Sort-purification of human CD34$^+$CD90$^+$ cells reduces target cell population and improves lentiviral transduction

Authors:

Stefan Radtke[1,2,#], Dnyanada Pande[1,2], Margaret Cui[1,2], Anai M. Perez[1,2], Yan-Yi Chan[1,2], Mark Enstrom[1,2], Stefanie Schmuck[1,2], Andrew Berger[2], Tom Eunson[2], Jennifer E. Adair[1,2,3], Hans-Peter Kiem[1,2,4,5,#]

Affiliations:

[1] Stem Cell and Gene Therapy Program, Fred Hutchinson Cancer Research Center, Seattle, WA, 98109, USA

[2] Clinical Research Division, Fred Hutchinson Cancer Research Center, Seattle, WA, 98109, USA

[3] Department of Medical Oncology, University of Washington School of Medicine, Seattle, WA 98195

[4] Department of Medicine, University of Washington School of Medicine, Seattle, WA, 98195, USA

[5] Department of Pathology, University of Washington School of Medicine, Seattle, WA, 98195, USA

#Correspondence should be addressed to:

S.R. (sradtke@fredhutch.org) or H.P.K. (hkiem@fredhutch.org)

Short Title: Characterization and purification of CD34 subsets for HSC gene therapy

ABSTRACT

Hematopoietic stem cell (HSC) gene therapy has the potential to cure many genetic, malignant and infectious diseases. We have shown in a nonhuman primate HSC gene therapy and transplantation model that the $CD34^+CD90^+CD45RA^-$ cell fraction was exclusively responsible for multilineage engraftment and hematopoietic reconstitution. Here we establish the translational potential of this HSC-enriched CD34 subset for lentivirus-mediated gene therapy.

Current alternative HSC-enrichment strategies purify $CD133^+$ cells or $CD38^{low/-}$ subsets of $CD34^+$ cells from human blood products. We directly compared these strategies to isolation of $CD90^+$ cells using a GMP-grade flow-sorting protocol with clinical applicability. We show that $CD90^+$ cell selection results in 40-fold fewer target cells in comparison to $CD133^+CD34^+$ or $CD38^{low/-}CD34^+$ subsets without compromising the engraftment potential *in vivo*. Single cell RNA sequencing confirmed nearly complete depletion of lineage committed progenitor cells in $CD90^+$ fractions compared to alternative selections. Importantly, lentiviral transduction efficiency in purified $CD90^+$ cells resulted in up to 3-fold higher levels of engrafted gene modified blood cells.

These studies should have important implications in manufacturing and clinical outcome, ultimately improving the safety and feasibility of patient-specific HSC gene therapy.

INTRODUCTION

Hematopoietic stem cell (HSC)-mediated gene therapy or editing for the treatment of hematologic disease is commonly performed with single-marker purified $CD34^+$ cells. However, it is well-established that $CD34^+$ cells are a heterogeneous mix of predominantly lineage-committed progenitor cells containing only very few true HSCs with long-term multilineage engraftment potential (1, 2). Moreover, the proportion of true HSCs is variable across individuals, requiring at least 2 x 10^6 viable $CD34^+$ cells per kg of body weight to meet safety and feasibility standards. This results in excessive use of costly reagents (e.g. lentiviral vectors or nucleases) for gene transfer to reliably target HSCs (3-9). Moreover, the efficiency of gene-modification seen *ex vivo* does not correlate with the *in vivo* engraftment of modified cells commonly stabilizing at a significantly lower frequency (4, 10). These major hurdles significantly limit the broad availability of this very promising technology as a potential cure for many patients.

HSC gene therapy and transplantation would benefit from the ability to isolate, target, and modify a more HSC-enriched subset that provides short-term reconstitution as well as long-term multilineage engraftment. Availability of a refined target would overcome all currently existing limitations simultaneously: 1) greatly reduce the amount of modifying reagents needed for manufacturing, 2) result in more reliable genetic-modification of HSCs, and 3) increase the predictability of transplant success *in vivo*.

Within the last four decades, a plethora of cell surface markers have been described for the identification and purification of human HSCs (1, 11-15). Similarly, a great variety of HSC-enrichment strategies has been proposed for gene therapy approaches. Some of the most promising candidate HSC-enrichment strategies include the purification of the $CD38^{low/-}$ (16, 17), $CD133^+$ (18), or $CD90^+$ (19) CD34 subsets. While each strategy claims to improve individual aspects of gene therapy and editing, a side by side analysis of these phenotypes has not been reported.

Here, we applied up-to-date read-outs (*in silico, in vitro*, and *in vivo*) for a systematic and objective side-by-side comparison of candidate HSC-enrichment strategies with the goal to identify the most refined HSC-enriched target for gene therapy and editing.

RESULTS

scRNA-seq objectively identifies committed and immature HSPCs

Single cell transcriptome analysis has proven useful in determining the heterogeneity of complex cell populations (20). Here, we used single-cell RNA sequencing (scRNA-seq) on the 10X Genomics chromium platform to assess the heterogeneity of $CD34^+$ hematopoietic stem and progenitor cells (HSPCs). To identify transcriptionally distinct CD34 subsets and generate a reference map of human HSPCs, we sequenced $CD34^+$ cells from steady-state BM (ssBM) of two healthy human individuals (**Supplemental Table 1**). The scRNA-seq data was analyzed using two different methods of dimensional reduction: t-distributed stochastic neighbor embedding (tSNE) and principle component analysis (PCA).

The tSNE plot showed multiple areas of cell-aggregated clusters (**Figure 1A**). The overlay of gene expression data onto the tSNE plot identified clusters of cells upregulating genes associated with lymphoid, myeloid, and/or erythroid commitment (**Figure 1B**). Cells with low or absent expression of lineage-specific genes demonstrated upregulation of genes associated with primitive HSPCs. Upregulation of cell cycle associated genes were found primarily in regions with transcriptionally lineage-committed cells.

Next, scRNA-seq data was visualized by PCA to preserve gene expression patterns and avoid the randomness of clustering. PCA analysis resulted in a two-prong hierarchical arrangement of cells (**Figure 1C**). In comparison to the clusters in the tSNE plot, cells upregulating genes for lymphoid, myeloid, and erythroid commitment were found along both arms, whereas more immature progenitor cells were located at the intersection in the apex of the scRNA-seq dataset (**Figure 1D**). Visualization of scRNA-seq data using either tSNE or PCA were highly reproducible among biological replicates with only minor differences in the quantitative representation of transcriptionally-distinct CD34 sub-clusters (**Supplemental Figure 1**).

Manual identification of scRNAseq clusters assessing the expression of representative genes was confirmed using the Seurat un-supervised graph-based clustering tool (21). Seven

transcriptionally distinct groups of single cells were identified within $CD34^+$ cell populations (**Supplemental Table 2**). Clusters were mapped onto both the tSNE (**Figure 1E and Supplemental Figure 2A**) and PCA (**Figure 1F and Supplemental Figure 2B**) visualizations. Almost identical to our manual assessment, genes associated with primitive HSPCs were predominantly expressed in cluster 1. Increased expression of genes for lymphoid, myeloid, and erythroid differentiation were observed in clusters 2 and 3, clusters 4 and 5, and clusters 6 and 7, respectively.

In summary, scRNA-seq allows for a reliable and reproducible assessment of the transcriptional heterogeneity within $CD34^+$ HSPCs from ssBM. Most importantly, visualization of single-cell data using either tSNE or PCA are interchangeable and provide a reliable reference map.

$CD90^+CD45RA^-$ cells are the most refined target for HSC gene therapy

Multilineage long-term engraftment and BM reconstitution is driven by $CD34^+CD90^+$ HSPCs in the pre-clinical nonhuman primate (NHP) stem cell transplantation model (19) or human $CD34^+CD38^{low/-}$ cells in the mouse xenograft model (22, 23). To evaluate the translational potential of the $CD34^+CD90^+$ phenotype, we compared this subset to alternative HSC-enrichment strategies including $CD34^+CD133^+$ (18) and $CD34^+CD38^{low/-}$ (16, 17) HSPCs.

We first analyzed the phenotype of sorted subpopulations for contaminating cell types. The $CD133^+$ subpopulation showed approximately 50% enrichment of $CD38^{low/-}$ but only 27% of the $CD90^+$ subset (**Figure 2A and Supplemental Figure 3**). Purified $CD38^{low/-}$ cells were markedly enriched for $CD133^+$ (78%) and $CD90^+$ (49%) cells but still contained a significant proportion of $CD90^-$HSPCs (51%). $CD90^+$ cells demonstrated more than 86% enrichment of $CD133^+$ as well as $CD38^{low/-}$ HSPCs. Across nine different healthy donors, enrichment of the $CD133^+$ subset would be predicted to reduce the target cell count for HSC gene therapy by 58%

(2.4-fold), purification of CD38$^{low/-}$ cells by approximately 83% (5.8-fold), and isolation of CD90^{+} HSPCs in average by about 92% (12.5-fold) (**Figure 2B, C**).

Next, we looked at the gene-expression of CD34, CD133, CD38, and CD90 in the reference map (**Figure 2D and Supplemental Figure 4A**). CD34 expression was detectable in all clusters, with a gradual increase towards the area at the top of the plot associated with more primitive HSPCs. In good agreement to previous publications (1, 24-27), CD133 mRNA expression was restricted to transcriptionally committed lympho-myeloid progenitors as well as immature HSPCs (clusters 1-5). CD38 expression was scattered throughout the reference map and showed no obvious association or absence in distinct clusters. Surprisingly, CD90 expression was undetectable in both donors.

Since sole assessment of mRNA expression was insufficient for an objective comparison of phenotypical CD34 subsets, we sort-purified CD133^{+}, CD38$^{low/-}$, and CD90^{+} cell fractions (**Supplemental Figure 3**), performed scRNA-seq, and overlaid the resulting data onto the CD34 PCA plot (**Figure 2E**). Similar to the observed expression pattern, sort-purified CD133 HSCPs exclusively colocalized with lympho-myeloid and more primitive HSPCs. Sort-purified CD38$^{low/-}$ cells predominantly overlaid with cluster 1 of the reference map, with individual cells spreading into both the lympho-myeloid as well as the erythroid branch. Surprisingly, mapping of sort-purified CD90^{+} cells was almost indistinguishable from CD38$^{low/-}$ cells (**Figure 2E and Supplemental Figure 4B**). Similar to CD38$^{low/-}$ HSPCs, CD90 sorted cells demonstrated a significant enrichment of immature HSPCs (cluster 1), as well as depletion of lympho-myeloid as well as erythroid primed progenitors (**Figure 4B, Supplemental Figure 4B**).

In summary, purification of CD90^{+} HSPCs yielded the greatest reduction of target cells for HSC gene therapy in comparison to CD133^{+} or CD38$^{low/-}$ subsets. While the phenotypical comparison of these HSC enrichment strategies demonstrated reproducible differences, CD38$^{low/-}$- and CD90^{+}-sorted subpopulations were transcriptionally indistinguishable by scRNAseq.

CD90$^+$ HSPCs are transcriptionally unique

Objective comparison of CD38$^{low/-}$ and CD90$^+$ HSPC subsets by scRNA-seq is significantly diminished by the mRNA capture efficiency (28-31). To increase the sensitivity for subtle transcriptional difference in between the CD90$^+$ and CD38$^{low/-}$ phenotypes, we performed bulk-RNA-seq on purified CD34 subsets from the ssBM of two healthy donors. Bulk RNA-seq data was plotted onto our CD34 scRNA-seq reference map to determine transcriptional similarities/differences as well as donor to donor variance of sorted subsets.

ssBM CD34$^+$ cells were initially subdivided and sorted based on the following gating strategy. For CD34$^+$CD90$^+$ HSPCs (Population *a*) the cells were gated on CD45RA and CD90 (19) (**Figure 3A and Supplemental Figure 5**). Assessing CD133 expression, CD34$^+$CD90$^-$ HSPCs were further subdivided into CD45RA$^-$CD133$^+$ (Population *b*), CD45RA$^+$CD133$^+$ (Population *c*) and CD45RA$^-$CD133$^{low/-}$ (Population *d*) HSPCs (26, 32, 33). For comparison with classical strategies using CD38, CD34$^+$CD38$^{low/-}$ cells (Population *e*) were subdivided into CD34$^+$CD38$^{low/-}$CD90$^+$ (Population *f*) and CD34$^+$CD38$^{low/-}$CD90$^-$ (Population *g*) subsets (16, 17, 27).

