

UNIVERSITA' VITA-SALUTE SAN RAFFAELE

**CORSO DI DOTTORATO DI RICERCA
INTERNAZIONALE
IN MEDICINA MOLECOLARE**

**CURRICULUM IN CELLULAR AND MOLECULAR
PHYSIOPATHOLOGY**

Unveiling the biological role of human
circulating Hematopoietic Stem and Progenitor
cells

DoS: Alessandro Aiuti



Second Supervisor: Marieke Essers

Tesi di DOTTORATO di RICERCA di Pamela Quaranta

matr. 013873

Ciclo di dottorato XXXIV

SSD BIO/11 BIO/13

Anno Accademico 2020/2021

CONSULTAZIONE TESI DI DOTTORATO DI RICERCA

Il/la sottoscritto/I Pamela Quranta
Matricola / *registration number* 013873
Nata a/ *born at* Acquaviva delle Fonti (BA)
il/on 06/04/1993

autore della tesi di Dottorato di ricerca dal titolo / *author of the PhD Thesis titled*
Unveiling the biological role of human circulating Hematopoietic Stem and Progenitor
cells

AUTORIZZA la Consultazione della tesi / *AUTHORIZES the public release of the thesis*

NON AUTORIZZA la Consultazione della tesi / *DOES NOT AUTHORIZE the public release of the thesis*

a partire dalla data di conseguimento del titolo e precisamente / *from the PhD thesis date, specifically*

Dal / *from*/...../.....

Poiché /*because*:

l'intera ricerca o parti di essa sono potenzialmente soggette a brevettabilità/ *The whole project or part of it might be subject to patentability;*

ci sono parti di tesi che sono già state sottoposte a un editore o sono in attesa di pubblicazione/ *Parts of the thesis have been or are being submitted to a publisher or are in press;*

la tesi è finanziata da enti esterni che vantano dei diritti su di esse e sulla loro pubblicazione/ *the thesis project is financed by external bodies that have rights over it and on its publication.*

Si rende noto che parti della tesi sono indisponibili in relazione all'utilizzo di dati tutelati da segreto industriale **(da lasciare solo se applicabile)** /*Please Note: some parts of the thesis are not available in relation to the norm of the use of information protected by trade secret (To leave only if relevant)*

E' fatto divieto di riprodurre, in tutto o in parte, quanto in essa contenuto / *Copyright the contents of the thesis in whole or in part is forbidden*

Data /Date 15/01/22

Firma/Signature

A handwritten signature in black ink, appearing to read "P. A. A. A.", written over a horizontal line.

DECLARATION

This thesis has been composed by myself and has not been used in any previous application for a degree. Throughout the text I use both 'I' and 'We' interchangeably.

All the results presented here were obtained by myself, except for:

1. **Collaboration:** scRNAseq analyses and figures of results (Figure 22, Figure 23A-C, Figure 24, Figure 25A-C, Figure 26D, Figure 30A,C,E, Figure 31, Figure 32, Figure 33) were performed in collaboration with Matteo Naldini e Matteo Barcella, from Dr. Gentner's lab (SR-Tiget). Dr Di Micco's lab (SR-Tiget) provided PB samples collected from aged subjects. Integration site analyses were performed in collaboration with Dr. Montini's lab (SR-Tiget).
2. **Fair Use Disclaimer.** The figures used in the introduction of this thesis are under fair use as defined under section 107 of the US Copyright Act of 1976 and according to Italian Law n. 68 of April 9 2003 «Attuazione della direttiva 2001/29/CE sull'armonizzazione di taluni aspetti del diritto d'autore e dei diritti connessi nella società dell'informazione». All rights and credit go directly to their rightful owners. No copyright infringement is intended.

All sources of information are acknowledged by means of reference.

ABSTRACT

Although mostly resident in the bone marrow (BM), few circulating HSPC (cHSPC) regularly traffic in the peripheral blood (PB) of un-mobilized subjects. Mainly descriptive studies have been published so far about this rare population in humans, and a complete evaluation of their composition, functional features and hierarchical relationship with respect to BM HSPC is still missing.

In the present study, we phenotypically characterized cHSPC composition during aging, by applying multi-parametric flow cytometry on 114 PB and, as control, 48 BM samples of healthy donors (HD) of diverse age. These analyses were integrated with single-cell transcriptome profiling (scRNAseq), and *ad hoc* designed *in vitro* and *in vivo* assays to investigate the transcriptional and functional properties of steady-state cHSPC with respect to BM counterpart. Moreover, to study circulating vs. resident HSPC relationships and differentiation potential *in vivo* in humans, we exploited integration site (IS) clonal tracking of cHSPC, BM HSPC and mature PB lineages isolated from patients treated with HSPC-gene therapy (GT).

We observed that cHSPC show a progressive reduction in number during aging and a different composition than BM counterpart, with Multi Lymphoid Progenitors (MLP) displaying the highest PB circulation capability. cHSPC are endowed with multilineage differentiation potential both *in vitro* and *in vivo*, with comparable BM homing capability but reduced long-term survival after transplantation in immune-deficient mice than BM HSPC. This latter finding can be explained by the low primitive HSC content and the transcriptional pre-activated state observed in steady-state PB HSC.

Indeed, applying scRNA-seq, we identified a unique transcriptional profile of both primitive and lineage-committed cHSPC subpopulations, characterized by lower replicative, metabolic and transcriptional activity, but increased differentiation-, adhesion- and immune response-priming than BM counterpart. The enrichment of lymphoid phenotypic and transcriptional signatures found in PB HSPC, together with their higher IS sharing with PB lymphoid than myeloid mature lineages suggest that cHSPC could have a role in seeding lymphoid organs. Moreover, the higher expression of erythroid marker genes detected in trafficking than resident HSPC was consistent with cHSPC erythroid differentiation bias observed after transplantation and in single-cell *in vitro* differentiation assay. These findings suggest cHSPC as a source of erythroid-committed progenitors, able to sustain stress-responsive extramedullary erythropoiesis.

Finally, our preliminary data on a cohort of HSPC-GT patients suggest that cHSPC may sustain clonal redistribution to distant BM sites, both during active hematopoietic reconstitution and, at a lower extent, during steady-state conditions.

Altogether, our findings indicate PB trafficking HSPC as a peculiar steady-state reservoir of low-cycling, pre-activated hematopoietic progenitors, which continuously recirculate among multiple BM sites and are poised for promptly sustaining activation and *in situ* local hematopoietic differentiation in case of demand.

TABLE OF CONTENT

ACRONYMS AND ABBREVIATIONS.....	4
LIST OF FIGURES AND TABLES.....	8
1 INTRODUCTION.....	12
1.1 The human hematopoietic system	12
1.2 Human hematopoietic stem and progenitor cells.....	13
1.2.1 Phenotypic definition of human HSPC	13
1.2.2 Cellular properties and functional features of human HSPC	15
1.2.3 HSPC differentiation: distinct models of human hematopoietic hierarchy	17
1.2.4 Assays for studying human hematopoiesis and HSPC functions.....	21
1.3 Studying human hematopoiesis through HSPC gene therapy.....	22
1.3.1 Ex vivo HSPC gene therapy.....	22
1.3.2 Investigating human HSPC dynamics by integration sites clonal tracking.....	26
1.4 The BM niche	27
1.4.1 Cellular and molecular components of the BM hematopoietic niche	28
1.4.2 Pivotal molecular axes controlling HSPC maintenance, retention, homing and egress in and out of the BM niche.....	32
1.4.3 Current strategies for HSPC mobilization.....	35
1.5 Circulating HSPC.....	37
1.5.1 Murine circulating HSPC.....	38
1.5.2 Human cHSPC content changes in association with multiple physiological and pathological factors: biomarker valence?	40
1.5.3 Functional features of human cHSPC: current knowledge and relevant gaps in the field.....	42
2 AIM OF THE WORK.....	45
3 RESULTS.....	46
3.1 cHSPC display a progressive quantitative change during physiological aging.....	46
3.2 Recirculation capability changes among the distinct cHSPC subsets and during physiological aging.....	49
3.3 cHSPC display multilineage differentiation potential <i>in vitro</i>	52

3.4	cHSPC show multilineage reconstitution capability but reduced long-term engraftment potential than BM HSPC after <i>in vivo</i> transplant	55
3.5	PB and BM HSPC display comparable BM homing potential and repopulation capability early after transplantation.....	59
3.6	cHSPC show enriched lymphoid and erythroid transcriptional signatures	63
3.7	Phenotypic circulating CMP and primitive HSPC subsets are transcriptionally and functionally committed toward erythroid lineages.....	69
3.8	cHSPC are poised for differentiation, adhesion, and immune activation with respect to BM counterpart.....	80
3.9	Investigating HSPC trafficking <i>in vivo</i> in humans through IS analyses.....	87
4	DISCUSSION.....	92
5	MATERIALS AND METHODS.....	100
5.1	Characteristics of healthy donors and gene therapy patients involved in the study.....	100
5.2	Flow cytometry analyses by WBD on healthy donors' and patients' samples.....	102
5.3	Isolation of human mononuclear cells and CD34⁺ cells from BM and PB samples.....	105
5.4	Single-cell RNA sequencing and multiparametric single-cell protein barcoding.....	105
5.4.1	<i>Sample preparation</i>	105
5.4.2	<i>Library preparation.....</i>	109
5.4.3	<i>Sequencing.....</i>	113
5.4.4	<i>Sample demultiplexing, barcode processing and UMI count.....</i>	113
5.4.5	<i>Gene expression data analysis.....</i>	114
5.4.6	<i>Cell type annotation</i>	115
5.4.7	<i>Downstream analysis</i>	116
5.5	Colony forming units (CFU) assay.....	117
5.6	<i>In vitro</i> multi-lineage differentiation assays	117
5.7	<i>In vivo</i> models.....	121
5.8	<i>In vivo</i> transplantation assays	122
5.9	IS retrieval and analysis	122
5.9.1	<i>Isolation of PB mature lineages, HSPC subpopulations and PB/BM CFC from GT-patients.....</i>	122

5.9.2	<i>DNA extraction and IS retrieval</i>	123
5.9.3	<i>IS mapping and filtering</i>	124
5.10	Statistical tests	127
REFERENCES	128

ACRONYMS AND ABBREVIATIONS

ADA-SCID: adenosine deaminase-severe combined immunodeficiency

ADT: antibody-derived tag

ATAC-seq: assay for transposase accessible chromatin with high-throughput sequencing

ANGPT-1: angiopoietin-1

APC: antigen presenting cell

BFU-E: burst-forming unit-erythroid

BM: bone marrow

BM L: bone marrow left

BM R: bone marrow right

BOP: N-(benzenesulfonyl)-l-prolyl-l-O- (1-pyrrolidinylcarbonyl)tyrosine

CAR cell: CXCL12-abundant reticular cell

CB: cord blood

CDK6: cyclin-dependent kinase 6

CFC: colony-forming cell

CFU: colony-forming unit

CFU-GM: colony forming unit-granulocytic and monocytic

CFU-GEMM: colony forming unit-granulocytic, erythroid, monocytic and megakaryocytic

CGD: chronic granulomatous disease

CI: circulation index

CITE-seq: cellular indexing of transcriptomes and epitopes by sequencing

CLL: chronic lymphocytic leukemia

CMP: committed myeloid progenitors

CO₂: carbon dioxide

CLOUD-HSPC: continuum of low-primed undifferentiated hematopoietic stem and progenitor cells

CXCL4: C-X-C motif chemokine ligand 4

CXCL12: C-X-C motif chemokine ligand 12

CXCR2: C-X-C motif chemokine receptor 2

CXCR4: C-X-C motif chemokine receptor 4

DC: dendritic cell
DCV-SP: DieCycle Violet-side population
DPP-4: dipeptidyl peptidase-4
EMA: European Medicines Agency
EP: erythroid progenitors
EPCR: endothelial protein C receptor
ETP: early T progenitors
FA: Fanconi Anemia
FACS: fluorescence-activated cell sorting
FDA: Food and Drugs Administration
FGF-1: fibroblast growth factor 1
FLT-3L: Fms-related tyrosine kinase 3 ligand
FOXP3: Forkhead box P3
G-CSF: granulocyte colony stimulating factor
GEX: gene expression
GM-CSF: granulocyte/monocytes colony stimulating factor
GMP: granulocyte/monocyte progenitors
GO-BP: gene ontology-biological processes
GT: gene therapy
HA: hyaluronic acid
HD: healthy donor
HDR: homologous end joining
HLA: human leukocyte antigen
HIF-1 α : hypoxia-inducible factor-1 alpha subunit
HSC: hematopoietic stem cell
HSPC: hematopoietic stem and progenitor cells
cHSPC: circulating HSPC
ICAM-1: intercellular adhesion molecule
IFN: interferon
IFITM3: interferon-inducible transmembrane 3
IL-7: interleukin-7
IL-10: interleukin-10

IS: integration site
LEC-1: liver endothelial cells-1
LFA-1: leukocyte function antigen 1
logFC: log fold change
LPS: lipopolysaccharide
LT: long-term
LV: lentiviral vector
MC: mononuclear cells
MDSC: myeloid-derived suppressor cells
MEP: megakaryocyte/erythroid progenitors
MLP: multi-lymphoid progenitors
MKP: megakaryocyte progenitors
MLD: metachromatic leukodystrophy
MMP-9: matrix metalloproteinase-9
MPB: mobilized peripheral blood
MPP: multipotent progenitors
MPSI: mucopolysaccharidosis-I
MSC: mesenchymal stromal cell
NES: normalized enrichment score
NK: natural killer
NLRs: nucleotide-binding oligomerization domain containing (NOD)-like-receptors
NOD/SCID: non-obese diabetic/severe combined immunodeficiency
NSG: NOD/SCID Il2rg^{-/-}
NSGW41: NOD/SCID Il2rg^{-/-}/Kit^{W41/W41}
OPN: osteopontin
OSM: oncostatin M
PAMPs: pathogen associated molecular patterns
PCs: principal components
PGE2: prostaglandin E2
PI: propidium iodide
PIDs: primary immunodeficiencies
PPAR γ : peroxisome proliferator-activator receptor γ

PreBNK: precursor of B/NK cells
PTH: parathyroid hormone
 β -THAL: β -thalassemia
TI: β -thalassemia intermedia
TM: β -thalassemia major
TN-C: tenascin-C
TPO: thrombopoietin
RBC: red blood cell
scRNAseq: single-cell RNA sequencing
ROS: reactive oxygen species
 γ -RV: γ -retroviral vectors
SCID-X1: X-linked severe combined immunodeficiency
SCD: sickle cell disease
SCF: stem cell factor
mbSCF: membrane SCF
SDF-1: stromal cell-derived factor-1
ST: short-term
 T_{reg} : regulatory T cell
TN-C: tenascin-C
TCR: T cell receptor
TGF β : transforming growth factor β
TL4: toll-like receptor 4
UMAP: uniform manifold approximation and projection
UMI: unique molecular identifier
VCAM-1: vascular cell adhesion molecule 1
VLA-4: very late activation antigen
WAS: Wiskott-Aldrich syndrome
WBD: Whole Blood Dissection

LIST OF FIGURES AND TABLES

Figure 1. Combination of markers used for the identification of human HSPC subsets.	15
Figure 2. Classical model of human hematopoietic differentiation.....	18
Figure 3. Human hematopoietic differentiation as a continuum process.....	20
Figure 4. Outline of <i>ex vivo</i> HSPC gene therapy protocol.....	23
Figure 5. The adult BM niche in homeostatic conditions.....	32
Figure 6. Molecular axes controlling BM HSPC homing, retention and egress.....	34
Figure 7. Hypotheses on circulating HSPC functions derived from studies in mice and humans.....	44
Figure 8. Phenotypic characterization of PB HSPC in distinct ranges of age.....	47
Figure 9. Count of HSPC single subsets from PB and BM during aging.....	48
Figure 10. Comparison of BM and PB HSPC composition and amount in the age-matched dataset.....	50
Figure 11. Circulation indexes (CI) of single cHSPC subsets belonging to the age-matched dataset.....	51
Figure 12. Clonogenic potential of cHSPC during aging.....	52
Figure 13. Experimental scheme of the <i>in vitro</i> multilineage differentiation protocol..	53
Figure 14. <i>In vitro</i> differentiation assay of human bulk CD34 ⁺ cells derived from BM or PB.....	54
Figure 15. Setting of long-term <i>in vivo</i> transplantation experiments.....	55
Figure 16. Long-term human PB reconstitution of major hematopoietic compartments in mice transplanted with BM or PB CD34 ⁺ cells.....	56
Figure 17. Human cell content in mice BM at 20 weeks after transplantation with BM or PB CD34 ⁺ cells.....	58
Figure 18. Setting of short-term <i>in vivo</i> experiments.....	59
Figure 19. Human cell content in mice PB and BM at 4 and 12 weeks after transplantation with BM or PB CD34 ⁺ cells.	60
Figure 20. Comparison of <i>in vivo</i> BM homing potential of BM and PB CD34 ⁺ cells...62	
Figure 21. Experimental workflow of our integrated approach based on FACS immunophenotyping and sorting, scRNAseq transcriptome analysis and multiparametric	

immunophenotyping through ADT protein barcoding (CITE-seq technology).....	64
Figure 22. Distribution of scRNAseq-processed cells by source and donor.....	65
Figure 23. Unsupervised clustering and cluster annotation of HSPC subpopulations in scRNAseq dataset.....	67
Figure 24. Detection of HSPC phenotypic subsets by ADT labeling strategy.....	70
Figure 25. Phenotypic assessment of scRNAseq-processed cells.....	72
Figure 26. Comparison of transcriptional and ADT-based phenotypic profiles in BM and PB scRNAseq datasets.....	75
Figure 27. Single cell <i>in vitro</i> differentiation assay of BM vs. PB HSPC.....	76
Figure 28. Expansion rate of BM and PB HSPC single-cell cultures.....	78
Figure 29. In vitro differentiation propensity of single-cell BM and PB HSPC.....	79
Figure 30. Analysis of cell cycle activity in scRNAseq dataset.....	81
Figure 31. Tile plot of the top 5 Gene Ontology-Biological Processes (GO-BP) macro-categories per cluster expressing NES<0 values (NESneg subgroup).....	83
Figure 32. Tile plot of the top 5 Gene Ontology-Biological Processes (GO-BP) macro-categories per cluster expressing NES>0 values (NESpos subgroup).....	84
Figure 33. Single cell expression heatmap of selected marker genes of cluster 1, which showed a differential expression between the two sources.....	86
Figure 34. Kinetics of cHSPC in WAS-, MLD- and MPSIH-GT patients before and after transplantation.....	88
Figure 35. Phenotypic composition of of WAS, MLD and MPSIH GT-patients before GT, at 30 days and 1 year after treatment.....	88
Figure 36. Circulation indexes (CI) in GT-treated patients before and after transplantation.....	89
Figure 37. Analyses of IS sharing between PB HSPC and BM HSPC derived from distinct BM sites (left and right) in 3 WAS patients at early and late phases after GT.....	90
Figure 38. Percentage of IS sharing between PB/BM CFC and PB myeloid and lymphoid mature lineages in 3 WAS patients at 1 year post-GT.....	91
Figure 39. Proposed functions of human cHSPC at steady state and during hematopoietic reconstitution.....	99

Table 1. Cohort of PB healthy donors of different ages included in the study.....	46
Table 2. Cohort of age-matched pediatric, adult and aged healthy donors of BM and PB samples.....	4
Table 3. Number of BM and PB cells composing transcriptional clusters identified in scRNAseq dataset.....	68
Table 4. Distribution of BM and PB cells among the diverse HSPC subsets identified by ADT barcoding.....	73
Table 5. Transcriptional and ADT phenotypic profiles detected for scRNA-seq single clusters.....	74
Table 6. Differentiation efficiency of BM- and PB-derived HSC, MPP, CMP, MEP, EP subsets at single cell level.....	77
Table 7. Cohort of GT-patients included in the study.....	101
Table 8. List of fluorescent antibodies for WBD phenotyping.....	103
Table 9. Hematopoietic populations identified by WBD analysis.....	103
Table 10. List of fluorescent antibodies used for cell sorting.....	106
Table 11. List of TotalSeq-A barcoded antibodies.....	106
Table 12. List of fluorescent antibodies used for HSPC staining of scRNAseq-processed samples upon thawing, before sorting and scRNAseq library preparation..	107
Table 13. Detailed list of donor ID, cell source, type of FACS staining applied before any processing, cellular fractions obtained after sorting, and TotalSeq-A antibody mix used for each scRNAseq-processed sample.....	108
Table 14. scRNAseq cell recovery and sequencing depth (median unique molecular identifier (UMI)/cells and median genes/cells) for Chromium 10x 3' GEX libraries..	109
Table 15. Median and interquartile values of UMI counts/cell for each ADT surface marker detected in Chromium 10x 3' ADT libraries.....	110
Table 16. Cytokine cocktail added to SFEM II medium for <i>in vitro</i> multilineage differentiation assay.....	118
Table 17. List of fluorescent antibodies used for HSPC sorting for single-cell multilineage differentiation assay.....	119
Table 18. List of fluorescent antibodies used for phenotypic characterization of cellular outputs at the end of multilineage differentiation assay.....	119
Table 19. Hematopoietic outputs identified by FACS analysis at the end of multilineage	

differentiation culture.....	120
Table 20. Number of IS retrieved from cHSPC, BM HSPC and PB mature lineages from GT-patients.....	125

1 INTRODUCTION

1.1 The human hematopoietic system

The hematopoietic system represents a complex and dynamic organization in which diverse cell types, differing from each other in terms of functions and properties, are in charge of providing multiple fundamental functions throughout the body, such as metabolite and oxygen transport, blood coagulation and immune surveillance. This integrated system displays a unique hierarchical structure in which differentiated short-term living cells are continuously produced from a pool of Hematopoietic Stem and Progenitor Cells (HSPC). The differentiation starts from multipotent primitive stem cells and, following a complex interplay of intrinsic and extrinsic factors, reaches its completion with the production of mature blood cells belonging to different lineages: myeloid, lymphoid, erythroid and megakaryocyte/platelet compartments. This process, called hematopoiesis, takes place in the bone marrow (BM) during adult life and provides constant turnover of blood cells to meet everyday demands and respond to potential increased needs, such as in case of injury or infections. In humans, the number of blood cells required every day for an adult individual is around one trillion (10^{12}) (Doulatov *et al*, 2012), therefore the production of mature cells needs to be finely regulated to maintain steady-state levels of the diverse blood cell lineages in the circulation.

