ARTICLE

Clonal tracking of erythropoiesis in rhesus macaques

Xing Fan,[1] Chuanfeng Wu,[1] Lauren L. Truitt,[1] Diego A. Espinoza,[1,2] Stephanie Sellers,[1] Aylin Bonifacino,[1] Yifan Zhou,[1,3] Stefan F. Cordes,[1] Allen Krouse,[1] Mark Metzger,[1] Robert E. Donahue,[1] Rong Lu[4] and Cynthia E. Dunbar[1]

[1]Translational Stem Cell Biology Branch, National Heart, Lung, and Blood Institute, National Institute of Health, Bethesda, MA, USA; [2]Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA; [3]Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, UK and [4]Eli and Edythe Broad Center for Regenerative Medicine and Stem Cell Research, University of Southern California, Los Angeles, CA, USA

Ferrata Storti Foundation

Haematologica 2020
Volume 105(7):1813-1824

ABSTRACT

The classical model of hematopoietic hierarchies is being reconsidered on the basis of data from *in vitro* assays and single cell expression profiling. Recent experiments suggested that the erythroid lineage might differentiate directly from multipotent hematopoietic stem cells / progenitors or from a highly biased subpopulation of stem cells, rather than transiting through common myeloid progenitors or megakaryocyte-erythrocyte progenitors. We genetically barcoded autologous rhesus macaque stem and progenitor cells, allowing quantitative tracking of the *in vivo* clonal output of thousands of individual cells over time following transplantation. $CD34^+$ cells were lentiviral-transduced with a high diversity barcode library, with the barcode in an expressed region of the provirus, allowing barcode retrieval from DNA or RNA, with each barcode representing an individual stem or progenitor cell clone. Barcode profiles from bone marrow $CD45^-CD71^+$ maturing nucleated red blood cells were compared with other lineages purified from the same bone marrow sample. There was very high correlation of barcode contributions between marrow nucleated red blood cells and other lineages, with the highest correlation between nucleated red blood cells and myeloid lineages, whether at earlier or later time points post transplantation, without obvious clonal contributions from highly erythroid-biased or restricted clones. A similar profile occurred even under stressors such as aging or erythropoietin stimulation. RNA barcode analysis on circulating mature red blood cells followed over long time periods demonstrated stable erythroid clonal contributions. Overall, in this non-human primate model with great relevance to human hematopoiesis, we documented continuous production of erythroid cells from multipotent, non-biased hematopoietic stem cell clones at steady-state or under stress.

Introduction

In the classical model of hematopoiesis, initially constructed from data obtained *via in vitro* colony assays and transplantation of populations of flow-sorted phenotypically-defined murine bone marrow (BM) cells, the top of the hematopoietic hierarchy is comprised of a pool of homogenous, self-renewing and always multipotent long-term hematopoietic stem cells (LT-HSC), producing downstream stem and progenitor cells *via* branching pathways passing through discrete intermediate stages. These processes were characterized by stepwise restriction of self-renewal and lineage potential, passing through short-term multipotent HSC (ST-HSC), multipotent progenitors (MPP), and lineage-restricted progenitors, bifurcating first into lymphoid *versus* myeloid progenitors, followed by common myeloid progenitors (CMP) branching towards granulocyte-monocyte progenitors (GMP) and megakaryocyte-erythrocyte progenitors (MEP) in both murine and human studies.[1-3] Optimized *in vitro* clonal assays, large-scale single cell murine transplantation assays, *in vivo* clonal tracking *via* genetic tags and single cell gene

Correspondence:

CYNTHIA DUNBAR
dunbarc@nhlbi.nih.gov

Received: July 8, 2019.

Accepted: October 3, 2019.

Pre-published: October 3, 2019.

doi:10.3324/haematol.2019.231811

Check the online version for the most updated information on this article, online supplements, and information on authorship & disclosures: www.haematologica.org/content/105/7/1813

expression profiling analyzed by computation algorithms predicting differentiation trajectories have challenged the classical branching hematopoietic model in both rodents and humans. Adolffson and co-workers reported direct differentiation of murine megakaryocytic-erythroid lineages from HSC/ MPP.[4] Notta and co-workers analyzed human MPP subpopulations and demonstrated almost exclusively uni-lineage potential of single cells *in vitro*, suggesting that both erythroid and megakaryocytic lineages differentiate directly and separately from HSC/MPP.[5] *In vitro* assays and single cell gene expression mapping of classical human MEP populations also suggested distinct erythroid and megakaryocytic pathways immediately downstream of multipotent progenitors, although other groups were able to purify rare bipotent progenitor cells.[6,7] Both murine and human single-cell RNA-seq profiling of hematopoietic stem and progenitor cells (HSPC) uncovered very early transcriptional lineage priming immediately downstream of HSC, imputing early branching towards individual hematopoietic lineages, and in some models the earliest branch being erythroid.[8-13]

In addition, large-scale optimized single cell murine transplantation assays have suggested that all long-term and self-renewing engrafting cells are not necessarily homogeneous or multipotent, with evidence for lineage-bias or even lineage-restriction. Dykstra and co-workers reported different classes of such cells with myeloid, or multipotent engraftment patterns long-term, maintained in secondary transplants, but did not examine erythroid or megakaryocytic lineages, given lack of expression of standard congenic markers on these lineages.[14] More recently, groups have devised strategies to allow tracking in all murine lineages, and uncovered megakaryocytic-restricted or highly-biased intermediate[15] or long-term engrafting/self-renewing single cells.[16] Use of an inducible transposon to create clonal tags in non-transplanted mice also uncovered a megakaryocyte-restricted differentiation pathway, and both clonal label propagation through various progenitor populations and gene expression profiling suggested that megakaryocyte-primed HSC are located at the top of the hematopoietic hierarchy.[17] These powerful *in vivo* approaches are dependent on methodologies such as single cell transplantation, transposon activation or lineage tracing that are not feasible in humans or large animals.

We have employed rhesus macaque (RM) HSPC autologous transplantation combined with lentiviral genetic barcoding to quantitatively track the *in vivo* clonal output of thousands of individual HSPC over time, in a model with great relevance to human hematopoiesis.[18] Macaques and humans have prolonged lifespans and similar HSPC cycling and dynamics.[19] We previously demonstrated early lineage-restricted engraftment of short-term progenitors for several months, followed by stable very long-term output from engrafted multipotent HSPC, analyzing DNA barcodes from nucleated neutrophils and lymphoid lineages, in the peripheral blood (PB) and BM.[20,21] Persistent myeloid or B-cell lineage bias, although not complete lineage restriction, could be appreciated,[20] and was increased in aged macaques.[22] Peripheral maintenance and expansion of T-cell and mature natural killer (NK) clones was documented.[23] We now apply this macaque model to examine the clonal ontogeny of the erythroid lineage at steady state post transplantation and under erythropoietic stimulation, employing both DNA and expressed RNA barcode analysis. Results in both young and aged macaques revealed closely shared clonal landscapes for erythropoiesis compared to myeloid and lymphoid lineages at both steady states following transplantation and under erythropoietic stress, and clonally-stable erythropoiesis over time.

Methods

Autologous rhesus macaque transplantation

All experiments were carried out on protocols approved by the National Heart, Lung and Blood Institute (NHLBI) Animal Care and Use Committee, following institutional and Department of Health and Human Services guidelines. Details of peripheral blood HSPC mobilization, CD34$^+$ purification, lentiviral transduction, and autologous transplantation following myeloablative (500 rads x2) total body irradiation have been published[24] including details for the specific animals included in the current paper.[20,21,23] Table 1 summarizes transplanted cell doses and length of follow up. Details of transplantation and the clonal patterns in non-erythroid lineages from animals ZH33, ZG66, ZJ31, ZK22, ZL40 and ZH19 have been previously reported.[20,23,25,26]

Barcoded library preparation, validation, transduction, and retrieval

The barcoded lentiviral vector consists of the backbone pCDH (Systems Biosciences) expressing the CopGFP marker gene followed by a 6 base pair (bp) library identifier and a 27 or 35bp highly diverse DNA barcode, flanked by polymerase chain reaction (PCR) amplification sites.[27] Lentiviral vectors were produced using the χHIV packaging system optimized for RM HSPC transduction.[28] CD34$^+$ HSPC were transduced with high-diversity barcoded libraries ensuring that the majority of transduced HSPC contain only one barcode per cell and that each barcode uniquely defines a single HSPC, as described and validated.[20,29] DNA from target cell

Table 1. Transplantation and follow-up characteristics of animals included in this study.