Consistent with the reported functional assessment of these phenotypically-defined CD34 subsets (16, 17, 19, 26, 27, 32, 33), we observed upregulation of lymphoid genes (*JCHAIN*, *DNTT*, *CD2*, *IGLL1*) in Population *c*, erythro-myelo-megakaryocytic genes (*HBB*, *HDC*, *GATA1*, *CD36*) enriched in Population *d*, and expression of more immature marker genes (*AVP*, *HES1*, *DLK1*) in Populations *a*, *e*, *f* and *g* (**Figure 3B, Supplemental Table 3**). Individual myeloid genes (*ELANE*, *MPO*, *CEBPD*) were simultaneously upregulated in Populations *c* and *d*. Interestingly, Population *b* did not show a unique cluster of differentially expressed genes but shared features of genes associated with lympho-myeloid and myeloid-primed as well as immature HSPCs (**Figure 3B, Supplemental Table 3**).

To confirm this manual assessment, we mapped the bulk-RNA-seq data onto our CD34 scRNAseq reference map (**Figure 3C**). Here we found similar distribution to previously defined clusters with minimal variability, likely attributed to differences in donors (**Figure 2F**). Lympho-myeloid primed cells (Population *c*) mapped within clusters 4 and 5 and erythro-myelo-megakaryocytic primed HSPCs (Population *d*) plotted within cluster 7. Population *b* showed greater heterogeneity (distance in between dots) and localized within clusters 2 and 3 of the lympho-myeloid arm. Populations *a*, *e*, *f*, and *g* were closely co-localized within cluster 1 at the top of the reference map. More detailed comparison of Populations *a*, *e*, *f*, and *g* revealed that $CD90^{+}$ HSPCs (Population *a*) demonstrated the lowest donor to donor variability. In contrast, $CD38^{low/-}$ (Population *e*) HSPCs showed the greatest heterogeneity despite nearly identical proportions of $CD38^{low/-}CD90^{+}$ (36/33%) and $CD38^{low/-}CD90^{-}$ (64/67%) subsets in both donors. $CD38^{low/-}CD90^{-}$ (Population *g*) subsets were similarly heterogeneous and shifted towards the erythroid-primed clusters matching the reported enrichment of erythro-myeloid progenitors in $CD133^{low/-}CD38^{-}CD90^{-}$ HSPCs (33, 34). Most importantly, sorting of the $CD38^{low/-}CD90^{+}$ subset (Population *f*) significantly reduced the donor to donor variability and led to closer co-localization with $CD90^{+}$ HSPCs (Population *a*).

In conclusion, bulk RNAseq increases the sensitivity for low-expressed genes and enables the comparison of CD34 subsets with only subtle differences in transcriptomics. Most importantly, bulk RNA-seq highlights the reproducibility of transcriptional patterns within $CD90^{+}$ HSPCs, whereas $CD38^{low/-}$ cells contain a significant proportion of erythro-myeloid primed $CD38^{low/-}CD90^{-}$ progenitor cells increasing the heterogeneity in the transcriptional signature.

GCSF-mobilized $CD90^{-}$ subsets are transcriptionally similar to immature $CD90^{+}$ HSPCs

Granulocyte-colony stimulating factor (GCSF)-mobilized $CD34^{+}$ HSPCs are the most frequently used cell source for HSC gene therapy approaches. To evaluate the conservation of transcriptional patterns in phenotypically-defined CD34 subsets across sources, bulk RNA-seq

data of sort-purified HSPC subsets (Populations *a*-*d*) from two GCSF-mobilized donors was projected onto our ssBM-based reference map.

Similar to ssBM, key regulators for lympho-myeloid and erythroid as well as immature GCSF-mobilized CD34 subsets were upregulated in Populations *c*, *d*, and *a*, respectively (**Supplemental Figure 6A, Supplemental Table 4**). As expected, GCSF-mobilized HSPCs from Population *a* ($CD90^+$) mapped within cluster 1, whereas Population *c* was located in the lympho-myeloid branch within cluster 2 and 3, and Population *d* fell within cluster 6 of the erythro-megakaryocytic clusters. Strikingly different to ssBM, Population *b* co-localized much closer with $CD90^+$ HSPCs (Population *a*) (**Figure 3D**). Pairwise comparison of $CD90^+$ (Population *a*) with $CD90^-$ (Population *b*) HSPCs from both sources confirmed that mobilized subsets demonstrate significantly less differentially expressed genes then ssBM-derived CD34 subpopulations (ssBM: 615 genes; GCSF: 131 genes; **Supplemental Figure 6B**, **Supplemental Tables 5–8**). Despite this low number of differentially expressed genes, expression patterns associated with primitive HSPCs such as *AVP*, *HLF*, and *KLF2* were found in Population *a*, whereas Population *b* showed upregulation of *CD38*, *HGF*, *HDC*, and other genes known for maturation and differentiation.

In summary, GCSF-mobilized CD34 subsets can be mapped onto the ssBM scRNA-seq reference. While bulk RNAseq data of immature as well as committed progenitor cells localized within the expected clusters, $CD34^+CD90^-CD133^+CD45RA^-$ HSPCs from GCSF-mobilized donors were transcriptionally almost identically to the HSC-enriched $CD34^+CD90^+$ subset.

BM reconstitution potential is restricted to $CD34^+CD90^+$ cell fractions

Pre-clinical experiments in the NHP and mouse showed that GCSF-primed $CD34^+CD90^+$ HSPCs were solely responsible for reconstitution of the BM stem cell compartment (19, 35). To validate the engraftment potential of this HSC-enriched phenotype in humans, $CD34^+$ cells from healthy human GCSF-mobilized donors were sort-purified into Populations *a*-*d* and transplanted into sublethally irradiated adult NSG mice.

The highest engraftment in the PB, BM, and thymus was observed in cohorts transplanted with $CD90^+$ cells (**Figure 4A,B, Supplemental Figure 7A-D, and Supplemental Figure 8A,B**). Lower level and less consistent multilineage engraftment of human cells was observed upon transplantation of Population *b*. Mice receiving Population *c* showed locally restricted human chimerism in the thymus, whereas Population *d* did not show any human engraftment in the analyzed tissues. Engraftment and reconstitution of the entire BM stem cell compartment including the recovery of phenotypically primitive human $CD34^+CD90^+$ HSPCs was exclusively observed after transplantation of $CD90^+$ cells (**Figure 4C, Supplemental Figure 7E,F, and Supplemental Figure 8A**). Similarly, erythroid, myeloid, and erythro-myeloid colony-forming cell (CFC) potentials were only detected in mice transplanted with $CD90^+$ cells (**Figure 4D, Supplemental Figure 7D**).

To confirm that human HSPCs from Population *b* do not contain primitive HSPCs with multilineage long-term engraftment potential, we performed limiting dilution experiments (**Supplemental Figure 9**). Gradually increasing the number of transplanted cells from Population *b* led to greater multilineage engraftment of human cells in all tissues including $CD34^+$ cells in the BM stem cell compartment (**Supplemental Figure 9A-E**). However, none of the mice demonstrated human $CD34^+CD90^+$ HSPCs *in vivo* after transplant with this population, and engrafted $CD34^+$ cells were restricted to erythroid and myeloid colony types lacking mixed CFU potentials (**Supplemental Figure 9F**). The number of SRCs (SCID-repopulating cells) in Population *b* was calculated to be 1 in 4.6×10^5 transplanted cells (**Supplemental Figure 9G**).

In summary, mouse xenograft experiments confirm enrichment of primitive human HSPCs with multilineage engraftment and BM reconstitution potential in the $CD34^+CD90^+$ phenotype (Population *a*). Furthermore, $CD34^+CD90^-CD133^+CD45RA^-$ HSPCs (Population *b*) contain low-level multilineage engraftment potential but lack the ability to recover the entire stem cell compartment.

Small molecules slightly enhance gene-modification of the CD90 subset

A variety of transduction enhancers have been reported for the use on human CD34$^+$ HSPCs to boost the efficiency of gene-modification in long-term engrafting HSCs (36-38). The majority of studies were performed with umbilical cord blood (UCB)-derived CD34$^+$ HSPCs, a less frequently used cell source for HSC gene therapy. Here, we tested some of the recently reported small molecules on GCSF-mobilized CD34$^+$ cells and specifically determined the frequency of GFP-expression in the HSC-enriched CD34$^+$CD90$^+$ subset characterized above (**Figure 5A**).

Bulk CD34$^+$ cells were purified on the CliniMACS Prodigy (day 1), stimulated in HSPC culture medium overnight, and gene-modified (MOI5x1) at day 2 in either regular culture medium without additional additives (Tsd Control), on Retronectin-coated plastic (Retronectin) or in the presence of PGE2, UM171, and a combination of SR1 (36, 39), UM171 (37), as well as Ly2228820 (38, 40, 41) (SULy). Virus and additives were removed after 2 hours (day 2), cells were cultured in HSPC medium for 3 days to allow GFP expression, and the phenotypical composition of cells then assessed flow-cytometrically on day 5. Short-term exposure of bulk CD34$^+$ cells with the small molecules did show no impact on the average frequency of HSC-enriched CD34$^+$CD90$^+$ cells in comparison to the non-transduced control (NT Control) (**Figure 5B,C and Supplemental Table 9**). Similarly, no obvious effect of the small molecules could be seen on the transduction efficiency in bulk CD34$^+$ cells. However, presence of UM171 and SULy did promote a slight increase in the average expression of GFP within HSC-enriched CD34$^+$CD90$^+$ HSPCs (**Figure 5C and Supplemental Table 9**).

To assess the impact of the gene-modification and culture conditions on the functional properties, bulk CD34$^+$ cells as well as HSC-enriched CD34$^+$CD90$^+$ HSPCs were introduced into CFC assays on day 3 (**Figure 5D and Supplemental Table 9**). Unexpectedly, a significant increase in the total CFC potential was seen across all gene-modified bulk CD34$^+$ cells in comparison to non-transduced cells. However, no impact on the overall colony-forming ability and

composition was seen comparing the individual conditions or specifically analyzing the $CD34^+CD90^+$ subset.

In summary, short-term exposure of adult GCSF-mobilized $CD34^+$ cells with PGE2, UM171, and the combination of multiple small molecules (SULy) does slightly enhance the transduction efficiency within $CD34^+CD90^+$ cells without impacting the functional properties. However, the reported improvement of transduction on UCB-derived $CD34^+$ cells could not be fully reproduced in GCSF-mobilized HSPCs.

Sort-purification increases the transduction efficiency in HSC-enriched $CD90^+$ cells

While transduction enhancers are proposed to increase the efficiency of lentivirus-mediated gene transfer in $CD34^+$ cells (36-38), these represent additional reagents which add to cost, complexity and preclinical safety determinations needed for gene therapy, while still requiring large amounts of expensive clinical-grade viral vector. To overcome this limitation, we next evaluated the feasibility of a flow-cytometry based sort-purification strategy for the large-scale enrichment and direct targeting of $CD90^+$ cells.