Different blood cell types compose the myeloid compartment, including granulocytes, monocytes, macrophages and dendritic cells. These cells are principally in charge of sustaining innate immune responses, given their capability to rapidly react to chemokine gradients and be promptly recruited to specific locations upon pathogen invasion or tissue injury (Abbas & Lichtman, 2013; Parkin & Cohen, 2001).

Within the lymphoid compartment, B and T lymphocytes are the principal cellular types sustaining adaptive immunity, which is characterized by two essential features: high antigen-specificity (the ability to target specific pathogens) and memory (the ability to quickly respond to pathogens in case of following expositions). Importantly, while B lymphocytes originate and mature in the bone marrow, T-cell precursors

originate in the bone marrow and then migrate to the thymus, where their terminal maturation takes place (Abbas & Lichtman, 2013; Day & Schultz, 2010).

Erythrocytes and megakaryocytes are both generated in the bone marrow and show distinct peculiar functions: while the first are responsible of delivering oxygen from lungs to peripheral tissues, as well as carbon dioxide (CO₂) as waste product from the tissues back to the lungs, the second have the role of producing and releasing thrombocytes, or platelets, into the circulation, which are essential for blood coagulation (Deutsch & Tomer, 2006; Dzierzak & Philipsen, 2013).

1.2 Human hematopoietic stem and progenitor cells

As mentioned above, lifelong maintenance and production of blood and immune cell components is supported by HSPC, a heterogeneous group of cells comprised by primitive multipotent hematopoietic stem cells (HSC) and lineage-committed progenitors, which show distinct phenotypes and functional properties in terms of self-renewal capability, differentiation potential and long-term (LT-) survival.

1.2.1 Phenotypic definition of human HSPC

The identification of human HSPC is historically based on the use of CD34 cell surface marker, which is commonly used to enrich for HSPC fraction for transplantation and gene therapy (GT) purposes (Doulatov *et al*, 2012; Ferrari *et al*, 2021). During the last years, the combinatorial use of novel phenotypic markers has shed light on the complexity of HSPC compartment, allowing the identification of diverse subsets with distinct specific differentiating potential and long-term survival.

Human HSPC are phenotypically characterized by the expression of CD34 molecule and the absence of any lineage markers (LIN⁻) (**Figure 1**). CD38 marker expression is acquired along differentiation and is useful to discriminate lineage-committed progenitors (CD34⁺ LIN⁻ CD38⁺) from the most primitive ones (CD34⁺ LIN⁻ CD38⁻). In the primitive compartment, CD90 (Thy) and CD45RA molecules are commonly employed to identify multipotent subsets with diverse long-term survival and

differentiation potential: primitive hematopoietic stem cells (HSC) ($CD34^+ LIN^- CD38^- CD90^+ CD45RA^-$), which are enriched in cells displaying long-term engraftment potential and the highest multipotent repopulation activity, and are considered to be at the top of the hematopoietic hierarchy; multipotent progenitors (MPP) ($CD34^+ LIN^- CD38^- CD90^- CD45RA^-$), defined by multilineage potential but short-term engraftment capability; immature multi-lymphoid progenitors (MLP) ($CD34^+ LIN^- CD38^- CD90^- CD45RA^+$), which can give rise to all lymphoid cell types, as well as myelo-monocytic lineages (monocytes, macrophages and dendritic cells) (Doulatov *et al*, 2010). Of note, besides CD90 molecule, other surface markers could be used to enrich for primitive multipotent HSC populations, such as CD49f (Notta *et al*, 2011) and endothelial protein C receptor (EPCR, also known as CD201) (Fares *et al*, 2017).

Within the $CD38^+$ cell compartment, the combination of CD7, CD10, CD45RA, CD71 and CD41 surface markers is used for the definition of 6 different lineage-committed hematopoietic progenitors: lymphoid-committed subsets include early T progenitors (ETP) ($CD34^+ LIN^- CD38^+ CD7^+$) and precursor of B/NK cells (PreBNK) ($CD34^+ LIN^- CD38^+ CD7^- CD10^+ CD45RA^+$), while among myeloid/erythroid-committed progenitors, it is possible to identify common myeloid progenitors (CMP) ($CD34^+ LIN^- CD38^+ CD7^- CD10^- CD45RA^- CD41/61^- CD71^-$), granulocyte/monocyte progenitors (GMP) ($CD34^+ LIN^- CD38^+ CD7^- CD10^- CD45RA^+$), megakaryocyte/erythroid progenitors (MEP) ($CD34^+ LIN^- CD38^+ CD7^- CD10^- CD45RA^- CD41/61^+ CD71^+$), erythroid progenitors (EP) ($CD34^+ LIN^- CD38^+ CD7^- CD10^- CD45RA^- CD41/61^- CD71^+$), and megakaryocyte progenitors (MKP) ($CD34^+ LIN^- CD38^+ CD7^- CD10^- CD45RA^- CD41/61^+ CD71^-$) (Doulatov *et al*, 2012, 2010; Basso-Ricci *et al*, 2017; Notta *et al*, 2016; Miyawaki *et al*, 2017).

Interestingly, among recently identified markers, the expression of CD71 and BAH1 surface molecules on phenotypically defined human MPP, CMP and MEP, identified novel subsets with higher erythroid differentiation potential, uncovering phenotypic and functional heterogeneity within HSPC populations previously considered to be homogeneous (Notta *et al*, 2016). Moreover, CD41 positive expression could be used to isolate from conventional CMP a subset of human megakaryocyte progenitors (MKP), which represents the 8% of total CMP and is characterized by exclusive expression of

MK gene modules, typical MK morphology and strong MK differentiation potential both *in vitro* and *in vivo* (Miyawaki *et al*, 2017).

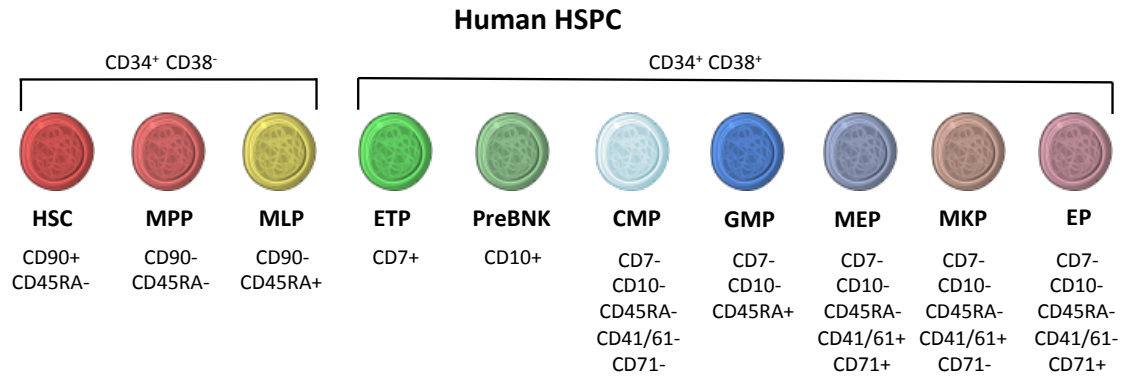


Figure 1. Combination of markers used for the identification of human HSPC subsets. All HSPC subpopulations are defined by the expression of CD34 and the absence of mature lineage (LIN) markers. Based on the expression of CD38 molecule, it is possible to distinguish between primitive subsets (CD38⁻) from committed progenitors (CD38⁺). HSC (Hematopoietic Stem Cells); MPP (Multipotent Progenitors); MLP (Multi-Lymphoid Progenitors); ETP (Early-T Progenitors), PreBNK (Precursors of B/NK cell); CMP (Common Myeloid Progenitors); GMP (Granulocyte-Monocyte Progenitors); MEP (Megakaryocyte-Erythroid Progenitors); MKP (Megakaryocyte Progenitors); EP (Erythroid Progenitors).

1.2.2 Cellular properties and functional features of human HSPC

The phenotypic heterogeneity within the HSPC pool is paralleled by distinct cellular properties in terms of self-renewal, replication rate and metabolism between the most primitive HSC and more committed progenitors.

HSC are historically defined on the basis of two essential properties: self-renewal, the ability to generate identical daughter cells without differentiating, and multipotency, the propensity of differentiating into all mature blood cell types. The balance between these two features allows HSC to continuously replenish the mature blood cell compartment, maintaining at the same time an adequate pool of stem cells throughout life. Based on their hematopoietic repopulation potential after transplantation into immune-deficient mice, the field now widely accept that HSC (from both humans and mice) able to repopulate for more than 16 weeks in a primary transplant and at least in a second round of transplantation can be defined as long-term (LT-) HSC. On the other

hand, HSC that can generate all lineages but are capable of transient engraftment are defined as short-term (ST-) HSC and MPP, characterized by progressive reduction of repopulation capacity and increasing cycling properties. On the other hand, progenitor cells are defined by the absence of extended self-renewal capability and a restricted differentiation potential (typically bi- or uni-lineage), and for these reasons they are generally lost early after transplantation (Doulatov *et al*, 2012; Laurenti & Göttgens, 2018).

Another key feature of HSC is quiescence, which is defined as a reversible absence of cell cycling or proliferation and corresponds to G_0 phase of the cell cycle. Consistently with their quiescent state, studies based on the detection of naturally occurring genomic modifications into human blood cells, such as X-chromosome inactivation ratio drift and telomere shortening with aging, allowed estimating that human HSC have ~ 1 replication per year during adult life (Scala & Aiuti, 2019). The time HSC spend in quiescence seems to be directly correlated with their heterogeneity in self-renewal capacity, since the most dormant cells display also the longest repopulation potential (Laurenti & Göttgens, 2018). Interestingly, the timing of exit from quiescence was described to be directly modulated by the expression levels of cyclin-dependent kinase 6 (CDK6), which regulates G_0 exit and early G_1 through the interaction with CyclinD. In particular, LT-HSC expressing low CDK6, both at mRNA and protein levels, are less prone to egress from G_0 than ST-HSC, which in turn express higher basal levels of CDK6 (Laurenti *et al*, 2015).

From a metabolic perspective, the increased quiescent state of HSC is associated with low levels of mitochondrial activity and strictly controlled rate of protein synthesis, while hematopoietic progenitors reside in a highly proliferative and metabolic state, which is dependent on oxidative metabolism and mitochondrial function (Laurenti & Göttgens, 2018). In line with this, it was shown that human cord blood (CB)-derived $CD34^+$ cells displaying a reduced mitochondrial mass ($CD34^+$ Mito^{low} fraction) were enriched in phenotypic $CD38^-$ cells with increased HSC function, as demonstrated by the high rates of *in vivo* hematopoietic reconstitution after transplantation into immune-deficient mice. Differently, high levels of mitochondrial mass ($CD34^+$ Mito^{high} fraction) functionally identified hematopoietic progenitors displaying reduced *in vivo* repopulation potential but increased *in vitro* proliferation capability (Romero-Moya *et*

al., 2013). Moreover, the hypoxic microenvironment of the BM niche can lead to the up-regulation of hypoxia-inducible factor-1 alpha subunit (HIF-1 α) in HSC, resulting in increased expression of glycolysis-stimulating targeted genes and reduced oxygen consumption and mitochondrial potential (Zhao & Li, 2015).

1.2.3 HSPC differentiation: distinct models of human hematopoietic hierarchy

Historically, blood production is believed to follow a strict hematopoietic “tree-like” hierarchy in which multipotency is progressively restricted. In the past years, the relationships among different hematopoietic populations have been object of many studies and several novel structures describing hematopoietic differentiation have been proposed.

The classical model of hematopoiesis, developed in mouse and applied also to the human system, sustains a stepwise hematopoietic development where a gradual progression from multipotent HSC to oligo-, bi- and uni-potent progenitors occurs (**Figure 2**). In this model HSC directly differentiate into MPP, which display a reduced self-renewal potential. Downstream MPP, the first branching point segregates common lineage precursors of myelopoiesis (CMP) and lymphopoiesis (MLP). In myeloid differentiation branch, oligopotent CMP undergo further restriction into bipotent GMP, which can generate granulocytes and monocytes, and MEP, that can differentiate into platelets and red blood cells, thus indicating CMP as a critical oligopotent progenitor from which myeloid, erythroid and megakaryocytic cells arise. On the other hand, in the lymphoid differentiation branch, MLP can give rise to bipotent PreBNK and unipotent ETP (**Figure 2**) (Seita & Weissman, 2010). However, although still largely used, this model is based on a general over-simplification of hematopoiesis, in which each subset is composed of a homogeneous set of cells with specific molecular properties.

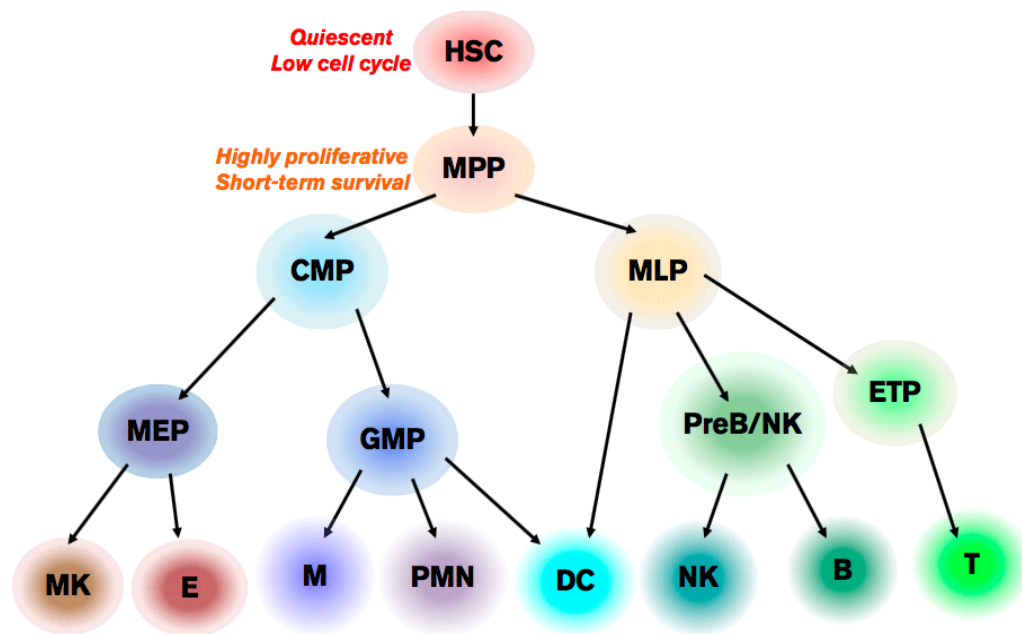


Figure 2. Classical model of human hematopoietic differentiation. In this model, hematopoiesis shows a tree-based hierarchical structure in which differentiation is depicted as a stepwise process with successive restrictions in differentiation potential. During this process, primitive multipotent subsets lose their “stem” properties to give rise to lineage-biased progenitors, and each HSPC subpopulation displays well-defined properties in terms of lineage differentiation potential and molecular features. MK (Megakaryocytes), E (Erythrocytes); M (Monocytes); PMN (Polymorphonucleates); DC (Dendritic Cells); NK (Natural Killers); B (B cells); T (T cells) (Scala & Aiuti, 2019).

Despite the differences existing in blood composition between human and mouse, it was initially assumed that human hematopoiesis would follow the same path of lineage specification described for mice. However, recent studies based on the integration of flow cytometric, transcriptional and functional data at single-cell level questioned the hierarchical models based on discrete populations for the human hematopoietic development (**Figure 3**). In particular, Velten and colleagues analyzed human BM HSPC compartment to quantitatively map early differentiation of human HSC. The authors proposed a model in which acquisition of lineage-specific fates is a continuous process, and unilineage-restricted cells emerge directly from a continuum of low-primed undifferentiated hematopoietic stem and progenitor cells (“CLOUD”)-HSPC within the $\text{Lin}^- \text{CD34}^+ \text{CD38}^-$ fraction. In the CLOUD, the developmental states downstream HSC, such as MLP and MPP, should be intended as the acquisition of transitory states within the HSPC continuum with higher probability of commitment to

specific lineages and cannot be explained by discrete cell populations at defined branching points. By contrast, once differentiation has progressed further, restricted progenitors with specific lineage potential start to appear, in association with the upregulation of the CD38 surface marker (Velten *et al*, 2017).