	ZH33	ZG66	ZJ31	ZH19	ZK22	ZL40	JD76	JM82	RQ3600
CD34$^+$ transplant dose (millions)	32	48	23	48	82	57	44.4	91	58.4
CD34$^+$ transplant dose/kg (millions)	6.9	8.5	4.1	7.1	7.2	8.1	4.1	7.2	15.9
% GFP$^+$ infused cells	35%	35%	35%	23%	31%	22.5%	27.1%	34%	35%
Infused GFP$^+$cells (millions)	11.1	16.7	8.0	11.0	25.2	12.8	12.0	30.6	20.4
Follow-up time points (months)	46, 54.5, 60, 61	48, 53, 55.5	3.5, 28, 31, 32, 33, 35	41, 44.5, 45.5, 48.5, 49	9, 11.5, 12.5, 15.5, 17.5	10.5, 12, 15.5	3.5	3.5	46, 48, 49

Details of transplantation and the clonal patterns in non-erythroid lineages from animals ZH33, ZG66, ZJ31, ZH19, ZK22, ZL40 and RQ3600 been previously reported.[20,23,25,26]

populations or cDNA reverse-transcribed from cellular RNA underwent low cycle PCR with primers bracketing the barcode followed by multiplex Illumina sequencing; see the *Online Supplementary Appendix* for details. Sequencing output was processed using custom Python and R code to retrieve and quantitate barcode contributions (available at: *www.github.com/dunbarlabNIH/*), previously validated to closely reflect fractional contributions of each clone within polyclonal cellular populations.[21]

Hematopoietic cell purification and phenotypic analyses

0-15 mL BM aspirates were obtained from the posterior iliac crests or ischial tuberosities. BM and PB samples were separated into mononuclear cell (MNC) and granulocyte (Gr) fractions via centrifugation over a Ficoll gradient (MP Biomedicals). Red blood cells (RBC) were removed from the Gr pellet via red cell lysis with ACK buffer (Quality Biological). BM MNC were passed over an immunoselection column to purify $CD34^+$ HSPC as described.[24] $CD34^-$ MNC flowing through the column or PB MNC were stained with lineage-specific antibodies and sorted for $CD45^-CD71^+$ nucleated red blood cells (NRBC) as described,[30] $CD3^+$ T cells, $CD20^+$ B cells, and $CD3^-CD20^-CD14^+$ monocytes (Mono) as reported previously,[23] using gating strategies shown in *Online Supplementary Figure S1*. The purity of sorted erythroid cells was validated by morphologic scoring of at least 500 cells on Wright's stained and benzidine-stained cytospins. Erythroid cells constituted at least 95% of sorted preparations. Monoclonal antibodies utilized are given in *Online Supplementary Table S1*.

Leukocyte depletion of peripheral blood was performed by filtration through a 10 mL syringe packed with 5 mL of cellulose fibers (Sigma-Aldrich) and fitted with two layers of Whatman™ lens paper (GE Healthcare Life Sciences) covering the outlet.[31] Before use, columns were rinsed with phosphate buffered saline (PBS) and then 5 mL whole blood was added and gently pressed to run through in droplets. Each product was checked for extent of $CD45^+$ cell depletion by flow cytometry, and by morphologic scoring of at least 500 cells on a Wright's stained smear. Each blood sample was >99%-depleted of non-erythroid cells.

Colony-forming unit assays

$CD34^+$ cells were plated for colony-forming unit (CFU) assays according to the manufacturer's (STEMCELL Technologies) instructions in two different methylcellulose formulations: one is MethoCult GF$^+$H4435 complete methylcellulose medium containing human IL-3, IL-6, SCF, G-CSF and GM-CSF to support formation of myeloid CFU, the other is MethoCult H4230 methylcellulose medium supplemented with 3 IU/mL human erythropoietin (EPO) (PeproTech), 5 ng/mL rhesus IL-3 (R&D) and 100 ng/mL human SCF (Miltenyi Biotec), to support erythroid colony formation.[32] Cells were plated at 1000 cells/mL in H4435 medium or at 10,000 cells/ mL in EPO-supplemented H4230 medium, incubated at 37°C and 5% CO_2. At day 12-14, colonies were enumerated and well-separated CFU were plucked individually for molecular analyses.

Erythropoietin treatment

Purified recombinant human EPO (PeproTech) was injected subcutaneously at 3000 U/kg for two doses 12 hours apart to stimulate macaque erythropoiesis[33-35] in barcoded animal ZL40 at 12 months post transplantation. Baseline BM and PB samples were collected five weeks before EPO stimulation, and reticulocyte concentrations in the PB were monitored every other day post EPO administration. When the reticulocytes rose to ≥8%, BM and PB samples were collected. Recovery samples were collected three months post EPO stimulation.

Results

Approaches for erythroid lineage tracking in the rhesus macaque model

In order to track clonal contributions to erythropoiesis in comparison to other lineages, $CD34^+$ HSPC from seven young RM and one aged RM were transduced with high diversity barcoded lentiviral libraries under conditions favoring a single unique barcode marking individual HSPC and reinfused into the autologous RM following ablative total body irradiation (TBI); see Table 1 for a summary of transplantation and transduction parameters. Following engraftment, samples were obtained from BM and PB (Figure 1). As reported previously, short-lived, lineage-restricted progenitors contributed for the first 1-2 months, followed by stable highly polyclonal contributions to Gr, monocytes, B cells, T cells and $CD56^{bright}$ NK cells from long-lived, stable, multipotent long-term repopulating HSPC clones.[21,23] In the current study, we designed assays to study clonal contributions to the erythroid lineage (Figure 1), with the anucleate state of mature circulating erythrocytes requiring design of alternative approaches for mapping of clonal contributions.

Clonal contributions to nucleated erythroid cells are shared with myeloid lineages

Nucleated RBC represent the final stage in marrow erythroid differentiation, before nuclear extrusion and exit from the BM into the PB. NRBC can be identified and sorted to high purity based on absent/low expression of the pan leukocyte marker CD45, and high expression of the transferrin receptor CD71[30,36] (Figure 2A). Unfortunately, RM-reactive antibodies recognizing other erythroid markers, such as glycophorin, do not currently exist. We sorted NRBC from BM MNC based on a $CD45^-CD71^+$ phenotype, with a starting population of 0.2-4.7%, and generated NRBC preparations with high purity (>95%) as confirmed by fluorescence-activated cell sorting (FACS), Wright-Giemsa staining and benzidine staining (Figure 2A and *Online Supplementary Figure S4*). Concurrent purified populations of NRBC, $CD34^+$ HSPC, T cells, B cells, monocytes and Gr were isolated from 10-15 mL BM aspirates from four young adult rhesus monkeys (JD76, ZK22, ZH19 and ZH33, 7-10 years of age) 3.5-46 months post transplantation, and one aged rhesus monkey (RQ3600, 23 years of age), 48 months post transplantation and DNA was obtained for barcode retrieval. Obtaining enough BM prior to three months post transplantation was not feasible due to low marrow cellularity during recovery from TBI.

As we reported recently,[23] there is marked clonal geographic segregation of the output from HSPC in the BM for at least six months post transplantation, followed by very gradual clonal mixing at different BM sites over subsequent months to years. Therefore, we analyzed the barcode clonal pattern of all lineages from the same BM sample, rather than comparing NRBC from one or a few BM sites to circulating myeloid and lymphoid cells. We visualized the contributions of the largest clones to each lineage mapped across all lineages in heat maps (Figure 2B) and analyzed Pearson correlations between all contributing clones (Figure 2C). At both early (3.5 months) and later time points up to several years post transplantation, clonal contributions were closely correlated between NRBC, monocytes and Gr (r values ranged from 0.68 to 0.95;

$P<0.05$) in both young and aged monkeys. Correlations were also high with B cells, known to be produced from HSPC locally in the BM, but clonal contributions to T cells were very distinct and poorly correlated, given that the final stages of T-cell development occur primarily in the thymus, thus T cells in the BM have returned back from the blood and represent the total body clonal T-cell landscape.[23] Figure 2D groups clonal contributions by degree of lineage bias for the young and aged macaques and demonstrates no appreciable contributions from HSPC clones contributing solely or in a highly biased way to NRBC but not to other hematopoietic lineages. The bias and relative size of barcoded clones in other lineages are shown in *Online Supplementary Figure S2*.