$CD34^+$ cells from six healthy GCSF-mobilized donors were enriched on the Miltenyi CliniMACS Prodigy according to our previously established protocol (**Figure 6A, Supplemental Table 10**) (42). Leukapheresis products yielded on average 2.56×10^8 $CD34^+$ cells with a purity greater than 93.5±1.9%. Donor #4 insufficiently mobilized $CD34^+$ cells and was excluded from the study.

Sort-purification of $CD90^+$ HSPCs was performed comparing the jet-in-air sorter FX500 from Sony with the cartridge-based closed-system sorter MACSQuant Tyto from Miltenyi Biotech (**Figure 6A**). Sorted $CD90^+$ cell fractions on both machines showed no significant phenotypical or quantitative differences. They both reached an average of 79.5±6.9% (Sony) and 77.1±5.5% (Tyto) purity with a yield of 52.3±9.6% and 49.9±11.1%, respectively (**Figure 6B,C**). Of special interest for HSC gene therapy, the average fold-enrichment of target cells was greater than

29.7±5.7-fold on the Sony sorter and up to 37.2±7.3-fold on the Tyto (**Figure 6C**). To determine whether the sort-purification impacted the differentiation potential of HSC-enriched CD34 subsets, cells were introduced into CFC assays. Sort-purified cell fractions did not demonstrate any differences in the total nor compositional CFC potential compared to unprocessed cells (**Figure 6D**).

Next, sorted $CD90^+$ as well as bulk $CD34^+$ cells were transduced with a lentivirus encoding for GFP according to our clinically approved CD34-mediated gene therapy protocol (43, 44). Five days post-transduction, the gene-modification efficiency in the $CD90^+$ subset of bulk-transduced CD34 cells reached 16.5±3.4% (SEM), whereas 2.3-fold (Tyto: 38.1±5.5%, SEM, p=0.012) and 3.1-fold (Sony: 51.9±5.8%, SEM, p=0.006) higher frequencies of GFP were seen in the sort-purified $CD90^+$ cell fractions (**Figure 6E**). Despite significant differences in the gene-modification efficiency, no significant increase in the overall MFI of GFP was observed transducing purified $CD90^+$ cells (**Figure 6E**). Gene-modified cells were introduced into CFC assay showing no significant quantitative and qualitative differences (**Figure 6F**). Finally, individual colonies were extracted to determine the gene-modification efficiency within erythro-myeloid colonies as well as to precisely quantify the vector copy number (VCN). No obvious differences were found in the expression of GFP within CFCs from all three conditions (**Figure 6G**). Similar to the MFI of GFP, almost identical VCNs were observed in all setting (**Figure 6H**).

In summary, the sort-purification of $CD90^+$ cells is technically feasible and does not impact the cells phenotypical and functional properties. Most importantly, purification of $CD90^+$ cells reduces the number of target cells and simultaneously improves the efficiency of lentivirus-mediated gene-transfer without increasing the MFI as well as VCN and without the need for additional transduction enhancers.

Improved in vivo engraftment of sort-purified and gene-modified CD90$^+$ HSPCs

Regardless of high gene-modification efficiency *ex vivo*, the frequency of long-term *in vivo* engraftment of modified cells is often both unpredictable and, in many studies, significantly lower compared to values determined during quality control of the infusion product (10, 45, 46). To compare the multilineage long-term engraftment potential of gene-modified bulk CD34$^+$ with sort-purified CD90$^+$ HSPCs, cells were transplanted into sublethally irradiated adult NSG mice. Of note, in order to mimic the average percentage of CD90$^+$ cells within bulk CD34$^+$ cell fractions (see **Figure 2B**) mice transplanted with sort-purified CD90$^+$ cells received 1/10th the cell number compared to mice transplanted with bulk CD34$^+$ cells.

Multilineage engraftment of human cells in the PB of transplanted mice was followed longitudinally by flow-cytometry (**Figure 7A**, **Supplemental Figure 10A**). Mice transplanted with bulk CD34$^+$ as well as Tyto-sorted CD90$^+$ cells showed persisting levels of human engraftment, whereas human chimerism in the vast majority of mice receiving Sony-sorted CD90$^+$ HSPCs gradually declined towards the end of study. Similar trends in the overall frequency of human multilineage engraftment were seen at the day of necropsy in the BM, spleen, and thymus. The greatest frequency of human chimerism was found in mice transplanted with bulk CD34$^+$ and Tyto-sorted CD90$^+$ HSPCs followed by Sony-sorted CD90$^+$ cells (**Figure 7B**, **Supplemental Figure 10B-D**).

Lentiviral-mediated delivery of GFP enabled us to flow-cytometrically follow the frequency of gene-modification longitudinally in the PB as well as at the day of necropsy in tissues. Consistent with the efficiency of gene-modification we saw *ex vivo* (**Figure 6E**), a 2- to 3-fold higher frequency of GFP$^+$ human cells was seen in the PB and tissues of animals receiving sort-purified CD90$^+$ cells compared to bulk CD34$^+$ cells (**Figure 7C,D**). The frequency of gene-modification in the PB of animals transplanted with sort-purified CD90$^+$ cells remained consistent throughout the follow-up, whereas in animals receiving gene-modified bulk CD34$^+$ cells two patterns were observed; human cells 1) either gradually lost the gene-modification leveling out at

about 10-15% or 2) started to show a rapid increase in GFP^+ human cells up to almost 80% towards the end of study.

Next, we comprehensively analyzed the BM stem cell compartment of engrafted mice. Mice transplanted with bulk $CD34^+$ and Tyto-sorted $CD90^+$ cells demonstrated significantly greater engraftment of human $CD34^+$ and $CD34^+CD90^+$ cells compared to Sony-sorted $CD90^+$ HSPCs (**Figure 7E**). Most importantly, the frequency of gene-modified $CD90^+$ cells was in average 2- to 3-fold higher in mice engrafted with sort-purified and gene-modified $CD90^+$ compared to bulk $CD34^+$ HSPCs. Finally, we introduced BM-resident human $CD34^+$ into CFC assays to test the maintenance of erythro-myeloid differentiation potentials (**Figure 7F**). After transplantation of either bulk $CD34^+$ or Sony-sorted $CD90^+$ in only two mice, each phenotypical $CD34^+$ cell gave rise to colonies, whereas in 8 out of 9 cases erythro-myeloid colonies were seen in mice receiving Tyto-sorted $CD90^+$ cells.

In summary, gene-modification of sort-purified $CD34^+CD90^+$ cells on the Tyto does not alter the multilineage long-term engraftment potential of human HSPCs in the mouse xenograft model. Most importantly, gene-editing efficiencies observed *ex vivo* successfully translated into the *in vivo* transplant and remained stable throughout the long-term follow-up.

DISCUSSION

Here, we show that the enrichment of the human CD34$^+$CD90$^+$ population has the potential to overcome current limitations of HSC gene therapy. Sort-purification of this HSC-enriched CD34 subset resulted in up to 40-fold reduction of target cells and consequently significant savings of costly reagents (here lentiviral vectors). Most importantly, sole purification of this phenotype improved the gene-modification efficiency of HSCs with long-term *in vivo* engraftment potential up to 3-fold perfectly reflecting the *ex vivo* assessment, enhancing the predictability of HSC gene therapy, and improving the long-term stability of modified cells at high levels. Thus, CD34$^+$CD90$^+$ HSCs demonstrate the most refined target population for HSC gene therapy and transplantation.

In order to systematic and objectively compare human candidate HSC-enriched target cell populations for HSC gene therapy, we here combined up-to-date transcriptional, phenotypical, and functional read-outs. We initially purified previously reported HSC-enriched cell populations as well as our recently defined CD34$^+$CD90$^+$ subset and performed high-throughput scRNA-seq analysis. The phenotypical heterogeneity, the relationship in between subsets, as well as the quantitative reduction of target cells for each CD34 subset was evaluated using multi-color flow-cytometry. In addition, comprehensive mouse xenograft assays were performed to evaluate the multilineage engraftment and BM reconstitution potential of human CD34$^+$ subsets. In this comprehensive side-by-side comparison, we observed the greatest reduction of target cells in the CD34$^+$CD90$^+$ cell fraction without compromising the multilineage engraftment and BM reconstitution potential in mouse xenograft transplants. Furthermore, scRNA-seq data showed that CD34$^+$CD90$^+$ cells are almost entirely depleted for lineage-committed progenitor cells, while phenotypically, transcriptionally, and functionally primitive HSPCs are markedly enriched.

After this comprehensive comparison, bulk CD34$^+$ cells from GCSF-mobilized PBSCs, the most common cell source for HSC gene therapy, were gene-modified in the presence of previously reported small molecule transduction enhancers to assess the gene-modification in the transcriptionally, phenotypically and functionally most primitive HSC-enriched CD34$^+$CD90$^+$

subset. In contrast to previous literature describing significantly enhanced transduction efficiency in umbilical cord blood (UCB)-derived CD34 HSPCs, only minor improvement of gene-modification could be observed in the $CD34^+CD90^+$ cell fraction from GCSF-mobilized PBSCs.

Finally, we compared two good manufacturing practice (GMP)-compatible flow-cytometry assisted cell sorting (FACS) platforms for the sort-purification and gene-modification of HSC-enriched CD34 subsets. Purity and yield of sorting, gene-modification efficiency, as well as maintenance of stem cell features was confirmed by flow-cytometry with *in vitro* colony-forming cell assays and mouse xenograft transplants. We demonstrate that direct transduction of sort-purified $CD34^+CD90^+$ cells significantly reduced gene-modifying reagents and, at the same time, enhanced the gene-modification efficiency without compromising the long-term multilineage engraftment potential.

These findings are all in line with our recently performed competitive reconstitution studies in the NHP to identify a refined target cell population for HSC gene therapy and transplantation (19). In our previous study, we reported the $CD34^+CD90^+$ phenotype to be exclusively responsible for both rapid short-term as well as robust long-term hematopoietic reconstitution. Here we show that human multilineage engraftment potential with full recovery of the BM stem cell compartment is similarly restricted to $CD34^+CD90^+$ HSPCs, whereas $CD34^+$ cells that lost CD90 expression lack full multilineage xenograft reconstitution potential. This finding mirrors previous publications describing the enrichment of human HSCs within CD90-expressing CD34 subsets. The first CD90-mediated isolation of human HSCs from fetal BM, adult BM, and cytokine-mobilized stem cells was reported in the 1990s (11, 47). Thereafter, $lin^-CD34^+CD90^+$ subsets were shown to be enriched for long-term culture-initiating cells (LTC-ICs) with thymus engraftment potential in the SCID mouse model (11, 47). However, alternative CD90/Thy-1 antibody clones (F15 421-5 and GM201) were used at this time that recognized a significantly higher expressed CD90 variant on up to 60% of cytokine-mobilized $CD34^+$ cells (11). Refining the phenotype with additional cell surface markers, switching to the currently used CD90 clone 5E10, and using an improved mouse

model, Majeti *et al*. later confirmed the enrichment of human HSCs in $CD34^+CD38^-CD90^+CD45RA^-$ cell fractions when using UCB- and BM-derived CD34 subsets (27). With the highest level of purification, Notta *et al*. demonstrated engraftment and reconstitution potential of single human $CD34^+CD38^-CD90^{+/-}CD45RA^-CD49f^+$ cells from UCB after intrafemoral injection into NSG mice (1). Despite significant differences in the phenotype, cell source, level of HSC purification, mouse model (NOD/SCID or NSG), and mode of transplantation (intravenous, intrafemoral), all groups including those in our study associate CD90 expression with human HSCs, whereas HSPCs lacking CD90 demonstrate limited engraftment potential and do not re-express or recover the HSC-enriched $CD90^+$ subset. Adding to this, we re-validated findings from the 1990s that employed the commonly used CD90 antibody clone 5E10, which showed that $CD34^+CD90^+$ HSPCs from GCSF-mobilized leukapheresis products, the most commonly used stem cell source for HSC gene therapy and transplantation, exclusively contain primitive HSCs with multilineage engraftment and BM reconstitution potential.