Importantly, the proposed “flow of differentiation” of the CLOUD-HSPC was sustained also by the existence of a continuous regulatory landscape at level of chromatin accessibility in early hematopoietic development leading to a broad range of allowable states, as described in the work by Buenrostro *et al*. (Buenrostro *et al*, 2018). In this study, combining single-cell epigenomic profiling through Assay for Transposase Accessible Chromatin with high-throughput sequencing (ATAC-seq) with single-cell RNA sequencing (scRNAseq) on 10 immunophenotypically defined BM HSPC subsets, the authors reconstructed a chromatin accessibility landscape of human hematopoiesis toward the diverse differentiation paths and correlated the changes in transcriptional factors expression to variations in chromatin accessibility. Of note, primitive HSC compartment showed higher expression of HOX motifs at chromatin level as well as HOX factors at transcriptional level, while motifs associated with master lineage regulators, such as ID3, CEBP and GATA1, displayed continuous gradients of activation toward lymphoid, myeloid and erythroid differentiation branches, respectively (Buenrostro *et al*, 2018). These data were in line with the transcriptional modules driving lineage commitment described in the study by Velten *et al*., where high expression of HOX transcriptional modules were associated to the least-primed low-cycling HSC compartment, while lineage-specific transcriptional signatures resulted to be gradually activated and reinforced along hematopoiesis to drive differentiation (Velten *et al*, 2017). Interestingly, genes from the first priming modules driving differentiation toward lymphoid/myeloid (FLT3/CEPB modules) and megakaryocyte/erythrocyte (GATA2/NFE2 modules) branches were shown to be expressed already at level of primitive HSC/MPP and to compete for ultimately directing them toward a specialized cell fate (Velten *et al*, 2017; Scala & Aiuti, 2019).

The continuum of hematopoietic differentiation process was sustained also by two further works focusing on specific groups of lineage-primed progenitor compartments (Karamitros *et al*, 2018; Psaila *et al*, 2016). Specifically, the combination of single-cell FACS index-sorting of human CB MLP ($\text{Lin}^- \text{CD34}^+ \text{CD38}^- \text{CD90}^{\text{neg-lo}} \text{CD45RA}^+$

CD10⁺), GMP (Lin⁻ CD34⁺ CD38⁺ CD123⁺ CD45RA⁺ CD10⁻) and lymphoid-myeloid primed progenitors (LMPP) (Lin⁻ CD34⁺ CD38⁻ CD90^{neg-lo} CD123⁺ CD45RA⁺ CD10⁻) with scRNAseq and functional assays, allowed the identification of a transcriptional continuum for the lympho-myeloid progenitors, where MLP tend to be positioned at one end, GMP at the second end and LMPP in the middle (Karamitros *et al*, 2018). A similar approach was applied to dissect the heterogeneity of erythroid/megakaryocytic differentiation branch and resulted in the identification of distinct subsets within human MEP population (Lin⁻ CD34⁺ CD38⁺ CD123⁻ CD45RA⁻), each characterized by unique phenotype and diverse degree of transcriptional priming toward specific lineage outputs: the “Pre-MEP/CMP” population (CD44^{high} CD71⁻ CD41⁻) enriched for bipotent progenitors with erythroid/megakaryocytic differentiation fate and residual myeloid potential, the “E-MEP” subset (CD71⁺ CD41⁻) with a strong bias toward erythroid specification, and the rare “MK-MEP” group (CD71⁺ CD41⁺), which was more committed toward MK differentiation (Psaila *et al*, 2016). Importantly, both studies (Karamitros *et al*, 2018; Psaila *et al*, 2016) abrogated the concept of discrete HSPC subpopulations and sustained for each cell a unique transcriptional profile along the multiplicity of states allowed in the hematopoietic differentiation trajectory.

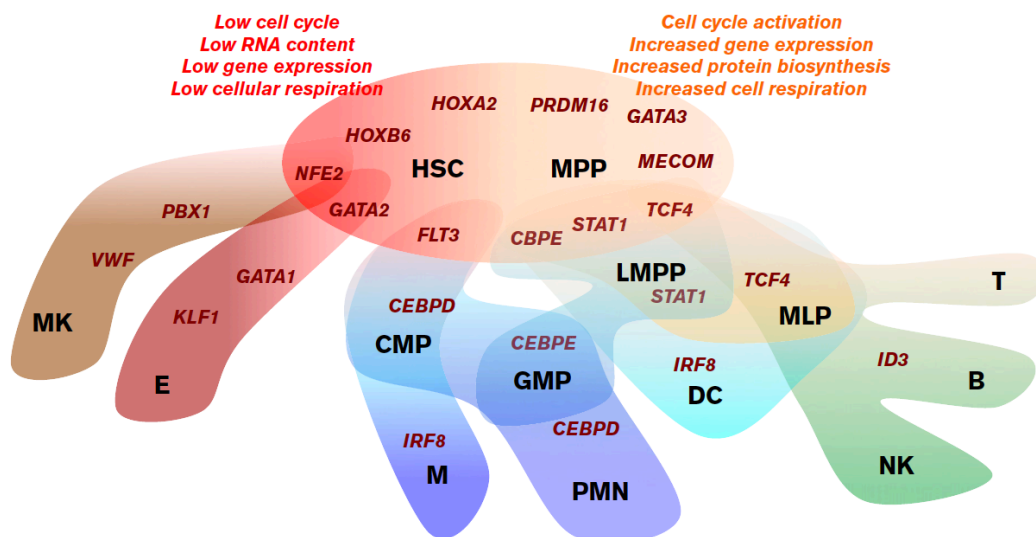


Figure 3. Human hematopoietic differentiation as a continuum process. Representation of the hematopoietic differentiation model based on recent literature. Hematopoiesis is described as a progressive process driven by gradual activation of key gene modules encoding for master lineage regulators. Lineage priming occurs already at level of primitive subsets and is enforced along differentiation. Major HSPC subpopulations are reported in black, while genes driving

specific hematopoietic differentiation are reported in dark red. LMPP (lymphoid-myeloid primed progenitor) (Scala & Aiuti, 2019).

1.2.4 Assays for studying human hematopoiesis and HSPC functions

Most of our understanding of the functional properties of human HSPC comes from studies based on *in vitro* assays and transplantation in animal models.

Colony-forming unit (CFU) assay, also known as colony-forming cell (CFC) assay, was one of the first to be developed for assessing the functionality of both murine and human HSPC (Bradley & Metcalf, 1966; Sarma et al, 2010). This assay is based on the ability of hematopoietic progenitors to proliferate and differentiate *in vitro* into colonies, both of myeloid and erythroid origin, in a semi-solid medium (methylcellulose) in response to cytokine stimulation. After two weeks of culture, the generated colonies can be enumerated and characterized according to their unique morphology, allowing the quantification of different types of progenitors, including the burst-forming unit-erythroid (BFU-E), colony forming unit-granulocytic and monocytic (CFU-GM) and colony forming unit-granulocytic, erythroid, monocytic and megakaryocytic (CFU-GEMM). However, despite being the most commonly used *in vitro* assay, it shows intrinsic limitations due to the impossibility of assessing the differentiation potential of lymphoid progenitors. Therefore, assays based on *in vitro* liquid culture of human HSPC, either alone or in co-culture with stromal cells, were developed to support differentiation into not only myelo/erythroid lineages (Notta *et al*, 2016), but also into B/NK (Yoshikawa *et al*, 1999) or T cells (La Motte-Mohs *et al*, 2005; Schmitt & Zúñiga-Pflücker, 2006). In these settings, depending on the combination of cytokines composing the medium and on the type of stromal cells used, it is possible to drive HSPC differentiation into specific cell fates, typically with an uni- or bi-lineage differentiation output.

However, given the over-simplification of *in vitro* assays in terms of setting, duration and components as compared to the complexity of a living organism, to date the best way for extensively studying human hematopoiesis is based on humanized xenotransplantation in immune-compromised mice. This setting provides a unique tool to assess the *in vivo* engraftment capability, repopulation potential and differentiation

outputs of human HSPC. One of the most currently used mouse strain is non-obese diabetic/severe combined immunodeficiency (NOD/SCID) model with either truncation or a deletion in the IL-2R common γ chain, known as NSG model, which is characterized by a complete loss of all lymphoid lineages in order to allow successful engraftment of human cells with poor risk of rejection (Ito *et al*, 2002). However, although showing good levels of engraftment of human CD45⁺ cells after transplant, there remain some limitations that prevent full exploitation of NSG model for the study of human hematopoiesis, including the poor production of human platelets, myeloid and erythroid cells observed overtime after transplantation (Brehm *et al*, 2013). Therefore, innovative mouse strains were generated to allow an improved myelo/erythroid reconstitution from engrafting cells in the host. One example is represented by KIT-deficient NOD/SCID Il2rg⁻/Kit^{W41/W41} (NSGW41) mouse model, generated by introducing a loss-of-function mutation in KIT receptor into NSG background. This model is capable of supporting efficient engraftment and differentiation of human HSPC (including myeloid and erythroid lineages) in a physiological-like setting, without any stress due to irradiation before transplant (Cosgun *et al*, 2014; Rahmig *et al*, 2016).

1.3 Studying human hematopoiesis through HSPC gene therapy

1.3.1 *Ex vivo* HSPC gene therapy

Ex vivo HSPC GT is based on the genetic modification of autologous HSPC to correct monogenic disorders or provide novel features to hematopoietic cells for treating infectious diseases or cancer (Naldini, 2011).

HSPC GT clinical protocol consists in a first step of collection of autologous HSPC, followed by gene transfer aimed at integrating a corrected copy of the altered gene in the recipient's genome. This method exploits either cell transduction through viral vectors as delivery platforms, or, more recently, gene editing approaches. After *ex vivo* manipulation, corrected cells are transplanted back into the patient, who previously underwent a conditioning regimen to deplete endogenous HSPC and make space in the

BM for the graft (**Figure 4**). Through this approach, the *ex vivo* genetically corrected HSPC would repopulate the recipient and reconstitute all blood lineages.

BM harvest and mobilized peripheral blood (MPB) are the most frequently employed biological sources for human HSPC GT. BM harvest is a procedure performed under general anesthesia and consists in multiple marrow aspirates along the posterior iliac crests (Panch *et al*, 2017). Over the last decades, leukapheresis following the administration of mobilizing agents has become, in most cases, the preferred HSPC source for patients undergoing autologous/allogeneic transplantation or gene therapy (de Kruijf *et al*, 2019; Panch *et al*, 2017), due to higher numbers of stem cells collected, rapid reconstitution of hematopoiesis, and reduced rates of morbidity and recovery time for the donor (Gulbas, 2018; Patel *et al*, 2015; de Kruijf *et al*, 2019).

Independently from the source, it is now well established that HSPC can be efficiently gene modified to continuously produce a cell progeny expressing the therapeutic gene, while maintaining the ability to engraft in the long term, for up to 13 years after treatment (Cicalese *et al*, 2016).

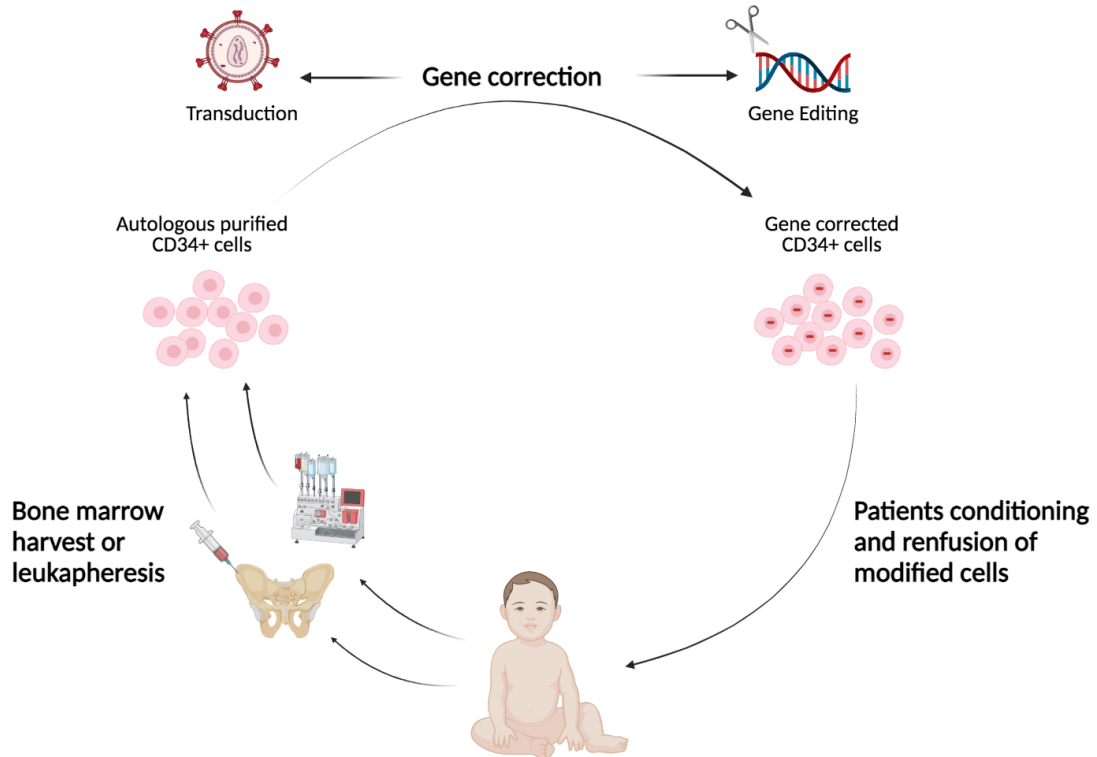


Figure 4. Outline of *ex vivo* HSPC gene therapy. Autologous HSPC are collected from the patient through bone marrow harvesting or leukapheresis following the administration of

mobilizing agents. Then, HSPC undergo gene-correction, by exploiting either transduction with integrating viral vectors containing the target gene or gene editing strategies. Before reinfusion of corrected cells, a conditioning regimen can be administered to the patient to deplete endogenous HSPC and make space in the BM for the graft.

Nowadays more than 400 patients have been treated in clinical trials applying *ex vivo* HSPC gene therapy, with robust and consistent results in terms of long-term durability, efficacy and safety of the treatment (Tucci *et al*, in press).

Primary immunodeficiencies (PIDs) are inherited disorders characterized by a compromised immune system and represent the first group of diseases that was targeted by vector-based GT curative approaches. Starting from early 90s, several clinical trials were developed, aiming at correcting a defective function in different PIDs, including X-linked severe combined immunodeficiency (SCID-X1) (Gaspar *et al*, 2004; De Ravin *et al*, 2016; Hacein-Bey-Abina *et al*, 2014), adenosine deaminase (ADA)-SCID (Aiuti *et al*, 2002; Cicalese *et al*, 2016), chronic granulomatous disease (CGD) (Ott *et al*, 2006) and Wiskott-Aldrich syndrome (WAS) (Aiuti *et al*, 2013; Ferrua *et al*, 2019; Braun *et al*, 2014). Of note, Strimvelis, the γ -retroviral vector (γ -RV)-GT medicinal product for ADA-SCID, was the first GT-based treatment to be approved by the European Medicines Agency (EMA) in 2016 (Aiuti *et al*, 2017).

HSPC-GT has also the ability to turn the hematopoietic progeny into hyper-functional enzyme factories capable of more effectively cross-correcting non-hematopoietic cells. This feature, together with the capability of corrected HSPC and myeloid blood cells to penetrate the central nervous system (CNS) and to locally differentiate into tissue-resident cells, was successfully exploited in the treatment of non-hematological inherited neurometabolic disorders (Ferrari *et al*, 2021). These diseases are caused by mutations in genes encoding for peroxysomal (in X-linked Adrenoleukodystrophy, X-ALD) or lysosomal (in metachromatic leukodystrophy, MLD, and mucopolysaccharidosis type I, MPSI) enzymes, resulting in the accumulation of toxic substrates in different organs, including bones and CNS. The first clinical application of lentiviral vector (LV)-mediated HSPC GT was performed in two patients affected by X-ALD, demonstrating neurological benefits after treatment (Cartier *et al*, 2009). In the context of MLD disorder, the first clinical trial based on LV HSPC-GT displayed safety and efficacy in most patients treated at a pre-symptomatic or very early

symptomatic stage (Biffi *et al*, 2013; Sessa *et al*, 2016; Fumagalli *et al*, in press), and the associated medicinal product has recently received positive opinion from EMA (Ferrari *et al*, 2021). Furthermore, promising results were recently reported for a phase 1-2 LV HSPC-GT trial for MPSIH, especially in terms of early signs of amelioration of patients' phenotype (Gentner *et al*, 2021; Bernardo ME *et al*, 2020).

HSPC-GT has shown its potential also when applied to β -hemoglobinopathies, including β -thalassemia (β -THAL) and sickle cell disease (SCD). These pathologies are caused by distinct types of mutations in β -globin (*HBB*) gene, that result in either a detrimental imbalance between α -globin and β -globin chains (in β -THAL), or in the production of a toxic hemoglobin variant (in SCD). Over the last decades, multiple LV-gene therapy trials have been started for both types of disease, demonstrating good levels of clinical efficacy and benefit for the patients (Markt *et al*, 2019; Thompson *et al*, 2018; Ribeil *et al*, 2017). Importantly, in 2019 the EMA approved the first LV-GT treatment (Zynteglo) for patients with transfusion-dependent β -THAL without severe β^0/β^0 -genotype (Ferrari *et al*, 2021).

Finally, another clinical target of HSPC-GT is Fanconi Anemia (FA). This genetic disease represents the most frequent inherited bone marrow failure syndrome and is caused by mutation in any of the 22 FA genes cooperating in pathways of DNA inter-strand cross-links repair (Bueren *et al*, 2020). In a recent clinical study, FA patients carrying mutations in *FANCA* gene were treated with LV-HSPC GT, without any conditioning to avoid drug toxicity. Patients displayed engraftment and proliferation advantage of gene-corrected HSPC, which were endowed with acquired resistance to DNA crosslinking reagents *in vitro*. Furthermore, an arrest of the BM failure progression was detected in patients with the highest levels of engraftment of corrected cells (Río *et al*, 2019).

Overall, gene therapy represents a promising clinical strategy showing good corrective capability. The development of novel GT-based therapeutic approaches, including gene editing, together with the improvement of HSPC collection and engineering, could allow to broaden the application of gene and cell therapies to genetic diseases with unmet medical need.

1.3.2 Investigating human HSPC dynamics by integration sites clonal tracking

Vector-based gene therapy approaches exploit the ability of stably integrating the therapeutic gene within the host genome, in order to permanently correct a genetic defect in the target cells. The insertion of exogenous DNA in the host genome could lead to aberrant transcriptional profile, an event known as insertional mutagenesis. For this reason, integration sites (IS) analyses has become a potent tool for monitoring the safety and efficacy of the treatment in the context of GT clinical trials (Ferrari *et al*, 2021). At the same time, since upon cell transduction a substantial fraction of the engrafting stem cells and their differentiated progeny is marked by a unique vector IS, it is also possible to exploit IS clonal tracking for addressing biological questions on the behavior of human HSPC after transplantation. In particular, tracing IS, it is possible to monitor the fate and clonal output of infused corrected HSPC in distinct phases after treatment (during early recovery and at steady state), as well as to model hematopoietic hierarchy by analyzing the sharing of IS among HSPC and multiple differentiated lineages. In this view, patients treated with genetically repaired HSPC may represent an extraordinary unique model to study hematopoiesis directly *in vivo* in humans.

Up to now, few cutting-edge studies exploited longitudinal analyses based on IS retrieval from GT-treated patients to unveil the complexity and the behavior of the hematopoietic system after transplantation (Six *et al*, 2020; Biasco *et al*, 2016; Scala *et al*, 2018). A recent analysis on 6 HSPC-GT patients with two different disease backgrounds (4 WAS patients and 2 patients affected by β -hemoglobinopathies), revealed the heterogeneity in differentiation programs activated in gene-corrected HSPC after transplantation. Specifically, by tracking clonal lineage output of human HSPC at long-term follow-ups (>2 years post-GT), the authors identified in each patient a coexistence of myeloid-dominant, lymphoid-dominant and multilineage corrected HSPC subsets, each contributing to long-term hematopoiesis after GT (Six *et al*, 2020).

