Cell Stem Cell
Article

Clonal Tracking of Rhesus Macaque Hematopoiesis Highlights a Distinct Lineage Origin for Natural Killer Cells

Chuanfeng Wu,[1,7] Brian Li,[1,7] Rong Lu,[2,7] Samson J. Koelle,[1] Yanqin Yang,[3] Alexander Jares,[1] Alan E. Krouse,[1] Mark Metzger,[1] Frank Liang,[4] Karin Loré,[4] Colin O. Wu,[5] Robert E. Donahue,[1] Irvin S.Y. Chen,[6] Irving Weissman,[2] and Cynthia E. Dunbar[1,*]

[1]Hematology Branch, National Heart, Lung and Blood Institute, National Institutes of Health, Bethesda, MD 20892, USA
[2]Institute for Stem Cell Biology and Regenerative Medicine, Stanford University School of Medicine, Palo Alto, CA 94305, USA
[3]DNA Sequencing and Genomics Core, National Heart, Lung and Blood Institute, National Institutes of Health, Bethesda, MD 20892, USA
[4]Vaccine Research Center, National Institutes of Health, Bethesda, MD 20892, USA
[5]Office of Biostatistics Research, National Heart, Lung and Blood Institute, National Institutes of Health, Bethesda, MD 20892, USA
[6]UCLA AIDS Institute, David Geffen School of Medicine at UCLA, Los Angeles, CA 90095, USA
[7]These authors contributed equally to this work
*Correspondence: dunbarc@nhlbi.nih.gov
http://dx.doi.org/10.1016/j.stem.2014.01.020

SUMMARY

Analysis of hematopoietic stem cell function in nonhuman primates provides insights that are relevant for human biology and therapeutic strategies. In this study, we applied quantitative genetic barcoding to track the clonal output of transplanted autologous rhesus macaque hematopoietic stem and progenitor cells over a time period of up to 9.5 months. We found that unilineage short-term progenitors reconstituted myeloid and lymphoid lineages at 1 month but were supplanted over time by multilineage clones, initially myeloid restricted, then myeloid-B clones, and then stable myeloid-B-T multilineage, long-term repopulating clones. Surprisingly, reconstitution of the natural killer (NK) cell lineage, and particularly the major $CD16^{+}/CD56^{-}$ peripheral blood NK compartment, showed limited clonal overlap with T, B, or myeloid lineages, and therefore appears to be ontologically distinct. Thus, in addition to providing insights into clonal behavior over time, our analysis suggests an unexpected paradigm for the relationship between NK cells and other hematopoietic lineages in primates.

INTRODUCTION

Understanding how hematopoietic stem and progenitor cells (HSPCs) produce a diversity of blood and tissue cell lineages at a high rate for the life of an individual, under homeostatic control, without exhaustion, and with only rare catastrophic neoplastic dysregulation has been a major investigative focus. Insights have been translated into a myriad of clinical advances, including hematopoietic stem cell (HSC) transplantation, cytokine treatment of cytopenias, and targeted therapies for hematopoietic neoplasms. Despite such progress, however, many issues remain to be investigated, including the degree of homogeneity of long-term self-renewing HSCs, the difficulty of modeling hematopoietic lineage ontogenies in vitro, the control of fate decisions, the patterns of the contributions of individual cells in vivo over time, and the number and lifespan of true HSCs (Copley et al., 2012). Murine HSCs have been characterized at close to a single-cell level in gold-standard transplantation assays via cell-surface phenotype and gene-expression profiling (Osawa et al., 1996; Smith et al., 1991; Spangrude et al., 1988; Weissman and Shizuru, 2008). Limit-dilution transplantation experiments define what one or a few stem cells can do, but whether HSC behavior in this setting of extraordinary replicative stress fully mimics normal steady-state hematopoiesis or even nonlimit-dilution transplantation is questionable.

Transduction of HSPCs with integrating viral vectors results in passage of the integrated provirus to every daughter cell. Each semirandom viral integration site (VIS) serves as a unique clonal "tag" for individual HSPCs and their progeny. Pioneering studies in mice identified VISs via Southern blot and provided evidence for multilineage clonal repopulation by single cells (Jordan and Lemischka, 1990); however, this insensitive approach detects only dominant clones and cannot resolve highly polyclonal patterns or investigate contributions to cells from less abundant lineages, such as natural killer (NK) cells. A number of more sensitive PCR-based VIS retrieval methods have been devised, but all have limitations regarding both efficiency and quantitation (Berry et al., 2012; Bushman et al., 2005; Schmidt et al., 2002; Wu et al., 2013). An alternative is to generate high-diversity lentiviral libraries to deliver a unique "barcode" to individual HSPCs, allowing quantitative clonal tracking. This concept was developed to study murine antigen-specific T cell dynamics and extended to proof-of-concept studies in murine HSPCs (Gerrits et al., 2010; Schepers et al., 2008). In our previous study (Lu et al., 2011), we combined lentiviral barcoding with PCR retrieval and high-throughput sequencing to elucidate the behavior of HSPCs in mice. The results showed that clonal output from individual cells can be assessed in a sensitive and quantitative manner, and yielded important insights into murine hematopoietic lineage bias.

Studying human HSPC clonal behavior is more challenging (Baum et al., 1992; Doulatov et al., 2012). Limit-dilution transplantation or vector tagging of human HSPCs in immunodeficient mice has provided valuable information regarding the phenotype of human engrafting cells and some information about their behavior in vivo (Doulatov et al., 2012; Guenechea et al., 2001; Larochelle et al., 1996). However, the behavior of human cells in mice is far from physiologic, given the reliance on a xenogeneic niche, with little release of fully differentiated progeny into the blood and very limited development of T or NK cells (Coulombel, 2004). Hematopoiesis in rodents is markedly different from that in large animals or humans with regard to cytokine utilization, location of stress hematopoiesis, HSPC phenotype and frequency, and lifetime hematopoietic demand (Abkowitz et al., 1996, 2002; Catlin et al., 2011; Larochelle et al., 2011; Shepherd et al., 2007).

We have utilized rhesus macaque autologous transplantation to investigate HSPC behavior in a model with direct relevance to humans (Donahue and Dunbar, 2001). Humans and macaques are phylogenetically close, with comparable lifespans, hematopoietic demand, and HSPC frequencies (Shepherd et al., 2007). We can efficiently mark rhesus HSPCs with lentiviral vectors and track marked progeny cells for many years (Kim et al., 2009). We now report the use of lentiviral HSPC barcoding to track and quantitate clonal contributions following autologous rhesus HSPC transplantation, and provide multiple insights into hematopoietic reconstitution and lineage hierarchies, most notably of NK cells.

RESULTS

Experimental Design and Validation

In order to track in vivo output from individual HSPCs in the rhesus macaque, we adapted our lentiviral cellular barcoding technology to rhesus $CD34^+$ HSPCs and followed clonal output following myeloablative autologous transplantation (Figure 1A). Highly diverse DNA barcode libraries were cloned into lentiviral vectors that also contained a GFP marker. Using mixtures of single-copy barcoded K562 cell clones, we showed that the retrieved barcode frequency quantitatively reflects the contribution of an individual clone within a polyclonal mixture over a range of ratios (Figure S1 available online).

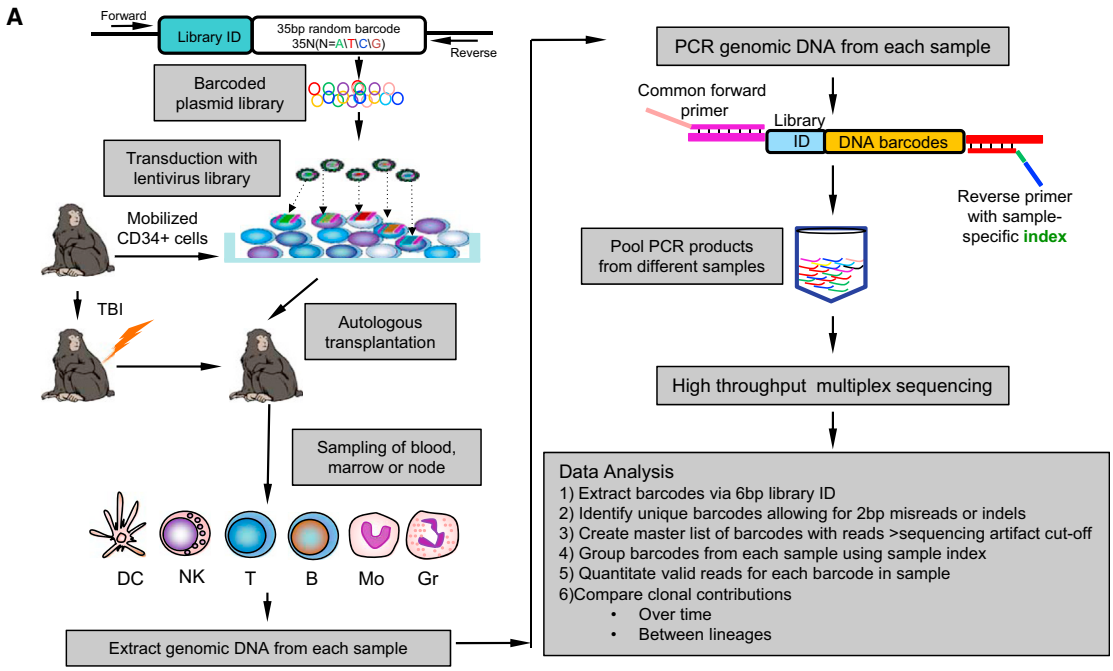
Mobilized peripheral blood (PB) $CD34^+$ cells from three macaques were transduced with barcoded vectors, and 7.8–16.7 million autologous GFP^+ cells were reinfused following ablative total body irradiation (Figure 1B). Vector barcode diversity was assessed following retrieval of barcodes from transduced rhesus cells at the end of transduction, and Monte Carlo simulations using these experimental data were performed to ask whether the library diversity was sufficient to ensure with high confidence that a single barcode labeled a single cell (Figure S2; Lu et al., 2011). Based on the best available prior estimate for the frequency of rhesus repopulating cells (i.e., 5 per 100,000 $CD34^+$ cells), the barcoded libraries chosen had at least 10-fold higher diversity than was required to be more than 95% certain that a single barcode marked only one HSPC clone at these transplanted doses (Kim et al., 2000; Shepherd et al., 2007).