Recent studies aiming to refine the target for HSC gene therapy and transplantation combined flow-cytometry with functional *in vitro* and/or *in vivo* read-outs (16, 17). Here, we additionally applied single-cell as well as bulk RNA sequencing to comprehensively compare the $CD34^+CD90^+$ phenotype with alternative HSC-enrichment strategies. In contrast to most approaches performing transcriptomics (48-50), we initially performed scRNA-seq on $CD34^+$ cells to build a detailed and comprehensive reference map rather than using historically defined human CD34 subsets for comparison. This strategy allowed us to define transcriptionally distinct clusters and perform an unbiased assessment of the expression profile in these clusters. Furthermore, we were able to determine the reproducibility of scRNAseq for $CD34^+$ HSPCs and ultimately establish a baseline for the comparison of phenotypically defined HSC-enriched cell fractions. Having this baseline further enabled us to objectively compare scRNA-seq data from different donors as well as reliably map sort-purified CD34 subsets. Most importantly, this strategy provided the foundation to overlay scRNA-seq with bulk RNAseq data from multiple donors without any

computational correction or manipulation of the dataset for donor-, batch-, source- or preparation-dependent variability.

Interestingly, we were not able to reliably determine transcriptional differences between $CD34^+CD90^+$ and $CD34^+CD38^{low/-}$ HSPCs by scRNA-seq. Using the 10x scRNA-seq V2 platform, we were unable to determine transcriptionally-distinct and meaningful clusters or define more detailed hierarchical structures within either subset. Similar to previous reports on $Lin^-CD34^+CD38^{low/-}$ cells describing a Continuum of LOw-primed UnDifferentiated (CLOUD) cells without hierarchical structures (48-51), $CD34^+CD90^+$ and $CD34^+CD38^{low/-}$ HSPCs were indistinguishable even though $CD38^{low/-}$ cells were displaying clear differences in their phenotypical composition. Platform-dependent limitations in the mRNA capturing efficiency (31), technical noise (52), and low-level expression of key genes currently limit the ability to reliably distinguish differences in RNA expression (49) required to identify cell subsets with short- and long-term engraftment potential within the $CD34^+CD90^+$ phenotype.

Despite the current limitation in scRNA-seq, our approach clearly shows that purified $CD34^+CD90^+$ as well as $CD34^+CD38^{low/-}$ HSPCs are similarly depleted for transcriptionally lineage-committed progenitor cells. However, combining phenotypical, quantitative, and transcriptional data, the $CD34^+CD90^+$ subset was determined to have the greatest overall reduction of target cells. With an average 12.5-fold reduction in total cells compared to a 5.8-fold for the $CD34^+CD38^{low/-}$ subset, isolation and targeting of $CD34^+CD90^+$ HSPCs significantly increases the feasibility of currently existing HSC gene therapy approaches. Most importantly, transplantation of purified $CD34^+CD90^+$ HSPCs did not compromise the multilineage engraftment potential and BM reconstitution capacity of CD34 subsets from GCSF-mobilized leukapheresis products.

Particularly important for the application of HSC gene therapy as a routine treatment option, we have successfully demonstrated that the large-scale sort-purification of our recently identified HSC-enriched $CD90^+$ subset is technically feasible and significantly reduces the target

cell number as well as costs for HSC gene therapy. In comparison to small molecule mediated approaches where bulk $CD34^+$ cells are used, only minor improvement in the transduction efficiency of $CD34^+CD90^+$ cells could be seen and therefor no reduction of modifying agents and improvement of feasibility would be achieved. $CD34^+CD90^+$ cells can be efficiently purified using either the Sony SH800S clean-room dependent droplet-cell sorter or the fully-closed, portable MACSQuant Tyto from Miltenyi without any significant differences in the purity and yield of cells. However, more comprehensive follow-up studies will be required to further investigate the biological impact of both cell-sorting strategies on the multilineage engraftment potential as well as the efficiency of gene-modification. Specific analysis of stress-mediated responses as well as activation of cell cycle pathways may help to explain the observed differences in the frequency of gene-modification and long-term engraftment seen comparing both systems.

Most surprisingly, we found that the sort-purification $CD34^+CD90^+$ cells significantly increased the transduction efficiency of HSCs with multilineage long-term engraftment potential compared to gold standard $CD34^+$ cells. In particular, the observation that the gene-modification efficiencies determined in culture directly translate into the frequency of gene-modification in vivo will help to increase consistency of quality in stem cell products and ultimately make HSC gene therapy applications more predictable. These findings should have important implications for currently available as well as future HSC gene therapy and gene editing protocols hence only minimal changes in procedures are require for the purification of $CD34^+CD90^+$ cells. Isolation of this HSC-enriched phenotype will allow more targeted gene modification, allow a more controlled vector copy number without the application of small molecule transduction enhancer, and thus likely reduce unwanted off target effects.

Application of this HSC-enriched subset will require the implementation of additional processing steps into existing clinical protocols. While the sort-purification will generate further expenses, costs performing these steps will be easily compensated by the 30- to 40-fold savings in expensive modifying reagents. Gene therapy vectors are currently estimated to be the biggest

hurdle in terms of large-scale production and by far the dominating cost factor limiting the routine application of HSC gene therapy (7-9). Our HSC-targeted gene therapy strategy will help to overcome both bottlenecks at the same time: a single large-scale virus production can be used for up to 30 patients preventing the shortage of GMP-grade vectors; at the same time, modifying reagents will no longer be the major factor for the price determination in HSC gene therapy. Availability of new sorting technology for GMP-grade closed-system cell sorting of human HSCs (53) in combination with our HSC-enriched target population will bring HSC gene therapy a significant step closer towards a feasible routing application.

In summary, this study describes a human HSC-enriched cell population with unique phenotypical, functional, and transcriptional features. Isolation of these HSC-enriched $CD34^{+}CD90^{+}$ HSPCs has the potential to improve the targeting efficiency of current clinical HSC gene therapy and editing applications, and also reduce toxicity from potential off-target modifications. Conservation of this HSC-enriched phenotype among different human stem cell sources and in the pre-clinical NHP model (19) further highlights its biological relevance. Species-independent similarities of engraftment patterns in mouse xenograft (human) and autologous NHP transplants (19, 27), as well as robust long-term multilineage engraftment with more than two years of follow-up in the NHP (19) are promising indicators for a successful clinical translation of this HSC-enrichment strategy. Finally, this stem cell selection strategy may also allow efficient and effective depletion of donor T cells in the setting of allogeneic stem cell or organ transplantation.

MATERIALS AND METHODS

Cell sources and CD34$^+$ enrichment. Fresh, whole BM in sodium heparin for single cell RNA sequencing was purchased from StemExpress (Folsom, CA). GCSF-mobilized leukapheresis collections were purchased from the Co-operative Center for Excellence in Hematology (CCEH) at the Fred Hutchinson Cancer Research Center. All human samples were obtained after informed consent according to the Declaration of Helsinki stating that the protocols used in this study have been approved by a local ethics committee/institutional review board of the Fred Hutchinson Cancer Research Center. Human CD34$^+$ cells from steady-state BM and GCSF-mobilized leukapheresis collections were harvested and enriched as previously described (19, 42). Enrichment of CD34$^+$ cell fractions was performed according to the manufacturer's instructions (Miltenyi Biotech, Bergisch Gladbach, Germany).

Flow cytometry analysis and FACS. Fluorochrome-conjugated antibodies used for flow cytometric analysis and FACS (fluorescence activated cell sorting) of human cells are listed in **Supplemental Table 11**. Dead cells and debris were excluded via FSC/SSC gating. Flow cytometric analysis was performed on an LSR IIu (BD, Franklin Lakes, NJ), Fortessa X50 (BD), and FACSAria IIu (BD). Cells for scRNA-seq, bulk RNA-seq, and *in vitro* assays were sorted using a FACSAria IIu cell sorter (BD). Large-scale clinical-grade purification of CD34$^+$CD90$^+$ cells was performed on the FX500 (Sony Biotechnology, San Jose, CA) and the MACSQuant Tyto (Miltenyi Biotech). Post-sort purity was assessed on the FACSAria IIu reanalyzing at least 500 cells for each sample. Data was acquired using FACSDiva^{TM} Version 6.1.3 and newer (BD). Data analysis was performed using FlowJo Version 8 and higher (BD).

RNA isolation for bulk RNA-seq. Total RNA from sort-purified CD34 subsets of GCSF-mobilized leukapheresis products was extracted with the Arcturus PicoPure RNA Isolation Kit (Thermo Fisher Scientific, Waltham, MA) according to the manufacturer's protocol. Total RNA

from sort-purified CD34 subsets of steady-state BM was extracted with the RNeasy Micro Kit (Qiagen, Hilden, Germany) according to the manufacturer's protocol. Detailed methods on the data analysis can be found in the Supplemental Methods.

RNA quality control for bulk RNA-seq. Total RNA integrity was analyzed using an Agilent 2200 TapeStation (Agilent Technologies, Inc., Santa Clara, CA) and quantified using a Trinean DropSense96 spectrophotometer (Caliper Life Sciences, Hopkinton, MA).

Single cell RNA sequencing. Steady-state BM-derived $CD34^+$ cells and CD34-subsets for single-cell RNA sequencing were sort-purified and processed using the Chromium Single Cell 3' (v2) platform from 10X Genomics (Pleasanton, CA). Separation of single cells, RNA extraction, and library preparation were performed in accordance with the 10X Chromium Single Cell Gene Expression Solution protocol. Detailed methods on the data analysis can be found in the Supplemental Methods.

Next-generation sequencing. Sequencing of GCSF-mobilized and steady-state BM bulk RNA-seq samples was performed using Illumina (San Diego, CA) HiSeq 2500 in rapid mode employing a paired-end, 50-base read length (PE50) sequencing strategy. Sequencing of single cell RNA-seq samples was performed using an Illumina HiSeq 2500 in rapid mode employing 26 base read length for read 1 (10x barcode and 10bp Unique Molecular Index (UMI)) and 98 base read length for read 2 (cDNA sequence). Image analysis and base calling was performed using Illumina's Real Time Analysis (v1.18) software, followed by 'demultiplexing' of indexed reads and generation of fastq files, using Illumina's bcl2fastq Conversion Software (v1.8.4).

Lentiviral vectors. The vector used in this study (pRSCSFFV.P140K.PGK.eGFP-sW) is a SIN LV vector produced with a third-generation split packaging system and pseudo-typed by the

vesicular stomatitis virus G protein (VSV.G). The vector for these studies was produced by our institutional Vector Production Core (PI H.-P.K.). Infectious titer was determined by flow cytometry evaluating eGFP protein expression following titrated transduction of HT1080 human fibrosarcoma-derived cells with research-grade LV vector preparations.

Small molecule transduction enhancer experiments. $CD34^+$ were thawed and cultured overnight in culture media (StemSpan ACF [Stemcell Technologies, Vancouver, Canada] with 1% Pen/Strep [Thermo Fisher Scientific] as well as SCF, TPO, and FLT-3 at 100ug/mL [all three PeproTech, Rocky Hill, NJ, USA]). The next day, cells were plated at a density of 1e6 per mL in the following conditions: 1) Non-transduced control, 2) Transduction control, 3) Retronectin-coated transduction control [Takara Bio Inc., Kusatsu, Shiga Prefecture, Japan], 4) PGE2 [10uM] [Cayman Chemical Company, Ann Arbor, MI, USA], 5) UM171 [40nM] [Stemcell Technologies], or 6) SULy [SR1 at 40nM [Stemcell Technologies], UM171 [Stemcell Technologies] at 750nM, and Ly [MedChemExpress LLC, Monmouth Junction, NJ, USA] at 100nM. Conditions 2 through 6 contained 4ug/mL protamine sulfate [Sigma-Aldrich, Saint Louis, MO, USA] and were transduced with lentiviral vector that encodes for GFP at an MOI of 5 for two hours. Virus was removed and cells kept in culture medium at 37°C, 5% CO_2.