Erythroid and myeloid colony-forming units share clonal contributions

Colony-forming unit assays are widely used to study HSPC output and differentiation at a single cell level. Previous publications reported erythroid-biased output

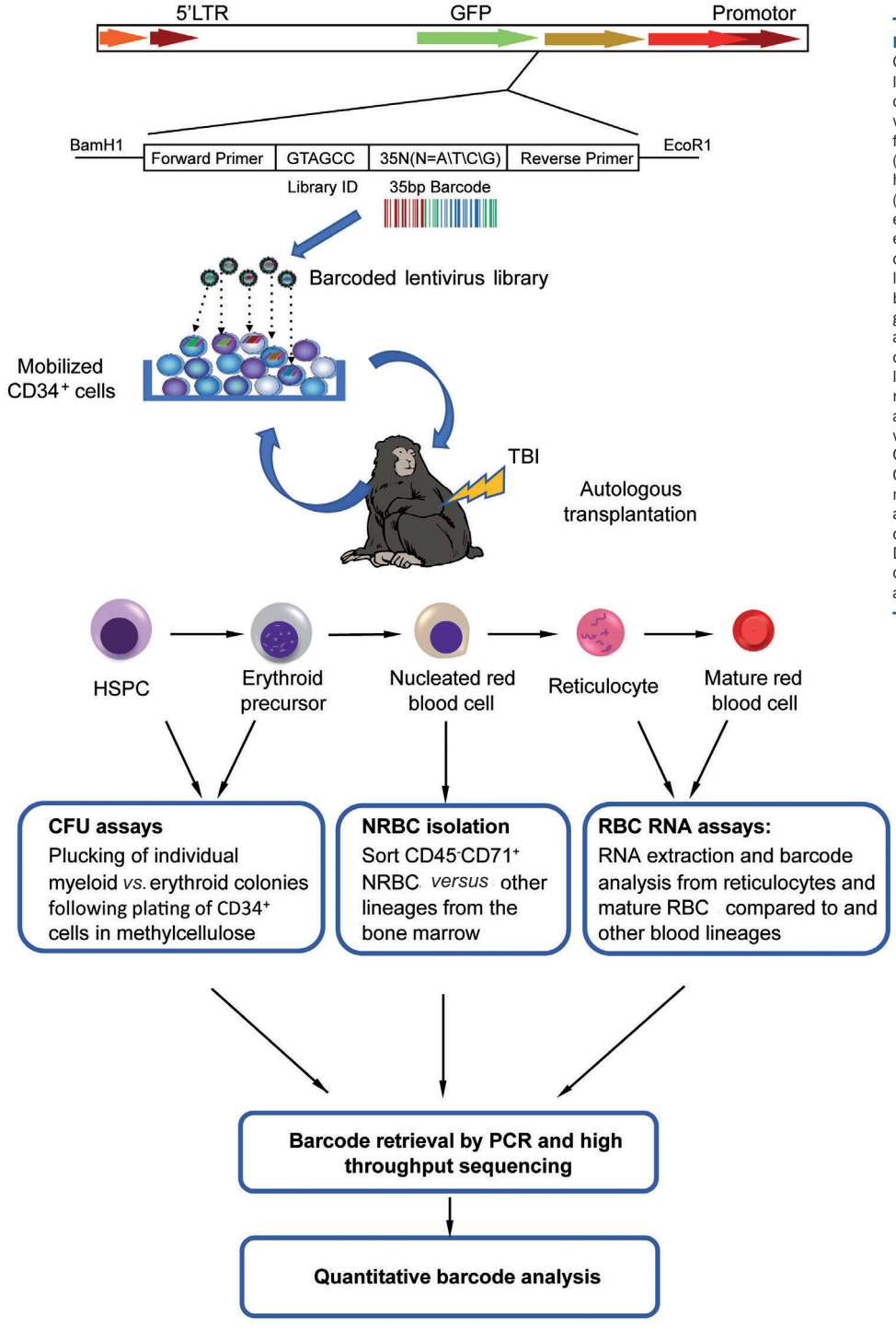


Figure 1. Experimental design. Oligonucleotides consisting of a 6bp library ID followed by a 27-35bp high diversity random sequence barcode were inserted into a lentiviral vector flanked by polymerase chain reaction (PCR) primer sites. RM CD34+ hematopoietic stem and progenitor cell (HSPC) were mobilized into the peripheral blood (PB), collected by apheresis, enriched *via* immunoselection, transduced with the barcoded lentiviral library, and infused back into the total body irradiation (TBI) irradiated autologous macaque. After engraftment, PB and bone marrow (BM) samples were obtained, and various hematopoietic lineages were purified for barcode retrieval and analyses. Lineage cells and nucleated red blood cell (NRBC) were purified from the BM and/or PB. Colony-forming unit (CFU) derived from CD34+ BM cells cultured in semi-solid media, and mature red blood cell (RBC) and reticulocytes were enriched *via* depleting nucleated cells from the PB. DNA and/or RNA were extracted for barcode PCR, high-throughput sequencing, and custom data analysis.

from individual HSPC based on CFU assays.[5,6] We conducted barcode analysis on CFU to investigate clonal output in this progenitor population. CD34+ cells were purified from BM samples post transplantation and cultured at low density under erythroid or myeloid cytokine conditions (Figure 3A). Individual CFU-E or CFU-GM were plucked and 450-600 of each type were pooled together for DNA extraction and barcode recovery. Cell populations of NRBC, T cells, B cells, monocytes and Gr were also purified from the remaining CD34- BM cells after CD34+ cell selection.