For the most straightforward tracking of individual HSPCs, the presence of only a single barcode per cell is desirable; thus, we targeted a maximum $CD34^+$ transduction efficiency of no more than 30%–35%, based on prior models (Kustikova et al., 2003; Verovskaya et al., 2013). The level of GFP positivity in circulating blood cells from myeloid and lymphoid lineages following transplantation is shown in Figure 1C. As expected, granulocytes (Gr) and monocytes (Mo) counts recovered rapidly, within 2 weeks of transplantation, followed by NK, B, and then T cells, and reached normal levels by 3–6 months. We found that 10%–30% of hematopoietic cells in all lineages were GFP^+ in ZH33 and ZG66, and stabilized by 2–3 months in Gr, Mo, B, and NK cells, with transduced T cell engraftment lagging behind (Figure 1C), reflecting slower replacement of endogenous T cells as we observed previously (An et al., 2007). ZH17 showed a similar pattern but overall much lower marking. It stabilized at levels of only 1%–2% despite receiving $CD34^+$ cells with the same transduction efficiency as the other two animals, likely reflecting greater competition from endogenous HSPCs that survived radiation, due to differences in radiation sensitivity between individual macaques.

Following engraftment, PB and lymph nodes were collected at various time points, and Gr, Mo, T, B, NK, and plasmacytoid and myeloid dendritic cells (DCs) were flow sorted (purity median 98.8%; Figure S3; Table S1). The mean barcode copy number per individual transduced HSPC in individual colonies (colony-forming units [CFUs]) derived from $CD34^+$ cells plated at the end of transduction in animal ZH33 was 1.2 ± 0.22 (SEM), and 1.07 ± 0.07 in CFUs grown from $CD34^+$ cells obtained from the bone marrow (BM) of ZH33 following engraftment, with only 7% of CFUs containing two copies and none containing more than two (Figure 2A). We also compared vector copy number and % GFP^+ cells in PB posttransplantation in all three animals (Figure 2B), and the mean copy number per GFP^+ cell was 1.04 ± 0.04. In total, these data confirm that the majority of transduced cells contained only one barcode. A minority of clones containing more than one barcode would skew calculations of the frequency of repopulating clones upward, but would not impact analysis of lineage contributions or kinetics.

Barcode retrieval by PCR, Illumina sequencing, and custom data analysis was performed on purified hematopoietic lineage samples monthly for 9.5 months (ZH33), 6.5 months (ZH17), and 4.5 months (ZG66) (Lu et al., 2011). Our algorithms extract barcodes by library IDs, account for sequencing misreads and indels, quantitate valid reads for each barcode, and create a master list of barcodes for each animal, consisting of barcodes present in at least one sample at a read number over cutoffs determined for each animal to exclude 99.5% of "false" artifactual barcodes (see the python code in Supplemental Experimental Procedures, and the master list of barcodes and contributions to each lineage and time point in Table S4). Very high reproducibility of barcode retrieval and quantitation was demonstrated via sequencing and PCR replicates on the same DNA samples (Figure S4).

We assayed independently processed replicate blood samples to identify a lower barcode read threshold that would result in 95% barcode retrieval between replicates. This method was designed to exclude false negatives arising from sampling error related to the number of cells (mean of 70,000 cells processed to yield 200 ng DNA) in each purified sample (Figure S4). For the highly marked animals ZH33 and ZG66, the majority of



B

	ZH33	ZH17	ZG66
Vector backbone	pCDH-MCS-T2A-copGFP-MSCV	pCDH-EF1α-MCS-T2A-copGFP	pCDH-MCS-T2A-copGFP-MSCV
Barcode library	MSCV #11	Ef1α #2	MSCV #11
Number of infused cells	32 x10^6	22x10^6	48x10^6
Transplant dose (CD34+ cells/kg)	6.9x10^6	4.1x10^6	8.5x10^6
GFP+ % infused cells	34.7%	35.3%	34.8%
Infused GFP+ cells	11.1 X 10^6	7.8 X 10^6	16.7 X 10^6

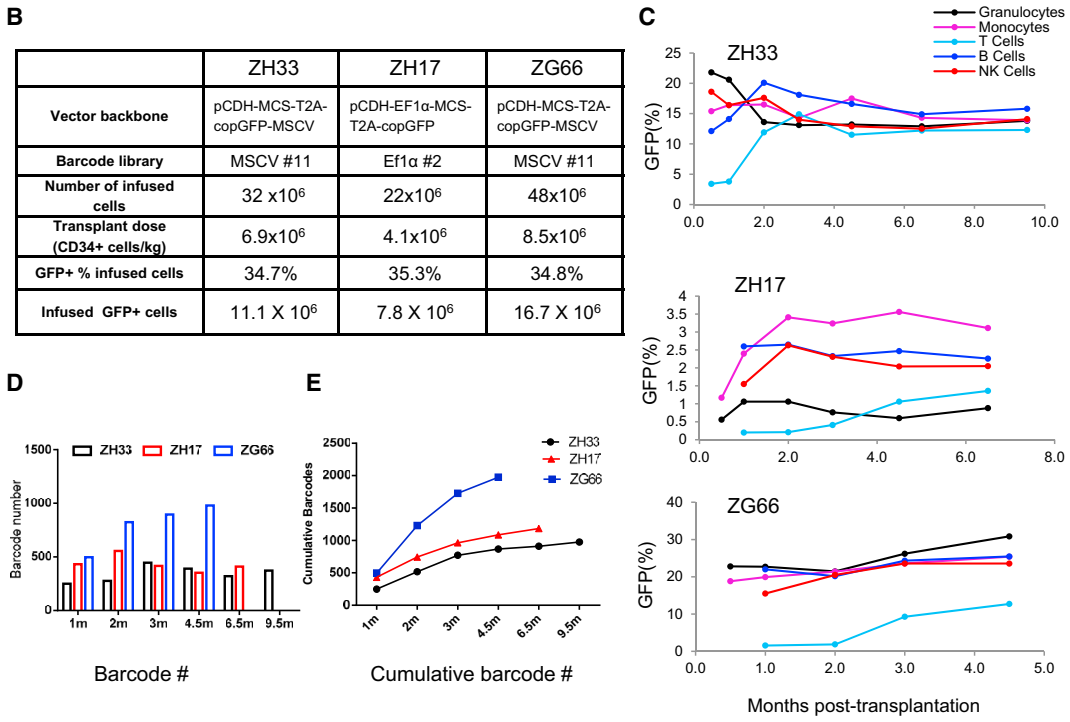


Figure 1. Experimental Design and Clonal Diversity

(A) Oligonucleotides consisting of a library ID followed by a random barcode were cloned into a lentiviral vector flanked by PCR primer sites. Mobilized PB autologous CD34$^+$ cells were transduced and infused back into the irradiated autologous macaque, and PB, BM, and lymph node samples were collected, purified by fluorescence-activated cell sorting (FACS), and barcode retrieval was performed via PCR and high-throughput sequencing. Information relevant to the methods used is presented in Figures S1–S4 and Tables S1–S3.

(B) Transduction and transplantation parameters for rhesus macaques ZH33, ZH17, and ZG66.

(C) Percentage of GFP$^+$ cells in PB lineages over time following transplantation.

(D) Number of independent barcoded clones detected in the PB (combined B, T, Gr, Mo, NK) at each time point (m, months).

(E) Cumulative number of independent barcoded clones detected in PB (combined B, T, Gr, Mo, NK) over time.

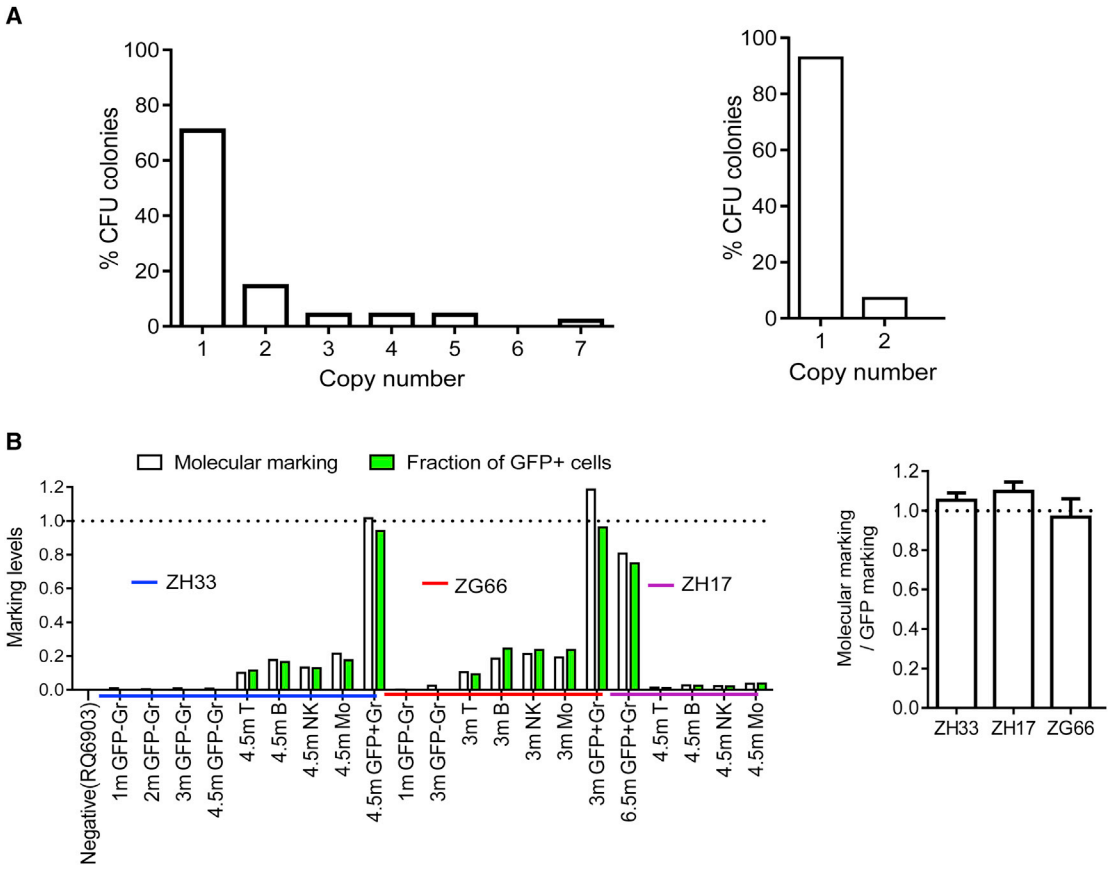


Figure 2. Barcode Copy Number per HSPC Clone

(A) Barcode copy number in individual CFUs obtained at the end of transduction (left, n = 48) or from marrow CFUs posttransplantation from ZH33 (right, n = 14). (B) Barcode copy number per transduced cell for PB samples. The molecular marking level (vector copy number) was determined by qPCR on sorted GFP-negative granulocytes (GFP$^-$ Gr), GFP-positive granulocytes (GFP$^+$ Gr), and sorted T, B, and NK cells from ZH33, ZH17, and ZG66, and plotted on the left panel (white bars) next to the fraction of GFP$^+$ cells in the same sample as assessed by FACS (green bars). A ratio of 1 implies one vector copy per each GFP$^+$ cell. On the right, the ratios of molecular marking level determined by copy number PCR/GFP marking level for all samples from each of the three animals (mean ± SEM) are shown.

barcoded hematopoiesis derived from clones above this threshold, and only clones above the threshold were included in pairwise comparisons between lineages or time points. ZH17, with a much lower overall in vivo marking level, was not used for pairwise comparisons because its barcode read threshold was at least 10-fold higher than in the other animals, which limited our ability to compare low-abundance clones between samples. However, individual abundant clones from this animal were tracked and overall clone numbers were scored.