Colony-forming cell (CFC) assay. For CFC assays, 1,000-1,200 sort-purified CD34- subpopulations were seeded into MethoCult H4435 (Stemcell Technologies) or H4230 supplemented with hIL-3, IL-5, G-CSF, SCF, TPO, and GM-SCF [all PeproTech], each at 100ug/mL as well as EPO [Amgen, Thousand Oaks, CA, USA] at 4U/mL for the large-scale clinical-grade experiments according to our established clinical protocols (43, 54). Colonies were scored after 12 to 14 days, discriminating colony forming unit- (CFU-) granulocyte (CFU-G),

macrophage (CFU-M), granulocyte-macrophage (CFU-GM) and burst forming unit-erythrocyte (BFU-E). Colonies consisting of erythroid and myeloid cells were scored as CFU-MIX.

Measurement of transduction efficiency and vector copy number (VCN). To measure the transduction efficiency and vector copy number, at least 80 colonies were picked for each condition. Genomic DNA was isolated by incubating tubes at 95°C for 2 h on a thermal cycler. Crude DNA preparations were then subjected to PCR using LV-specific primers (Fwd: 5' AGAGATGGGTGCGAGAGCGTCA and Rev: 5'-TGCCTTGGTGGGTGCTACTCCTAA [Integrated DNA Technologies; IDT, Coralville, IA, USA]) and, in a separate reaction, actin-specific primers were used (human Fwd: 5' TCCTGTGGCACTCACGAAACT and Rev: 5'-GAAGCATTTGCGGTGGACGAT [IDT]). Colonies containing expected bands for both lentivirus and actin were scored as transduced. Reactions which did not yield actin products were considered non-evaluable.

Vector copy number (VCN) per genome equivalent was determined as previously described (38). Briefly, VCNs were assessed by a multiplex TaqMan 5' nuclease quantitative real-time PCR assay in triplicate reactions. Colony gDNA samples were subjected to a lentivirus-specific primer/probe combination (Fwd, 5'-TGAAAGCGAAAGGGAAACCA; Rev, 5'-CCGTGCGCGCTTCAG; probe, 5'- AGCTCTCTCGACGCAGGACTCGGC [IDT]) as well as an endogenous control (TaqMan Copy Number Reference assay RNaseP, Thermo Fisher Scientific, Pittsburgh, PA, USA) using TaqMan GTXpress Master Mix (Applied Biosystems Foster City, CA). Samples with an average VCN ≥0.5 were considered transduced.

Mouse xenograft transplantation. Adult (8 to 12 weeks) NSG mice (NOD.Cg-*Prkdc*^{scid}*Il2rg*^{tm1Wjl}/SzJ) received a radiation dose of 275 cGy followed 4 hours later by a 200 µL intravenous injection of sort-purified human CD34^{+} cells or CD34-subpopulations. Beginning at 8 weeks post-injection, blood samples were collected every 2 to 4 weeks and analysed by flow

cytometry. After 16 to 20 weeks, animals were sacrificed, and tissues were harvested and analysed. All animal studies were carried out at Fred Hutchinson Cancer Research Center in compliance with the approved Institutional Animal Care and Use Committee (IACUC) protocol #1483.

Statistics. Data analysis of limiting dilution experiments was performed as previously described (55). Statistical analysis of data was performed using GraphPad Prism Version 5. Significance analyses were performed with the unpaired, two-sided Student's t-test (*: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$).

AUTHOR CONTRIBUTIONS

SR and HPK designed the study. SR, YYC and AMP performed RNAseq. DP and ME analyzed RNAseq data. SR, YYC, AMP, and MC performed longitudinal mouse follow-up, necropsy and data analysis. MC performed small molecule transduction experiments. SR, AMP, YYC, MC, SS, AB and TE performed clinical-grade sorting experiments, follow-up experiments, and data analysis. SR, DP and ME generated the figures. HPK and JEA funded the study. SR and HPK wrote the manuscript. All authors reviewed and edited the final manuscript.

COMPETING INTERESTS

H.P.K is a consultant to and has ownership interests with Rocket Pharma and Homology Medicines. H.P.K. is a consultant to CSL Behring and Magenta Therapeutics. S.R., J.E.A. and H.-P.K. are inventors on patent applications (#62/351,761, #62/428,994 and #PCT/US2017/037967) submitted by the Fred Hutchinson Cancer Research Center that covers the selection and use of cell populations for research and therapeutic purposes as well as strategies to assess and/or produce cell populations with predictive engraftment potential. The other authors declare that they have no competing interests.

ACKNOWLEDGEMENTS

Human PB stem cells and steady-state BM were kindly provided by Shelly Heimfeld. We thank Helen Crawford for help in preparing this manuscript and figures.

FUNDING

This work was supported in part by grants to HPK from the Evergreen Fund from the Fred Hutchinson Cancer Research Center, the National Institutes of Health (R01 AI135953-01) and the Immunotherapy Integrated Research Center as well as by funds to JEA from the Fred Hutchinson Cancer Research Center, the Cuyamaca Foundation and from the NHI/NCI Cancer

Center Support Grant P30 CA015704. HPK is a Markey Molecular Medicine Investigator and received support as the inaugural recipient of the José Carreras/E. Donnall Thomas Endowed Chair for Cancer Research and the Fred Hutch Endowed Chair for Cell and Gene Therapy.

DATA AND MATERIALS AVAILABILITY

All original RNAseq data were uploaded to the NCBI database. The BioProject Accession code is available upon request.

REFERENCES

as of 31-Oct-2019

1. Notta F, Doulatov S, Laurenti E, Poeppl A, Jurisica I, and Dick JE. Isolation of single human hematopoietic stem cells capable of long-term multilineage engraftment. *Science.* 2011;333(6039):218-21.
2. Doulatov S, Notta F, Eppert K, Nguyen LT, Ohashi PS, and Dick JE. Revised map of the human progenitor hierarchy shows the origin of macrophages and dendritic cells in early lymphoid development. *Nat Immunol.* 2010;11(7):585-93.
3. Peterson CW, Haworth KG, Burke BP, Polacino P, Norman KK, Adair JE, et al. Multilineage polyclonal engraftment of Cal-1 gene-modified cells and in vivo selection after SHIV infection in a nonhuman primate model of AIDS. *Molecular therapy Methods & clinical development.* 2016;3:16007.
4. Peterson CW, Wang J, Norman KK, Norgaard ZK, Humbert O, Tse CK, et al. Long-term multilineage engraftment of autologous genome-edited hematopoietic stem cells in nonhuman primates. *Blood.* 2016;127(20):2416-26.
5. Wang J, Exline CM, DeClercq JJ, Llewellyn GN, Hayward SB, Li PW, et al. Homology-driven genome editing in hematopoietic stem and progenitor cells using ZFN mRNA and AAV6 donors. *Nat Biotechnol.* 2015;33(12):1256-63.
6. Genovese P, Schiroli G, Escobar G, Di TT, Firrito C, Calabria A, et al. Targeted genome editing in human repopulating haematopoietic stem cells. *Nature.* 2014;510(7504):235-40.
7. Naldini L. Gene therapy returns to centre stage (Review). *Nature.* 2015;526(7573):351-60.
8. Morrison C. $1-million price tag set for Glybera gene therapy. *Nat Biotechnol.* 2015;33(3):217-8.
9. Melchiorri D, Pani L, Gasparini P, Cossu G, Ancans J, Borg JJ, et al. Regulatory evaluation of Glybera in Europe - two committees, one mission. *Nat Rev Drug Discov.* 2013;12(9):719.

10. Hoban MD, Cost GJ, Mendel MC, Romero Z, Kaufman ML, Joglekar AV, et al. Correction of the sickle cell disease mutation in human hematopoietic stem/progenitor cells. *Blood.* 2015;125(17):2597-604.
11. Murray L, Chen B, Galy A, Chen S, Tushinski R, Uchida N, et al. Enrichment of human hematopoietic stem cell activity in the CD34+Thy-1+Lin- subpopulation from mobilized peripheral blood. *Blood.* 1995;85(2):368-78.
12. Yin AH, Miraglia S, Zanjani ED, Almeida-Porada G, Ogawa M, Leary AG, et al. AC133, a novel marker for human hematopoietic stem and progenitor cells. *Blood.* 1997;90(12):5002-12.
13. Civin CI, Strauss LC, Brovall C, Fackler MJ, Schwartz JF, and Shaper JH. Antigenic analysis of hematopoiesis. III. A hematopoietic progenitor cell surface antigen defined by a monoclonal antibody raised against KG-1a cells. *Journal of Immunology.* 1984;133:157-65.
14. Lansdorp PM, Sutherland HJ, and Eaves CJ. Selective expression of CD45 isoforms on functional subpopulations of CD34+ hemopoietic cells from human bone marrow. *Journal of Experimental Medicine.* 1990;172:363-6.
15. Terstappen LW, Huang S, Safford M, Lansdorp PM, and Loken MR. Sequential generations of hematopoietic colonies derived from single nonlineage-committed CD34+CD38- progenitor cells. *Blood.* 1991;77(6):1218-27.
16. Zonari E, Desantis G, Petrillo C, Boccalatte FE, Lidonnici MR, Kajaste-Rudnitski A, et al. Efficient Ex Vivo Engineering and Expansion of Highly Purified Human Hematopoietic Stem and Progenitor Cell Populations for Gene Therapy. *Stem cell reports.* 2017;8(4):977-90.
17. Masiuk KE, Brown D, Laborada J, Hollis RP, Urbinati F, and Kohn DB. Improving gene therapy efficiency through the enrichment of human hematopoietic stem cells. *Mol Ther.* 2017;25(9):2163-75.

18. Gordon PR, Leimig T, Babarin-Dorner A, Houston J, Holladay M, Mueller I, et al. Large-scale isolation of CD133+ progenitor cells from G-CSF mobilized peripheral blood stem cells. *Bone Marrow Transplant.* 2003;31(1):17-22.
19. Radtke S, Adair JE, Giese MA, Chan YY, Norgaard ZK, Enstrom M, et al. A distinct hematopoietic stem cell population for rapid multilineage engraftment in nonhuman primates. *Sci Transl Med.* 2017;9(414):[Epub ahead of print; doi: 10.1126/scitranslmed.aan45].
20. Povinelli BJ, Rodriguez-Meira A, and Mead AJ. Single cell analysis of normal and leukemic hematopoiesis. *Mol Aspects Med.* 2018;59:85-94.
21. Butler A, Hoffman P, Smibert P, Papalexi E, and Satija R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol.* 2018;36(5):411-20.
22. Bhatia M, Wang JC, Kapp U, Bonnet D, and Dick JE. Purification of primitive human hematopoietic cells capable of repopulating immune-deficient mice. *Proc Natl Acad Sci U S A.* 1997;94(10):5320-5.
23. Hogan CJ, Shpall EJ, and Keller G. Differential long-term and multilineage engraftment potential from subfractions of human CD34+ cord blood cells transplanted into NOD/SCID mice. *Proceedings of the National Academy of Sciences.* 2002;99(1):413-8.
24. Görgens A, Radtke S, Möllmann M, Cross M, Dürig J, Horn PA, et al. Revision of the human hematopoietic tree: granulocyte subtypes derive from distinct hematopoietic lineages. *Cell reports.* 2013;3(5):1539-52.
25. Radtke S, Görgens A, Kordelas L, Schmidt M, Kimmig KR, Köninger A, et al. CD133 allows elaborated discrimination and quantification of haematopoietic progenitor subsets in human haematopoietic stem cell transplants. *British journal of haematology.* 2015;169(6):868-78.