Moreover, by combining phenotypic and functional characterization with high-throughput IS clonal tracking, in the last years our group put major efforts in the dissection of the dynamics of distinct HSPC subpopulations overtime after GT (Biasco *et al*, 2016; Scala *et al*, 2018). In particular, it was observed that while in the first 6 months after GT a consistently high number of clones is able to sustain hematopoietic

recovery, at steady state only a few thousands of clones are in charge of hematopoietic maintenance and about 1 in 10^5 - 10^6 gene-corrected HSPC were estimated to be endowed with long-term engraftment potential. These data, sustaining two distinct clonal waves, may suggest an active role of short-term progenitors at early phases, followed by the gradual takeover of primitive HSC in the long term (Biasco *et al*, 2016). Indeed, this model was in line with recent results from Scala *et al.*, who by isolating 7 distinct HSPC subsets from six WAS patients both at early (<12 months) and late (>24 months) phases after GT treatment, generated unprecedented information on the role of HSPC subpopulations overtime up to 5 years of follow up (Scala *et al*, 2018). Specifically, we described that early after transplantation, committed myeloid progenitors contained in the initial graft (CMP/GMP) mainly contribute to the production of mature myeloid lineages and then get exhausted, while multilineage reconstitution is sustained by short-term HSC and MPP components. Differently, around 1-2 years after transplant, once steady-state hematopoiesis is established, long-living HSC become in charge of maintaining long-term hematopoietic production. These results also suggested that long-term HSC, which were activated *in vitro*, were characterized by a delayed period of activation after re-infusion and homing into the BM (Scala *et al*, 2018).

From a clinical point of view, the information generated from these studies suggested that efficacious transplantation protocols should preserve both primitive and committed progenitors, in order to achieve long-term graft maintenance and sustain early hematopoietic recovery by highly proliferating progenitors, thus preventing rapid exhaustion of primitive HSC.

1.4 The BM niche

Bone marrow is a peculiar organ with a unique microanatomical structure and is composed of several hematopoietic and non-hematopoietic cell types. These populations are interconnected by vascular and innervated networks within the bone cavities, defining the BM hematopoietic niche. From a functional point of view, the BM niche provides molecular cues and physical interactions for the regulation of HSC

properties, including localization, quiescence, maintenance and lineage differentiation. While in mice all bones support hematopoiesis, in humans the axial skeleton (composed by cranium, sternum, ribs, vertebrae and ilium) is the principal site of hematopoiesis (Yu & Scadden, 2016; Pinho & Frenette, 2019; Birbrair & Frenette, 2016).

In the murine setting, HSC with different cell cycle activity have been found associated to diverse local BM anatomical compartments. In particular, quiescent HSC with higher *in vivo* homing, lodgment and reconstitution potential, were described to be preferentially associated with *periarteriolar niche*, which consists of small arterioles located in the endosteal region. By contrast, upon cell cycle activation, HSC tend to redistribute away from arterioles reaching the *perisinusoidal niche*, which is closer to the center of the marrow cavity and is connected to the central venous sinus (Kunisaki *et al*, 2013). This organization is not casual: while dormant HSC need a finely regulated microenvironment, proliferating HSPC, which will eventually become mature blood cells, need to be in closer contact with the permeable barrier of sinusoids to facilitate cell egress from the BM niche (Birbrair & Frenette, 2016; Itkin *et al*, 2016).

1.4.1 Cellular and molecular components of the BM hematopoietic niche

As mentioned above, BM niche is a heterogeneous mixture of both hematopoietic and non-hematopoietic cell types, as well as multiple niche factors that are essential regulators of HSC functions (**Figure 5**).

Among the non-hematopoietic cellular components, ***osteoblasts*** were described as critical modulators of HSPC properties in the BM, since they are capable of supporting their expansion *in vitro* and of increasing their engraftment rate *in vivo* in co-transplantation settings. Osteoblasts are located in the endosteal region and can produce a wide range of molecules implicated in HSC maintenance, including osteopontin (OPN), a negative regulator of HSC pool size, thrombopoietin (TPO) and angiopoietin-1 (ANGPT-1), both regulating HSPC quiescence (Birbrair & Frenette, 2016; Pinho & Frenette, 2019).

Mesenchymal stromal cells (MSC) are rare primitive self-renewing cells that are able to differentiate into osteolineage cells, chondrocytes and adipocytes and show hematopoiesis-supporting functions, by providing both direct physical support and

different soluble factors that tightly control HSPC fate. MSC are commonly defined as perivascular cells since they are generally located in close proximity of BM blood vessels, and can be classified in different subtypes according to their anatomical localization and expression of distinct surface markers. In particular, $CD271^+$ and $CD271^+/CD146^{/low}$ MSC have been described as bone-lining cells associated with LT-HSC in low oxygen areas, whereas $CD146^+$ and $CD271^+/CD146^+$ MSC are located in the sinusoidal region in proximity to activated and fast-proliferating HSC. *Nestin⁺ perivascular MSC*, which are associated with parasympathetic nerve fibers, release several key factors involved in HSC maintenance and regulation, such as stromal cell-derived factor-1 (SDF-1) (also known as C-X-C Motif Chemokine Ligand, CXCL12), stem cell factor (SCF), ANGPT-1, OPN, Interleukin-7 (IL-7) and vascular cell adhesion molecule-1 (VCAM-1), and can be further classified in two distinct subsets based on the level of expression of Nestin: while $Nestin^{dim}$ cells are ubiquitously distributed near sinusoids, $Nestin^{bright}$ cells are enriched in periarteriolar regions and represent an important source of SDF-1, but not SCF, in the niche. *CXCL12-abundant reticular cells (CAR cells)* are adipo-osteogenic progenitors that are mainly distributed around sinusoids and are major producers of SDF-1 and SCF, essential for HSC maintenance *in vivo*, as well as IL-7, required for the maintenance of lymphoid progenitors and mature B cells. Furthermore, *leptin receptor-expressing MSC (Lepr⁺ MSC)* identify primitive BM MSC regulating HSC maintenance in BM, since their depletion was shown to be paralleled by reduction in quiescent HSC. Of note, approximately 90% of $Lepr^+$ MSC phenotypically overlap with CAR cells, while nearly 80% of them express also the Nestin marker, thus pointing out a clear redundancy among the different molecular markers which are commonly employed to identify the distinct MSC populations (Crippa & Bernardo, 2018; Pinho & Frenette, 2019; Birbrair & Frenette, 2016; Yu & Scadden, 2016).

Adipocytes appear dispersed within the BM and have a smaller size than visceral and subcutaneous counterparts. Their number and dimension tend to increase with aging, resulting in progressive replacement of hematopoietic sites with fatty yellow marrow with reduced hematopoietic activity (Pinho & Frenette, 2019; Yu & Scadden, 2016). It was observed that the frequency of HSPC is reduced in adipocyte-rich vertebrae of the mouse tail with respect to the adipocyte-poor thoracic vertebrae (Naveiras *et al*, 2009).

Moreover, in lipotrophic A-ZIP/F1 ‘fatless’ mice, which are genetically deficient in adipogenesis, and in mice treated with a peroxisome proliferator-activator receptor γ (PPAR γ) antagonist, which pharmacologically inhibits adipogenesis, BM engraftment and recovery after lethal irradiation and BM transplantation was enhanced as compared to wild-type or untreated mice (Naveiras *et al*, 2009). Altogether, these data suggest adipocytes as predominantly negative regulators of HSC functions and BM microenvironment.

BM **endothelial cells** compose the inner cellular lining of blood vessels and release several paracrine growth factors and cytokines regulating HSPC activity and maintenance, such as Notch ligand, SDF-1, SCF and pleiotrophin (Pinho & Frenette, 2019). Moreover, they express a broad range of adhesion molecules, such as E-selectin, P-selectin and VCAM-1 (Yu & Scadden, 2016), that are responsible for the retention of HSPC within the BM niche. In addition, arterioles versus sinusoids can modulate the levels of reactive oxygen species (ROS) in neighboring HSPC, thus influencing their activity and localization in the BM space. In particular, while HSPC in proximity to less permeable arteriolar endothelial cells show low ROS content and are quiescent, HSPC residing close to more permeable sinusoid endothelial cells have increased ROS levels, that result in enhanced HSPC activation, differentiation and migration (Itkin *et al*, 2016).

Sympathetic nerves penetrate within BM tissue and BM stromal cells were described as the main targets of sympathetic stimulation. However, how the adrenergic signaling could directly influence HSC function still remains poor understood. It has been shown that pharmacological or genetic ablation of BM adrenergic neurotransmission resulted in suppression of HSC migratory response to granulocyte colony stimulating factor (G-CSF) treatment, suggesting an indirect regulatory role in HSC mobilization and trafficking (Pinho & Frenette, 2019; Yu & Scadden, 2016). Moreover, **Schwann cells**, unmyelinated glial cells that insulate sympathetic and sensory nerve fibers within the BM, can contribute to the maintenance of HSC quiescence and hibernation through the release of transforming growth factor- β (TGF β) and the downstream activation of SMAD signaling. BM denervation resulted in reduced number of cells producing active TGF β , thus leading to HSC loss due to increased proliferation (Yamazaki *et al*, 2011).

Megakaryocytes were described to promote HSC quiescence by releasing TPO and

TGF β factors, and to negatively regulate HSC proliferation through the production of chemokine C-X-C Motif Ligand 4 (CXCL4). At the same time, under stress conditions megakaryocytes are able to release fibroblast growth factor-1 (FGF-1), which can induce HSC expansion and BM repopulation. Importantly, megakaryocyte ablation was shown to reduce HSC engraftment and proliferation (Pinho & Frenette, 2019; Birbrair & Frenette, 2016).

Macrophages promote HSC retention in the BM by directly expressing very late antigen 4 (VLA-4) and VCAM-1 adhesion molecules, or through indirect mechanisms involving the induction of SDF-1 expression by Nestin⁺ MSC. The crosstalk between BM macrophages and MSC may be mediated by the soluble factor oncostatin M (OSM). Moreover, a rare population of macrophages expressing high levels of α -smooth muscle actin was shown to localize next to HSC in the BM and to actively contribute to the reduction of ROS production under stress, resulting in protection of HSC from exhaustion (Pinho & Frenette, 2019; Birbrair & Frenette, 2016).

Serine proteases produced and released by **neutrophils** are capable of cleaving cytokines and receptors essential for HSPC retention *in vitro*, including CXCL12, chemokine C-X-C Motif receptor 4 (CXCR4), VCAM-1, c-KIT and SCF, suggesting that activated neutrophils can create a proteolytic microenvironment that may contribute to HSPC release from the BM (Birbrair & Frenette, 2016).

Regulatory T cells (T_{reg}) promote the survival of allogeneic HSC after transplantation by secreting the immunoregulatory cytokine interleukine-10 (IL-10). Depletion of T_{reg} expressing forkhead box P3 (FOXP3) transcription factor can lead to rapid loss of allogeneic HSC, suggesting that T_{reg} cells are capable of endowing the HSC niche with immune privilege (Pinho & Frenette, 2019).

In summary, multiple cell populations and soluble molecules regulate HSPC functional properties within the BM niche. Nevertheless, the combinatorial effects of all these factors remain still not fully elucidated. Future clarifications on the interactions between HSPC and the complex BM microenvironment are thus required, as they could lead to improved methods for extensively exploiting the clinical potential of HSPC.

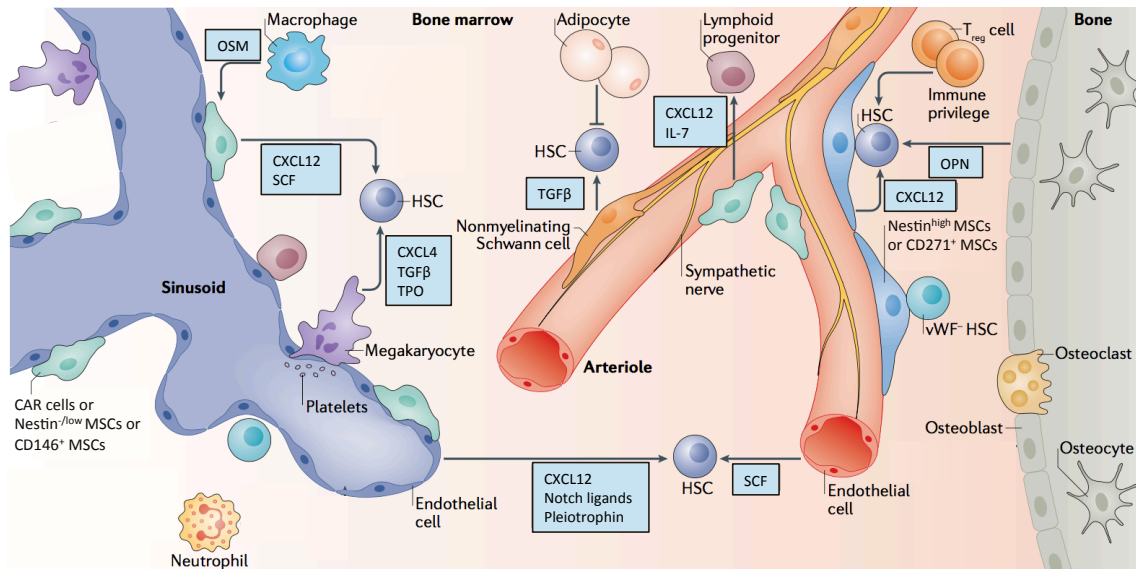


Figure 5. The adult BM niche in homeostatic conditions. Schematic representation of adult BM niche microenvironment. Sinusoid and arteriole structures define a compartmentalization into perisinusoidal and periarteriolar niches, which are respectively populated by cycling or quiescent HSC. The main molecular axes and cellular components that play direct or indirect roles in the regulation of HSC functions are reported. OSM (Oncostatin M); SCF (Stem Cell Factor); CXCL12 (C-X-C Motif Chemokine Ligand 12); CXCL4 (C-X-C Motif Chemokine Ligand 4); TGFβ (transforming growth factor β); TPO (Thrombopoietin); IL-7 (Interleuchine-7); OPN (Osteopontin); T_{reg} (T regulatory cell); vWF (Von Willebrand Factor); MSC (Mesenchymal Stromal Cell). Adapted from Pinho & Frenette, 2019.

1.4.2 Pivotal molecular axes controlling HSPC maintenance, retention, homing and egress in and out of the BM niche

A fine combination of chemokines, growth factors, adhesion molecules and signaling pathways is required to maintain HSPC physiological homeostasis and to strictly regulate their retention, recruitment and egress in and out of the BM niche (**Figure 6**).

In the endosteal region, the expression of many adhesion molecules by both HSC and stromal cells is fundamental to allow a tight retention of quiescent HSC. The interaction between SCF and c-Kit receptor (also known as CD117), expressed by HSPC, has a crucial role for HSPC long-term maintenance and self-renewal. Of note, the membrane-bound form of SCF (membrane SCF, mbSCF) is expressed by osteoblasts and can engage a more sustained interaction with c-Kit receptor than soluble SCF (Suárez-Álvarez *et al*, 2012). ANGPT-1, expressed by osteoblasts, can interact with Tie2

tyrosine kinase receptor present on LT-HSC, inducing a quiescent state and a firm adhesion to the bone (Suárez-Álvarez *et al*, 2012).

A critical molecular axis playing a pivotal role in HSPC retention, mobilization and chemotactic recruitment into the BM is dependent on the interaction between CXCR4 receptor, expressed on HSPC surface, and its ligand SDF-1 (Zhao & Li, 2015; Suárez-Álvarez *et al*, 2012). SDF-1 was historically identified as the first chemoattractant reported for human CD34⁺ cells (Aiuti *et al*, 1997) playing an essential role in HSPC homing and retention. Of note, during HSPC homing to the BM, the interaction of CXCR4 with endothelium-bound SDF-1 can induce a high affinity state of membrane-borne HSPC integrins, such as VLA-4 (also known as integrin- $\alpha_4\beta_1$ or CD49d/CD29) and leukocyte function antigen 1 (LFA-1, also known as integrin- $\alpha_L\beta_2$ or CD11a/CD18), which are able to respectively recognize and bind to the cellular adhesion molecules VCAM-1 and ICAM-1, expressed at high levels on the BM endothelium. These bindings are essential for promoting transmigration, recruitment and successive retention of HSPC within the BM through strong adhesive interaction (Buffone *et al*, 2018).

HSPC also express integrin $\alpha_9\beta_1$, which plays a role in human HSPC adhesion to BM osteoblasts through the interaction with OPN, tenascin-C (TN-C) and VCAM-1 molecules (Schreiber *et al*, 2009). Moreover, integrin $\alpha_6\beta_1$ was demonstrated to affect human HSPC adhesion to laminin-10/11 and laminin-8, expressed by sinusoidal subendothelial basement membranes (Gu *et al*, 2003), as well as murine HSPC homing to BM after transplantation (Qian *et al*, 2006).

Interestingly, the expression of the multifunctional receptor CD44 on HSPC can induce diverse functional effects depending on the specific ligand recognized: 1) the interaction with E-selectin, expressed by BM sinusoidal cells, is essential to allow selectin-mediated braking and rolling during the initial steps of homing from the circulation to the BM niche (Buffone *et al*, 2018); 2) the binding of OPN, released by osteoblasts, can result in a negative regulation of HSPC pool size within the BM (Suárez-Álvarez *et al*, 2012); 3) the interaction with BM hyaluronic acid (HA), broadly expressed in the endosteum and by endothelium lining sinusoids, was shown to positively regulate not only HSPC BM homing and engraftment, but also SDF-1-induced HSPC adhesion within the niche (Avigdor *et al*, 2004).

In summary, multiple axes cooperate in regulating HSPC maintenance and retention in the BM. Further dissections of the diverse molecular factors involved would be fundamental for better elucidating the complex cellular properties of HSPC, both inside and outside the BM niche.

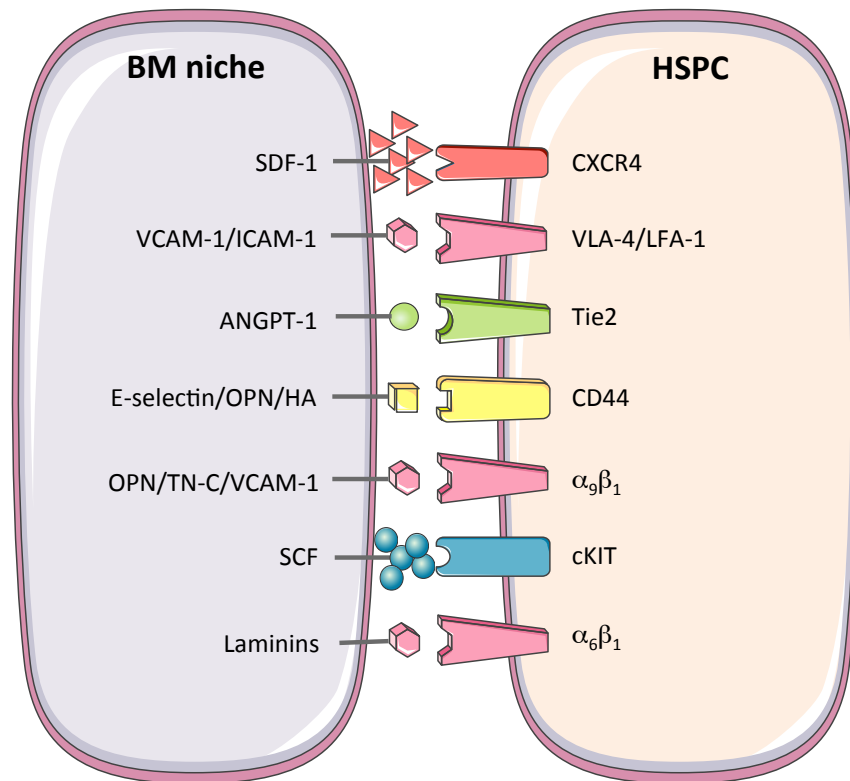


Figure 6. Molecular axes controlling BM HSPC homing, retention and egress. Schematic representation of the principal molecular axes tuning HSPC retention or egress into or from the BM niche. Distinct receptors expressed on HSPC surface interact with secreted factors or adhesion molecules produced by diverse niche components. The integration of all these signals maintains HSPC within the BM niche, where they can exert their hematopoietic supporting function, while disruption of one or multiple signaling can drive HSPC egress from the BM and recirculation. SDF-1 (stromal-cell derived factor 1); CXCR4 (Chemokine (C-X-C Motif) ligand 4); VCAM-1 (Vascular Cell Adhesion Molecule 1); ICAM-1 (Intercellular Adhesion Molecule); VLA-4 (Very Late Activation Antigen); LFA-1 (Leukocyte Function Antigen-1); ANGPT-1 (Angiopoietin-1); HA (Hyaluronic Acid); TN-C (Tenascin-C).