Overall HSPC Clonal Diversity Posttransplantation

Figures 1D and 1E plot the number of unique barcodes that contributed at each time point, and the cumulative number of unique clones that were identified over time. A total of 976 individual clones in ZH33 (9.5 months), 1,184 in ZH17 (6.5 months), and 1,977 in ZG66 (4.5 months) were detected at one or more time points, at a read number above the cutoff determined for each animal set to exclude sequencing artifacts. Based on the number of GFP$^+$ cells transplanted (Figure 1B), ~1 in 9,000 CD34$^+$ cells contributed at any time point (range 1 in 6,500–11,300). This is a lower limit, given that low-contributing clones were missed due to sampling constraints or excluded by sequencing background cutoffs. The cumulative number of clones detected began to plateau as early as 3–4 months posttransplantation, with fewer new clones appearing after that time point. At the longest follow-up in each animal (4.5–9.5 months), the frequency of clones that were still contributing was ~2.5-fold lower: 1 in 22,000 CD34$^+$ cells (range 1 in 17,000–30,000).

Lineage and Longitudinal Clonal Contributions

We assessed the fractional contributions of individual clones to different hematopoietic lineages over time. As shown in Figure 3 (ZH33) and Figure S5 (ZG66), the fractional contribution of each individual barcoded clone can be directly compared between lineages and time points, and the distribution of lineage bias in the population of clones can be assessed. The frequency of highly biased clones (defined as >10-fold fractional representation of a barcode in one lineage or time-point sample versus another), as well as pairwise Pearson correlations between all lineages and time points, can be quantified (Figure 4). Spearman analysis

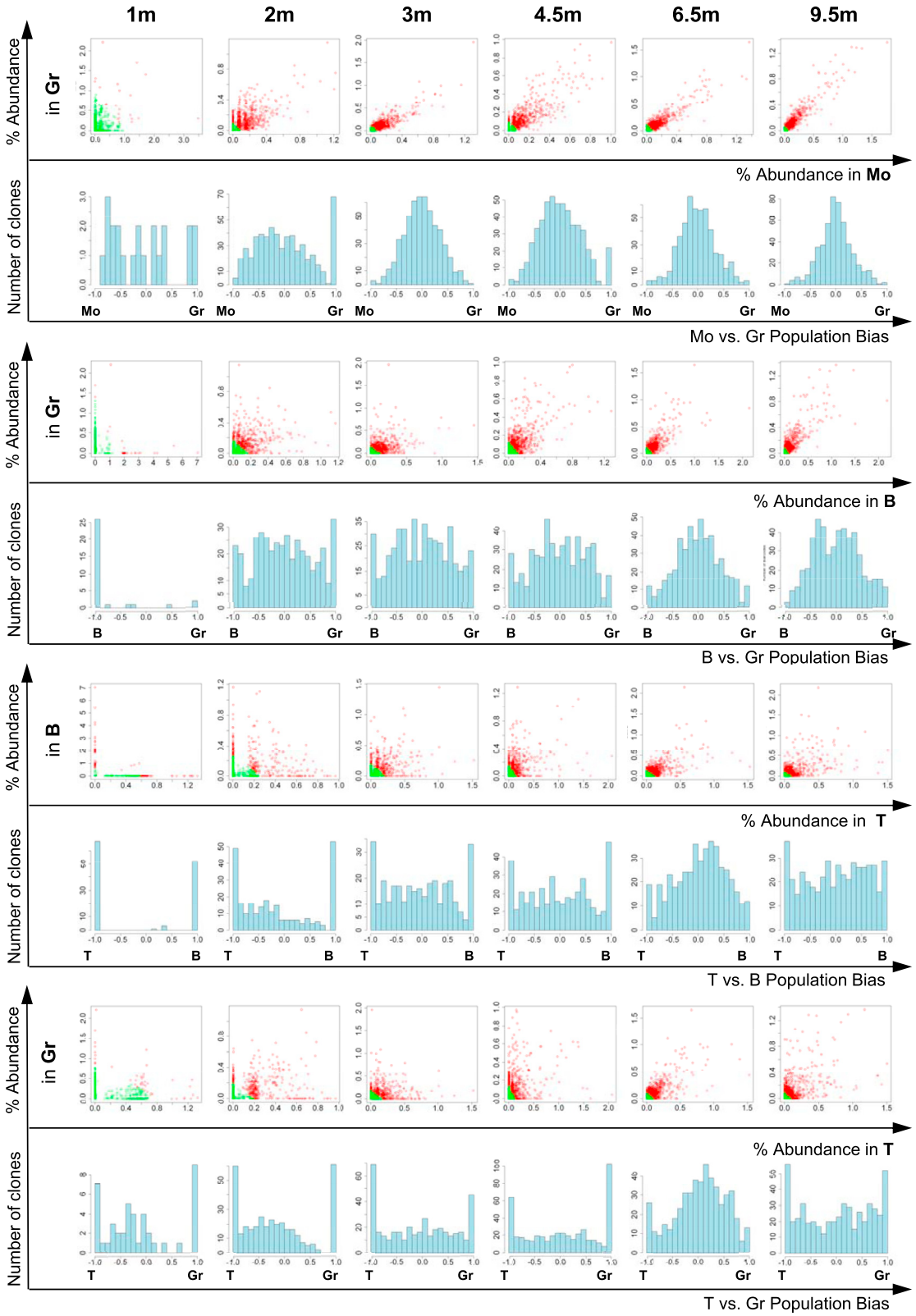


Figure 3. Clonal Comparisons between Lineages

Scatterplots compare sequencing read fractions in ZH33 between pairs of lineages over time. Each dot represents an individual barcode/clone from the master list, detected in either lineage. Green dots are barcodes with reads falling below the sampling threshold of 1,144 reads combined between the two lineages, and

(legend continued on next page)

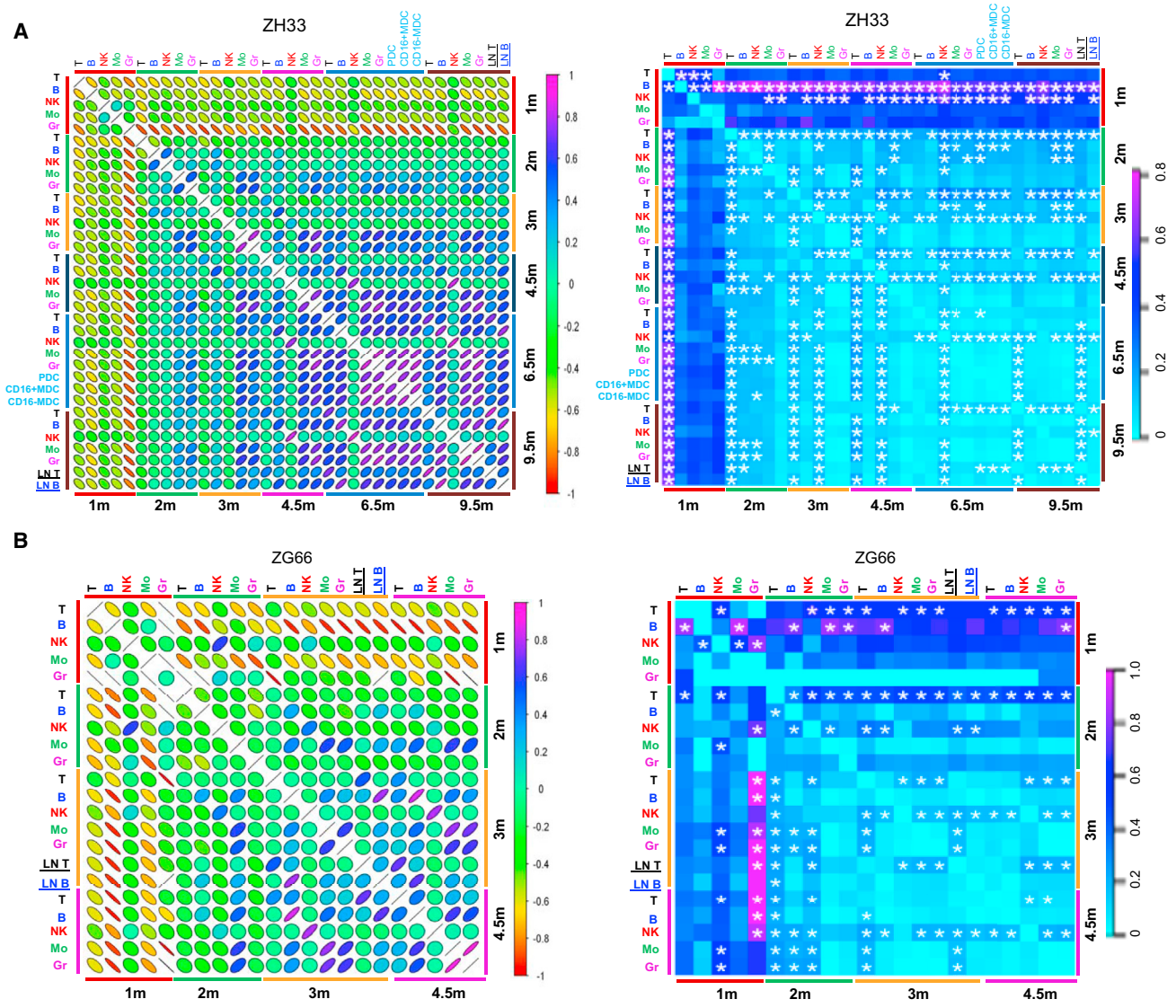


Figure 4. Summary of Clonal Lineage and Kinetic Relationships over Time

(A) Pearson correlation coefficients comparing barcode contributions for all barcodes that fall above the sampling threshold of 1,144 for pairwise comparisons, between all lineages and time points, for animal ZH33 and ZG66 (r values and their associated p values and 95% confidence intervals are given in Table S5). The color scale bar for r values is on the right. Slope indicates negative or positive correlation; the shape and the color signify the strength of the correlation. MDCs, myeloid dendritic cells; PDCs, plasmacytoid dendritic cells; LN, lymph node. An alternative analysis utilizing Spearman correlation is shown in Figure S6 and Table S6.