26. Radtke S, Haworth KG, and Kiem HP. The frequency of multipotent CD133(+)CD45RA(-)CD34(+) hematopoietic stem cells is not increased in fetal liver compared with adult stem cell sources. *Exp Hematol.* 2016;44(6):502-7.
27. Majeti R, Park CY, and Weissman IL. Identification of a hierarchy of multipotent hematopoietic progenitors in human cord blood. *Cell Stem Cell.* 2007;1(6):635-45.
28. Saliba AE, Westermann AJ, Gorski SA, and Vogel J. Single-cell RNA-seq: advances and future challenges. *Nucleic acids research.* 2014;42(14):8845-60.
29. Shalek AK, Satija R, Adiconis X, Gertner RS, Gaublomme JT, Raychowdhury R, et al. Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. *Nature.* 2013;498(7453):236-40.
30. Islam S, Zeisel A, Joost S, La Manno G, Zajac P, Kasper M, et al. Quantitative single-cell RNA-seq with unique molecular identifiers. *Nature methods.* 2014;11(2):163-6.
31. Svensson V, Natarajan KN, Ly LH, Miragaia RJ, Labalette C, Macaulay IC, et al. Power analysis of single-cell RNA-sequencing experiments. *Nat Methods.* 2017;14(4):381-7.
32. Dmytrus J, Matthes-Martin S, Pichler H, Worel N, Geyeregger R, Frank N, et al. Multi-color immune-phenotyping of CD34 subsets reveals unexpected differences between various stem cell sources. *Bone Marrow Transplant.* 2016;51(8):1093-100.
33. Görgens A, Ludwig AK, Möllmann M, Krawczyk A, Dürig J, Hanenberg H, et al. Multipotent hematopoietic progenitors divide asymmetrically to create progenitors of the lymphomyeloid and erythromyeloid lineages. *Stem cell reports.* 2014;3(6):1058-72.
34. Gorgens A, Radtke S, Horn PA, and Giebel B. New relationships of human hematopoietic lineages facilitate detection of multipotent hematopoietic stem and progenitor cells. *Cell Cycle.* 2013;12(22):3478-82.
35. Radtke S, Chan YY, Sippel TR, Kiem HP, and Rongvaux A. MISTRG mice support engraftment and assessment of nonhuman primate hematopoietic stem and progenitor cells. *Exp Hematol.* 2019;70:31-41 e1.

36. Boitano AE, Wang J, Romeo R, Bouchez LC, Parker AE, Sutton SE, et al. Aryl hydrocarbon receptor antagonists promote the expansion of human hematopoietic stem cells. *Science.* 2010;329(5997):1345-8.

37. Fares I, Chagraoui J, Gareau Y, Gingras S, Ruel R, Mayotte N, et al. Cord blood expansion. Pyrimidoindole derivatives are agonists of human hematopoietic stem cell self-renewal. *Science.* 2014;345(6203):1509-12.

38. Zou J, Zou P, Wang J, Li L, Wang Y, Zhou D, et al. Inhibition of p38 MAPK activity promotes ex vivo expansion of human cord blood hematopoietic stem cells. *Ann Hematol.* 2012;91(6):813-23.

39. Wagner JE, Jr., Brunstein CG, Boitano AE, DeFor TE, McKenna D, Sumstad D, et al. Phase I/II Trial of StemRegenin-1 Expanded Umbilical Cord Blood Hematopoietic Stem Cells Supports Testing as a Stand-Alone Graft. *Cell Stem Cell.* 2016;18(1):144-55.

40. Ishitsuka K, Hideshima T, Neri P, Vallet S, Shiraishi N, Okawa Y, et al. p38 mitogen-activated protein kinase inhibitor LY2228820 enhances bortezomib-induced cytotoxicity and inhibits osteoclastogenesis in multiple myeloma; therapeutic implications. *British journal of haematology.* 2008;141(5):598-606.

41. Psatha N, and Papayannopoulou T. Specific Companion Cells Enhance the Engraftment of Ex-Vivo Expanded HSCs (CD34+/CD38−/CD90+). *Molecular therapy Methods & clinical development.* 2016;24(Supplement 1):pS210.

42. Adair JE, Waters T, Haworth KG, Kubek SP, Trobridge GD, Hocum JD, et al. Semi-automated closed system manufacturing of lentivirus gene-modified haematopoietic stem cells for gene therapy. *Nat Commun.* 2016;7:13173.

43. Adair JE, Beard BC, Trobridge GD, Neff T, Rockhill JK, Silbergeld DL, et al. Extended survival of glioblastoma patients after chemoprotective HSC gene therapy. *Science translational medicine.* 2012;4(133):133ra57.

44. Adair JE, Chandrasekaran D, Sghia-Hughes G, Haworth KG, Woolfrey AE, Burroughs LM, et al. Novel lineage depletion preserves autologous blood stem cells for gene therapy of Fanconi anemia complementation group A. *Haematologica.* 2018;103(11):1806-14.
45. Peterson CW, Wang J, Norman KK, Norgaard ZK, Humbert O, Tse CK, et al. Long-term multi-lineage engraftment of genome-edited hematopoietic stem cells after autologous transplantation in nonhuman primates. *Blood.* 2016.
46. Humbert O, Radtke S, Samuelson C, Carrillo RR, Perez AM, Reddy SS, et al. Therapeutically relevant engraftment of a CRISPR-Cas9-edited HSC-enriched population with HbF reactivation in nonhuman primates. *Science translational medicine.* 2019;11(503).
47. Baum CM, Weissman IL, Tsukamoto AS, Buckle AM, and Peault B. Isolation of a candidate human hematopoietic stem-cell population. *Proceedings of the National Academy of Sciences of the United States of America.* 1992;89:2804.
48. Karamitros D, Stoilova B, Aboukhalil Z, Hamey F, Reinisch A, Samitsch M, et al. Single-cell analysis reveals the continuum of human lympho-myeloid progenitor cells. *Nat Immunol.* 2018;19(1):85-97.
49. Velten L, Haas SF, Raffel S, Blaszkiewicz S, Islam S, Hennig BP, et al. Human haematopoietic stem cell lineage commitment is a continuous process. *Nature cell biology.* 2017;19(4):271-81.
50. Notta F, Zandi S, Takayama N, Dobson S, Gan OI, Wilson G, et al. Distinct routes of lineage development reshape the human blood hierarchy across ontogeny. *Science.* 2016;351(6269):aab2116.
51. Wilson NK, Kent DG, Buettner F, Shehata M, Macaulay IC, Calero-Nieto FJ, et al. Combined Single-Cell Functional and Gene Expression Analysis Resolves Heterogeneity within Stem Cell Populations. *Cell Stem Cell.* 2015;16(6):712-24.

52. Brennecke P, Anders S, Kim JK, Kolodziejczyk AA, Zhang X, Proserpio V, et al. Accounting for technical noise in single-cell RNA-seq experiments. *Nature methods.* 2013;10(11):1093-5.

53. Czechowska K, Lannigan J, Wang L, Arcidiacono J, Ashhurst TM, Barnard RM, et al. Cyt-Geist: Current and Future Challenges in Cytometry: Reports of the CYTO 2018 Conference Workshops. *Cytometry Part A : the journal of the International Society for Analytical Cytology.* 2019;95(6):598-644.

54. Adair JE, Johnston SK, Mrugala MM, Beard BC, Guyman LA, Baldock AL, et al. Gene therapy enhances chemotherapy tolerance and efficacy in glioblastoma patients. *The Journal of clinical investigation.* 2014;124(9):4082-92.

55. Hu Y, and Smyth GK. ELDA: extreme limiting dilution analysis for comparing depleted and enriched populations in stem cell and other assays. *Journal of immunological methods.* 2009;347(1-2):70-8.

56. Bray NL, Pimentel H, Melsted P, and Pachter L. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol.* 2016;34(5):525-7.

57. Love MI, Huber W, and Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014;15(12):550.

58. Hu Y, and Smyth GK. ELDA: extreme limiting dilution analysis for comparing depleted and enriched populations in stem cell and other assays. *Journal of Immunological Methods.* 2009;347(2-Jan):70-8.

FIGURES AND FIGURE LEGENDS

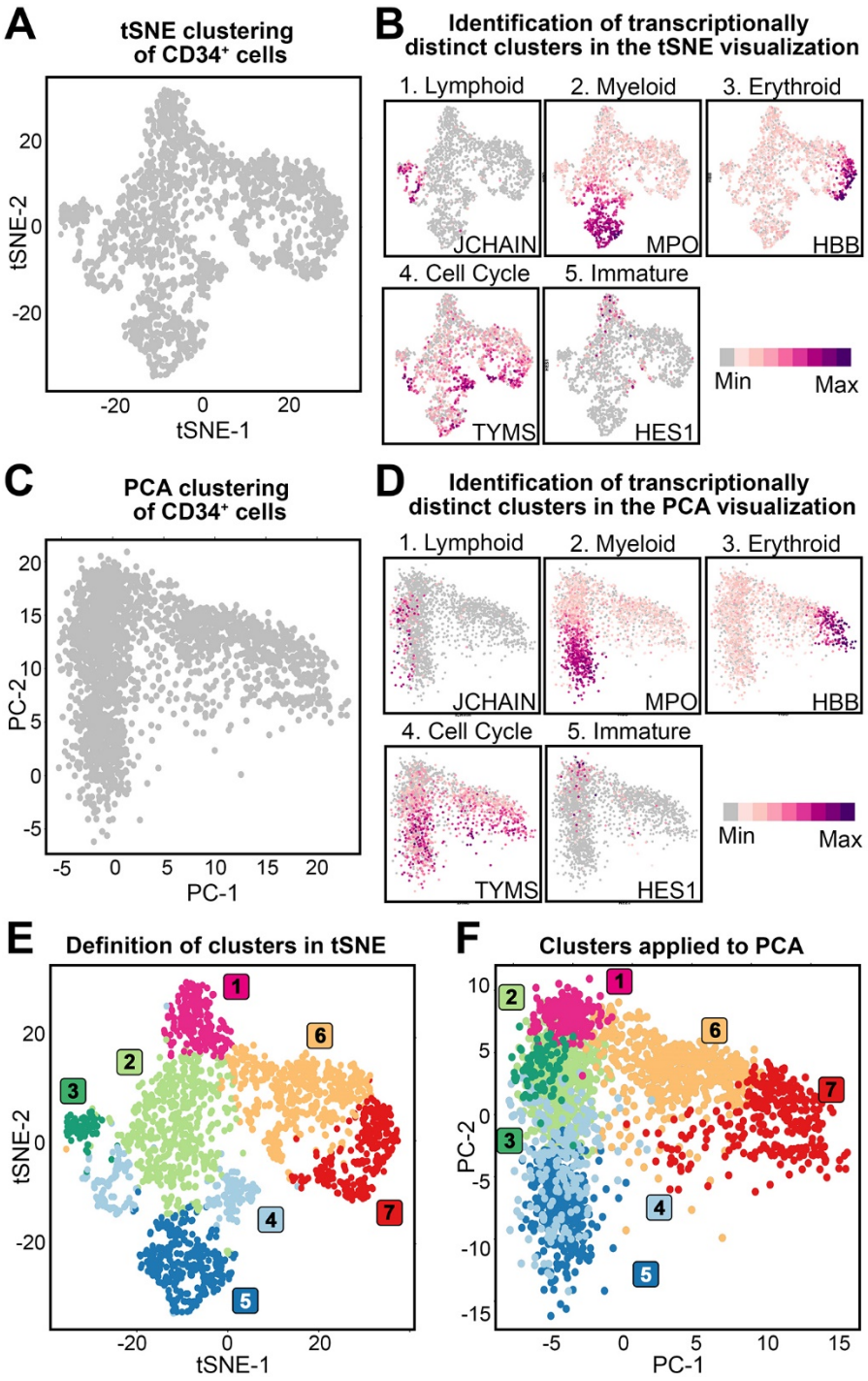

Figure 1. A scRNA-seq reference map of ssBM-derived CD34⁺ HSPCs. (**A**) Dimensional reduction (tSNE) of scRNAseq data from ssBM-derived CD34⁺ cells. (**B**) Feature plots showing the expression of representative genes associated with lymphoid-, myeloid-, erythroid-primed, proliferating, and immature HSPCs. Level of expression is color coded as shown in the legend. (**C**) PCA based transformation and (**D**) expression of representative genes for the same dataset shown in panel A. (**E**) Graph-based clustering of ssBM-derived CD34⁺ cells. Transcriptionally distinct CD34 clusters were color-coded and numbered as indicated. (**F**) Clusters defined in panel A projected onto the PCA analysis.