1.4.3 Current strategies for HSPC mobilization

From a clinical standpoint, mobilization may be defined as an enforced pharmacologically-induced release of HSPC from the BM into the PB stream and can be achieved through the disruption of adhesive interactions in the BM niche between stromal cells and HSPC (Bozdag *et al*, 2015).

Recombinant G-CSF is the most common mobilizing agent currently used in the clinical practice. G-CSF mechanism of action is based on the engagement of G-CSF receptor expressed on the surface of myeloid cells, which triggers both cell expansion and massive release of active neutrophils proteases, such as Cathepsin G, elastase, matrix metalloproteinase-9 (MMP-9) and cell surface protease dipeptidyl peptidase-4 (DPP-4, also known as CD26), thus generating a highly proteolytic environment in the BM. Proteases additively cleave multiple anchorage molecules, including VCAM-1, c-Kit, CXCR4 and SDF-1, enabling HSPC dislodgement from the BM stroma and subsequent egress into the circulation. In addition, G-CSF was shown to exert a protease-independent mechanism by reducing BM osteoblast activity, as well as SDF-1 mRNA expression (Albakri *et al*, 2020; Suárez-Álvarez *et al*, 2012). During mobilization procedures, G-CSF is administered daily for up to 6 days and apheresis is performed not before the fifth day after four consecutive days of treatment, when CD34⁺ levels start to peak (Domingues *et al*, 2017; Bozdag *et al*, 2015). However, despite being well tolerated, approximately 5-30% of G-CSF-treated patients fails to mobilize adequate numbers of HSPC to reach the minimum dose for an efficient transplantation (Domingues *et al*, 2017), thus calling the need of improved alternative mobilizing agents. One of these is Plerixafor (also known as AMD3100), a selective and reversible CXCR4 antagonist approved by the Food and Drugs Administration (FDA). Plerixafor is administered in combination with G-CSF (5 days of G-CSF at 10ug/kg plus AMD3100 at 240ug/kg on day 5) in order to synergistically enhance HSPC mobilization in poor mobilizers or in the context of gene therapy trials (Hübel, 2019; Eichler *et al*, 2017; Markt *et al*, 2019; Thompson *et al*, 2018), showing a 2-3 fold increase in MPB CD34⁺ cell concentration than G-CSF alone (de Kruijff *et al*, 2019; DiPersio *et al*, 2019).

Since the introduction of Plerixafor, several other promising classes of CXCR4 antagonists have been identified, such as the peptides POL6326, BKT-140 and LY2510924 and the small molecules TG-0054 and ALT-1188. These agents are currently at different stages of clinical and pre-clinical development and were shown to induce a mobilization response either alone or, in some cases, in combination with G-CSF (Domingues *et al*, 2017).

As extensively discussed in the previous section, HSPC express a broad range of integrins that have key roles in modulating their retention in the BM microenvironment, thus highlighting integrin antagonists as promising mobilization agents. Among these, VLA-4-antagonists have been revealed as good candidates for enhancing the mobilization response. In particular, the small molecule BIO5192 was described as one of the first VLA-4-inhibitor able to stimulate HSPC mobilization, either alone or in combination with G-CSF or Plerixafor (Ramirez *et al*, 2009). In addition, the dual $\alpha_4\beta_1$ - $\alpha_9\beta_1$ antagonist N-(benzenesulfonyl)-l-prolyl-l-O-(1-pyrrolidinylcarbonyl)tyrosine (BOP) was reported as an efficient, rapid, single dose-based mobilizer of long-term HSPC (de Kruijf *et al*, 2019; Domingues *et al*, 2017).

Another promising target for HSPC mobilization is C-X-C Motif Chemokine Receptor (CXCR2), which is expressed on the surface of neutrophils and BM endothelial cells, but not HSPC (DiPersio *et al*, 2019). The chemokine Gro- β , a ligand of the chemokine receptor CXCR2, and its truncated human recombinant analogue SB-251353, were shown to rapidly mobilize HSPC when used alone or in synergistic combination with G-CSF in mice and rhesus monkeys (Domingues *et al*, 2017). Of note, the combined administration of Gro β with a VLA-4 antagonist was shown to induce a rapid, synergistic and efficient mobilization of murine long-term HSPC as compared to the single stimuli alone (Karpova *et al*, 2019). As mechanism of action, it was proposed that upon stimulation, CXCR2 activation can trigger the release of neutrophils proteases in the BM microenvironment in association with a remarkable increase in BM endothelial layer permeability, that are both critical events for boosting the mobilization response induced by VLA-4 antagonist administration (DiPersio *et al*, 2019).

Finally, parathyroid hormone (PTH) is another factor that can indirectly enhance HSPC mobilization since, by stimulating osteoblasts to release hematopoietic growth

factors, it can induce a quantitative expansion of BM resident HSPC available for pharmacological mobilization (Calvi *et al*, 2003). Of note, the daily administration of PTH for 5 weeks followed by G-CSF treatment was shown to increase the mobilization capacity of the recipient as compared to G-CSF alone. However, the reduced clinical efficacy of PTH + G-CSF mobilized-grafts have prevented further clinical development of PTH-based mobilizing regimens (Domingues *et al*, 2017).

Despite the recent advances in the clinical applications of soluble molecules capable of inducing HSPC egress from the bone marrow, in certain pathological conditions or in heavily treated patients mobilization can still fail and result in poor HSPC yield (poor mobilizers), thus limiting its applicability. Therefore, clinical practice could benefit from the implementation of current mobilization procedures or the identification of novel mechanisms responsible of HSPC trafficking, in order to broaden the application of HSPC-based therapies.

1.5 Circulating HSPC

Thanks to the interplay of all the signals described above, adult HSPC mainly reside in the BM. However, few of them can be also found in the peripheral blood circulation and within many peripheral tissues (lung, liver, spleen, intestines, kidneys and thymus) at steady state, and the mechanisms and the significance of the spontaneous HSPC release and trafficking still remain not completely understood. HSPC egress from the BM may be driven by the loss of cell-cell contacts upon the down-regulation of multiple adhesion molecules, as well as by the desensitization of chemotactic signals. In this regard, some pieces of evidence have been collected so far about the impact of SDF-1/CXCR4 interaction on the steady-state release of BM HSPC into the circulation. In particular, the adrenergic signals delivered in the BM niche by sympathetic nerves through the β_3 -adrenergic receptor were shown to lead to rhythmic fluctuations in local SDF-1 concentration, thus inducing cycling circadian oscillation in the number of circulating HSPC (cHSPC) (Méndez-Ferrer *et al*, 2008). Moreover, SDF-1 was observed to be produced by not only BM stromal cells, but also other cell types, including tissue resident-dendritic and endothelial cells (Müller *et al*, 2001; Pablos *et al*,

1999), and its release could be enhanced by local hypoxia after transient ischemic insult in peripheral tissues, thus driving selective migration of CXCR4-expressing HSPC into the injured site (Ceradini *et al*, 2004). Nevertheless, although several hypotheses about the pathways and molecular axes involved have been proposed, definitive answers on the trafficking of HSPC in humans are yet to be provided. Moreover, very little information has been collected so far about human cHSPC role in physio-pathological conditions and their correlation with BM-resident HSPC.

1.5.1 Murine circulating HSPC

The regular traffic of HSPC through the blood stream was proposed to be important for sustaining hematopoietic homeostasis at steady state since, by facilitating stem cell redistribution and competition for different BM niches, allows to maintain a balance between blood cell production and differentiation/removal (Wright *et al*, 2001; Mazo *et al*, 2011; Flidner, 1998). *In vivo* experiments based on distinct genetically marked parabiotic mice, which were surgically conjoint to obtain a common cross circulation, were exploited to investigate the relevance of physiological HSPC circulation in homeostatic condition. In particular, Wright *et al*. showed that mice established a stable cross circulation by day 3 after surgical joining, reaching about 50% of hematopoietic chimerism at day 7, and maintained long-term partner-derived hematopoiesis after surgical separation. Moreover, it was estimated that 100 to 400 LT-HSC were present in the PB of a mouse at any time (Wright *et al*, 2001).

More recently, novel pieces of evidence have shed light on the role of murine cHSPC not only at steady state but also in stressed contexts, such as infection and sepsis (Massberg *et al*, 2007; Burberry *et al*, 2014). In particular, Massberg and colleagues demonstrated that murine migratory HSPC regularly reach and proliferate within extramedullary tissues in physiological conditions, in order to locally foster the homeostatic replenishment of tissue-resident myeloid immune cells (e.g. dendritic cells, macrophages and granulocytes). This phenomenon could be exacerbated after the release of pro-inflammatory signals during local infection or tissue damage, when cHSPC might act as an immediate and highly adaptive stem cell source able to promptly migrate and *in situ* differentiate into innate immune effector cells (Massberg *et al*,

2007).

The immune-surveillance role of cHSPC was also suggested by an additional work by Burberry *et al.* In this case, they observed that during systemic *E. coli* sepsis, cHSPC tend to accumulate within the mouse spleen in order to sustain a prompt extramedullary production of myeloid cells to hamper the infection. This mechanism could be activated by diverse pathogen associated molecular patterns (PAMPs), such as lipopolysaccharide (LPS) and peptidoglycan, which are able to respectively trigger the activation of Toll-like receptor 4 (TLR4) and nucleotide-binding oligomerization domain containing (NOD)-like-receptors (NLRs) NOD1 and NOD2 signaling pathways. These events synergistically lead to an increased production of G-CSF and a lower local concentration of CXCL12 in the BM, resulting in a massive mobilization of HSPC to the spleen and intra-splenic production of neutrophils and monocytes to fight against the infection (Burberry *et al.*, 2014).

Not only the spleen but also other peripheral organs in mouse have been described as reservoirs of HSPC, such as the lungs (Lefrançois *et al.*, 2017), the muscles (McKinney-Freeman *et al.*, 2002) and the liver (Cardier & Barberá-Guillem, 1997). Studies based on 2-photon intravital microscopy showed that lungs contain a population of BM-like MK and represent a primary site of platelet biogenesis, contributing to approximately 50% of total platelet production at steady state. Furthermore, immature MK together with hematopoietic progenitors were detected in the extravascular spaces of lungs and were shown to sustain stable platelet count restoration, as well as multilineage hematopoietic production after transplantation into thrombocytopenic mice, thus demonstrating the contribution of lung HSPC to hematopoietic reconstitution (Lefrançois *et al.*, 2017). Along the same line, murine muscle-derived CD45⁺ cells showed to have hematopoietic clonogenic potential *in vitro* and long-term *in vivo* multilineage repopulation capacity after mice transplantation, thus indicating murine skeletal muscles as reservoirs of HSC (McKinney-Freeman *et al.*, 2002). Finally, HSPC were found also in the liver, where an extramedullary hematopoietic microenvironment within hepatic sinusoids was found to sustain their proliferative and differentiation capability. Importantly, liver endothelial cells-1 (LEC-1) were shown to play a key role in supporting liver hematopoietic niche (Cardier & Barberá-Guillem, 1997).

1.5.2 Human cHSPC content changes in association with multiple physiological and pathological factors: biomarker valence?

In the last years, many descriptive studies have detected consistent variations in the count/frequency of human cHSPC in association with different biological and pathological factors, suggesting this trafficking population as a conceivable biomarker of not only genetic and health-related conditions (Cohen *et al*, 2013), but also disease prognosis (Tsaganos *et al*, 2006; Skirecki *et al*, 2019), treatment efficacy (Napolitano *et al*, 2016) and BM status after disease onset (Abdellatif, 2018; Pizarro *et al*, 2014; Wu *et al*, 2014; Santoro *et al*, 2020).

In a large study on a cohort of 1595 subjects without any cardiovascular diseases, Cohen and colleagues found significant correlations between the frequency of steady-state PB CD34⁺ cells and diverse genetic and clinical factors. Specifically, circulating CD34⁺ cells frequency was inversely correlated to older age, female sex and smoking habit, while showed a positive correlation with weight, serum total cholesterol and statin therapy (Cohen *et al*, 2013).

In two prospective observational studies performed respectively on 44 (Tsaganos *et al*, 2006) and 33 (Skirecki *et al*, 2019) patients with septic syndrome in comparison to healthy controls, a 100-fold increase of PB CD34⁺ CD45⁺ cells (Tsaganos *et al*, 2006) and 3-fold increase of PB CD34⁺ CD38⁻ cells (Skirecki *et al*, 2019) was observed in the days following sepsis onset with respect to healthy individuals. Interestingly, in both studies, the count of CD34⁺ cells in PB inversely correlated with the patients' time of survival (the highest the circulating CD34⁺ amount, the worst the patient's outcome), suggesting the evaluation of PB CD34⁺ cells as a valuable prognostic marker during sepsis (Tsaganos *et al*, 2006; Skirecki *et al*, 2019).

In the context of hematological disorders, a prospective study performed on a cohort of 69 β -thalassemic patients revealed a higher count of PB CD34⁺ cells in patients affected by β -thalassemia major (TM, mean $6.9 \pm 4.5/\text{ul}$) and β -thalassemia intermedia (TI, mean $11.8 \pm 14.8/\text{ul}$) than in healthy volunteers (mean $3.5 \pm 2.9/\text{ul}$). Considering that TM is a highly severe clinical condition treated with regular blood transfusion, while patients with TI are usually transfusion-independent, the observed raise in PB CD34⁺ amount in non-transfused than transfused patients may suggest circulating

CD34⁺ quantity as a valuable marker of clinical response to transfusion and treatment efficacy (Napolitano *et al*, 2016).

The count of cHSPC was described to change also in association to non-hematological pathologies, representing a valuable biomarker of disease progression. In patients with chronic hepatitis B, an approximately 450-fold increase of trafficking CD34⁺ quantity was described in comparison to healthy controls and their amount positively correlated with patients' hepatitis B viral load. Despite no functional characterization was provided for this phenomenon, the authors speculated that in chronic liver disorder, when the regenerative capacity of hepatocytes is severely compromised, hematopoietic progenitors can be mobilized from the BM to the liver in order to contribute to liver repair (Abdellatif, 2018). Differently, in a cohort of 62 patients affected by chronic obstructive pulmonary disease, a significant reduction in CD34⁺ cell frequency was observed in affected patients than healthy individuals. This trend could be explained either by a disease-induced BM impairment in producing and releasing progenitor cells, or by a massive recruitment of circulating HSPC in injured lung vessels (Pizarro *et al*, 2014).

Finally, in patients with both solid and blood tumors an increased content of cHSPC associated with a skewing toward the myeloid lineage was detected (Wu *et al*, 2014; Santoro *et al*, 2020). Specifically, a higher frequency of PB Lin⁻ CD34⁺ cells was observed in a cohort of 90 patients with diverse types of solid tumors as compared to 63 age-matched healthy controls. Of note, the composition of cHSPC was consistently altered in cancer patients than healthy subjects, with a four- to seven-fold increased percentage of GMP resulting in higher production of CFU-GM, CFU-M and CFU-G. These results suggested a general myeloid bias of PB HSPC in cancer-affected patients, which was positively correlating to tumor progression and associated with high plasma levels of pro-inflammatory cytokines, such as G-CSF, granulocyte/monocytes colony stimulating factor (GM-CSF) and Interleukin-6. Interestingly, the increased release of tumor-derived pro-inflammatory cytokines was shown to induce not only a prominent HSPC myeloid differentiation bias, but also the generation and expansion of myeloid-derived suppressor cells (MDSC), which, by negatively regulating immune responses, could enhance cancer cells growth and lead to adverse outcomes for the patient (Wu *et al*, 2014). Furthermore, focusing on hematological tumors, a study conducted on a

cohort of 69 patients affected by chronic lymphocytic leukemia (CLL) revealed an increased relative frequency of circulating ST-HSC and CMP-MEP in CLL patients in comparison to healthy controls, which also in this case was associated with a skewed myeloid differentiation *in vitro*. Importantly, these observations could represent a first sign of BM failure and exhaustion of normal HSPC (Santoro *et al*, 2020).

1.5.3 Functional features of human cHSPC: current knowledge and relevant gaps in the field

While multiple studies have tried to elucidate the role of murine cHSPC both in physiological and altered conditions, few investigations have been accomplished so far regarding the characterization of functional features of cHSPC in humans.

Over the last decades, several studies have shown that ST- and LT-HSPC are present within human intestine and liver, and are able to contribute to long-lasting hematopoietic chimerism in the recipient after organ allotransplantation, giving proof of principle of the BM engraftment capability of allograft-passenger HSPC *in vivo* in humans (Alexander *et al*, 2008; Wang *et al*, 2012; Fu *et al*, 2019). Importantly, some graft-derived HSPC were shown of being able to seed recipient lymphoid organs and contribute to the *de novo* generation of donor-derived recipient-tolerant T cells after organ allotransplantation (Alexander *et al*, 2008; Fu *et al*, 2019).

Of interest, Bourdieu *et al*. identified an enrichment of HSPC with an immature phenotype (CD34⁺ CD38^{med/low} CD133⁺ CD90⁻) within the DieCycle Violet-side population (DCV-SP) of steady-state peripheral blood. By comparing the functional properties of SP vs. non-SP cells isolated from steady-state PB, they observed a 4.5-fold higher proliferation rate, a 3.1-fold increased CFC production *in vitro* and a more robust long-term *in vivo* engraftment capability (up to 12 weeks) after mice xenotransplantation sustained by SP-cells as compared to non-SP fraction, thus implying SP phenotype as a valuable criterion for primitive PB HSPC isolation. However, despite providing a first demonstration of cHSPC functional features in terms of both *in vitro* differentiation capability and *in vivo* engraftment potential, no functional comparison with steady-state BM-derived HSPC was performed in this study (Bourdieu *et al*, 2018).

Recently, Mende *et al.* performed a transcriptional characterization of splenic and PB extramedullary HSPC in comparison to their BM resident counterpart (Mende *et al.*, 2020) By applying single-cell RNA sequencing on matched BM, PB and spleen tissues derived from 2 deceased adult donors, they showed that PB- and spleen-derived HSPC are less proliferative than BM counterpart, with a larger proportion of cells in G1 phase of the cell cycle. Interestingly, splenic HSPC displayed high expression levels of early erythroid differentiation-associated genes together with increased erythroid differentiation *in vitro*, suggesting their pre-activated state to enable fast extramedullary emergency erythropoiesis in case of stress or injury. Moreover, they identified a subset of extramedullary HSC/MPP, that are not present in the BM and are composed by the sum of PB- and spleen-derived HSPC, expressing unique transcriptional modules related to both hematopoietic lineage commitment (typical of CMP, earlyB, MEP state) and actomyosin cytoskeleton (*ACTB*, *ACTG1*, *TPM1/3/4*, *MYL6*, *MYL12A*, *RHOA*, *RAC1*, *CDC42*). By contrast, pan-tissue HSC/MPP, which can be indistinctly found in the BM, PB and splenic tissues, were enriched in gene signatures typical of LT-HSC. These results indicated PB- and spleen-derived HSC/MPP as a unique population with a high degree of lineage priming and distinct properties linked to adhesion and motility than pan-tissue component. Although providing a frame of reference for extramedullary HSPC transcriptional properties, this study showed some limitations. First, the transcriptional analyses performed were mainly focused on the sum of PB and splenic HSPC, without investigating the transcriptional state of cHSPC only and, in particular, of the distinct cHSPC subsets. Furthermore, since they analyzed HSPC isolated from two deceased organ donors, the transcriptional patterns detected by scRNAseq could have been affected by the post-mortem state of the donors. Finally, no extensive comparison on the functional features of HSPC isolated from the different sources was performed, thus leaving a lot of pending questions about the effective differentiation and repopulation potentials of human PB and splenic HSPC with respect to BM counterpart (Mende *et al.*, 2020).