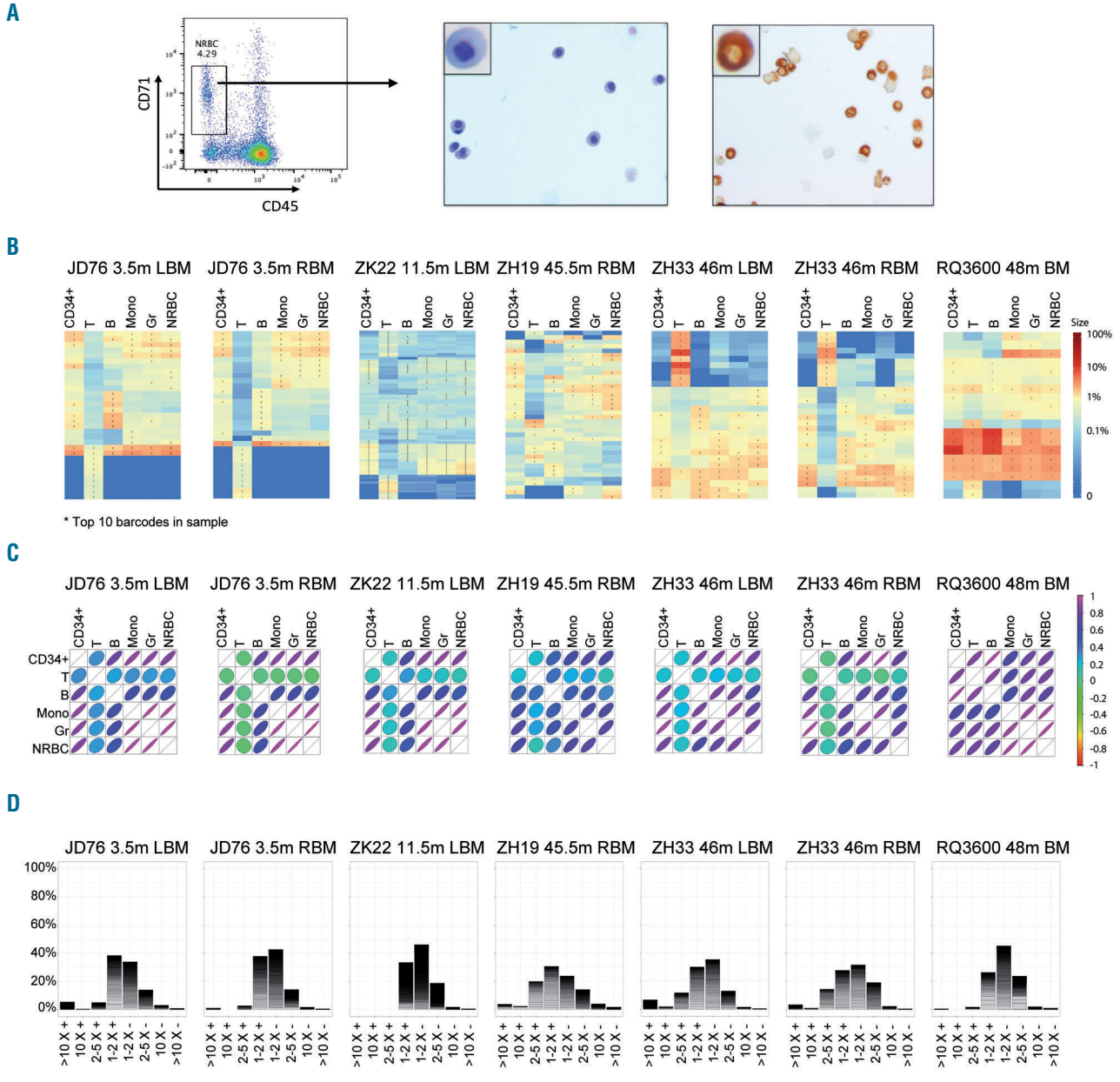


Figure 2. Clonal relationships between erythroid and other hematopoietic lineages. (A) CD45-CD71+ nucleated red blood cell (NRBC) were FACS-purified from six bone marrow (BM) mononuclear cell (MNC) samples obtained from four macaques post transplantation. A representative flow plot and corresponding cytospins of Wright-Giemsa staining and Benzidine staining of purified NRBC are shown. The percentage of NRBC was 94.3±0.8% by Wright-Giemsa staining, and 95.2% by Benzidine staining, counting at least 500 cells. (B) Heat maps representing the log fractional contributions of the top ten most abundant contributing clones retrieved from each different bone marrow (BM) cell lineage population plotted over all BM cell populations. Each individual row represents the fractional contributions from an individual barcode (clone), and each individual column represents a sample. *A barcode is one of the top ten contributing clones in that cell sample (column). Since the top ten barcoded clones in each sample are plotted across all samples, each row in the heat map has at least one*, and each column has exactly ten*. The rows are ordered by unsupervised hierarchical clustering using Euclidean distances to group barcoded clones together that manifest similar patterns of clonal contributions. The color scale on the right depicts the log fractional contribution size. Samples include CD34+ hematopoietic stem and progenitor cell (HSPC), T cells, B cells, monocytes (Mono), granulocytes (Gr), and NRBC. (C) Pearson correlation coefficients comparing all barcoded clonal contributions between different lineages in the same six BM samples shown in (A). The color scale bar for r values is on the right, the shape and the color signify the strength of the correlation. (D) Stacked histograms displaying the degree of erythroid lineage bias and relative size of barcoded clonal contributions (largest to smallest, with smallest clonal contributions appearing only as overlapping lines). Erythroid lineage bias is calculated *via* a ratio of the fractional contribution of a barcode to NRBC *versus* the fractional contribution of the barcode to another lineage (Gr, Mono, CD34+, T, or B), using the largest fractional contribution among the other lineages for each barcode to calculate the ratio. The positive sign (+) indicates bias towards the NRBC lineage and the negative sign (-) indicates bias away from the NRBC lineage.

We analyzed CFU from two animals at 3.5 months post-transplantation (ZJ31 and JD76). The clone heatmaps from these two monkeys documented that the majority of barcodes retrieved from pooled CFU-E were also detected in pooled CFU-GM, along with purified monocytes and Gr, suggesting a shared unbiased myeloid-erythroid HSPC pool (Figure 3B). In addition, we observed unique groups of clones contributing to either CFU-E or CFU-GM but not to both, and those unique clones were also detected as low-contributing clones in circulating blood lineages (Figure 3B), suggesting that the apparent lineage restriction may be due to sampling limitations of the pooled CFU approach.

RNA barcoding tracing of reticulocytes and mature red blood cells from peripheral blood

In an attempt to analyze contributions to the erythroid lineage more comprehensively, sampling the entire blood compartment instead of localized marrow sites, we designed an approach to barcode retrieval from circulating erythroid cells. During the process of erythropoiesis, maturing erythroid cells excluded their nuclei and exited the marrow as reticulocytes containing ample RNA, with loss of RNA over time during the approximately 100-day lifespan of circulating erythrocytes, given lack of ongoing transcription in anucleate cells. Since the lentiviral vectors utilized located the barcodes in a transcribed region of the integrated provirus, we asked whether barcode retrieval from cellular RNA could be used to quantitatively study clonal contributions to various lineages, allowing comparisons between circulating RBC and other blood lineages (Figure 4A). We purified RBC depleted of >99% of nucleated leukocytes as confirmed *via* flow cytometry for CD45 expression and by morphologic scoring (Figure 4B). Polychromatophilic reticulocytes containing ample RNA could be identified by larger size and a grayish color on Wright's-stained blood smears (Figure 4B).

To ask whether RNA barcode contributions match DNA clonal contributions for nucleated blood lineages, we first compared both DNA and RNA barcodes retrieved from the same blood or marrow samples of Gr, monocytes, B cells, T cells and marrow NRBC, specifically from animal ZK22 15.5 months post transplantation; a time point late enough to reach clonal equilibration and homogeneity between different BM sites and PB. In these nucleated cells, the fractional contributions of barcoded clones to DNA *versus* RNA from the same sample were overall well-correlated, with Pearson correlation coefficients (r values) of 0.76±0.04 (Figure 4C), although it is apparent and not surprising that some clones in all lineages contribute primarily at a DNA but not an RNA level, likely due to insertions in chromatin at sites inhospitable to pro-viral gene expression.

However, in comparing major barcode contributions in RNA from circulating RBC *versus* BM NRBC (Figure 4D), most of the barcodes found in NRBC and myeloid samples were also present in the RBC RNA barcodes. But quantitatively the relative levels of individual barcode contributions could vary markedly between immature NRBC and more mature anucleate circulating cells,

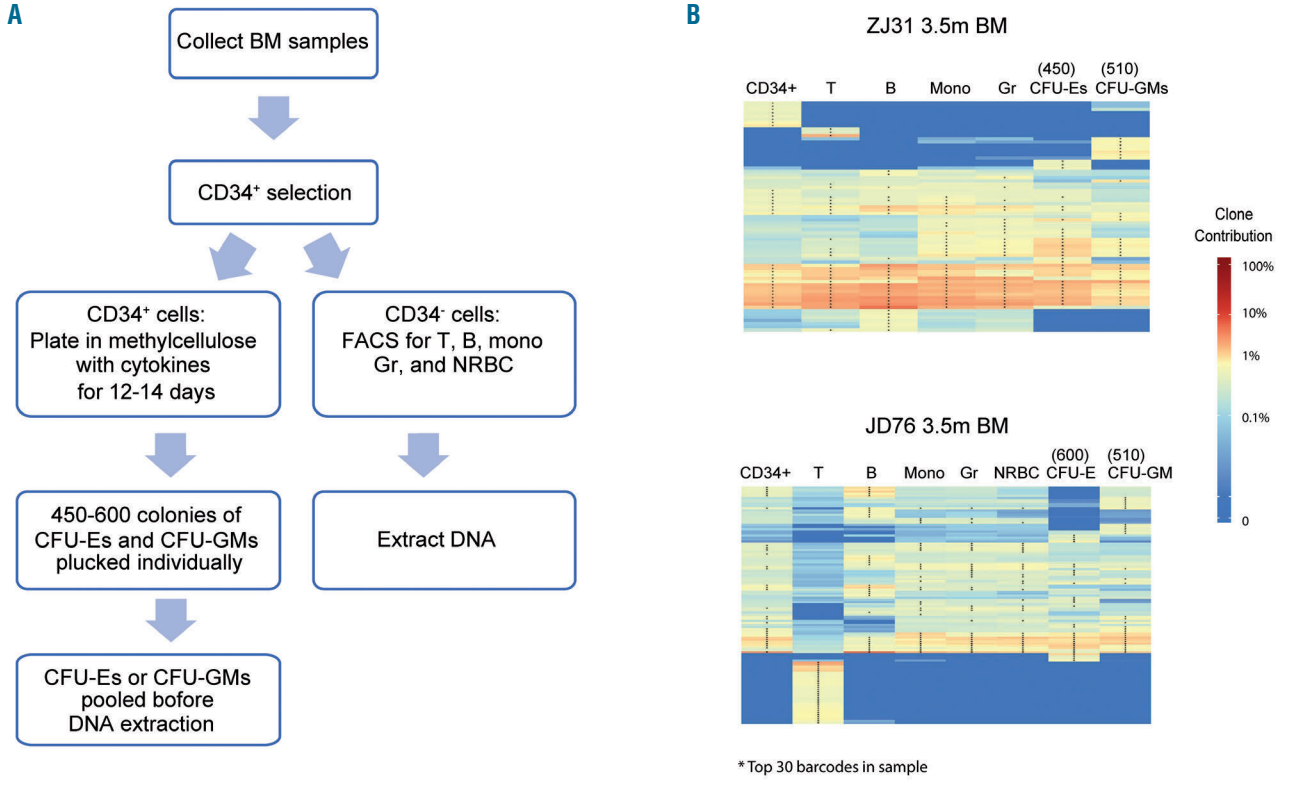


Figure 3. Barcode analysis of bone marrow (BM) colony-forming units (CFU). (A) Flowchart for CFU collection and barcode retrieval. (B) Heatmap of top 30 clones in BM CD34+ cells, T, B, Mono, Gr and nucleated red blood cell (NRBC), and pooled myeloid and erythroid CFU samples from ZJ31 (3.5m) and JD76 (3.5m). The colony number of CFU-E and CFU-GM pooled for DNA extraction and analysis are given on top of each CFU column. Heatmaps were constructed as described in Figure 2B.