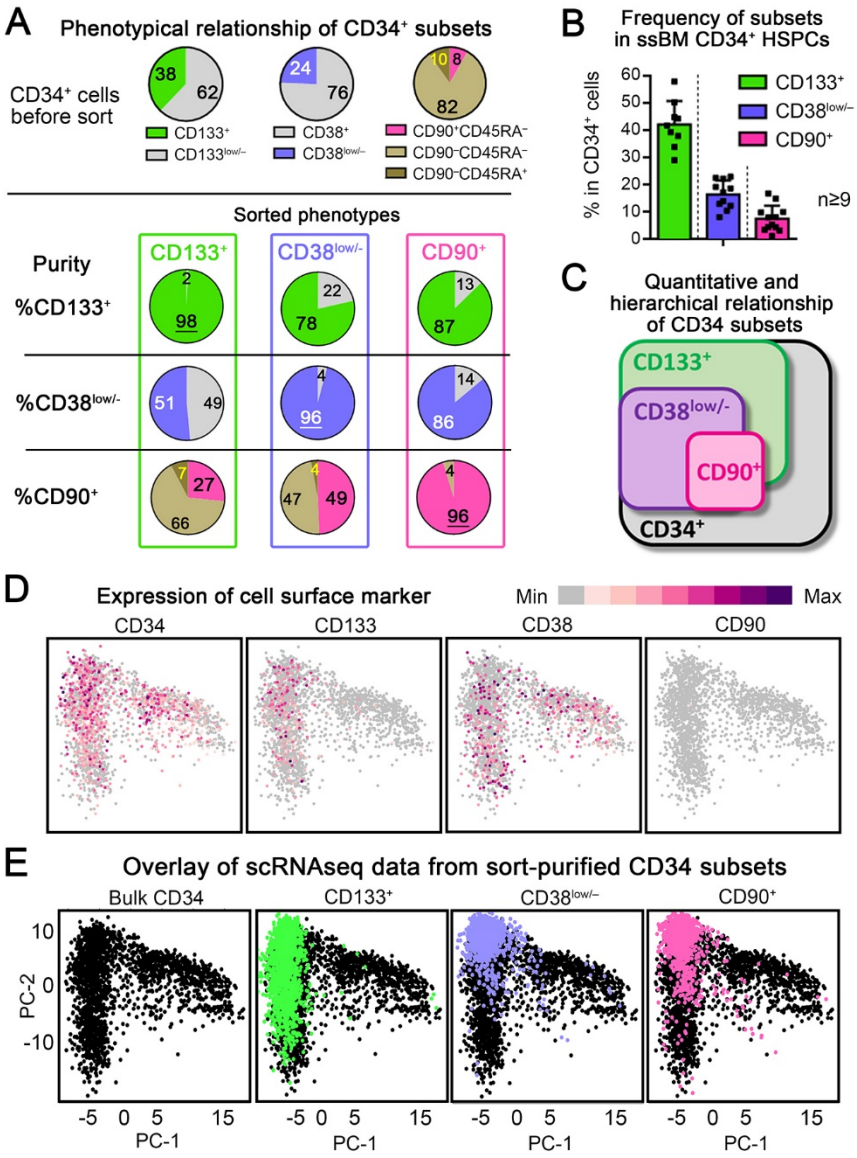

Figure 2. Quantitative comparison of phenotypically defined target CD34 subsets. (**A**) Frequency of phenotypically defined subsets in bulk CD34$^+$ cells before sort (top row) and cross-contamination of phenotypical subsets within sort-purified CD133$^+$ (2nd row), CD38$^{low/-}$ (3rd row) and CD90$^+$ (4th row) HSPCs. (**B**) Average frequency of CD133$^+$, CD38$^{low/-}$ and CD90$^+$ HSPCs within bulk CD34$^+$ cells (mean ± SEM, n≥9 independent healthy donors). (**C**) Schematic of the quantitative and hierarchical relationship of phenotypical CD34 subsets. (**D**) Expression of CD34, CD133, CD38, and CD90 in ssBM-derived CD34$^+$ cells. Level of expression is color coded as shown in the legend. (**E**) Overlay of scRNA-seq data from CD34$^+$ cells (black, 1st plot) with sort-purified CD133$^+$ (green, 2nd plot), CD38$^{low/-}$ (purple, 3rd plot), and CD90$^+$ (pink, 4th plot) HSPCs.

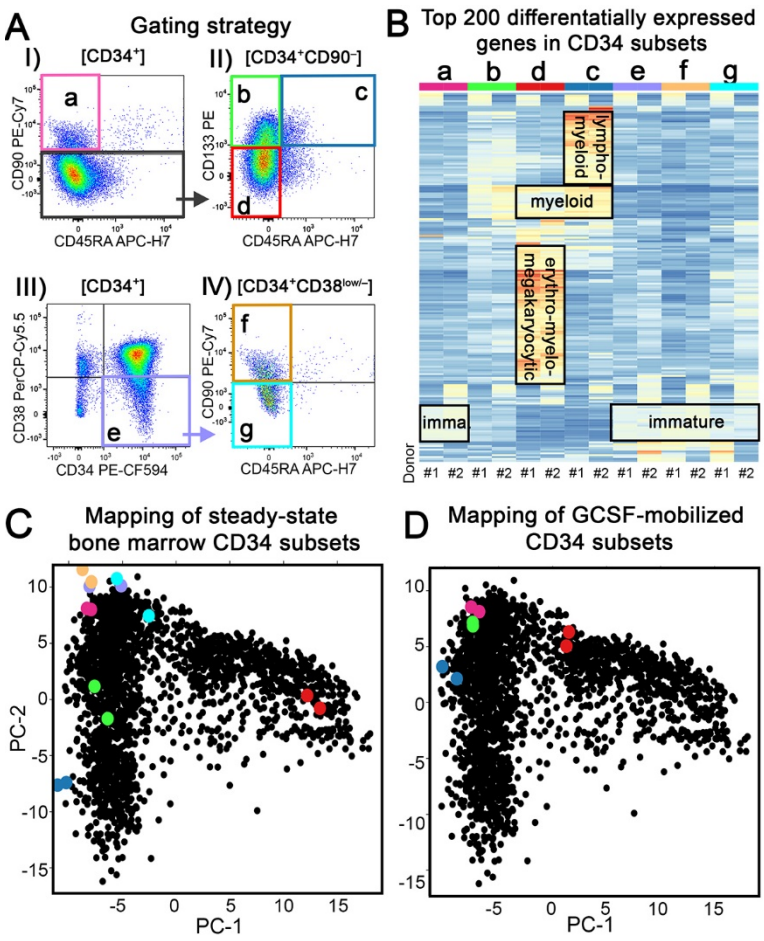

Figure 3. Bulk RNAseq of ssBM-derived CD34 subsets. (**A**) Gating strategy for phenotypically-defined human CD34 subpopulations. (**B**) Heat map of the Top 200 differentially expressed genes in phenotypically-defined CD34 subsets listed in panel B from two independent human donors. A detailed list of the Top 200 genes can be found in **Supplemental Table 3**. (**C**) Overlay of bulk RNAseq data from sort-purified ssBM subsets (Populations *a*–*g* color coded as defined in panel A) on the ssBM scRNAseq PCA reference map (black dots) (**D**) Overlay of bulk RNAseq data from sort-purified GCSF-mobilized CD34 subsets (Population *a*–*d* color coded as defined in panel A) on the ssBM scRNAseq PCA reference map (black dots).

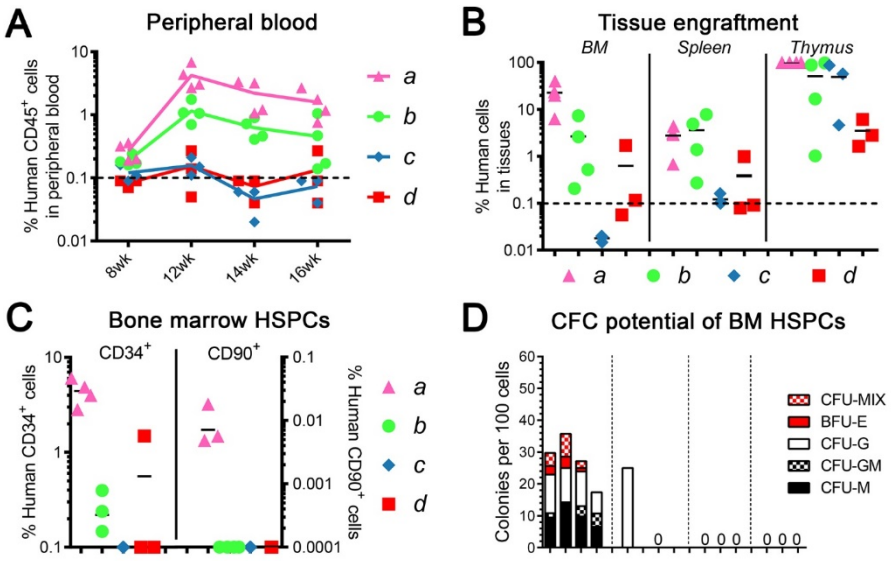

Figure 4. Multilineage engraftment potential of human CD34 subpopulations. Frequency of human chimerism in the (**A**) PB and (**B**) BM, spleen, and thymus after transplantation of sort-purified CD34 subpopulations (1×10^5 cells per mouse). (**C**) Frequency of engrafted human $CD34^+$ and $CD90^+$ HSPCs. $CD34^+$ frequency use left y-axis, $CD90^+$ frequency right y-axis. (**D**) Erythroid, myeloid and erythro-myeloid colony-forming potential of engrafted human $CD34^+$ cells. CFU, colony-forming unit; CFU-M, macrophages; CFU-G, granulocytes; CFU-GM, granulocyte/macrophage; and BFU-E, erythroid.