(B) Fraction of highly biased (>10-fold greater fractional representation in one sample versus the other) barcodes between all lineages and time points for ZH33 and ZG66. Bias is toward the sample shown in the row and away from the sample in the column. The color scale is given on the right. The frequency of highly biased clones in these comparisons differed significantly from the total frequency of highly biased clones at the time point combination (*$p < 0.05$). A hierarchical clustering analysis for fractional barcode contributions in ZH33 and ZG66 is given in Figure S7.

of the data set gave essentially the same results, although because of the different nature of the two tests, some patterns of biological interest are not as apparent in the Spearman analyses (Figure S6).

At 1 month posttransplantation, there was no correlation between lineages, with unilineage or extremely biased low-abundance clones predominating. Clones that were contributing at 1 month did not persist in contributing substantially to any

red dots are barcodes with reads above the threshold. The bias histograms are generated from the scatterplots for barcodes that fall above the threshold by measuring the angle from the axis for each barcode, and renormalizing from −1 to 1 before placing each barcode into a bias bin. The number of clones in each bias bin is on the y axis, and the degree of bias is on the x axis, with clones having equal lineage contributions defined as zero. Each clone is counted equally for the bias histograms without regard to its overall level of contribution. Analogous data from animal ZG66 are shown in Figure S5.

lineage at 3 month or later (Figures 3 and 4), and thus were defined as short-term, lineage-restricted progenitors. The first bilineage Gr/Mo clones began to emerge at 2 months and rapidly dominated. By 3 months, virtually all Gr and Mo clones were shared and there were few highly biased clones, directly demonstrating in vivo output from common Gr/Mo progenitors. By 3 months, clones shared between myeloid (Gr/Mo) and B lineages began to dominate, with stronger correlations as well as a decrease in highly biased Gr/Mo or B clones, before clones contributing to T and B lineages appeared (Figures 3 and 4). By 4.5–6.5 months, multilineage, less biased Gr/Mo/B/T clones became dominant. $CD11c^+CD16^+$ and $CD11c^+CD16^-$ myeloid DCs, as well as $CD11c^-CD123^+$ plasmacytoid DCs collected at 6.5 months from ZH33, were most closely related to each other and to the Gr/Mo myeloid lineages.

To further depict the complexities of the clonal contributions to all lineages, we performed hierarchical clustering analysis for ZH33 and ZG66 (Figure S7), which demonstrated grouping of the 1 month samples separate from later time points for all lineages, confirmed the close relationship between Gr and Mo, and showed the lack of a clear grouping of B and T lineages separate from myeloid lineages at any time point.

We also compared the clonal composition of blood versus lymph node T and B cells in ZH33 (9.5 months) and ZG66 (3 months), and found shared clonal patterns of these lineages in blood versus node (Figure 4), with a single expanded T cell clone outlier in the PB of ZG66 accounting for almost 20% of marked cells not found at a similar level in the one node sampled.

NK Cell Ontogeny

We were surprised to find a very different pattern of clonal contributions in NK cells (Figures 4, 5, and S6). Between 3 and 6.5 months, when many shared Gr/Mo/T/B clones were contributing, NK clonal composition remained distinct, and then began to converge toward the other lineages by 9.5 months (Figures 4 and 5). Highly biased clones predominated compared with the other four lineages at the same time points. Triangle plots (Figure 6) confirmed the presence of highly skewed NK contributing clones distinct from shared T and B contributing clones at 6.5 months (ZH33) or 4.5 months (ZG66), versus clusters of clones contributing to T, B, and Gr at the same time points. The extent of NK clonal diversity was similar to that in other lineages in terms of the total number of clones; however, a number of very high high-contributing clones dominated the PB at individual time points.

We fractionated NK cells into distinct populations (the major $CD16^+/CD56^-$ PB fraction and the minor blood $CD16^-/CD56^+$ fraction) at 6.5 months (ZH33) or 4.5 months (ZG66) (Figure 5B). We also obtained lymph node NK cells from ZG66, which, in contrast to PB NK cells, were primarily $CD16^-/CD56^+$. The $CD16^+/CD56^-$ PB NK cells accounted for the clones that were distinct from T/B/Gr/Mo lineages, as shown by bias plots, Pearson correlation coefficients, the fraction of highly biased (>10-fold) clones, and both K-mean and hierarchical clustering (Figure 5). Spearman correlation coefficients also supported this result (Figure S6), although the Pearson correlation is more biologically relevant. In contrast, the clonal patterns in the $CD16^-/CD56^+$ subsets of NK cells from blood and lymph nodes strongly correlated with T/B/Gr/Mo lineages.

Analysis of Abundant Clones

In highly polyclonal samples, it is likely that some clones may still intermittently fall below sampling detection limits despite the use of thresholds set to exclude false negatives, complicating analysis of lineage contributions. Since we did not have sufficient DNA to run duplicate samples on every lineage at every time point, the thresholds that were set to exclude false-negative bias due to sampling error were based on limited information. Therefore, we also individually mapped the ten most abundant clones identified in each lineage at 1 month, 3 months, and longest follow-up (9.5 months [ZH33], 6.5 months [ZH17], and 4.5 months [ZG66]), tracking contributions from each clone over all time points and in all other lineages (Figure 7). These abundant clones are inherently interesting because they account for a disproportionate fraction of ongoing hematopoiesis and are less likely to be misinterpreted due to sampling error. Even the most dominant clones at 1 month were primarily unilineage, and ceased or greatly decreased their relative contributions by 2–3 months. T and B cells produced during initial lineage recovery even from these highly productive clones disappeared or greatly decreased at later time points.

The patterns were very different for the top clones at 3 months and the top clones at longest follow-up, with the top Gr, Mo, T, and B clones detected in the same lineage tracing back through 2 months, but not 1 month, and increasing in their relative contributions over time between 2 months and 4.5–9.5 months as short-term progenitor contributions decreased. A number of individual clones fell into the top ten clones for multiple lineages at the later time points, with the most overlap among B, Gr, and Mo; some overlap with T; and the least overlap with NK lineages. This suggests that at least some clones were contributing at relatively balanced and high levels to multiple lineages, confirming the patterns discerned at the population level. Some of these clones also contributed to NK cells, but at much lower levels compared with the relative contributions of the top ten NK clones. The top NK clones from 3 months or latest follow-up were not detected at equivalent or appreciable levels in the other four lineages, further supporting the unique characteristics of this lineage. In two of the animals, there was transient marked expansion of one or more T cell clones at single time points, suggesting a clonal response to an infection or other environmental insult.

DISCUSSION

The application of HSPC barcoding in the rhesus model has provided a number of new insights into large-mammal hematopoiesis. This approach is highly quantitative and requires only small amounts of DNA that can be obtained from multiple lineages repeatedly over time from the same animal, in contrast to the less efficient and quantitative VIS-retrieval-dependent approaches (Bushman et al., 2005). Barcoding relies on vector integration, which raises concerns about the impact on HSPC behavior via activation of adjacent proto-oncogenes (Calmels et al., 2005; Kustikova et al., 2005; Nienhuis et al., 2006). However, lentiviral vectors are much less likely to transform HSPCs compared with murine retroviral vectors, due to lentiviral integration patterns and the deletion of endogenous enhancer sequences in these vectors (Modlich et al., 2009; Montini et al.,

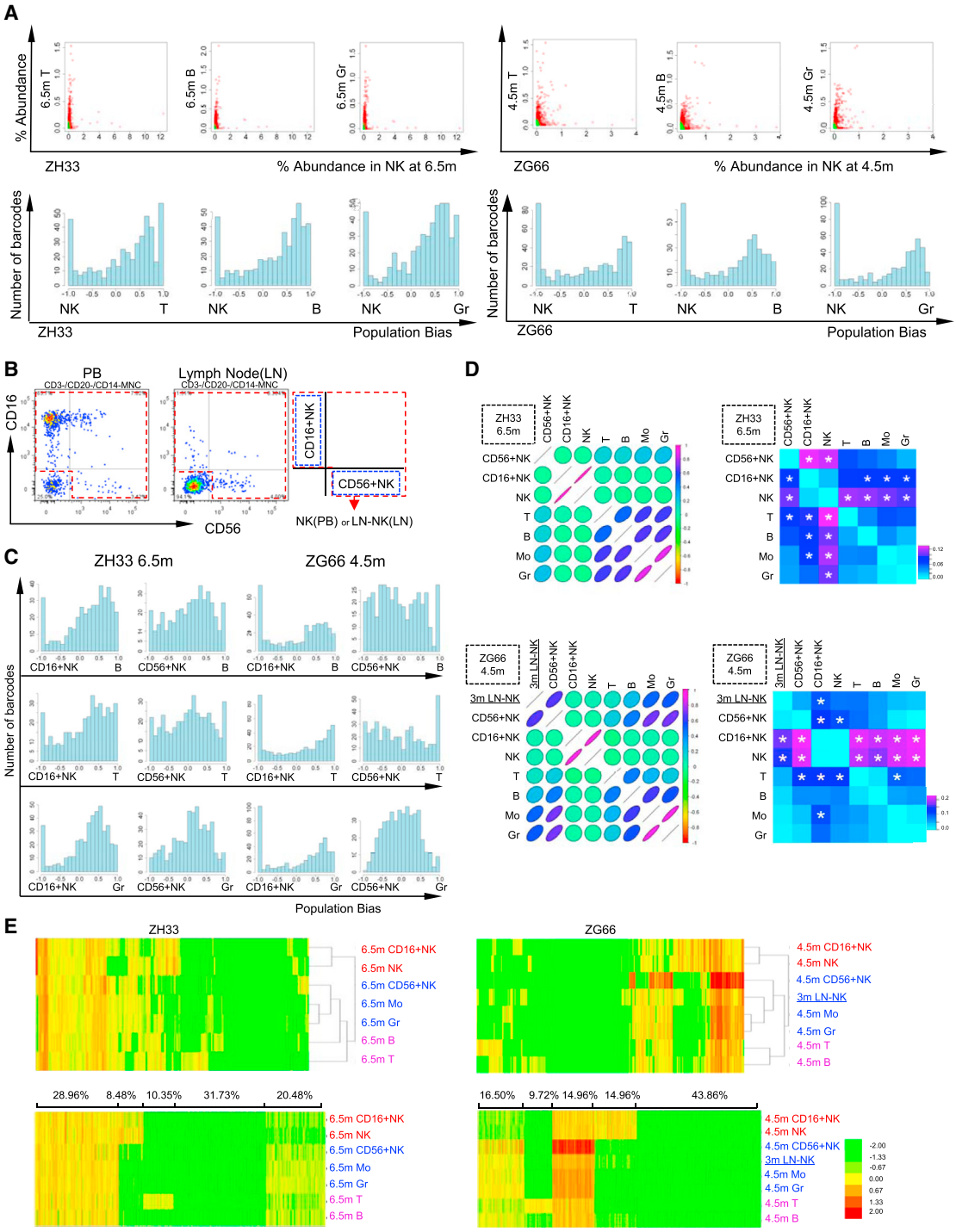


Figure 5. Clonal Contributions to NK Cells

(A) Scatterplots and bias plots for NK versus T, B, and Gr from ZH33 (6.5 months) and ZG66 (4.5 months).