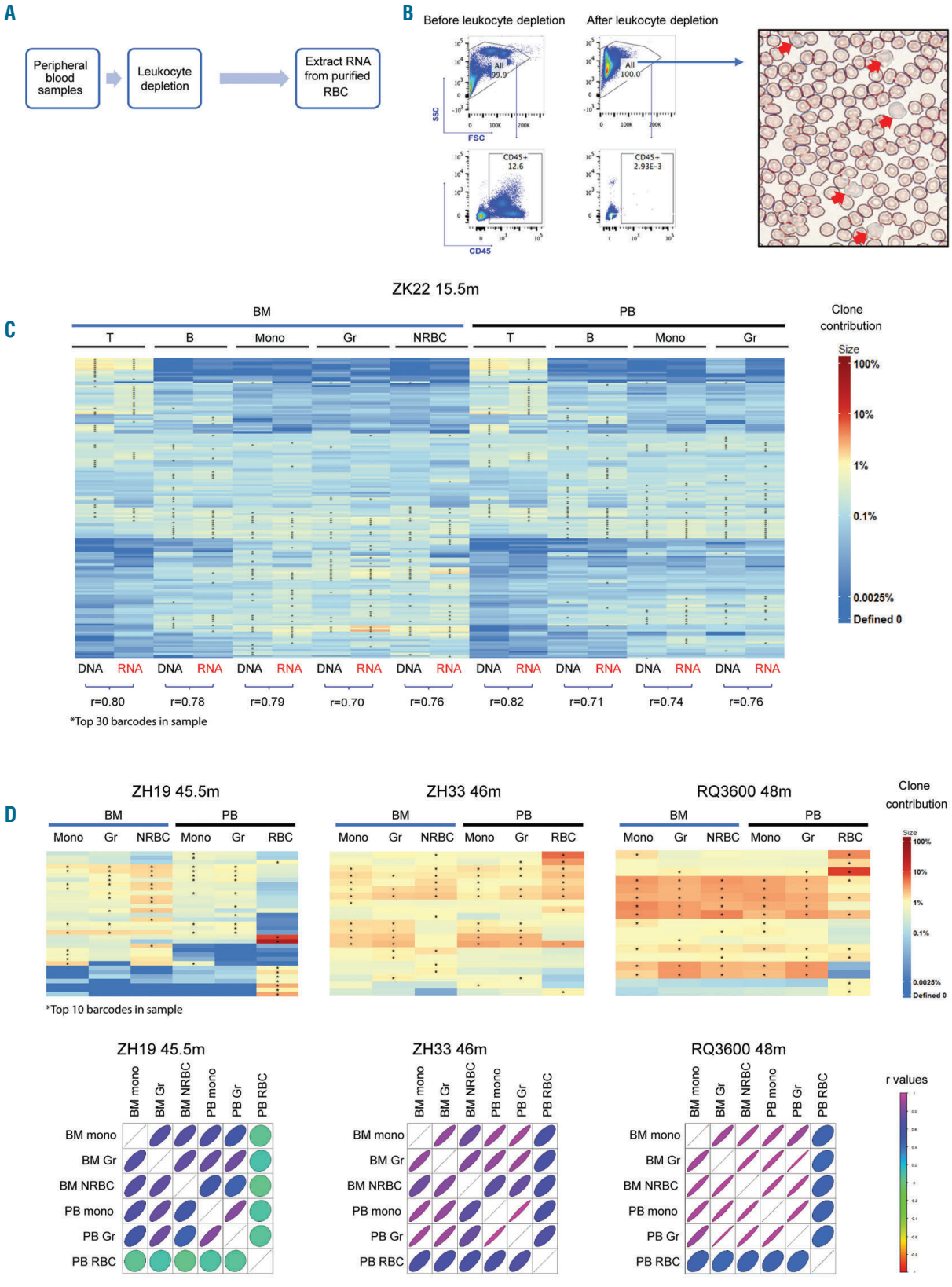


Figure 4. Clonal contributions to circulating erythrocytes. (A) Flowchart for RNA barcode retrieval from circulating anucleate mature red blood cells (RBC) and reticulocytes. (B) FACS plots (left panel) showing CD45 expression on whole blood cells before and after leukocyte depletion, and a Wright's stained blood smear post leukocyte depletion (right panel). The red arrows indicate polychromatophilic reticulocytes with a blue-gray color due to increased RNA content. (C) Heatmap plotting contributions from the top 30 clones in each sample of DNA or RNA obtained from ZK22 15.5m post transplantation, plotted across all samples; heatmap was made as explained in Figure 2B. The paired Pearson correlations between DNA and RNA global barcode contributions to the same sample are given on the bottom of the heatmap. (D) The heatmaps (upper panel) and Pearson correlation plots (lower panel) show peripheral blood (PB) RBC RNA barcodes and the DNA barcode from PB Mono and granulocyte (Gr) and BM Mono, Gr, nucleated red blood cell (NRBC) at the same time point from three rhesus macaque (RM). The color scale is on the right of each panel.

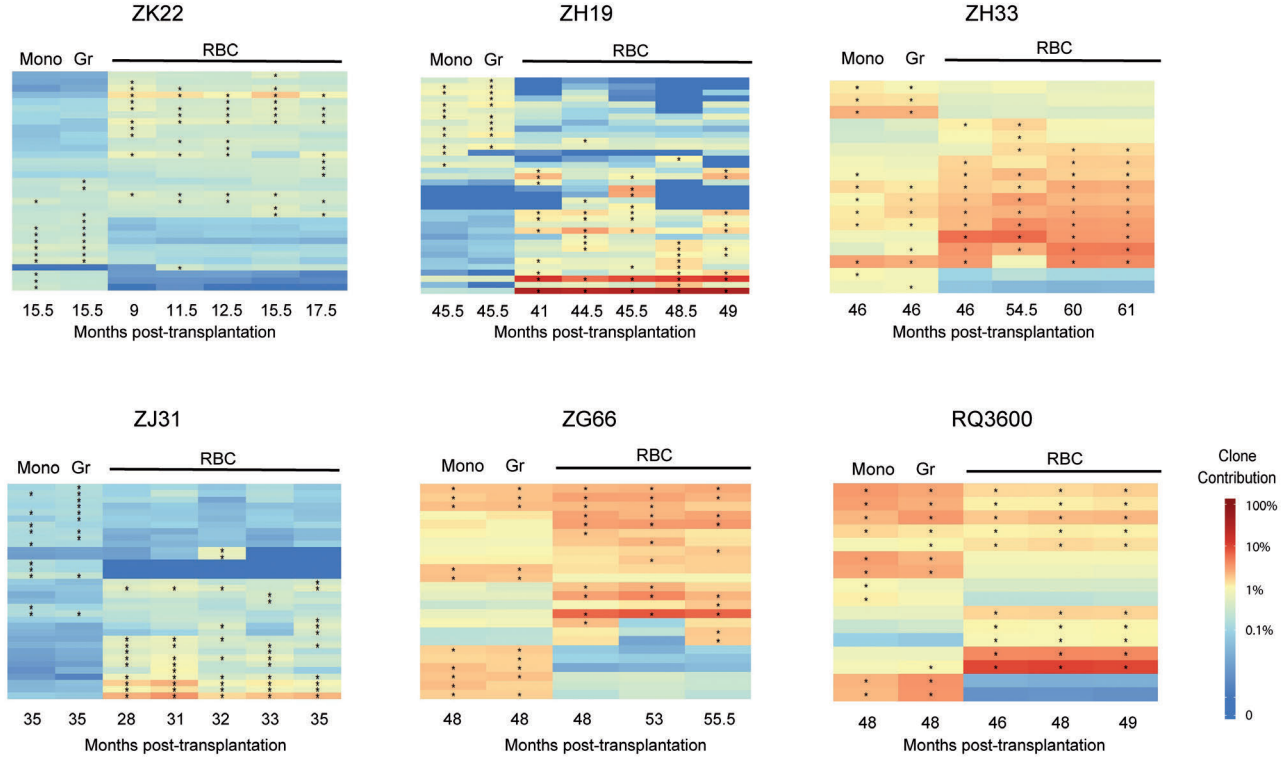
including some barcodes found contributing at high relative levels to RBC compared to NRBC or to other lineages. Conversely, many clones found contributing to NRBC as well as BM or circulating myeloid cells (Gr and Mono) were undetectable or contributing at a fractionally much lower level in RBC. Overall, the Pearson correlations between RNA barcodes contributing to RBC *versus* myeloid cells were much lower than between the other lineages, including between NRBC and myeloid cells, and low even between NRBC and RBC (Figure 4D). These findings may result from legitimate HSPC clonal bias, i.e. HSPC clones that contribute in a highly-biased manner to the erythroid *versus* other lineages. But this seems unlikely, given the discrepancy in clonal patterns between NRBC and anucleate RBC. Instead, we suspect this reflects the marked transcriptional restriction and differential RNA stability occurring in maturing RBC as they transition towards producing massive amounts of only a few proteins, most notably hemoglobin.[37] A large fraction of vector insertion sites may be silenced transcriptionally or translationally during the final stages of RBC enucleation and release, making extrapolating clonal contributions from RNA barcode expression problematic.