In summary, based on studies performed in mouse and on few pieces of evidence in humans, we are now able to draw diverse hypotheses regarding the role of human cHSPC in physiological and stressed conditions, such as infections or sepsis, as summarized in **Figure 7**. However, given the fragmentary nature of the data collected

so far and the lack of comprehensive studies on the topic, there is still plenty of open biological questions about the functional properties of cHSPC in humans, both at steady state and in pathological contexts. Moreover, the phenotypic and transcriptional states of human cHSPC, the mechanisms controlling their trafficking, as well as their hierarchical relationship with respect to BM-resident counterpart remain still to be elucidated.

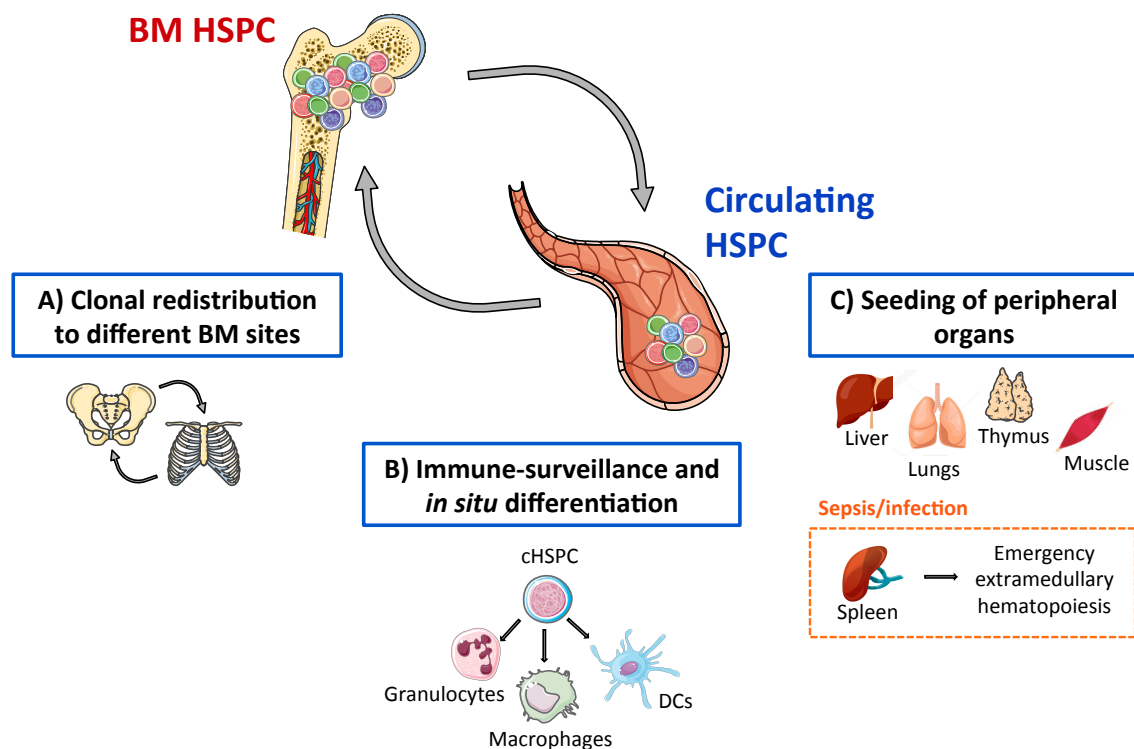


Figure 7. Hypotheses on circulating HSPC functions derived from studies in mice and humans. (A) In physiological conditions, circulating HSPC (cHSPC) may exert a homeostatic role to ensure both full occupancy of all BM cavities and clonal redistribution to distant niches. (B) cHSPC are able to actively patrol the circulation and peripheral tissues in order to *in situ* differentiate into myeloid cells, both at steady state and during local infections, thus participating in immune-surveillance. (C) Multiple extramedullary organs were shown to be reservoir of murine HSPC, including liver, lungs and muscles. Furthermore, some evidence from murine and human studies has shown the presence of HSPC in the spleen, where they might contribute to emergency extramedullary hematopoiesis during sepsis or infections.

2 AIM OF THE WORK

Although mostly resident in the bone marrow, few circulating HSPC regularly traffic in the peripheral blood of un-mobilized individuals and their role in humans is today poorly characterized. Indeed, mainly descriptive studies have been published so far about this rare trafficking population in humans and a complete evaluation of their composition, functional properties and hierarchical relationship with respect to BM resident counterpart is still missing. Furthermore, little information is currently available on the molecular mechanisms tuning physiological HSPC recirculation in humans.

Here we exploited a combination of state-of-the-art multi-parametric flow cytometry, single-cell transcriptome profiling (scRNAseq), *ad hoc* designed *in vitro* and *in vivo* assays, and IS clonal tracking to comprehensively investigate the phenotypic composition, the transcriptional features and the functional properties of the 10 cHSPC subpopulations in relationship to their BM counterparts, both at steady state and in stressed conditions, such as after gene therapy.

The investigation of the biological functions of circulating progenitors in both physiological and pathological conditions could help to develop novel therapeutic and diagnostic tools that could be exploited in the clinical setting. Specifically, unveiling the molecular mechanisms controlling physiological HSPC migration may provide novel targets for the implementation and optimization of HSPC mobilization strategies. Moreover, by assessing the relationship between PB and BM resident HSPC, we would evaluate the role of cHSPC as biomarker of BM state, thus avoiding patients' periodical BM samplings during follow-ups encompassed in many clinical protocols, through the collection of HSPC directly from the PB. Finally, since HSPC collection through leukapheresis during drug induced-mobilization represents a procedure hampered by intrinsic complications, especially in low body-weight pediatric subjects, cHSPC could represent a readily accessible and safe source of hematopoietic stem cells for expansion and genetic modification.

3 RESULTS

3.1 cHSPC display a progressive quantitative change during physiological aging

In order to phenotypically characterize and quantify cHSPC during physiological hematopoietic development, we collected peripheral blood samples from 114 healthy donors (HD) of different ranges of age: at birth (neonates, NEON), during pediatric (PED 1, PED 2, PED 3), adolescent (PED 4), adult (AD) and old (AGED) age (**table 1**). Samples were analyzed through the multi-parametric “Whole Blood Dissection” (WBD) flow-cytometry protocol, that can unambiguously identify and quantify 24 different blood cell types, including 10 HSPC subsets (Basso-Ricci *et al*, 2017). For HSPC subsets phenotypic identification, we exploited surface markers reported in **Figure 1** and **Table 8**.

Table 1. Cohort of PB healthy donors of different ages included in the study. Samples are classified in 7 distinct categories based on age.

	Number	Range of age
NEON	26	0-1 day
PED 1	7	0-1 years
PED 2	13	1-6 years
PED 3	15	6-12 years
PED 4	11	12-18 years
AD	27	18-65 years
AGED	15	>70 years

We found a progressive reduction of cHSPC count and changes in their composition throughout hematopoietic maturation (**Figure 8A-B**). In particular, we observed a higher frequency of primitive HSC and MPP populations in newborns as compared to older groups. Moreover, while no major differences were evident in the relative frequencies of myeloid-committed progenitors, we detected a progressive enrichment of

erythroid progenitors over aging. We observed differential behavior of the diverse lymphoid populations. The fraction of ETP was higher in the NEON group and then reduced and stabilized overtime. On the other hand, the proportions of MLP and PreBNK were lower in neonates, enriched between 1 and 18 years, and then reduced in the adults (**Figure 8B**).

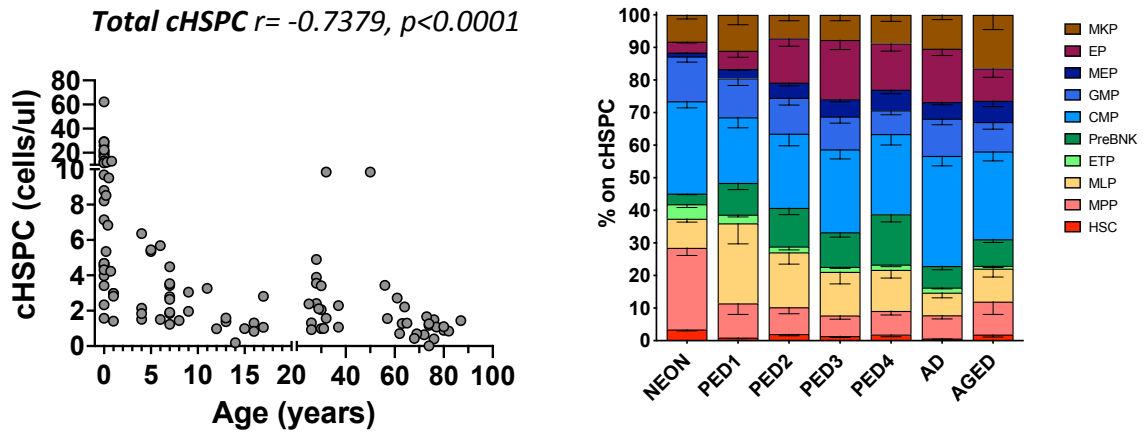


Figure 8. Phenotypic characterization of PB HSPC in distinct ranges of age. (A) Graph showing correlation between age and absolute counts of cHSPC. Statistical test for correlation: Spearman r . Spearman's correlation coefficient (r) and p value are reported. (B) Stacked bar graph displaying the relative composition of PB HSPC subsets in the healthy subjects included in the study. Data are shown as mean with Standard Error Mean (SEM).

To understand the contribution of each HSPC subpopulation to the observed decline of total PB HSPC count, we measured the amount of the distinct circulating HSPC subsets overtime (**Figure 9A-D**). All HSPC subtypes were quantitatively increased in newborns and in very young subjects, while reducing their count in PB overtime. However, we observed slight differences in subset-specific cell kinetics. In particular, primitive HSC (**Figure 9A, left**) and MLP and ETP lymphoid progenitors (**Figure 9B**) displayed a progressive decrease starting from pediatric time of life, while the counts of circulating MPP (**Figure 9A, right**), myeloid-committed (**Figure 9C**), PreBNK (**Figure 9B, right**), and erythroid/MK (except MEP) (**Figure 9D**) precursors showed a marked reduction after birth and then remained almost stable overtime.

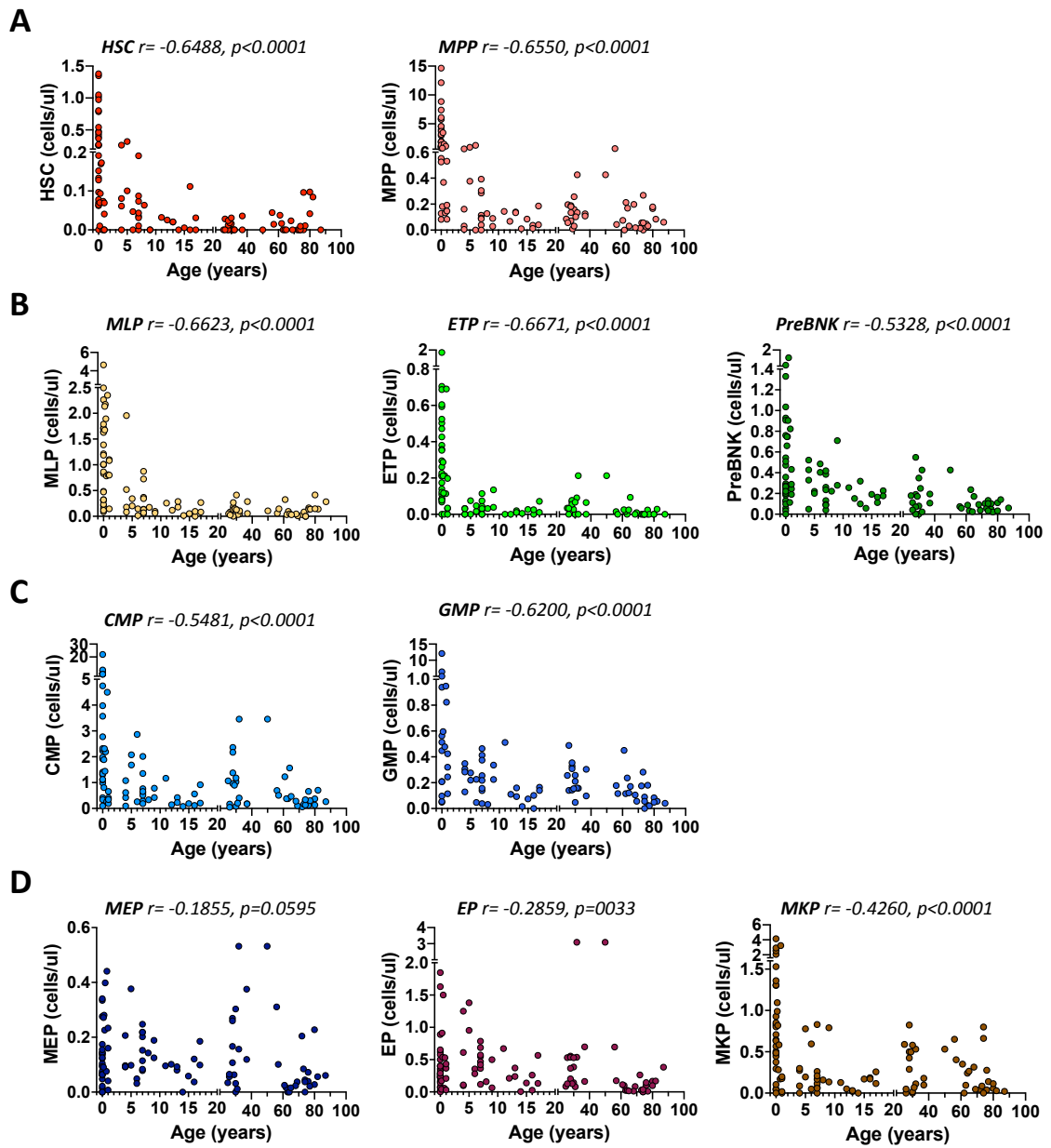


Figure 9. Counts of HSPC single subsets from PB and BM during aging. Graphs showing correlations between age and absolute counts of primitive (A), lymphoid (B), myeloid (C) and erythroid/MK (D) HSPC subsets. Statistical test for correlation: Spearman r . For each correlation, respective spearman's correlation coefficient (r) and p values are reported.

3.2 Recirculation capability changes among the distinct cHSPC subsets and during physiological aging

In order to understand if the overtime changes detected in the amount and composition of cHSPC subsets were a reflection of the previously described variations in BM HSPC content occurring during physiological aging (Nilsson *et al*, 2016), we collected 48 BM samples from pediatric, adult and aged healthy subjects. Given the rarity in retrieving donor-matched BM and PB samples, we compared HSPC composition and count in a selected cohort of age-matched PB-BM donors (**Table 2**).

Table 2. Cohort of age-matched pediatric, adult and aged healthy donors of BM and PB samples. The exact number of donors, the range of age and mean age of each group are reported.

		Number	Range of age (years)	Mean age (years)
BM	PED	9	3-17	9.8
	ADULT	20	18-64	41.4
	AGED	19	65-87	75.6
PB	PED	15	7-11	7.9
	ADULT	27	25-64	38.7
	AGED	15	65-87	75.1

As expected, we detected a higher total HSPC count in BM with respect to PB samples from the same age-matched group (**Figure 10A**). By comparing PB and BM HSPC relative compositions, we observed a reduced frequency of primitive HSC and PreBNK, but an increased fraction of MPP and MLP in PB than BM HSPC (**Figure 10B**). In addition, cHSPC showed an inverted GMP/CMP ratio and a higher proportion of megakaryocytic/erythroid progenitors (MEP, EP, MKP) as compared to age-matched BM samples.

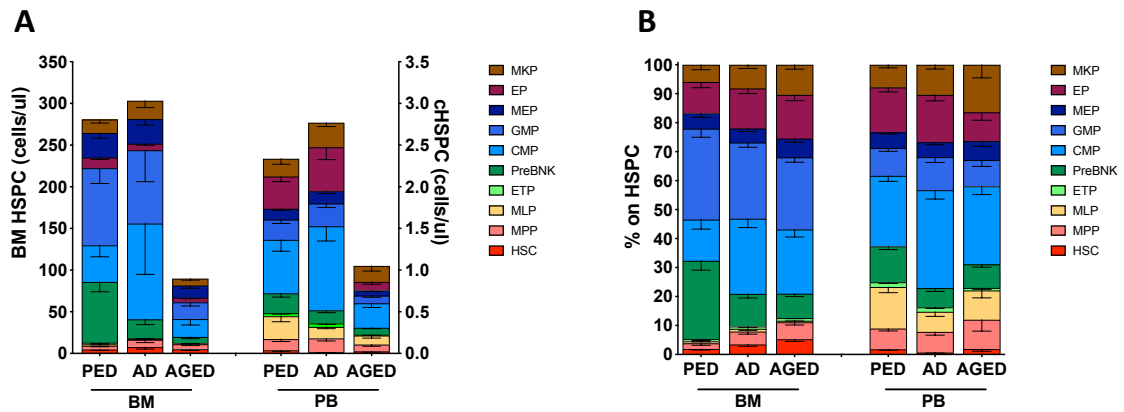


Figure 10. Comparison of BM and PB HSPC compositions and amounts in the age-matched dataset. Stacked bar graph displaying absolute counts (A) and relative compositions (B) of BM and PB HPSC subsets derived from age-matched individuals (PED, AD, AGED) included in the study. Data are shown as mean with Standard Error Mean (SEM).

The finding that the differences in HSPC composition observed between the two sources are consistent among the 3 groups may suggest diverse intrinsic capabilities of the distinct HSPC subsets in egressing from the BM and recirculating, rather than being the result of age-related physiological changes of BM HSPC composition. To measure the propensity of recirculation of HSPC populations, we calculated the Circulation Index (CI), by normalizing the count of each HSPC subset in each PB donor on the mean count in age-matched BM:

$$CI = \frac{PB \text{ HSPC subset count}}{\text{mean BM HSPC subset count}}$$

Within each group, we found a subset-specific capability of recirculating in the PB. We also observed that MLP had the highest CI among all HSPC subsets and across the different groups of age, while HSC, MPP and ETP CI were increased in pediatric group and then progressively reduced throughout aging (Figure 11).