(B) Example of FACS gating for sorting of NK subsets in blood (ZH33, 6.5 months) and node (ZG66, 3 months).

(C) Bias plots comparing $CD16^+$ and $CD56^+$ NK with B, T, and Gr from ZH33 at 6.5 months (left) and ZG66 at 4.5 months (right).

(D) Graphical summary of Pearson correlation coefficients (left) (Table S7 gives r values, p values, and 95% confidence intervals) and fractions of highly biased (>10-fold difference in fractional representation) clones (right; *$p < 0.05$), comparing NK cells and subsets from the PB (ZH33, 6.5 months, upper panel) and from ZG66 blood (4.5 months) and node (3 months, lower panel). An alternative analysis utilizing Spearman correlation is shown in Figure S6 and Table S7.

(E) Euclidian distance hierarchical clustering (upper panel) and K-mean clustering with k = 5 (lower panel), ZH33 (6.5 months), and ZG66 (4.5 months blood, 3 months node). For each animal, all barcodes on the master clone list are included. For K-mean clustering, the % of clones in each cluster is shown above the plots. NK, NK cells in PB; LN-NK, lymph node NK; $CD16^+$NK, $CD16^+CD56^-$; $CD56^+$NK, $CD16^-CD56^+$.

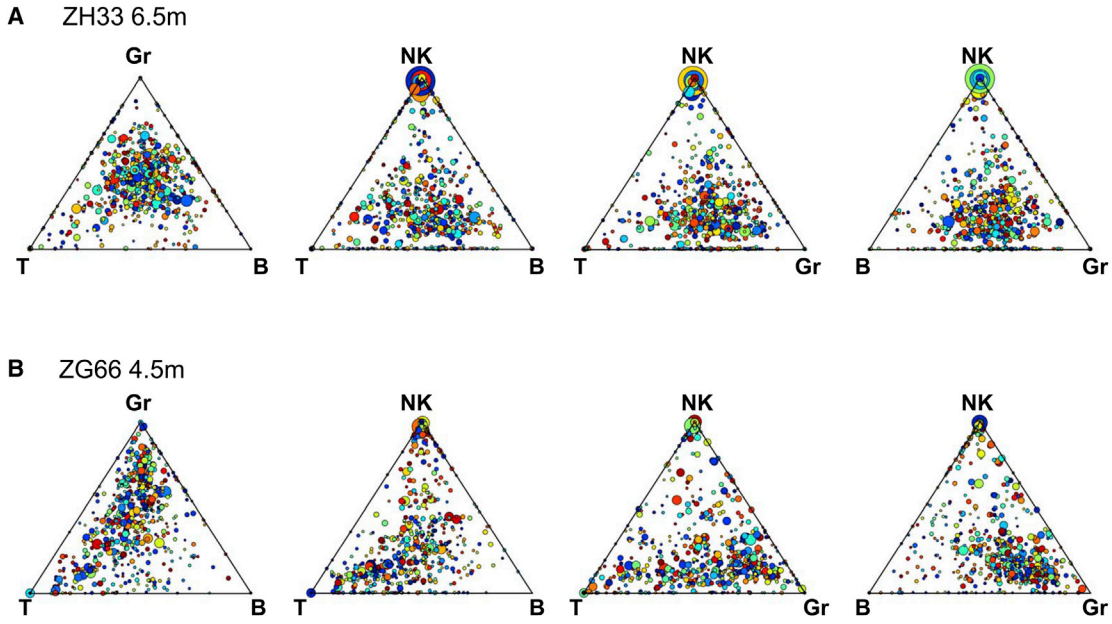


Figure 6. Triangle Plots of Clonal Contributions to T, B, NK, and Gr Lineages

Triangle plots of the clonal contributions to B, T, NK, and Gr in ZH33 at 6.5 months (A), and in ZG66 at 4.5 months (B). Each dot represents an individual barcoded clone, with the size of the dots corresponding to the total fractional contribution of the individual barcode to the total barcodes retrieved at the time point. The color of the dot is assigned randomly to help distinguish adjacent dots. The location of the barcode represents the ratio of barcode reads in each of the three cell lineages. The presence of large dots clustering at the vertex of plots including NK cells suggests a unique population of highly biased NK clones in both animals, in contrast to clusters of dots in the center of B/T/Gr triangles, suggesting common clones that contribute in a balanced manner to these lineages, particularly by 6.5 months in ZH33.

2009). Follow-ups of 5–10 years in rhesus macaques (Kim et al., 2009, 2014) and over 3 years in human clinical trials revealed stable polyclonal engraftment of lentiviral-transduced HSPC without expansion of dominant clones (Biffi et al., 2011; Cartier et al., 2009).

We believe that nonhuman primate models provide information that is highly relevant to human hematopoiesis and difficult to obtain in any other system. Immunodeficient mice lack a physiologic niche for human hematopoiesis, exhibit inefficient engraftment with adult HSPCs, show little maturation of T or NK cells, and do not release human cells normally into the blood, precluding most quantitative analyses of clonal HSPC output or lineage hierarchies. In patients enrolled in clinical gene therapy trials, hematopoiesis can be tracked long term via VIS identification, but most of these patients will experience perturbations due to activation of proto-oncogenes by murine retroviral vectors, dysregulated expression of the therapeutic gene, or underlying hematopoietic abnormalities (Deichmann et al., 2011; Deichmann et al., 2007; Gabriel et al., 2009; Schwarzwaelder et al., 2007). Our rhesus approach is not limited by these constraints.

In our studies, $CD34^+$ cells were selected prior to barcoding. Virtually all HSPC engrafting activity in adult BM and mobilized PB in humans and monkeys is within the $CD34^+$ fraction (Andrews et al., 2000; Doulatov et al., 2012). Long-term HSCs are a minority of HSPCs, and human HSCs assayed in xenografts are $CD34^+38^{lo}90^+117^+Lin^-$ (Majeti et al., 2007; Park et al., 2008). Human $CD34^+$ subfractions have been placed in a hierarchy of multipotent progenitors (MPPs), lymphomyeloid-biased progenitors (LGMPs), common lymphoid progenitors (CLPs), common myeloid progenitors (CMPs), GM progenitors (GMPs), megakaryocyte-erythroid progenitors (MEPs), and unilineage-restricted progenitors of the myeloerythroid lineages, as assayed in vitro and in xenografts (Galy et al., 1995; Majeti et al., 2007; Manz et al., 2002). Cell-surface markers that delineate each of these subpopulations in the rhesus macaque have not been rigorously identified; however, our model will allow prospective purification, barcoding, and autologous engraftment of putative progenitor populations to define each subset in an autologous, physiologic setting.

We calculate a minimum frequency for rhesus engrafting HSPCs as ~1 in 9,000 $CD34^+$ cells, corresponding to ~1 per 4.5 million mobilized PB mononucleated cells (PBMNCs). At the longest follow-up (4.5–9.5 months), ~1 in 22,000 $CD34^+$ cells continued to contribute. Limit-dilution studies measured the "SCID repopulating" HSPC frequency to be 1 in 6 million mobilized human PBMNCs (Wang et al., 1997). The lower frequency of primate or human repopulating cells in xenografts as compared with direct autologous measurement may reflect the inefficiency of xenogeneic engraftment. The majority of rhesus clones found as early as 3 months were still contributing at 9.5 months, and only a small number of new clones appeared for the first time after 3 months, suggesting that relatively stable hematopoiesis may be achieved quite rapidly in a physiologic autologous model, perhaps derived from the rhesus equivalent of "intermediate" mouse repopulating HSCs (Benveniste et al., 2010). This rapid clonal stability is in contrast to the much more heterogeneous results that were recently obtained using