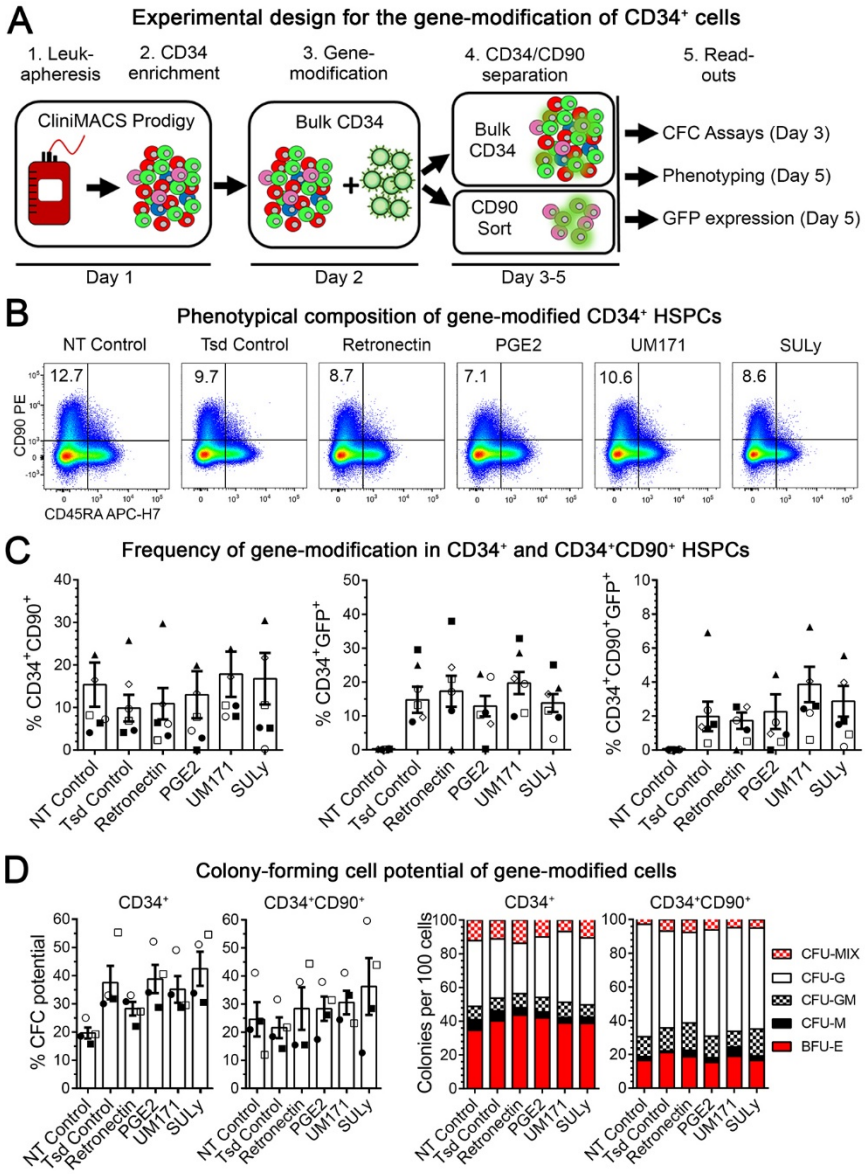

Figure 5: Effect of small molecules on the transduction of GCSF-mobilized bulk CD34⁺ HSPCs. (**A**) Schematic of the experimental design. **(B)** Flow-cytometric assessment of bulk CD34⁺ cells after lentiviral gene-modification (MOI5) in the presence of small molecules at day 5 post-transduction. (abbreviations: NT = non-transduced; Tsd = transduction) **(C)** Flow-cytometric quantification of CD34⁺CD90⁺, CD34⁺GFP⁺, and CD34⁺CD90⁺GFP⁺ cells at day 5 post-transductions. **(D)** Total colony-forming cell potential (left two graphs) and composition of colony types (right two graphs) of bulk CD34⁺ cells and the CD34⁺CD90⁺ subset. (statistics: means ± SEM; significance values for C and D summarized in **Supplemental Table 10**).

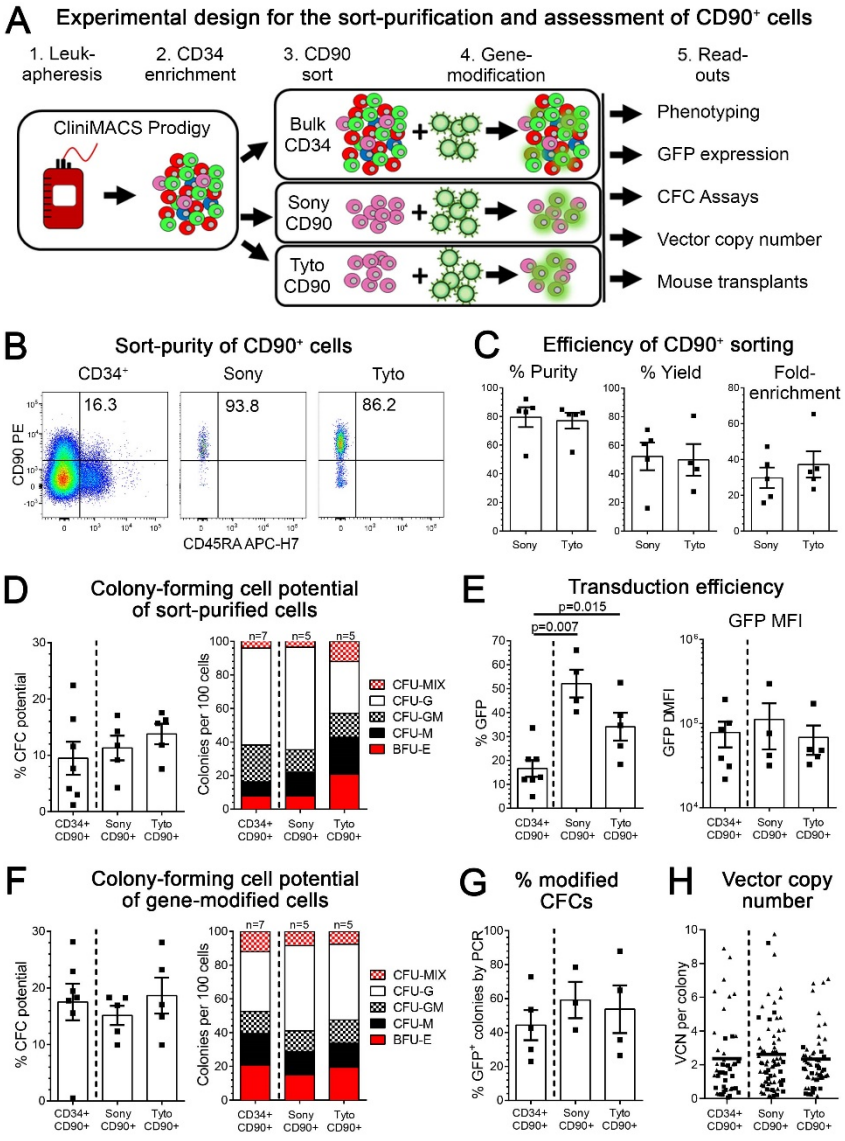

Figure 6: Sort-purification and quality control of CD90⁺ HSPCs. (**A**) Schematic of the experimental design. (**B**) Flow-cytometric assessment of cells before sort (CD34⁺, 1st plot) and purified CD90⁺ cells after sorting on the Sony (2nd plot) and Tyto (3rd plot). (**C**) Comparison of the purity, yield and fold-enrichment of CD90⁺ HSPCs on the Sony and Tyto sorter. (**D**) Colony-forming cell potential of CD90⁺ cells within bulk CD34⁺ HSPCs (CD34⁺CD90⁺) and sort-purified CD90⁺ subsets. (**E**) Flow-cytometric quantification of GFP-expressing CD90⁺ cells within bulk CD34⁺ cell and sort-purified CD90⁺ subsets (left) and the delta-MFI (mean fluorescence intensity) of GFP expression in gene-modified cells (right). (**F**) Erythroid, myeloid and erythro-myeloid

colony-forming cell potential of gene-modified $CD34^+CD90^+$ and sort-purified $CD90^+$ subset. Individual colonies from all three conditions in (F) were picked and (**G**) the gene-modification efficiency in CFCs determine by PCR as well as (**H**) the vector copy number (VCN) in modified CFCs quantified by qPCR. (statistics: means ± SEM; significance values: two-tailed paired t test)

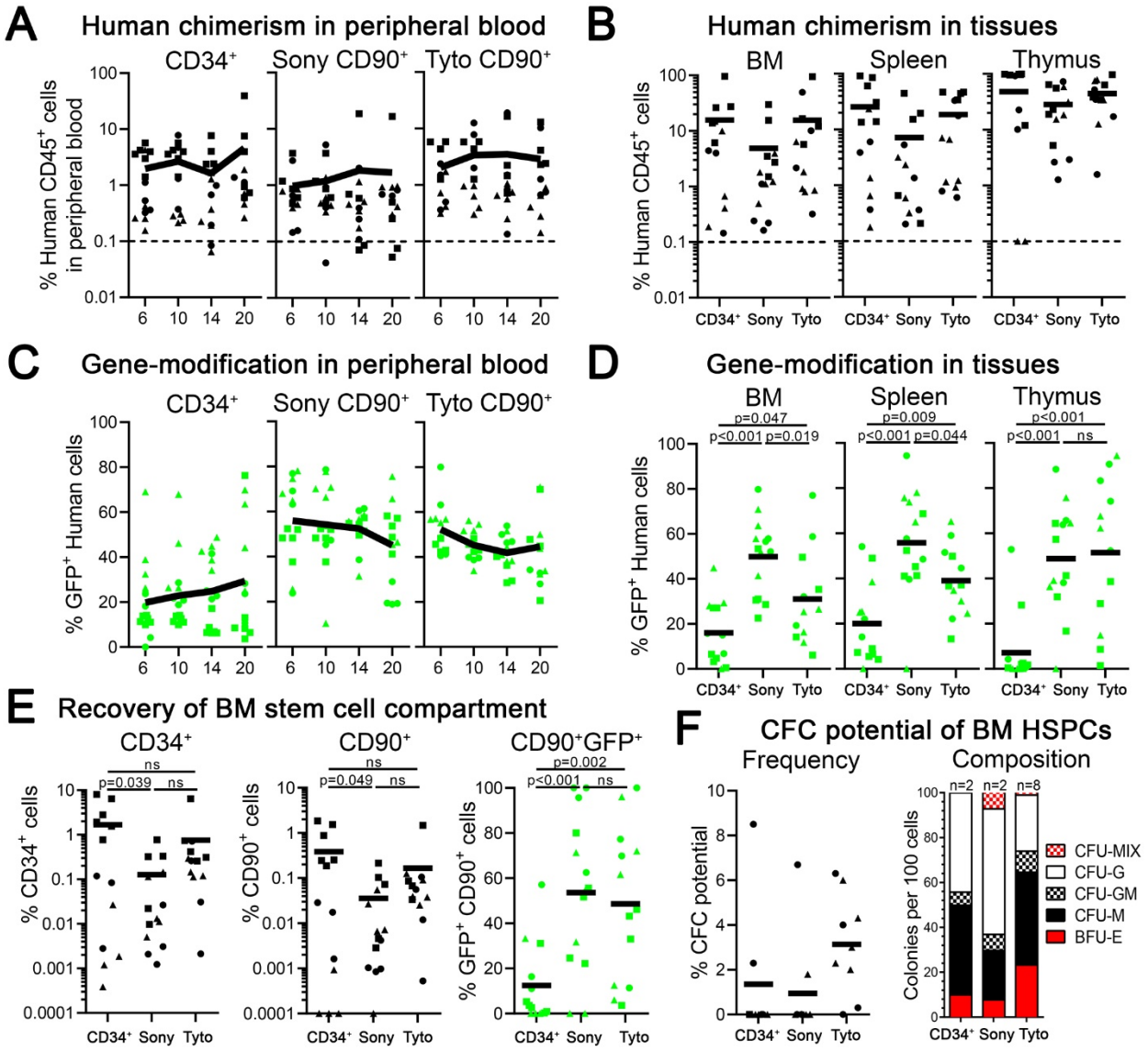

Figure 7: Transplantation of sort-purified and gene-modified CD90⁺ cells in NSG mice. (**A**) Frequency of human chimerism in the PB over time. (**B**) Human chimerism in the BM, spleen and thymus at 20 weeks post-transplant. Frequency of GFP⁺ human cells in (**C**) the PB over time and (**D**) tissues at 20 weeks post-transplant. (**E**) Engraftment of human CD34⁺, CD90⁺ and CD90⁺GFP⁺ HSPCs in the BM at necropsy. (**F**) Erythroid, myeloid and erythro-myeloid colony-forming potential of engrafted human CD34⁺ cells.