 5th floor, I was lucky to work with **Annet** and **Gerrie**, and once we moved to Eriba, also with **Sylvia, Maria, Joke, Kim, Henk** and **Helena**. Thank you for your help!

I would like to thank the members of **Department of Experimental Hematology** for the critical comments and discussions during Monday morning seminars, and for the nice interactions at national hematology meetings. In particular, **Marta** and **Susan**, good luck finishing your PhDs!

It was great to meet a lot of new people upon the move to Eriba (and working in open space environment is especially instrumental for that), and I would like to thank the members of the labs of **Prof. Peter Lansdorp,** and **Dr. Floris Foyer** for gezelligheid on the first floor.

My science journey in Groningen started with **Topmaster program in Medical and Pharmaceutical Drug Innovation**. Dear **Han, Anke** and **Harrie**, thank you for organizing this amazing introduction into research.

I would like to thank the lab of **Prof. Rainer Bischoff** for hosting me for the first Topmaster rotation and for proteomics studies in the beginning of my PhD. **Natalia**, thank you for all the time, help and nice discussions! **Theo**, you have been a great master project supervisor, and I also learnt a lot about *carbonade* buffers and properly assembled *platjes*.

With the end of my PhD, I am moving to UCSF to join the lab of **Dr. Emmanuelle Passegué**. I would like to thank the whole lab for the warm welcome, and I am looking forward to working you!

Of course, there is life outside of work, and I was lucky to meet a lot of wonderful people in Groningen, within and outside PhD community. **Vitya** and **Ania K**, your every visit is a lot of fun! I am happy to see you so happy with each other! **Vladimir, Lora, Zmaray, Katya P, Georg** – thank you for all the good time spent together! People from **Department of Biomedical Engineering** – you are a great team, thank you for always being so welcoming! My fellow board

members from HUTAC and GOPHER, members of PhD Day Organizing Committee – it was great to share the exciting experience of setting up organizations and planning events with you!

My biggest gratitude goes to **my closest friends and family** – people who supported me throughout this long journey. I wouldn't manage without your encouragement! **Ксюша, Аня, Оля, Катя**, спасибо за вашу поддержку, мне очень повезло найти таких прекрасных подруг, я всегда могу на вас положиться! **Ксюшочек**, ты всегда можешь найти правильные слова и показать правильный пример. Для меня ты часть семьи. Семья! Для меня вы стержень, который помогает выдержать любые испытания. **Папа**, я очень рада, что следую по твоим научным стопам! **Саша**, помимо твоих рабочих успехов, ты успел стать папой. Как здорово, что я недавно стала тетей! Я очень жду встречи с Надюшкой. **Митюша**, я очень горжусь твоими спортивными достижениями. **Серый**, мне очень приятно, что в нашей семье появился еще один фармацевт! **Евгений Александрович**, Ваше чувство юмора незаменимо! **Тетечка**, твой оптимизм и доброта всегда вдохновляют! **Тетя Света**, Вы для меня прекрасный образец мудрости. И наконец, **мама**. Ты мой самый близкий и родной человек, мой герой и пример. Каждый раз, когда передо мной стоит выбор, я думаю о том, как бы поступила ты. Спасибо за всё, что ты делала и делаешь для меня.

CURRICULUM VITAE

Evgenia Valerianovna Verovskaya

Experience

2013 - 2014 **University Medical Centre Groningen. European Research Institute for the Biology of Ageing (ERIBA).** Groningen, The Netherlands.
Researcher in a joint project of Laboratory of Stem Cell Regulation (supervised by Prof. E. Berezikov) and Laboratory of Ageing Biology and Stem Cells (supervised by Prof. G. de Haan).

2008 - 2013 **University Medical Centre Groningen. European Research Institute for the Biology of Ageing (ERIBA).** Groningen, The Netherlands.
PhD student in Laboratory of Ageing Biology and Stem Cells under supervision of Prof. G. de Haan and Dr. L. Bystrykh. Supervised 5 master and bachelor student projects.

Education

2006 – 2008 **University of Groningen**. Groningen, The Netherlands.
Topmaster program in Medical and Pharmaceutical Drug Innovation.
Master of Science *cum laude*.

2001 - 2006 **Sechenov's Moscow Medical Academy.** Moscow, Russia.
Degree in Pharmacy. Diploma *with Honors*.

2000 – 2001 **Banks High School.** Banks, OR, USA.
Diploma *with Honors*.

1991 - 2001 **Gymnasium #3.** Volgograd, Russia.
Graduated with Gold Medal for outstanding results at studies.

Membership in boards and committees

2012 **Groningen PhD Day.**
Head of organizing committee (www.phd-day.nl/2012/)

2011 **Groningen Organization for PhD Education and Recreation (GOPHER).** Founding board member, treasurer, PR manager.

2010 – 2011 **Groningen Association for PhD Students (GRASP)** (over 400 members). Board member, treasurer, PR manager, event manager.