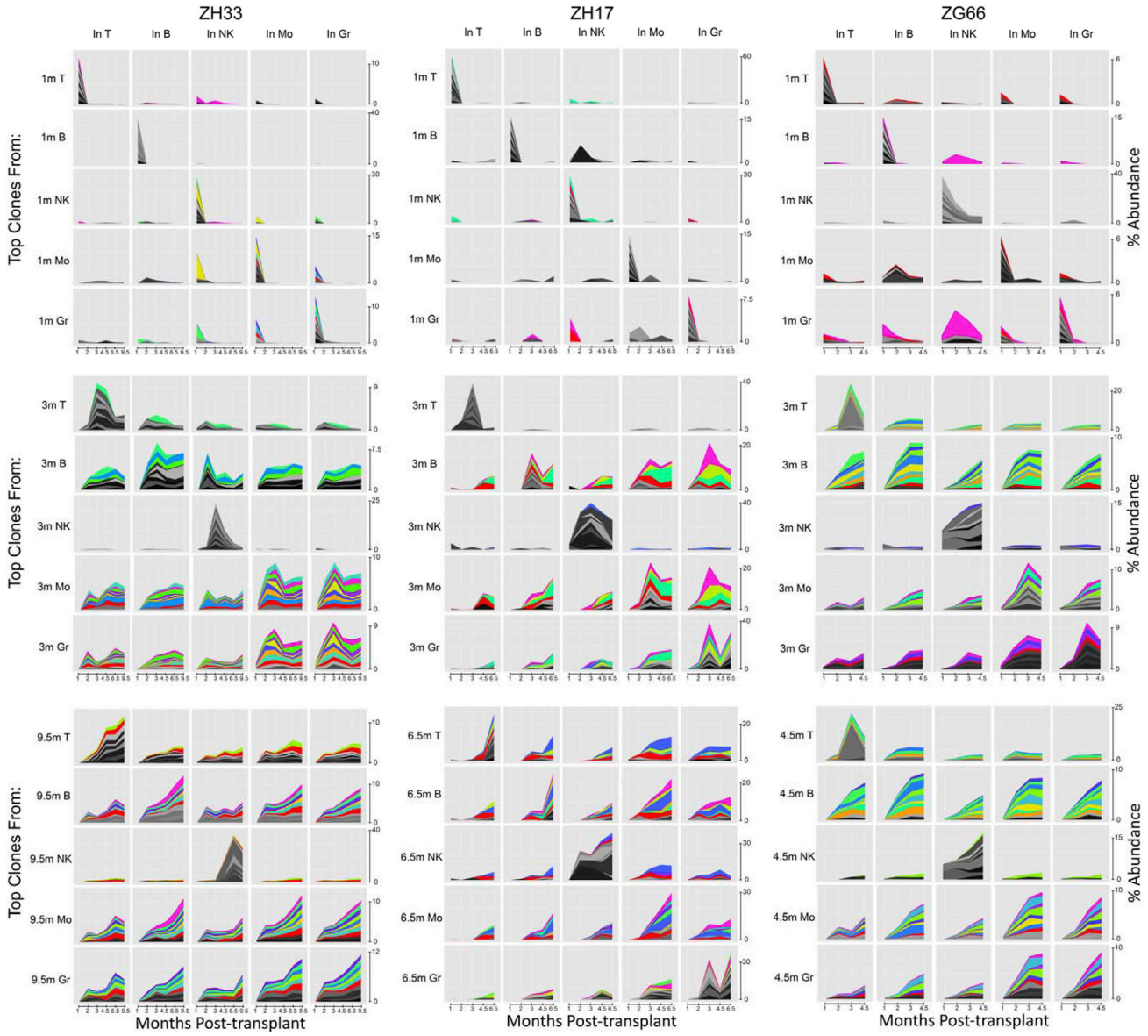


Figure 7. Tracking of Individual High-Contributing Clones

The fractional contribution of each barcode to the total barcodes retrieved from each lineage at each time point were ranked, and the top ten contributing barcodes in each lineage at 1, 3, and 9.5 months from ZH33 (left column); 1, 3, and 6.5 months from ZH17 (middle column); and 1, 3, and 4.5 months from ZG66 (right column) were tracked for their contributions in all lineages over time. Each graph shows the time posttransplantation on the x axis, and the percentage of total reads each clone contributes to the lineage is shown as a stacked graph on the y axis. Each y axis scale is linear and identical across each row of graphs, with the scale on the far right of each row, but note that scale is adjusted between rows, in particular when one very dominant clone is being depicted. Clones shown in grayscale are top ten in only one lineage. Clones shown in color are top ten in more than one lineage, and each clone has a unique color, allowing comparisons between plots for each animal.

barcoding of human cord blood cells in a murine xenograft, likely resulting from inefficiencies of engraftment and hematopoietic output in that model (Cheung et al., 2013). The highly polyclonal nature of reconstitution in the barcoded macaques bypasses stochastic variations that arise in the setting of limiting stem cell doses, which is predicted to result in very prolonged periods of hematopoietic instability (Abkowitz et al., 1996). However, hematopoietic tracking for much longer time periods is required to determine the frequency and clonal output of true long-term repopulating cells, and to distinguish models of HSC clonal exhaustion/succession versus clonal stability. A related study from our primate program, published in this issue (Kim et al., 2014), utilized very-long-term lentiviral VIS retrieval to address some of these issues. Longer follow-up combined with barcoding will allow additional highly quantitative analyses of whether lineage-skewed classes of HSCs exist in nonhuman primates, analogous to recent findings in murine models regarding lymphoid versus myeloid biased HSCs and our

companion long-term primate study (Copley et al., 2012; Kim et al., 2014).

Our studies demonstrate the presence of unilineage engrafting progenitors in all lineages, supporting prior results from murine and xenograft models, and a recent paper utilizing barcoding that demonstrated unilineage B, granulocytic, and dendritic cell output from lymphoid primed multipotent progenitors (LMPPs) (Guenechea et al., 2001; Jones et al., 1990; Jordan and Lemischka, 1990; Naik et al., 2013). Highly purified ex vivo cultured $CD34^+$ cells do not contain viable mature T, B, or NK cells at the time of reinfusion; thus, reestablishment of these lineages must occur from $CD34^+$ progenitor cells rather than by peripheral expansion of mature barcoded cells. The ability to at least transiently engraft lineage-restricted progenitors is of interest for human transplantation applications, and our model can be applied to identify the phenotype of relevant transplantable progenitor cells. Output from these unilineage short-term repopulating clones largely disappeared from the blood and lymph node by 2–3 months, suggesting that the half-life of newly produced mature B and T cells and their direct progenitors is surprisingly short, at least in the posttransplantation setting.

As early as 2 months, Gr and Mo clonal architecture became highly correlated, with shared and unbiased clones contributing, confirming in vivo the close relationship between these lineages revealed by in vitro assays. Beginning at 3 months and continuing through 9.5 months, myeloid and lymphoid lineages began to share clonal patterns. However, the relationship between B and myeloid cells was closer and was established more rapidly than that between B and T cells. A recent analysis of the lineage potential of putative human HSPC populations suggested that the majority of previously characterized multi-lymphoid progenitors (MLPs), the human equivalent of CLPs, retain the ability to produce myeloid lineage cells in vitro and in xenografts (Doulatov et al., 2010). A recent murine barcoding analysis did not find a higher correlation between contributions of clones to T and B than to myeloid and B cells (Verovskaya et al., 2013). We did not document an appreciable set of clones contributing to T and B, but not myeloid, lineages at any time during early hematopoietic reconstitution, supporting the existence of an MLP versus a CLP in primates. Multilineage clones appeared by 3 months and continued to contribute to Gr/Mo/B/T lineages through 9.5 months, with faster stabilization than reported in xenografts, perhaps reflecting a more polyclonal and efficient engraftment process in autologous macaques as compared with human cells in murine niches (McKenzie et al., 2006).

Our most interesting findings concern the putative ontogeny of NK cells. NK cells are defined by target cell killing and cytokine production independent of priming or antigen/MHC restriction (Caligiuri, 2008; Vivier et al., 2011). Rhesus NK cells have been well characterized and share most (but not all) cell-surface markers with human NK cells (Mavilio et al., 2005; Webster and Johnson, 2005). The ability of both T and NK cells to lyse targets, as well as some shared cell-surface receptors, has suggested that they are ontologically related. In vitro, a human NK-B precursor has been identified that is able to produce both lineages under certain culture conditions. However, NK output from this progenitor, or from single MLPs able to produce T, B, NK, and myeloid lineages in vitro, was not detected in xenografts (Doulatov et al., 2010; Galy et al., 1995). The original studies that identified a mouse CLP able to produce T, B, and NK cells in vitro did not show clonal production of all three lineages in vivo (Kondo et al., 1997). In mice, CLPs, as a population, do give rise to T, B, and NK cells (Fathman et al., 2011; Karsunky et al., 2008).

In our model, as late as 6.5 months, the most abundant NK clones made virtually no contributions to the other four lineages. Analysis of NK subsets demonstrated that blood $CD16^+/CD56^-$ NK cells accounted for the unique clonal pattern, distinct from the other four lineages, and that the less abundant $CD16^-/CD56^+$ cells were more closely related to B/T/Gr/Mo lineages. In vitro culture assays and comparisons of node versus blood human NK cells have led to the hypothesis that lymph node $CD16^{-/dim}/CD56^+$ noncytotoxic cytokine-producing NK cells are precursors to circulating cytotoxic $CD16^+/CD56^{-/dim}$ NK cells (Caligiuri, 2008; Reeves et al., 2010). Our results instead suggest separate origins for these two populations. It is of interest that patients with *GATA2* mutations and profoundly disordered hematopoiesis characterized by abnormalities of myeloid cells and B cells have an almost complete lack of $CD16^{dim}/CD56^+$ NK cells, but preservation of the $CD16^+/CD56^{dim}$ NK population, which goes against the hypothesis that $CD16^{dim}/CD56^+$ cells are precursors of $CD16^+/CD56^{dim}$ cells (Mace et al., 2013).

Alternatively, HSCs or other multipotent progenitors may land in "NK niches," specifically supporting NK development or proliferation. Self-renewing NK precursors may exist; however, by 9.5 months, the NK cell clonal architecture began to be more closely related to that of T/B/Gr/Mo, with shared clones contributing to all five lineages at appreciable levels. Finally, the most abundant $CD16^+/CD56^-$ clones in the PB might represent a massive expansion of clones in response to specific environmental stimuli, although the lack of contribution from most major NK clones in the T/B/Gr/Mo lineages, at any time point, makes this possibility less likely. The ability to track NK clones via barcoding in vivo and in vitro opens new avenues of investigation of this poorly understood lineage.

EXPERIMENTAL PROCEDURES

Generation of Barcoded Vectors

DNA barcodes consisting of a 6 bp library identifier followed by a 35 bp (ZH33 and ZG66) or 30 bp (ZH17) random barcode were produced from synthesized oligonucleotides and cloned into the nonexpressing region of the HIV-derived, replication-defective pCDH vector (System Biosciences; Lu et al., 2011; Figure 1A). High-diversity libraries suitable for efficient transduction of rhesus macaque $CD34^+$ cells were produced by calcium phosphate transfection of 293T cells with the barcoded lentiviral vector plasmid and χHIV helper plasmids (Uchida et al., 2009, 2012).

The diversity of the library was assessed via transduction of the rhesus cell line LCL8864 and $CD34^+$ primary cells, low-cycle PCR, Illumina sequencing, custom python code processing, and quantitation of unique barcodes and copy number per barcode (Figure S2). Monte Carlo simulation (Supplemental Experimental Procedures) utilizing these experimental data was used to determine the number of target cells that could be transduced with this vector barcode library diversity, with various probabilities that >95% of retrieved barcodes represent single cells (Lu et al., 2011).

Collection, Transduction, and Transplantation of $CD34^+$ HSPCs

All animal studies were approved by the NHLBI Animal Care and Use Committee. Mobilization with granulocyte colony-stimulating factor and stem cell factor (SCF), $CD34^+$ cell collection and immunoselection, transduction, conditioning, and transplantation were performed as previously described (Donahue et al., 2005). $CD34^+$ cells were cultured overnight on RetroNectin-coated

(Takara) plates in X-VIVO 10 media (Lonza) supplemented with 100 ng/ml each of human Flt3 ligand, SCF, and thrombopoietin (R&D Systems), and 1% human albumin, and transduced with barcode-containing lentiviral vectors at a multiplicity of infection (moi) of 25 in the presence of 4 μg/ml protamine sulfate (Sigma). Transduced $CD34^+$ cells were reinfused 24 hr later into the irradiated (500 cGy/day for 2 days) autologous macaque.