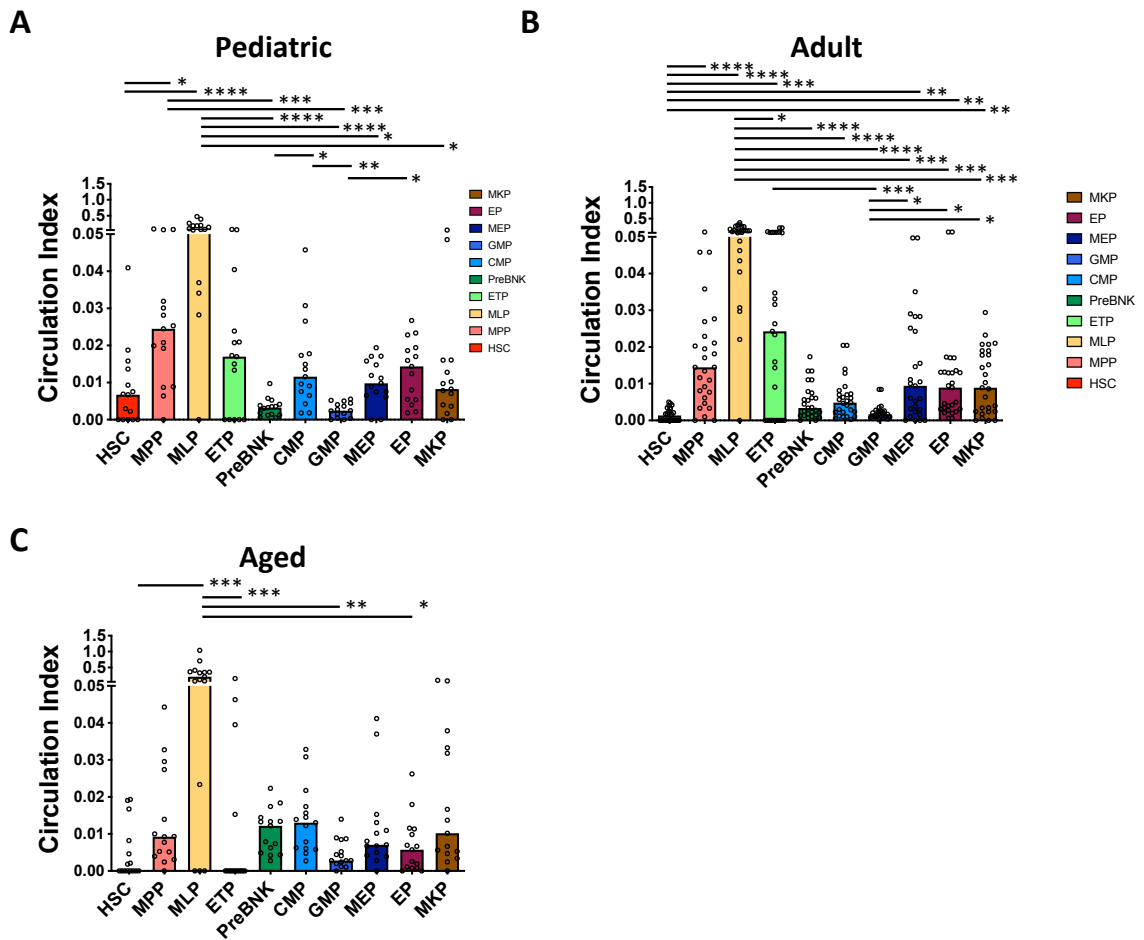


Figure 11. Circulation indexes (CI) of single cHSPC subsets belonging to the age-matched dataset. CI estimated for pediatric (A), adult (B) and aged (C) subjects-derived HSPC subpopulations. Statistical tests: Kruskal-Wallis test with Dunn's multiple comparisons test. $*=p<0.05$; $=p<0.01$; $***=p<0.001$; $****=p<0.0001$. Medians are shown.**

Overall, our phenotypic characterization shows that cHSPC count and composition change during aging, with each HSPC population displaying subset-specific circulation capability. These findings might suggest diverse physiological functions of each trafficking subpopulation at steady state.

3.3 cHSPC display multilineage differentiation potential *in vitro*

In order to functionally characterize cHSPC, we firstly assessed their clonogenic capability by performing CFC assay. With this aim, we plated total cells retrieved upon lysis of 250 μ l whole PB collected from our cohort of HD, and we characterized the generated colonies after 14 days of culture. Consistently with the observed gradual decrease of cHSPC count during aging (**Figure 8A**), we detected a progressive reduction of the number of produced colonies during hematopoietic maturation (**Figure 12A**), indicating that HSPC circulating in the peripheral blood are endowed with clonogenic potential. In line with the progressive changes of primitive and megakaryocytic/erythroid HSPC subsets fractions observed in PB during hematopoietic development (**Figure 8B**), colonies from newborns and young subjects were mainly composed by GEMM-CFU and GM-CFU, while BFU-E production was more pronounced starting from 1 year-old group (**Figure 12B**).

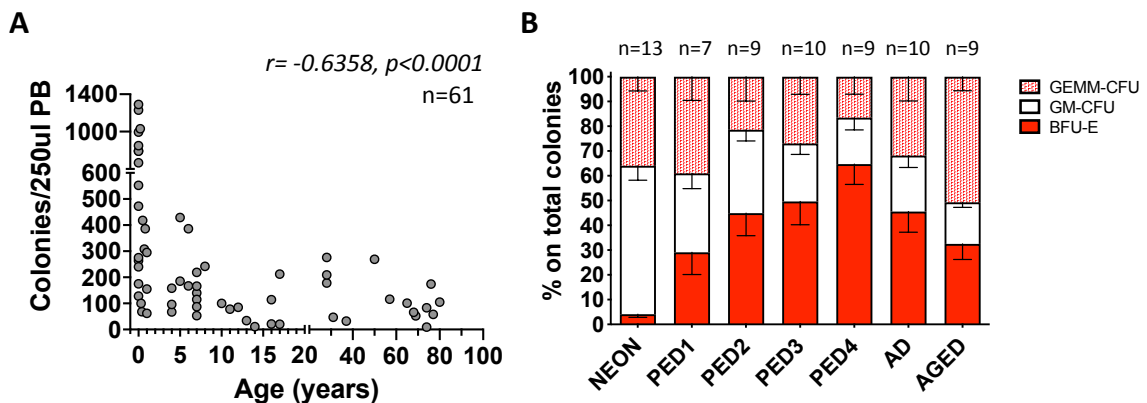


Figure 12. Clonogenic potential of cHSPC during aging. (A) Graph showing correlation between age and total count of colonies generated through CFC assay starting from 250 μ l of whole PB. Statistical test for correlation: Spearman r . Spearman's correlation coefficient (r) and p values are reported. (B) Distribution of colonies subtypes generated from PB cells of healthy subjects of diverse cohorts of age. Data are shown as mean with SEM. CFU-GEMM: Granulocyte-Erythroid-Monocytes-Megakaryocyte colonies; CFU-GM: Granulocyte-Monocyte colonies; BFU-E: Burst Forming Units-Erythroid colonies.

To further investigate the differentiation potential of cHSPC, we set up a novel *in vitro* multilineage differentiation assay to comprehensively test the output of cHSPC toward all the hematopoietic compartments. Our protocol was built on published *in*

in vitro differentiation assays for human myeloid/erythroid/MK cells, B/NK cells and T cells differentiation (Doulatov *et al*, 2010; Notta *et al*, 2016; Miyawaki *et al*, 2017), and is based on culturing CD34⁺ cells in presence of a cocktail of soluble cytokines optimized for multilineage differentiation. We tested the differentiation potential of 500 CD34⁺ cells isolated from either PB (n=5) or, as control, BM (n=8) samples of adult healthy donors. Cells were cultured for 3 weeks and their differentiation output was analyzed by flow cytometry at the end of culture (**Figure 13**).

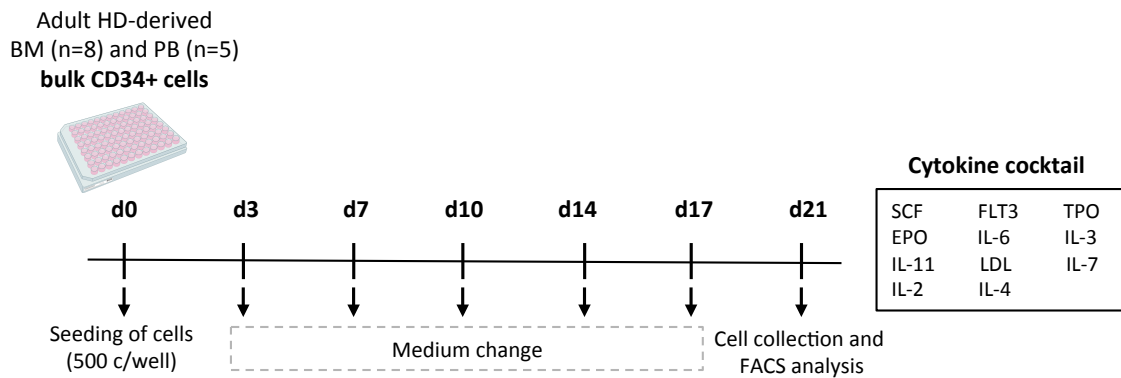


Figure 13. Experimental scheme of the *in vitro* multilineage differentiation protocol. 500 CD34⁺ cells from HD adult BM (n=8) or PB (n=5) samples were seeded in a 96 well plate in presence of a cytokine cocktail optimized for multilineage differentiation. Cellular output was analyzed after 3 weeks of culture through FACS analysis. See the materials and methods section for more details.

After 3 weeks of culture we detected a comparable expansion rate starting from CD34⁺ cells derived from the two sources, with no statistically significant differences between the two groups (**Figure 14A**). Furthermore, PB and BM HSPC displayed a very similar differentiation output in all the diverse hematopoietic lineages, both in terms of quality (**Figure 14B**) and quantity (**Figure 14C**) of retrieved myeloid, lymphoid, erythroid and megakaryocytic cells. Importantly, no statistically significant differences were observed in the frequency and absolute amount of undifferentiated CD34⁺ cells retrieved at the end of BM and PB cultures (**Figure 14D**), indicating that a fraction of the most undifferentiated compartment from the two sources was preserved and/or expanded without differentiating during the culture.

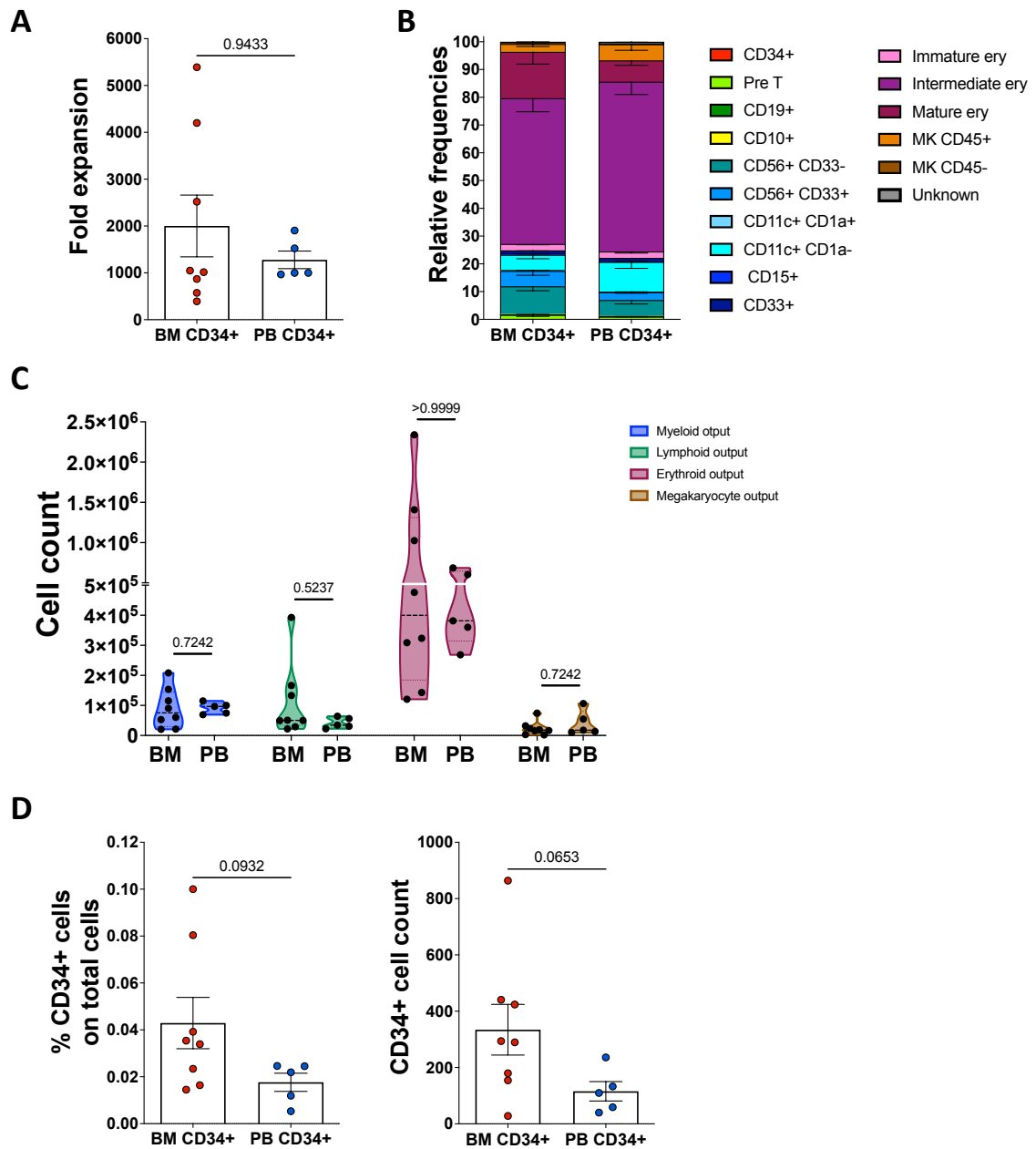


Figure 14. In vitro differentiation assay of human bulk CD34⁺ cells derived from BM or PB. (A) Fold expansion of 500 BM (red) (n=8) and PB (blue) (n=5) CD34⁺ cells after 3 weeks of culture. Each dot represents a distinct donor. Data are shown as mean with SEM. (B) Stacked bar graph displaying the composition of HSPC cell output after 3 weeks of culture. Data are shown as mean with SEM. (C) Violin plot showing the number of myeloid, lymphoid, erythroid, and megakaryocytic differentiated cells derived from 500 BM and PB CD34⁺ cells after 3 weeks of culture. Each dot shows a distinct donor. Data are shown as median with interquartile ranges. (D) Frequency (left) and absolute number (right) of undifferentiated CD34⁺ cells retrieved from BM (red) and PB (blue) CD34⁺ cultures at the end of the assay. Each dot represents a distinct donor. Data are shown as mean with SEM. For (A, C, D) Mann-Whitney statistical test was applied for groups' comparison and single p values are reported within the graphs.

In summary, our results show that cHSPC are endowed with clonogenic and multilineage differentiation potential *in vitro*, with no major discrepancies in their multilineage differentiation capability compared to BM counterpart.

3.4 cHSPC show multilineage reconstitution capability but reduced long-term engraftment potential than BM HSPC after *in vivo* transplant

To study the homing and differentiation potential of cHSPC *in vivo*, we transplanted human CD34⁺ cells derived from either BM (n=3) or PB (n=3) adult HD (**Figure 15A**) into NSGW41 mice, and we compared the human hematopoietic engraftment and blood lineages reconstitution overtime, up to 20 weeks post-transplant, as described in **Figure 15B**.

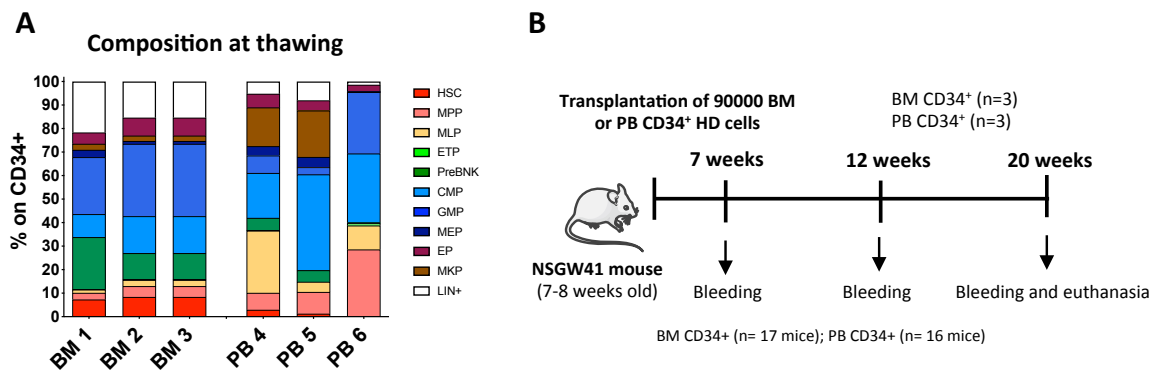


Figure 15. Setting of long-term *in vivo* transplantation experiments. (A) Phenotypic composition of CD34⁺ cells derived from BM (n=3) and PB (n=3) sources at thawing before transplantation into NSGW41 recipients. (B) Experimental scheme for *in vivo* characterization of long-term engraftment and repopulation potential of BM vs. PB CD34⁺ cells. 90,000 CD34⁺ cells from BM or PB sources were transplanted into 7-8 weeks-old NSGW41 mice. A total of n=17 and n=16 mice were transplanted with BM and PB CD34⁺ cells, respectively. PB bleedings were performed at 7, 12 and 20 weeks after transplantation and mice were euthanized after 20 weeks. Human PB and BM cell content was analyzed by WBD to assess engraftment and *in vivo* differentiation of BM- and PB-derived CD34⁺ cells.

We observed that mice transplanted with PB CD34⁺ cells displayed a significantly lower engraftment than BM CD34⁺-transplanted group (**Figure 16A**). Consistently with this result, the counts of PB myeloid and lymphoid cell lineages were remarkably

reduced in mice transplanted with PB source (**Figure 16B-C**). Of note, no statistically significant differences were found in the absolute amounts of PB erythroblasts between the two experimental groups (**Figure 16D**). By analyzing the composition of differentiated human cells present in the PB overtime, no major differences were observed in the lymphoid lineage between the two experimental groups, while increased frequencies of myeloid cells (at 20 weeks) and erythroblasts (at 12 and 20 weeks) were detected in PB than BM CD34⁺-transplanted mice (**Figure 16E**).