We were, however, able to use RNA barcode retrieval to track the stability of clonal contributions to circulating RBC over time, studying six macaques for intervals of up to 15 months. PB RBC RNA barcode analysis showed very stable barcode RBC contribution patterns over time in both young and old animals (Figure 5).

Erythropoietin stimulation does not alter erythroid clonal patterns

We investigated the impact of erythropoietic stress on erythroid clonal patterns. EPO is the key lineage-specific humoral regulator of mammalian erythropoiesis, and in the setting of anemia, levels increase to expand the erythroid compartment. To assess whether lineage-biased clones could be recruited *via* lineage-specific stimulation, we administered a short course of high-dose EPO to achieve significant erythroid proliferation[33-35] in barcoded monkey ZL40. BM and PB samples were collected from animal ZL40 at baseline 10.5 months following transplantation, at the peak of reticulocytosis on day 6 of EPO administration, and at recovery several months later, once blood counts and reticulocyte numbers had returned to baseline (Figure 6A and B).

The clonal patterns in BM NRBC were stable over time whether at baseline, peak or recovery following EPO, and matched those of BM monocytes and Gr (Figure 6C). The RNA barcode analysis of PB RBC RNA also showed an unchanged clonal pattern in response to EPO (Figure 6C). The results indicated that the EPO administration did not change the erythroid clonal output, it only stimulated the existing erythroid progenitor pool to produce more RBC, but without recruiting previously-quiescent HSPC clones to generate new erythroid cells.



*Top 10 barcodes in sample

Figure 5. Stability of erythroid clonal contributions over time. Heatmaps plotting barcode contributions to peripheral blood (PB) red blood cell (RBC) RNA over time in five young macaques and one aged (RQ3600) macaque, compared to DNA barcode from PB monocytes (Mono) and granulocytes (Gr). Each heatmap plots the top ten contributing clones in each sample across all samples, heatmaps were made as explained in Figure 2B. The color scale is on the right.

Discussion

The cellular differentiation pathways supporting ongoing adult erythropoiesis from primitive marrow HSPC are still not completely clear, with some contradictory findings in murine and human studies, particularly regarding whether erythroid production is supported by multipotent then bipotent intermediate steps downstream from the most primitive HSC, as in the classical model.[1-3] Recent findings in murine model human cells studied in robust single cell *in vitro* assays, and pseudotemporal ordering of single cell RNA-Seq gene expression patterns have raised the possibility that the erythroid and megakaryocytic lineages, along with eosinophils and basophils in some studies, may represent the earliest branch point during hematopoiesis from self-renewing LT-HSC.[5,8-10,12,13,17,38,39] The RNA-Seq study most relevant to erythropoiesis came from Tusi and co-workers, enriching for cell populations previously linked to erythropoiesis to study differentiation trajectories, reporting that HSC/MPP first bifurcate towards erythroid-basophil-megakaryocytes *versus* myeloid-lymphoid pathways before constricting to erythroid or towards myeloid and lymphoid fates. Murine lineage tracing studies using platelet lineage-specific promoters have uncovered evidence for long-lasting and self-renewing megakaryocyte-restricted HSC, but to date, erythroid-restricted engrafting HSPC have not been uncovered, despite the suggestion, based on gene expression studies, that they may exist.[15-17,40] These observations led us to ask whether we could detect long-lived erythroid-restricted or highly erythroid-biased HSPC in our RM barcoded clonal tracking model.

Techniques allowing long-term tracking of hematopoiesis *in vivo* at a single cell level *via* fate mapping or clonal tags are powerful tools to provide answers to these questions and have been utilized primarily in murine models over the past decade. Yamamoto *et al.*[15] used single cell murine HSPC transplants along with lineage-specific markers and demonstrated lineage-restricted engrafting progenitors producing platelets, platelets and red cells, or all myeloid/erythroid lineages, and single cells able to produce both megakaryocyte and multipotent engrafting daughter cells, but no other uni-lineage daughter HSC. Using a non-transplant "naïve hematopoiesis" murine transposon-tagging model, Rodriguez *et al.* came to similar conclusions regarding LT-HSPC heterogeneity, showing megakaryocytic-restricted long-term contributing HSPC clones, but did not examine erythroid output at a clonal level.[17] The Jacobsen laboratory also reported lineage-restricted long-term self-renewing potential initially only for the platelet lineage in mice,[40] more recently adding an erythroid-specific tracer that did not reveal long-term erythroid-restricted HSPC, but did uncover very rare platelet-erythroid