2009 – 2010 **Huygens Talent Circle (HUTAC).** Association for recipients and alumni of Huygens Scholarship Program.
Founding member, board member, secretary.

2007 - 2009 **University of Groningen. Topmaster program in Medical and Pharmaceutical Drug Innovation.** Educational committee.

2007 - 2008 **Erasmus student network – Groningen (ESN-Groningen).**
Erasmus magazine (EM) committee. Editor.

Language skills

Russian (native), English (fluent, TOEFL CBT – 273/300), Dutch (fluent, NT2 II diploma).

Honors and awards

2008 – 2012 **Toptalent grant** from the Netherlands Organization for Scientific Research (NWO) for implementation of self-designed PhD project.

2011 **Best poster presentation prize** at Hydra VII Summer School on Stem Cells and Regenerative Medicine.

2010 – 2013 **Travel grants** for participation in international conferences.
- V. EKF Symposium on Adult Stem Cells in Aging, Diseases, and Cancer in Eisenach (Germany).
- StemCellMathLab dedicated to quantitative tracking of clonality in Dresden (Germany).
- Annual Meetings of Society for Hematology and Stem Cells (ISEH) in Melbourne (Australia), Vancouver (Canada) and Amsterdam (the Netherlands).
- Congress of European Hematology Association (EHA) in London (the United Kingdom).

2007 – 2008 **HSP Huygens scholarship** from the Netherlands Organization for Cooperation in higher education (Nuffic) for excellent students covering 1-year study in the Netherlands.

2006 – 2007 **Scholarship of Graduate School of Drug Exploration** (GUIDE) for students selected for participation in Topmaster programs.

2006 **University medal "For the succeeding one"** from Moscow Medical Academy for students with all excellent grades.

2000 – 2001 **Future Leaders EXchange (FSA FLEX) scholarship** from US Department of State covering 1-year study in a high school in the USA.

Oral presentations

International conferences

2010, 2011, 2012 - Annual Meetings of Society for Hematology and Stem Cells (ISEH) – *Melbourne, Australia; Vancouver, Canada; Amsterdam, the Netherlands*
2012 – StemCellMathLab – *Dresden, Germany*
2011 - Congress of European Hematology Association (EHA) – *London, the United Kingdom*
2009 - Workshop on Innovative Mouse Models – *Leiden, the Netherlands*
2009 - EuroSyStem consortium meeting – *Cambridge, the United Kingdom*

National conferences

2010, 2011 - Annual meeting of Dutch Society for Stem Cell Research (DSSCR)
2010, 2011, 2014 - Dutch Hematology Congress

LIST OF PUBLICATIONS

Bystrykh, L.V., de Haan, G., Verovskaya, E. Barcoded vector libraries and retro- or lentiviral barcoding of hematopoietic stem cells. In: Hematopoietic stem cell protocols, 3rd edition (in press).

Wójtowicz, E.[1], Verovskaya, E.[1], de Haan, G. Aging of hematopoietic stem cells in the mouse. In: Stem cell aging: Mechanisms, Consequences, Rejuvenation (in press). [1]Equal contribution.

Verovskaya, E., Broekhuis, M.J., Zwart, E., Weersing, E., Ritsema, M., Bosman, L.J., van Poele, T., de Haan, G., and Bystrykh, L.V. Asymmetry in skeletal distribution of mouse hematopoietic stem cell clones and their equilibration by mobilizing cytokines. J. Exp. Med. *211,* (2014), 487-497.

Verovskaya, E., and de Haan, G. Noncanonical Wnt comes of age in hematopoietic stem cells. Cell. Stem Cell. *13,* (2013), 642-643.

Verovskaya, E., Broekhuis, M.J., Zwart, E., Ritsema, M., van Os, R., de Haan, G., and Bystrykh, L.V. Heterogeneity of young and aged murine hematopoietic stem cells revealed by quantitative clonal analysis using cellular barcoding. Blood *122,* (2013), 523-532.

Bystrykh, L.V., Verovskaya, E., Zwart, E., Broekhuis, M., and de Haan, G. Counting stem cells: methodological constraints. Nat. Methods *9,* (2012), 567-574.

Verovskaya, E. and de Haan, G. The power of diversity: hematopoietic stem cell heterogeneity and its clinical relevance. Hematology Education *5(1),* (2011), 132-139.

Gerrits, A., Dykstra, B., Kalmykowa, O.J., Klauke, K., Verovskaya, E., Broekhuis, M.J., de Haan, G., and Bystrykh, L.V. Cellular barcoding tool for clonal analysis in the hematopoietic system. Blood *115,* (2010), 2610-2618.

Ilyasov, I.R., Beloborodov, V.L., Tyukavkina, N.A., Verovskaya, E.V. Use of ABTS·+ cationic radicals to evaluate the antiradical activity of flavonoids. Pharmacy *6*, (2008), 15-18 (In Russian).