Hematopoietic Cell Purification and Analysis

Mononuclear cells from lymph nodes were isolated by gentle mechanical disruption of tissues and collagenase dissociation (Seggewiss et al., 2007). PB cells were separated by Ficoll; stained with antibodies (Table S2) for Gr, T, B, Mo, NK cells, NK subsets, and myeloid and plasmacytoid dendritic cells (Gujer et al., 2011); and sorted on a FACSAria II instrument (Figure S3; for the purities obtained, see Table S1). DNA was extracted with the DNeasy Blood & Tissue Kit (QIAGEN) and quantified by NanoDrop. Quantitative PCR (qPCR) was performed to determine vector copy number with an ABI PRISM 7700 (Applied Biosystems) using the primers and probes listed in Table S3.

Barcode Extraction, Sequencing, and Analysis

We subjected 200 ng (ZH33 and ZG66) or 500 ng (ZH17) of DNA to a first round of 14-cycle PCR with primers bracketing the barcode, and a second round of 14 cycles with primers adding a sample ID to facilitate multiplex sequencing (Table S3). Equal amounts of gel-purified product from individual samples were pooled together for multiplex sequencing (Illumina HiSeq 2000). We processed the sequencing output using custom python code (Supplemental Experimental Procedures) to extract barcodes via library IDs, combining barcodes to allow for sequencing indels and mismatches up to 2 bp, and setting a lower cutoff for read numbers to eliminate 99.5% of background "false barcodes" for each animal (Lu et al., 2011). Barcodes with total read numbers smaller than these cutoffs (1,845 [ZH33], 1,686 [ZH17], and 541 [ZG66]) were eliminated, and a cumulative master list of barcodes with total reads over these cutoffs was generated for each animal (Table S4). The raw sequencing data from all time points and lineages were then reanalyzed versus this master list to pull out sequences representing these barcodes, allowing for mismatches and indels up to 2 bp, and determining the final read number for each barcode in each sample. The fractional representation of each barcode in a sample was then calculated from the ratio of an individual barcode's read number divided by the total number of valid barcode reads (Table S4).

For all pairwise comparisons between samples collected from different lineages or at different time points for ZH33 and ZG66, a combined read number "sampling" threshold was set to exclude 95% of clones that scored as falsely "biased" (>10× barcode ratio) between the replicate samples (Figure S4), and was determined to be 1,144 based on the median value calculated for all available replicate samples (n = 10).

Computational and Statistical Analysis

Data analysis and plot generation were done using R code, ggplot2, corrplot, and reshape2 software. Pearson and Spearman correlation coefficients were calculated using R (R Foundation for Statistical Computing, Vienna, Austria; ISBN 3-900051-07-0, http://www.R-project.org/). We calculated p values for highly biased clones using Fisher's exact test, comparing the frequency of highly biased clones against the overall frequency of highly biased clones at the time-point combination being considered.

For global analyses of clonal contributions via hierarchical clustering, we applied three steps of normalization to the total data set consisting of the valid read number for all barcodes: the fractional contribution of each barcode to the overall reads in each individual sample, 75^{th} percentile normalization, and log10 transformation. Unsupervised two-way hierarchical clustering by a Euclidean distance algorithm was applied to all samples. K-mean clustering with a k = 5 was applied to all lineages at specific time points, filtering out any barcodes that were not present in any sample prior to K-mean clustering.

SUPPLEMENTAL INFORMATION

Supplemental Information for this article includes python code, seven figures, and seven tables and can be found with this article online at http://dx.doi.org/10.1016/j.stem.2014.01.020.

AUTHOR CONTRIBUTIONS

C.W., B.L., A.J., S.J.K., F.K. and K.L. performed and analyzed the experiments; C.W., B.L., S.J.K., R.L., I.W., and C.E.D. wrote the paper. C.E.D., C.W., R.L., and B.L. designed the experiments. A.E.K., M.M., and R.E.D. transplanted the macaques. R.L. wrote the python code and ran data analyses. C.O.W., S.J.K., Y.Y., R.L., and B.L. performed computational and statistical analyses. I.S.Y.C. provided important insights.

ACKNOWLEDGMENTS

We thank Naoya Uchida for the χHIV packaging system, Danny Douek for helpful comments, and Ann Williams (NHLBI FACS Core) and Jun Zhu (DNA Sequencing and Genomics Core) for technical assistance. This research was supported by the Division of Intramural Research of the NHLBI (grants NIH-UO1-HL099999 to I.W., NIH-K99-HL11304 to R.L., and CIRM-TG2-01159 to R.L.).

Received: March 18, 2013
Revised: December 9, 2013
Accepted: January 30, 2014
Published: April 3, 2014

REFERENCES

Abkowitz, J.L., Catlin, S.N., and Guttorp, P. (1996). Evidence that hematopoiesis may be a stochastic process in vivo. Nat. Med. *2*, 190–197.

Abkowitz, J.L., Catlin, S.N., McCallie, M.T., and Guttorp, P. (2002). Evidence that the number of hematopoietic stem cells per animal is conserved in mammals. Blood *100*, 2665–2667.

An, D.S., Donahue, R.E., Kamata, M., Poon, B., Metzger, M., Mao, S.H., Bonifacino, A., Krouse, A.E., Darlix, J.L., Baltimore, D., et al. (2007). Stable reduction of CCR5 by RNAi through hematopoietic stem cell transplant in non-human primates. Proc. Natl. Acad. Sci. USA *104*, 13110–13115.

Andrews, R.G., Peterson, L.J., Morris, J., Potter, J., Heyward, S., Gough, M., Bryant, E., and Kiem, H.P. (2000). Differential engraftment of genetically modified CD34(+) and CD34(-) hematopoietic cell subsets in lethally irradiated baboons. Exp. Hematol. *28*, 508–518.

Baum, C.M., Weissman, I.L., Tsukamoto, A.S., Buckle, A.M., and Peault, B. (1992). Isolation of a candidate human hematopoietic stem-cell population. Proc. Natl. Acad. Sci. USA *89*, 2804–2808.

Benveniste, P., Frelin, C., Janmohamed, S., Barbara, M., Herrington, R., Hyan, D., and Iscove, N.N. (2010). Intermediate-term hematopoietic stem cells with extended but time-limited reconstitution potential. Cell Stem Cell *8*, 48–58.

Berry, C.C., Gillet, N.A., Melamed, A., Gormley, N., Bangham, C.R., and Bushman, F.D. (2012). Estimating abundances of retroviral insertion sites from DNA fragment length data. Bioinformatics *28*, 755–762.

Biffi, A., Bartolomae, C.C., Cesana, D., Cartier, N., Aubourg, P., Ranzani, M., Cesani, M., Benedicenti, F., Plati, T., Rubagotti, E., et al. (2011). Lentiviral vector common integration sites in preclinical models and a clinical trial reflect a benign integration bias and not oncogenic selection. Blood *117*, 5332–5339.

Bushman, F., Lewinski, M., Ciuffi, A., Barr, S., Leipzig, J., Hannenhalli, S., and Hoffmann, C. (2005). Genome-wide analysis of retroviral DNA integration. Nat. Rev. Microbiol. *3*, 848–858.

Caligiuri, M.A. (2008). Human natural killer cells. Blood *112*, 461–469.

Calmels, B., Ferguson, C., Laukkanen, M.O., Adler, R., Faulhaber, M., Kim, H.J., Sellers, S., Hematti, P., Schmidt, M., von Kalle, C., et al. (2005). Recurrent retroviral vector integration at the Mds1/Evi1 locus in nonhuman primate hematopoietic cells. Blood *106*, 2530–2533.

Cartier, N., Hacein-Bey-Abina, S., Bartholomae, C.C., Veres, G., Schmidt, M., Kutschera, I., Vidaud, M., Abel, U., Dal-Cortivo, L., Caccavelli, L., et al. (2009). Hematopoietic stem cell gene therapy with a lentiviral vector in X-linked adrenoleukodystrophy. Science *326*, 818–823.

Catlin, S.N., Busque, L., Gale, R.E., Guttorp, P., and Abkowitz, J.L. (2011). The replication rate of human hematopoietic stem cells in vivo. Blood *117*, 4460–4466.

Cheung, A.M., Nguyen, L.V., Carles, A., Beer, P., Miller, P.H., Knapp, D.J., Dhillon, K., Hirst, M., and Eaves, C.J. (2013). Analysis of the clonal growth and differentiation dynamics of primitive barcoded human cord blood cells in NSG mice. Blood *122*, 3129–3137.

Copley, M.R., Beer, P.A., and Eaves, C.J. (2012). Hematopoietic stem cell heterogeneity takes center stage. Cell Stem Cell *10*, 690–697.

Coulombel, L. (2004). Identification of hematopoietic stem/progenitor cells: strength and drawbacks of functional assays. Oncogene *23*, 7210–7222.

Deichmann, A., Hacein-Bey-Abina, S., Schmidt, M., Garrigue, A., Brugman, M.H., Hu, J., Glimm, H., Gyapay, G., Prum, B., Fraser, C.C., et al. (2007). Vector integration is nonrandom and clustered and influences the fate of lymphopoiesis in SCID-X1 gene therapy. J. Clin. Invest. *117*, 2225–2232.

Deichmann, A., Brugman, M.H., Bartholomae, C.C., Schwarzwaelder, K., Verstegen, M.M., Howe, S.J., Arens, A., Ott, M.G., Hoelzer, D., Seger, R., et al. (2011). Insertion sites in engrafted cells cluster within a limited repertoire of genomic areas after gammaretroviral vector gene therapy. Mol. Ther. *19*, 2031–2039.

Donahue, R.E., and Dunbar, C.E. (2001). Update on the use of nonhuman primate models for preclinical testing of gene therapy approaches targeting hematopoietic cells. Hum. Gene Ther. *12*, 607–617.

Donahue, R.E., Kuramoto, K., and Dunbar, C.E. (2005). Large animal models for stem and progenitor cell analysis. Curr. Protoc. Immunol. *Chapter 22*, Unit 22A.21.

Doulatov, S., Notta, F., Eppert, K., Nguyen, L.T., Ohashi, P.S., and Dick, J.E. (2010). Revised map of the human progenitor hierarchy shows the origin of macrophages and dendritic cells in early lymphoid development. Nat. Immunol. *11*, 585–593.

Doulatov, S., Notta, F., Laurenti, E., and Dick, J.E. (2012). Hematopoiesis: a human perspective. Cell Stem Cell *10*, 120–136.

Fathman, J.W., Bhattacharya, D., Inlay, M.A., Seita, J., Karsunky, H., and Weissman, I.L. (2011). Identification of the earliest natural killer cell-committed progenitor in murine bone marrow. Blood *118*, 5439–5447.

Gabriel, R., Eckenberg, R., Paruzynski, A., Bartholomae, C.C., Nowrouzi, A., Arens, A., Howe, S.J., Recchia, A., Cattoglio, C., Wang, W., et al. (2009). Comprehensive genomic access to vector integration in clinical gene therapy. Nat. Med. *15*, 1431–1436.