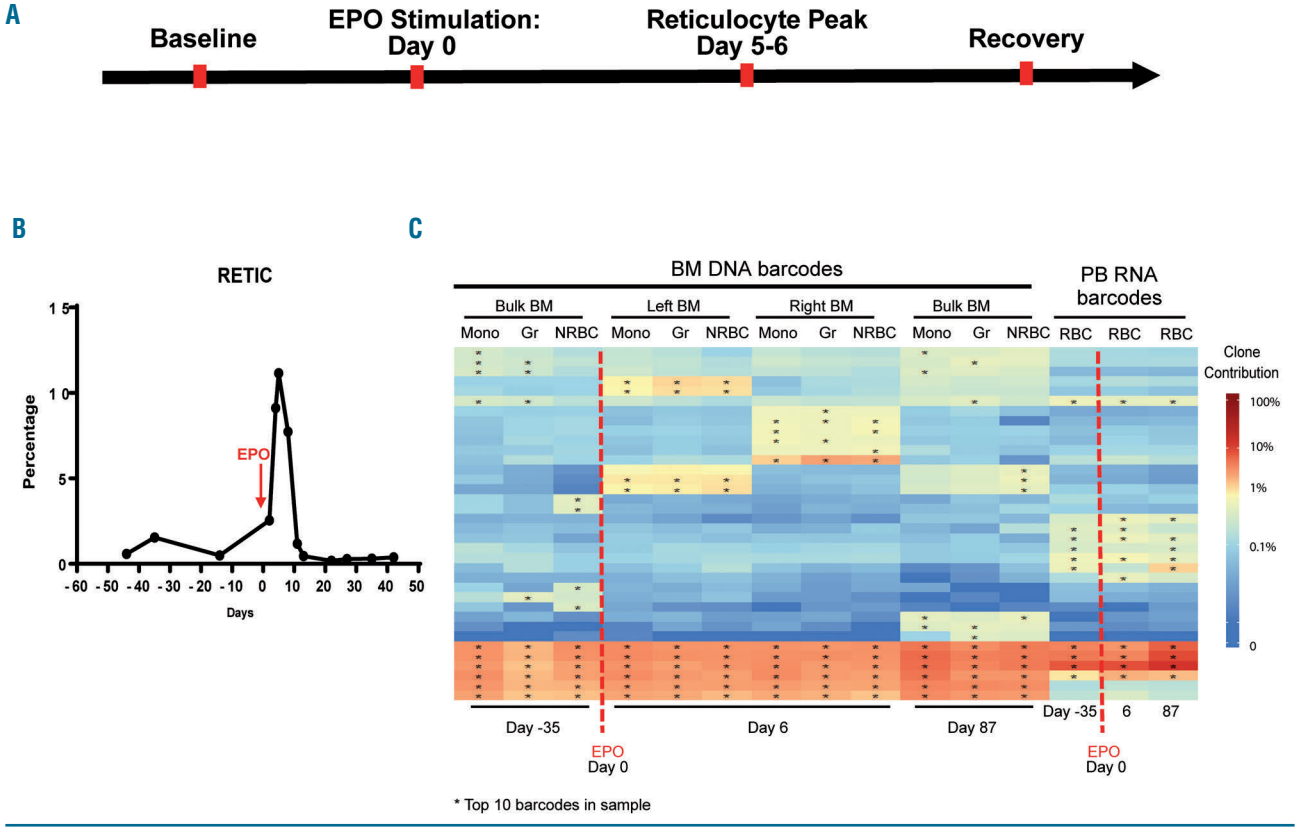


Figure 6. Impact of erythropoietin (EPO) administration on erythroid clonal patterns. (A) Timeline of EPO stimulation and sampling in ZL40, EPO was administrated at 11.5m post transplantation. (B) The reticulocyte (RETIC) percentage during EPO stimulation. (C) Heatmap of the top ten contributing clones from bone marrow (BM) monocytes (Mono), granulocytes (Gr) and nucleated red blood cell (NRBC) DNA barcodes and peripheral blood (PB) red blood cell (RBC) RNA barcodes samples before and post EPO stimulation. Heatmap was constructed as described in Figure 2B. day -35: baseline, day 6: reticulocyte peak, and day 87: recovery after EPO stimulation are shown.

restricted LT-HSPC.[16] However, the majority of single HSPC were multipotent.

Similar *in vivo* studies of human LT-HSPC properties *via* single cell transplantation or naïve hematopoietic tagging are not feasible, but the significant differences between human and rodent hematopoiesis make extrapolation difficult.[41] *In vitro* single cell CFU assays are the classic approach to study lineage relationships and search for intermediate progenitors in hematopoiesis.[1,3] Notta *et al.*[5] found that populations of human CD34$^+$ HSPC sorted by previously-described erythroid and megakaryocytic lineage markers and grown in single cell culture did not reveal single progenitors with both erythroid and megakaryocytic or erythroid and myeloid potential, in contrast to the presence of such progenitors in fetal liver and cord blood, thereby supporting the concept that both erythroid and megakaryocytic lineages emerge directly from multipotent HSC, at least postnatally. In contrast, several other groups did find appreciable numbers of single human progenitors with both erythroid and myeloid potential using a different culture system and lineage-defining antibodies.[6,7]

Our macaque barcoding approach provides a robust platform to quantitatively track the output of thousands of individual HPSC clones long-term over multiple lineages *in vivo*. The clonal landscape in NRBC closely matched that of other hematopoietic lineages sampled from the same BM locus at same time point, with a particularly close relationship between NRBC and Gr and monocytes, which are continuously produced from HSPC, in contrast to T cells that can clonally expand and renew peripherally. Our analysis of thousands of individual clones contributing to purified nucleated erythroid precursors *versus* other lineages at the same time point and at the same marrow location did not uncover a measurable population of markedly erythroid-biased LT-HSPC in this post-transplantation model. Likewise, analysis of barcodes from hundreds of pooled myeloid and erythroid CFU grown from marrow post transplantation led to similar conclusions, although a small fraction of clones unique to CFU-E were identified; however, CFU-GM analyses also revealed unique clones. Given that engraftment in macaques results from many thousands of individual LT-HSPC,[20] the apparent presence of lineage-restricted CFU was likely due to sampling bias or potentially differential clonal *in vivo versus* in CFU *in vitro*. However, we cannot rule out that these lineage-restricted clones detected only *via* CFU analysis represent true myeloid or erythroid-restricted clones not detected in our NRBC analyses.

Our findings were confirmed at various time points from months to years post transplantation. Even in an aged macaque, previously shown to have long-term persistence of highly myeloid and lymphoid biased clones,[22] we found no evidence for erythroid-biased clones not also biased towards myeloid output. A recent human lentiviral gene therapy study in patients with an immunodeficiency disorder used insertion site retrieval to map clonal relationships between lineages, and also reported primarily shared clones contributing to erythroid and myeloid lineages.[42] Of note, our findings do not contradict the concept, derived from recent single cell gene expression studies, that erythroid pathways diverge very early from myeloid and lymphoid lineages, representing one of the first branches from differentiating LT-HSC. A multipotent LT-HSC marked by a barcode could produce daughters going down both erythroid and non-erythroid pathways. Our results do suggest that no significant contributions from self-renewing or engrafting erythroid-restricted or highly biased progenitors could be detected, at least in this transplantation model. However, in several murine studies, the stress of transplantation or *in vitro* culture was shown to drive megakaryocytic-biased HSC to contribute to additional lineages,[16,17] thus our transplantation model may not reflect physiologic naïve hematopoiesis.

Since we have discovered that clonal output from individual HSPC can remain highly geographically restricted within the BM for up to years post transplantation,[23] it may be difficult to study the entire clonal composition of NRBC or erythroid progenitors *via* marrow sampling alone. Thus we also analyzed the barcodes in circulating mature RBC and reticulocytes *via* retrieval of RNA barcodes, given the lack of DNA in enucleated cells. Fractional contributions of DNA and RNA barcodes retrieved from the same nucleated sample of each lineage were compared and showed high correlation, suggesting the differentiation pathway for these lineages does not impact significantly on expression level of barcodes from our vector regardless of insertion site, due to expression of the marker gene and the barcode RNA from a strong constitutive viral promoter. However, RNA barcodes in circulating reticulocytes/mature RBC revealed a less diverse clonal pattern and major differences from clonal contribution patterns revealed in RNA of circulating myeloid and lymphoid cells, and even from NRBC concurrently sampled from the BM in animals many years post transplantation. Rather than implying erythroid lineage bias, we suspect these findings resulted from a major constriction of gene expression in anucleate erythroid cells, with marked silencing of many endogenous genes other than hemoglobin and red cell structural proteins. Bonafoux *et al.*[43] reported globally skewed transcriptional activity follow erythroid differentiation, with major differences in gene expression between end stage anucleate erythroid cells and leukocytes, and even in comparison with earlier stage erythroid progenitors. More than 50% of transcripts in these end stage erythroid cells encoded globin. The fact that many viral integrations were likely also silenced and thus did not express the barcode is supported by the observation that the GFP percentage in RBC was much lower than in Gr in all our macaques (*Online Supplementary Figure S3*). Despite the difficulties in comparing clonal contributions in RBC RNA to other lineages, sampling of PB RBC over prolonged periods of time allowed us to establish the long-term clonal stability of contributing HSPC to erythropoiesis for as long as four years post transplantation.

Given the advantage of our experimental models, we sought to answer some additional questions, including whether erythropoiesis at a clonal level will be stable under lineage-specific stimulation, or if lineage-specific clones might be recruited under this proliferative stress, as has been suggested for lineage-restricted megakaryocytopoiesis.[44] Thus, we tracked the clonal characteristics of hematopoiesis under stimulation by EPO, a key lineage-specific humoral regulator responsible for erythroid progenitor proliferation and accelerated maturation. We found no detectable impact on clonal contributions under EPO stimulation, suggesting that the pool of HSPC contributing to erythropoiesis does not shift between steady state and proliferative stress resulting from EPO stimula-

tion. It would be of interest to examine megakaryocytes/platelet output in our model, at steady state or under stress, but the size and lack of nucleus in platelets present challenges, as does the rarity of megakaryocytes in the BM. We are developing an optimized vector allowing concurrent single cell RNA-Seq and barcode retrieval to further investigate platelet and other lineage relationships.[17]

In conclusion, this study is the first to quantitatively track erythropoiesis at a clonal level *in vivo* in a translationally-relevant model. Overall, our analyses indicate long-term shared ontogeny with other lineages, particularly myeloid cells, without measurable contributions from highly erythroid-biased or restricted long-term engrafting HSPC, even under stressors such as aging or erythroid lineage-specific cytokine stimulation. A better understanding of erythropoiesis at this level has relevance for further development of new therapies targeting erythroid disorders.

Funding

This research was supported by the NHLBI Division of Intramural Research.

Acknowledgments

We thank Naoya Uchida for the χHIV plasmid, and Keyvan Keyvanfar for cell sorting. We acknowledge the support of the NHLBI FACS Core, the NHLBI DNA Sequencing and Genomics Core, the NHLBI animal care and veterinary staff, and the NIH Biowulf High-Performance Computing Resource.