Galy, A., Travis, M., Cen, D., and Chen, B. (1995). Human T, B, natural killer, and dendritic cells arise from a common bone marrow progenitor cell subset. Immunity *3*, 459–473.

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

Guenechea, G., Gan, O.I., Dorrell, C., and Dick, J.E. (2001). Distinct classes of human stem cells that differ in proliferative and self-renewal potential. Nat. Immunol. *2*, 75–82.

Gujer, C., Sundling, C., Seder, R.A., Karlsson Hedestam, G.B., and Loré, K. (2011). Human and rhesus plasmacytoid dendritic cell and B-cell responses to Toll-like receptor stimulation. Immunology *134*, 257–269.

Jones, R.J., Wagner, J.E., Celano, P., Zicha, M.S., and Sharkis, S.J. (1990). Separation of pluripotent haematopoietic stem cells from spleen colony-forming cells. Nature *347*, 188–189.

Jordan, C.T., and Lemischka, I.R. (1990). Clonal and systemic analysis of long-term hematopoiesis in the mouse. Genes Dev. *4*, 220–232.

Karsunky, H., Inlay, M.A., Serwold, T., Bhattacharya, D., and Weissman, I.L. (2008). Flk2+ common lymphoid progenitors possess equivalent differentiation potential for the B and T lineages. Blood *111*, 5562–5570.

Kim, H.J., Tisdale, J.F., Wu, T., Takatoku, M., Sellers, S.E., Zickler, P., Metzger, M.E., Agricola, B.A., Malley, J.D., Kato, I., et al. (2000). Many multipotential gene-marked progenitor or stem cell clones contribute to hematopoiesis in nonhuman primates. Blood *96*, 1–8.

Kim, Y.J., Kim, Y.S., Larochelle, A., Renaud, G., Wolfsberg, T.G., Adler, R., Donahue, R.E., Hematti, P., Hong, B.K., Roayaei, J., et al. (2009). 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 *113*, 5434–5443.

Kim, S., Kim, N., Presson, A.P., Metzger, M.E., Bonifacino, A.C., Sehl, M., Chow, S.A., Crooks, G.M., Dunbar, C.E., An, D.S., et al. (2014). Dynamics of HSPC repopulatoin in nono-human primate revealed by a decade-long clonal tracking study. Cell Stem Cell *14*, this issue, 473–485.

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

Kustikova, O.S., Wahlers, A., Kuhlcke, K., Stahle, B., Zander, A.R., Baum, C., and Fehse, B. (2003). Dose finding with retroviral vectors: correlation of retroviral vector copy numbers in single cells with gene transfer efficiency in a cell population. Blood *102*, 3934–3937.

Kustikova, O., Fehse, B., Modlich, U., Yang, M., Düllmann, J., Kamino, K., von Neuhoff, N., Schlegelberger, B., Li, Z., and Baum, C. (2005). Clonal dominance of hematopoietic stem cells triggered by retroviral gene marking. Science *308*, 1171–1174.

Larochelle, A., Vormoor, J., Hanenberg, H., Wang, J.C., Bhatia, M., Lapidot, T., Moritz, T., Murdoch, B., Xiao, X.L., Kato, I., et al. (1996). Identification of primitive human hematopoietic cells capable of repopulating NOD/SCID mouse bone marrow: implications for gene therapy. Nat. Med. *2*, 1329–1337.

Larochelle, A., Savona, M., Wiggins, M., Anderson, S., Ichwan, B., Keyvanfar, K., Morrison, S.J., and Dunbar, C.E. (2011). Human and rhesus macaque hematopoietic stem cells cannot be purified based only on SLAM family markers. Blood *117*, 1550–1554.

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

Mace, E.M., Hsu, A.P., Monaco-Shawver, L., Makedonas, G., Rosen, J.B., Dropulic, L., Cohen, J.I., Frenkel, E.P., Bagwell, J.C., Sullivan, J.L., et al. (2013). Mutations in GATA2 cause human NK cell deficiency with specific loss of the CD56(bright) subset. Blood *121*, 2669–2677.

Majeti, R., Park, C.Y., and Weissman, I.L. (2007). Identification of a hierarchy of multipotent hematopoietic progenitors in human cord blood. Cell Stem Cell *1*, 635–645.

Manz, M.G., Miyamoto, T., Akashi, K., and Weissman, I.L. (2002). Prospective isolation of human clonogenic common myeloid progenitors. Proc. Natl. Acad. Sci. USA *99*, 11872–11877.

Mavilio, D., Benjamin, J., Kim, D., Lombardo, G., Daucher, M., Kinter, A., Nies-Kraske, E., Marcenaro, E., Moretta, A., and Fauci, A.S. (2005). Identification of NKG2A and NKp80 as specific natural killer cell markers in rhesus and pigtailed monkeys. Blood *106*, 1718–1725.

McKenzie, J.L., Gan, O.I., Doedens, M., Wang, J.C., and Dick, J.E. (2006). Individual stem cells with highly variable proliferation and self-renewal properties comprise the human hematopoietic stem cell compartment. Nat. Immunol. *7*, 1225–1233.

Modlich, U., Navarro, S., Zychlinski, D., Maetzig, T., Knoess, S., Brugman, M.H., Schambach, A., Charrier, S., Galy, A., Thrasher, A.J., et al. (2009). Insertional transformation of hematopoietic cells by self-inactivating lentiviral and gammaretroviral vectors. Mol. Ther. *17*, 1919–1928.

Montini, E., Cesana, D., Schmidt, M., Sanvito, F., Bartholomae, C.C., Ranzani, M., Benedicenti, F., Sergi, L.S., Ambrosi, A., Ponzoni, M., et al. (2009). The genotoxic potential of retroviral vectors is strongly modulated by vector design and integration site selection in a mouse model of HSC gene therapy. J. Clin. Invest. *119*, 964–975.

Naik, S.H., Perié, L., Swart, E., Gerlach, C., van Rooij, N., de Boer, R.J., and Schumacher, T.N. (2013). Diverse and heritable lineage imprinting of early haematopoietic progenitors. Nature *496*, 229–232.

Nienhuis, A.W., Dunbar, C.E., and Sorrentino, B.P. (2006). Genotoxicity of retroviral integration in hematopoietic cells. Mol. Ther. *13*, 1031–1049.

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

Park, C.Y., Majeti, R., and Weissman, I.L. (2008). In vivo evaluation of human hematopoiesis through xenotransplantation of purified hematopoietic stem cells from umbilical cord blood. Nat. Protoc. *3*, 1932–1940.

Reeves, R.K., Gillis, J., Wong, F.E., Yu, Y., Connole, M., and Johnson, R.P. (2010). CD16- natural killer cells: enrichment in mucosal and secondary lymphoid tissues and altered function during chronic SIV infection. Blood *115*, 4439–4446.

Schepers, K., Swart, E., van Heijst, J.W., Gerlach, C., Castrucci, M., Sie, D., Heimerikx, M., Velds, A., Kerkhoven, R.M., Arens, R., and Schumacher, T.N. (2008). Dissecting T cell lineage relationships by cellular barcoding. J. Exp. Med. *205*, 2309–2318.

Schmidt, M., Zickler, P., Hoffman, G., Haas, S., Wissler, M., Muessig, A., Tisdale, J.F., Andrews, R.G., Wu, T., Kiem, H.-P., et al. (2002). Polyclonal long-term repopulating stem cell clones in a primate model. Blood *100*, 2737–2743.

Schwarzwaelder, K., Howe, S.J., Schmidt, M., Brugman, M.H., Deichmann, A., Glimm, H., Schmidt, S., Prinz, C., Wissler, M., King, D.J., et al. (2007). Gammaretrovirus-mediated correction of SCID-X1 is associated with skewed vector integration site distribution in vivo. J. Clin. Invest. *117*, 2241–2249.

Seggewiss, R., Loré, K., Guenaga, F.J., Pittaluga, S., Mattapallil, J., Chow, C.K., Koup, R.A., Camphausen, K., Nason, M.C., Meier-Schellersheim, M., et al. (2007). Keratinocyte growth factor augments immune reconstitution after autologous hematopoietic progenitor cell transplantation in rhesus macaques. Blood *110*, 441–449.

Shepherd, B.E., Kiem, H.P., Lansdorp, P.M., Dunbar, C.E., Aubert, G., LaRochelle, A., Seggewiss, R., Guttorp, P., and Abkowitz, J.L. (2007). Hematopoietic stem-cell behavior in nonhuman primates. Blood *110*, 1806–1813.

Smith, L.G., Weissman, I.L., and Heimfeld, S. (1991). Clonal analysis of hematopoietic stem-cell differentiation in vivo. Proc. Natl. Acad. Sci. USA *88*, 2788–2792.

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

Uchida, N., Washington, K.N., Hayakawa, J., Hsieh, M.M., Bonifacino, A.C., Krouse, A.E., Metzger, M.E., Donahue, R.E., and Tisdale, J.F. (2009). Development of a human immunodeficiency virus type 1-based lentiviral vector that allows efficient transduction of both human and rhesus blood cells. J. Virol. *83*, 9854–9862.

Uchida, N., Hargrove, P.W., Lap, C.J., Evans, M.E., Phang, O., Bonifacino, A.C., Krouse, A.E., Metzger, M.E., Nguyen, A.D., Hsieh, M.M., et al. (2012). High-efficiency transduction of rhesus hematopoietic repopulating cells by a modified HIV1-based lentiviral vector. Mol. Ther. *20*, 1882–1892.

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

Vivier, E., Raulet, D.H., Moretta, A., Caligiuri, M.A., Zitvogel, L., Lanier, L.L., Yokoyama, W.M., and Ugolini, S. (2011). Innate or adaptive immunity? The example of natural killer cells. Science *331*, 44–49.

Wang, J.C., Doedens, M., and Dick, J.E. (1997). 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 *89*, 3919–3924.

Webster, R.L., and Johnson, R.P. (2005). Delineation of multiple subpopulations of natural killer cells in rhesus macaques. Immunology *115*, 206–214.

Weissman, I.L., and Shizuru, J.A. (2008). The origins of the identification and isolation of hematopoietic stem cells, and their capability to induce donor-specific transplantation tolerance and treat autoimmune diseases. Blood *112*, 3543–3553.

Wu, C., Jares, A., Winkler, T., Xie, J., Metais, J.Y., and Dunbar, C.E. (2013). High efficiency restriction enzyme-free linear amplification-mediated polymerase chain reaction approach for tracking lentiviral integration sites does not abrogate retrieval bias. Hum. Gene Ther. *24*, 38–47.