References

1. Akashi K, Traver D, Miyamoto T, Weissman IL. A clonogenic common myeloid progenitor that gives rise to all myeloid lineages. Nature. 2000; 404(6774):193-197.
2. Vannucchi AM, Paoletti F, Linari S, et al. Identification and characterization of a bipotent (erythroid and megakaryocytic) cell precursor from the spleen of phenylhydrazine-treated mice. Blood. 2000; 95(8):2559-2568.
3. Manz MG, Miyamoto T, Akashi K, Weissman IL. Prospective isolation of human clonogenic common myeloid progenitors. Proc Natl Acad Sci U S A. 2002; 99(18):11872-11877.
4. Adolfsson J, Mansson R, Buza-Vidas N, et al. Identification of Flt3+ lympho-myeloid stem cells lacking erythro-megakaryocytic potential a revised road map for adult blood lineage commitment. Cell. 2005;121(2):295-306.
5. Notta F, Zandi S, Takayama N, et al. Distinct routes of lineage development reshape the human blood hierarchy across ontogeny. Science. 2016;351(6269): aab2116.
6. Psaila B, Barkas N, Iskander D, et al. Single-cell profiling of human megakaryocyte-erythroid progenitors identifies distinct megakaryocyte and erythroid differentiation pathways. Genome Biol. 2016;17:83.
7. Sanada C, Xavier-Ferrucio J, Lu YC, et al. Adult human megakaryocyte-erythroid progenitors are in the CD34+CD38mid fraction. Blood. 2016;128(7):923-933.
8. Athanasiadis EI, Botthof JG, Andres H, Ferreira L, Lio P, Cvejic A. Single-cell RNA-sequencing uncovers transcriptional states and fate decisions in haematopoiesis. Nat Commun. 2017;8(1):2045.
9. Tusi BK, Wolock SL, Weinreb C, et al. Population snapshots predict early haematopoietic and erythroid hierarchies. Nature. 2018;555(7694):54-60.
10. de Graaf CA, Choi J, Baldwin TM, et al. Haemopedia: An Expression Atlas of Murine Hematopoietic Cells. Stem Cell Reports. 2016;7(3):571-582.
11. Hamey FK, Nestorowa S, Kinston SJ, Kent DG, Wilson NK, Gottgens B. Reconstructing blood stem cell regulatory network models from single-cell molecular profiles. Proc Natl Acad Sci U S A. 2017;114(23):5822-5829.
12. Paul F, Arkin Y, Giladi A, et al. Transcriptional Heterogeneity and Lineage Commitment in Myeloid Progenitors. Cell. 2015;163(7):1663-1677.
13. Velten L, Haas SF, Raffel S, et al. Human haematopoietic stem cell lineage commitment is a continuous process. Nat Cell Biol. 2017;19(4):271-281.
14. Dykstra B, Kent D, Bowie M, et al. Long-term propagation of distinct hematopoietic differentiation programs in vivo. Cell Stem Cell. 2007;1(2):218-229.
15. Yamamoto R, Morita Y, Ooehara J, et al. Clonal analysis unveils self-renewing lineage-restricted progenitors generated directly from hematopoietic stem cells. Cell. 2013; 154(5):1112-1126.
16. Carrelha J, Meng Y, Kettyle LM, et al. Hierarchically related lineage-restricted fates of multipotent haematopoietic stem cells. Nature. 2018;554(7690):106-111.
17. Rodriguez-Fraticelli AE, Wolock SL, Weinreb CS, et al. Clonal analysis of lineage fate in native haematopoiesis. Nature. 2018; 553(7687):212-216.
18. Donahue RE, Dunbar CE. Update on the use of nonhuman primate models for preclinical testing of gene therapy approaches targeting hematopoietic cells. Hum Gene Ther. 2001; 12(6):607-617.
19. Shepherd BE, Kiem HP, Lansdorp PM, et al. Hematopoietic stem-cell behavior in nonhuman primates. Blood. 2007;110(6):1806-1813.
20. Koelle SJ, Espinoza DA, Wu C, et al. Quantitative stability of hematopoietic stem and progenitor cell clonal output in rhesus macaques receiving transplants. Blood. 2017;129(11):1448-1457.
21. Wu C, Li B, Lu R, et al. Clonal tracking of rhesus macaque hematopoiesis highlights a distinct lineage origin for natural killer cells. Cell Stem Cell. 2014;14(4):486-499.
22. Yu KR, Espinoza DA, Wu C, et al. The impact of aging on primate hematopoiesis as interrogated by clonal tracking. Blood. 2018;131(11):1195-1205.
23. Wu C, Espinoza DA, Koelle SJ, et al. Geographic clonal tracking in macaques provides insights into HSPC migration and differentiation. J Exp Med. 2018;215(1):217-232.
24. Donahue RE, Kuramoto K, Dunbar CE. Large animal models for stem and progenitor cell analysis. Curr Protoc Immunol. 2005;Chapter 22:Unit 22A.1.
25. Wu C, Li B, Lu R, et al. Clonal tracking of rhesus macaque hematopoiesis highlights a distinct lineage origin for natural killer cells. Cell Stem Cell. 2014;14(4):486-499.
26. Yu KR, Espinoza DA, Wu C, et al. The impact of aging on primate hematopoiesis as interrogated by clonal tracking. Blood. 2018;131(11):1195-1205.
27. 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;29(10):928-933.
28. Uchida N, Washington KN, Hayakawa J, et al. Development of a human immunodeficiency virus type 1-based lentiviral vector that allows efficient transduction of both human and rhesus blood cells. J Virol. 2009; 83(19):9854-9862.
29. 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;29(10):928-933.
30. Fornas O, Domingo JC, Marin P, Petriz J. Flow cytometric-based isolation of nucleated erythroid cells during maturation: an approach to cell surface antigen studies. Cytometry. 2002;50(6):305-312.
31. Sriprawat K, Kaewpongsri S, Suwanarusk R, et al. Effective and cheap removal of leukocytes and platelets from Plasmodium vivax infected blood. Malar J. 2009;8:115.
32. Wong S, Keyvanfar K, Wan Z, Kajigaya S, Young NS, Zhi N. Establishment of an erythroid cell line from primary CD36+ erythroid progenitor cells. Exp Hematol. 2010;38(11):994-1005.e1-2.
33. Ramakrishnan R, Cheung WK, Farrell F, Joffee L, Jusko WJ. Pharmacokinetic and pharmacodynamic modeling of recombinant human erythropoietin after intravenous and subcutaneous dose administration in cynomolgus monkeys. J Pharmacol Exp Ther. 2003;306(1):324-331.
34. Lavelle D, Molokie R, Ducksworth J, DeSimone J. Effects of hydroxurea, stem cell factor, and erythropoietin in combination on fetal hemoglobin in the baboon. Exp Hematol. 2001;29(2):156-162.
35. Umemura T, al-Khatti A, Donahue RE, Papayannopoulou T, Stamatoyannopoulos G. Effects of interleukin-3 and erythropoi-

etin on in vivo erythropoiesis and F-cell formation in primates. Blood. 1989;74(5):1571-1576.
36. Dzierzak E, Philipsen S. Erythropoiesis: development and differentiation. Cold Spring Harb Perspect Med. 2013; 3(4):a011601.
37. Thiadens KA, von Lindern M. Selective mRNA translation in erythropoiesis. Biochem Soc Trans. 2015;43(3):343-347.
38. Wilson NK, Kent DG, Buettner F, et al. Combined Single-Cell Functional and Gene Expression Analysis Resolves Heterogeneity within Stem Cell Populations. Cell Stem Cell. 2015;16(6):712-724.
39. Pellin D, Loperfido M, Baricordi C, et al. A comprehensive single cell transcriptional landscape of human hematopoietic progenitors. Nat Commun. 2019;10(1):2395.
40. Sanjuan-Pla A, Macaulay IC, Jensen CT, et al. Platelet-biased stem cells reside at the apex of the haematopoietic stem-cell hierarchy. Nature. 2013;502(7470):232-236.
41. Doulatov S, Notta F, Laurenti E, Dick JE. Hematopoiesis: a human perspective. Cell Stem Cell. 2012;10(2):120-136.
42. Biasco L, Pellin D, Scala S, et al. In vivo tracking of human hematopoiesis reveals patterns of clonal dynamics during early and steady-state reconstitution phases. Cell Stem Cell. 2016;19(1):107-119.
43. Bonafoux B, Lejeune M, Piquemal D, et al. Analysis of remnant reticulocyte mRNA reveals new genes and antisense transcripts expressed in the human erythroid lineage. Haematologica. 2004;89(12):1434-1438.
44. Haas S, Hansson J, Klimmeck D, et al. Inflammation-induced emergency megakaryopoiesis driven by hematopoietic stem cell-like megakaryocyte progenitors. Cell Stem Cell. 2015;17(4):422-434.