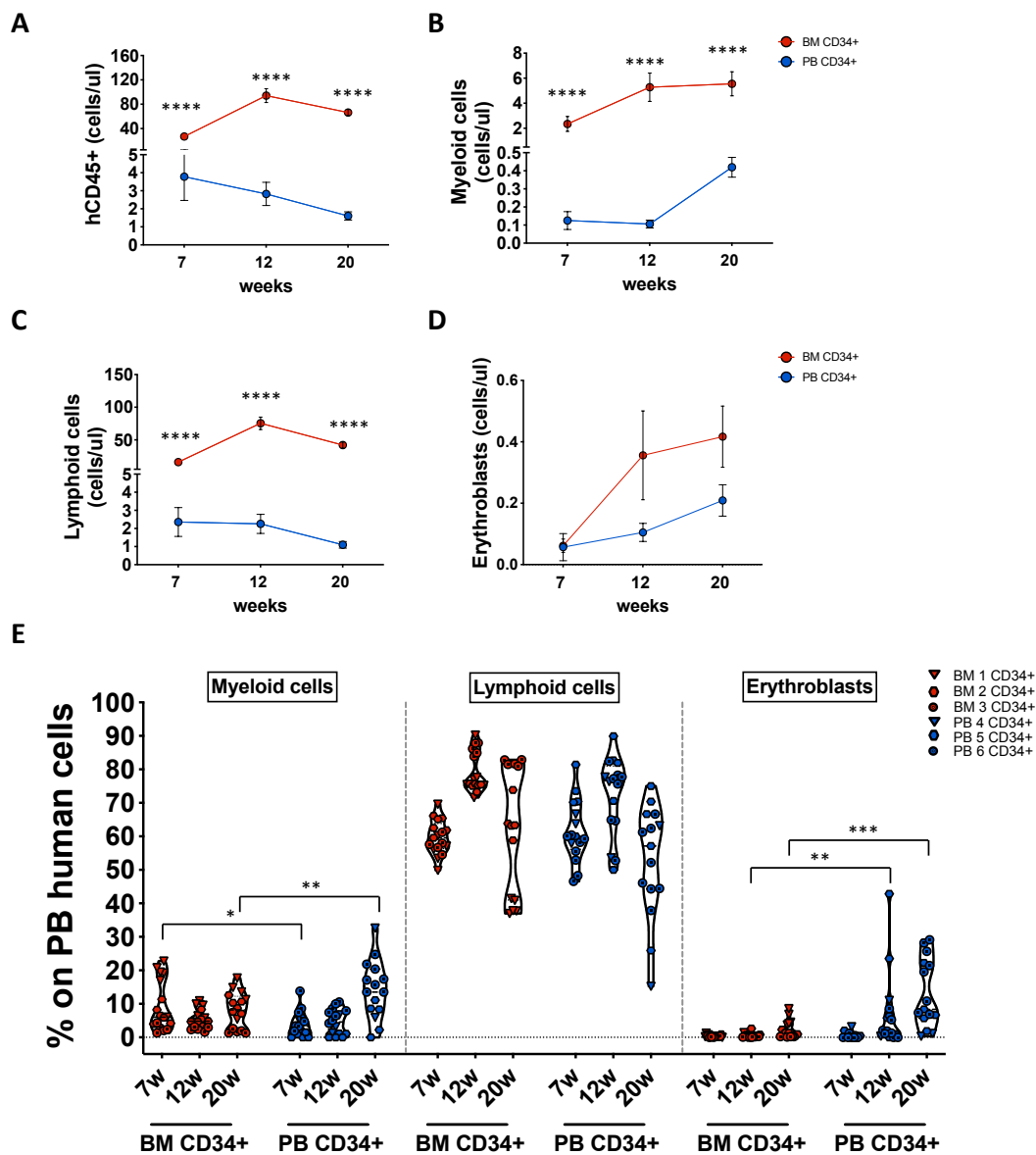


Figure 16. Long-term human PB reconstitution of major hematopoietic compartments in mice transplanted with BM or PB CD34⁺ cells. (A-D) Graphs showing absolute cell count of human CD45⁺ cells (A), myeloid cells (B), lymphoid cells (C) and erythroblasts (D) detected at

7, 12 and 20 weeks in mice transplanted with BM (red) or PB (blue) CD34⁺ cells. Data of BM and PB CD34⁺ experimental groups are shown as mean with SEM. **(B)** Violin plots showing the frequencies on PB total human cells of PB myeloid, lymphoid and erythroblast components in mice transplanted with BM (red) or PB (blue) CD34⁺ cells at 7, 12 and 20 weeks after transplantation. Data are shown as median with interquartile range. Mann-Whitney statistical test was applied for groups' comparison. Only statistically significant *p* values are reported within the graphs. *=*p*<0.05; **=*p*<0.01; ***=*p*<0.001; ****=*p*<0.0001.

At 20 weeks after transplant, human CD45⁺ cells were still detectable in the BM of PB CD34⁺-transplanted mice, despite their count was lower than in the BM CD34⁺-transplanted group (**Figure 17A**). Consistently with these data, human myeloid, lymphoid and erythroid cells, as well as HSPC were quantitatively reduced in the BM of mice transplanted with PB source (**Figure 17B**). Furthermore, the human HSPC found in the BM of PB CD34⁺-transplanted mice displayed a complete lack of primitive stem cells at 20 weeks (**Figure 17C**).

The reduced human cell content in the BM of PB CD34⁺-transplanted mice could be explained by the lower fraction of primitive population in the adult PB CD34⁺ cells before transplantation (**Figure 15A**). Indeed, we found a significant positive correlation (*r*=0.7887, *p*<0.0001) between the count of infused primitive HSC and the long-term human cell content in BM at sacrifice, independently from the source used (**Figure 17D**). Interestingly, when we normalized the amount of human CD45⁺ cells retrieved in BM at sacrifice by the number of infused HSC, the difference of BM human cell engraftment between the two groups of mice still persisted (**Figure 17E**), even though it was reduced in comparison to not normalized data (**Figure 17B, right**) (8.4 (normalized data)- vs. 100 (not normalized data)-fold increase between BM and PB CD34⁺ cells). These results could suggest that not only the number of injected HSC, but also other factors related to the intrinsic quality of infused stem cells derived from the two sources might impact on the human hematopoietic engraftment in the long term. In particular, a lower BM homing capability and/or a faster exhaustion overtime of PB- than BM-HSPC could explain the trends observed in our long-term *in vivo* experimental data.

Taken together, our data suggest that, although at a lower extent than BM HSPC, adult cHSPC are able of engrafting in murine BM and supporting multilineage output *in vivo*, with a skewing toward erythroid differentiation, up to 20 weeks after transplant.

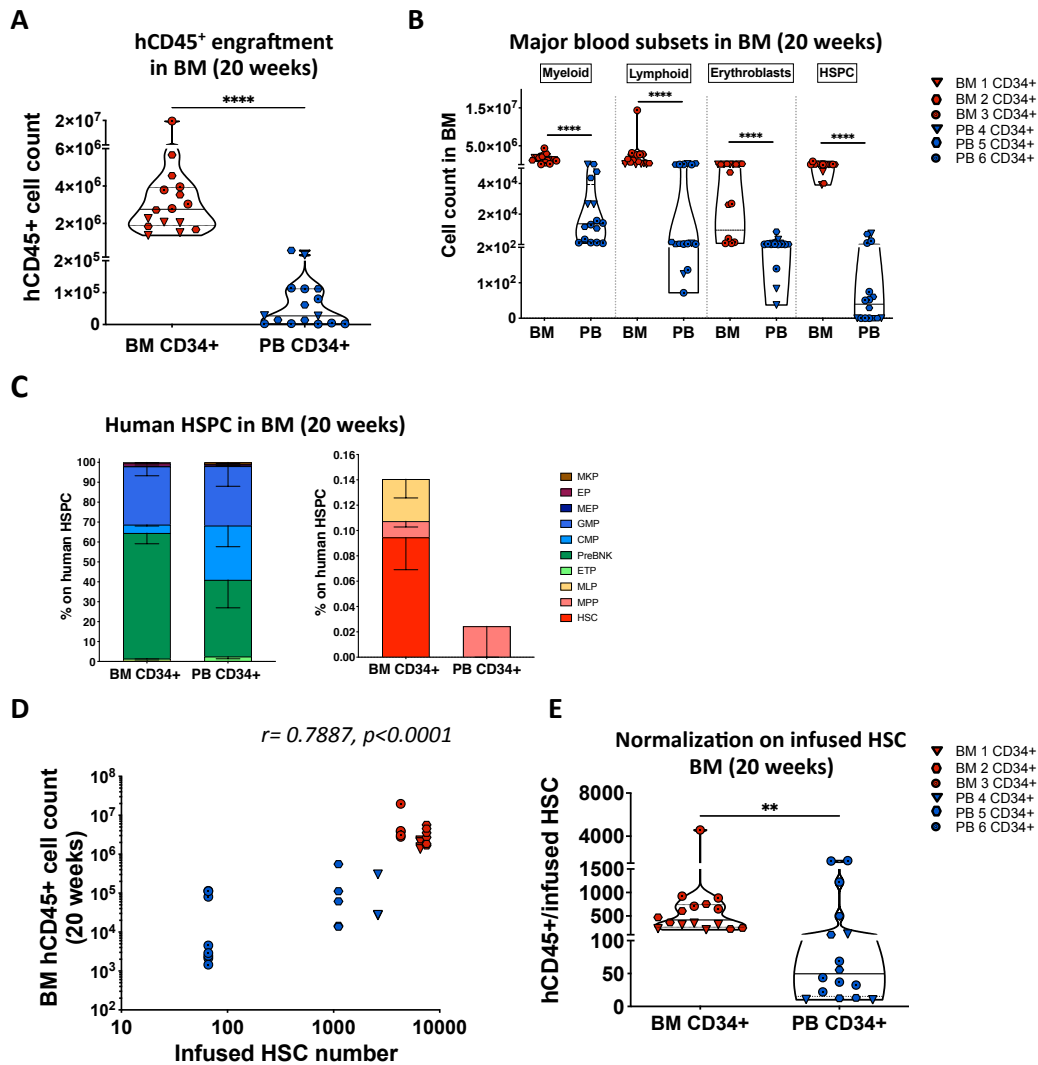


Figure 17. Human cell content in murine BM at 20 weeks after transplantation with BM or PB CD34⁺ cells. (A) Graph showing absolute cell count of human CD45⁺ cells detected at 20 weeks in the BM of mice transplanted with BM (red) or PB (blue) CD34⁺ cells. Each dot shows one animal. Data are shown as median with interquartile range. (B) Violin plots displaying the total cell count of human myeloid cells, lymphoid cells, erythroblasts and HSPC detected at 20 weeks in the BM of mice transplanted with BM (red) or PB (blue) CD34⁺ cells. Each dot shows one animal. Data are shown as median with interquartile range. (C) Stacked bar graphs displaying phenotypic composition of HSPC progenitors (left) and primitive subsets (right) detected in murine BM at 20 weeks after transplantation with BM and PB CD34⁺ cells. Data are shown as mean with SEM. (D) Graph showing correlation between the number of infused HSC and the amount of human CD45⁺ cells detected in the BM of euthanized mice at 20 weeks after transplant. Each dot shows one animal. Red and blue dots represent animals transplanted with BM and PB CD34⁺ cells, respectively. Statistical test for correlation: Spearman *r*. Spearman's correlation coefficients (*r*) and *p* value are reported. (E) Graph showing the absolute cell count of human BM CD45⁺ cells detected after 20 weeks normalized by the number of infused HSC. Each dot shows one animal. Red and blue dots represent animals transplanted with BM and PB CD34⁺ cells, respectively. Data are shown as median with interquartile range. For (A, B, E) Mann-Whitney statistical test was applied for groups'

comparison. Only statistically significant *p* values are reported within the graphs. *=*p*<0.05; **=*p*<0.01; ***=*p*<0.001; ****=*p*<0.0001.

3.5 PB and BM HSPC display comparable BM homing potential and repopulation capability early after transplantation

To investigate the behavior of cHSPC early after transplant, we performed an additional *in vivo* experiment by transplanting 44 NSGW41 mice with either BM (n=3) or PB (n=4) human CD34⁺ cells. Transplanted mice were divided into two groups (A and B), which were respectively euthanized at 4 and 12 weeks after transplant to evaluate human cell content in the murine BM (**Figure 18**).

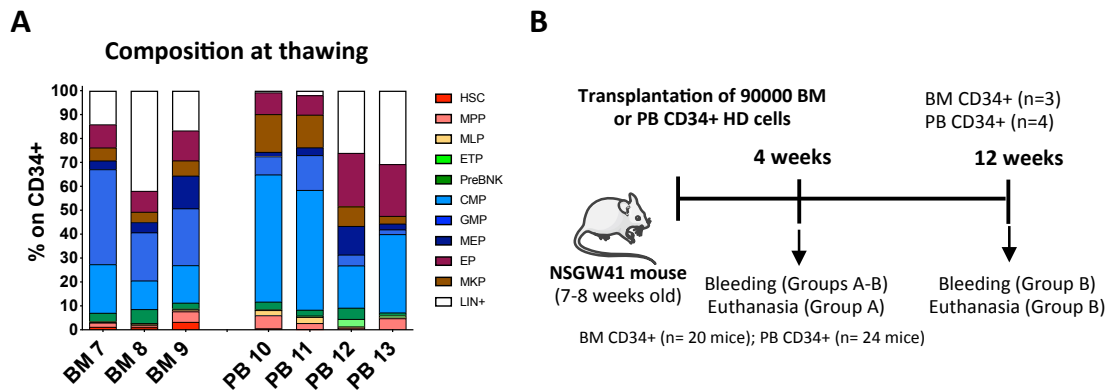


Figure 18. Setting of short-term *in vivo* experiments. (A) Phenotypic composition of CD34⁺ cells derived from BM (n=3) and PB (n=4) sources at thawing before transplantation into NSGW41 recipients. (B) Experimental scheme for *in vivo* characterization of short-term engraftment and repopulation potential of BM vs. PB CD34⁺ cells. 90,000 CD34⁺ cells from BM or PB sources were transplanted in 7-8 weeks-old NSGW41 mice. A total of n=20 and n=24 mice were transplanted with BM and PB CD34⁺ cells, respectively. Mice were divided into two groups (A and B), which were euthanized at 4 (group A) and 12 (group B) weeks after transplantation. PB bleedings were performed at 4 (group A and B) and 12 (group B) weeks post-transplant. Human PB and BM cell content was analyzed by WBD to assess engraftment and *in vivo* differentiation of BM- and PB-derived CD34⁺ cells.

Despite at 4 weeks after transplant we detected a comparable amount of human CD45⁺ cells in PB and BM in the two experimental groups, BM CD34⁺-transplanted mice showed a higher total human CD45⁺ cell content in PB and BM at 12 weeks

(Figure 19A). At 4 weeks we found an increased myeloid production and a reduced lymphoid output in the PB of PB CD34⁺-transplanted mice (Figure 19B, left), while no statistically significant difference was observed in the murine BM between the two groups (Figure 19C, left). At 12 weeks, consistently with the overall higher human CD45⁺ cell content detected (Figure 19A), we observed higher myeloid and lymphoid outputs in both PB and BM of BM CD34⁺-transplanted mice (Figure 19B, right and Figure 19C, right).

In line with the previous long-term *in vivo* experiment, cHSPC-transplanted mice displayed an increased production of erythroblasts, initially detected in the murine BM (4 weeks) (Figure 19C, left) and later in the PB (12 weeks) (Figure 19B, right). Moreover, we found a reduced human HSPC amount, both at 4 and 12 weeks after transplant, in mice transplanted with PB CD34⁺ cells (Figure 19C).

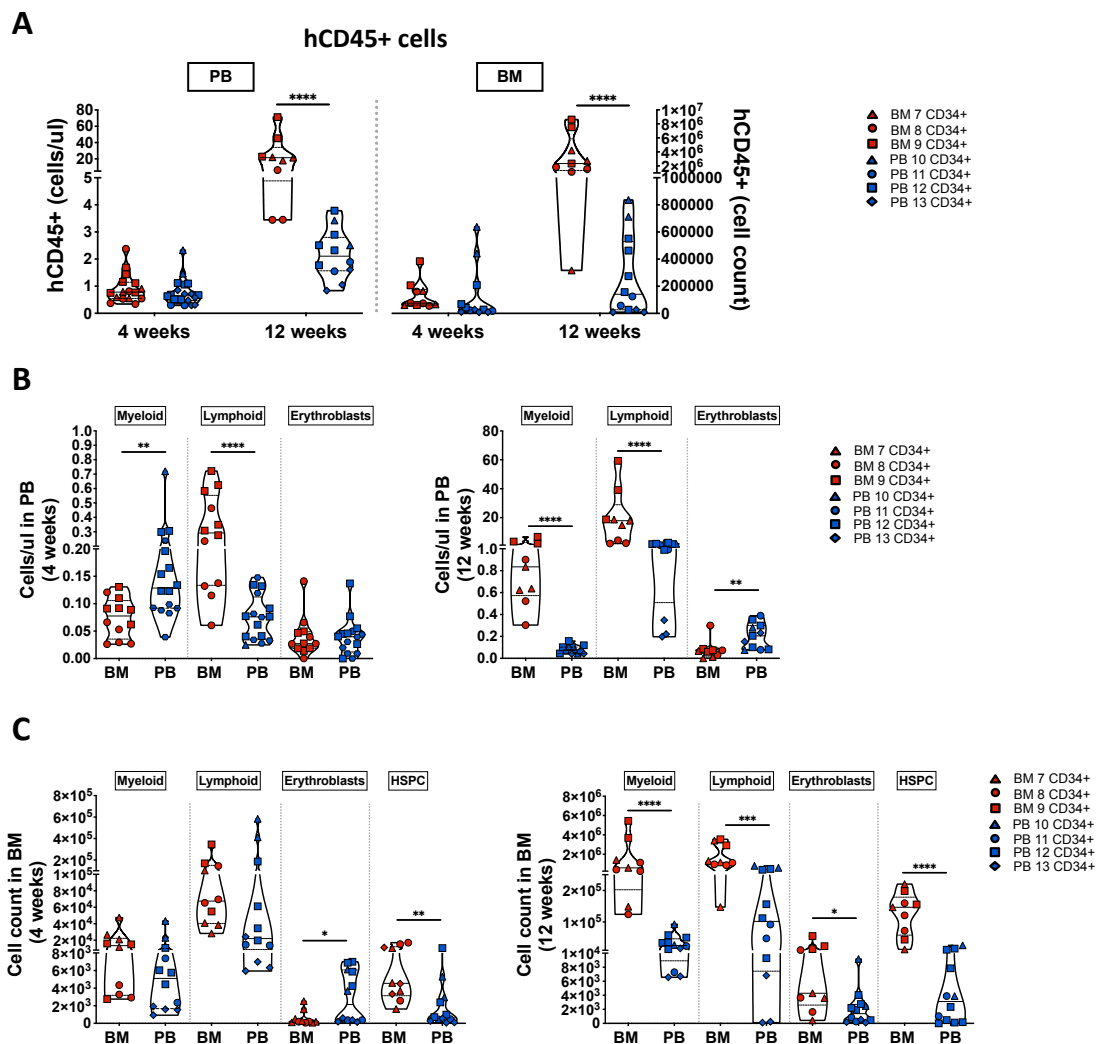


Figure 19. Human cell content in mice PB and BM at 4 and 12 weeks after transplantation with BM or PB CD34⁺ cells. (A) Graphs showing absolute count of human CD45⁺ cells detected at 4 and 12 weeks in PB (left) and BM (right) of mice transplanted with BM (red) or PB (blue) CD34⁺ cells. Each dot shows one animal. Data are shown as median with interquartile range. (B-C) Graphs showing absolute count of human myeloid cells, lymphoid cells, erythroblasts, and HSPC (only for C) detected at 4 (left) and 12 (right) weeks in PB (B) and BM (C) of mice transplanted with BM (red) or PB (blue) CD34⁺ cells. Each dot shows one animal. Data are shown as median with interquartile range. Statistical test used for groups' comparison: Mann-Whitney. Only statistically significant *p* values are reported within the graphs. *=*p*<0.05; **=*p*<0.01; ***=*p*<0.001; ****=*p*<0.0001.

To rule out diverse homing potential between PB and BM HSPC, we performed an *in vivo* homing assay, by assessing the human HSPC content in the murine BM of mice transplanted with the two CD34⁺ cell sources at 72 hours post-injection (**Figure 20A**). We detected a similar human HSPC count in the two experimental groups (**Figure 20B**), indicating similar homing capability of the two sources.

Of note, by comparing the composition of injected HSPC (**Figure 20C**) with the ones present in murine BM at 72 hours (**Figure 20D**), we observed an enrichment of HSC, MPP and CMP subpopulations in the engrafted fraction.

To compare the homing capability of PB- and BM-derived primitive and myeloid HSPC subpopulations, we calculated their homing index by normalizing the count of each engrafted population with their respective transplanted amount:

$$\text{Homing Index \%} = \frac{\text{engrafted cells}}{\text{infused cells}} * 100$$

As shown in **Figure 20E**, no statistically significant differences were detected in HSC, MPP and CMP populations, indicating comparable homing potential of the two HSPC sources.

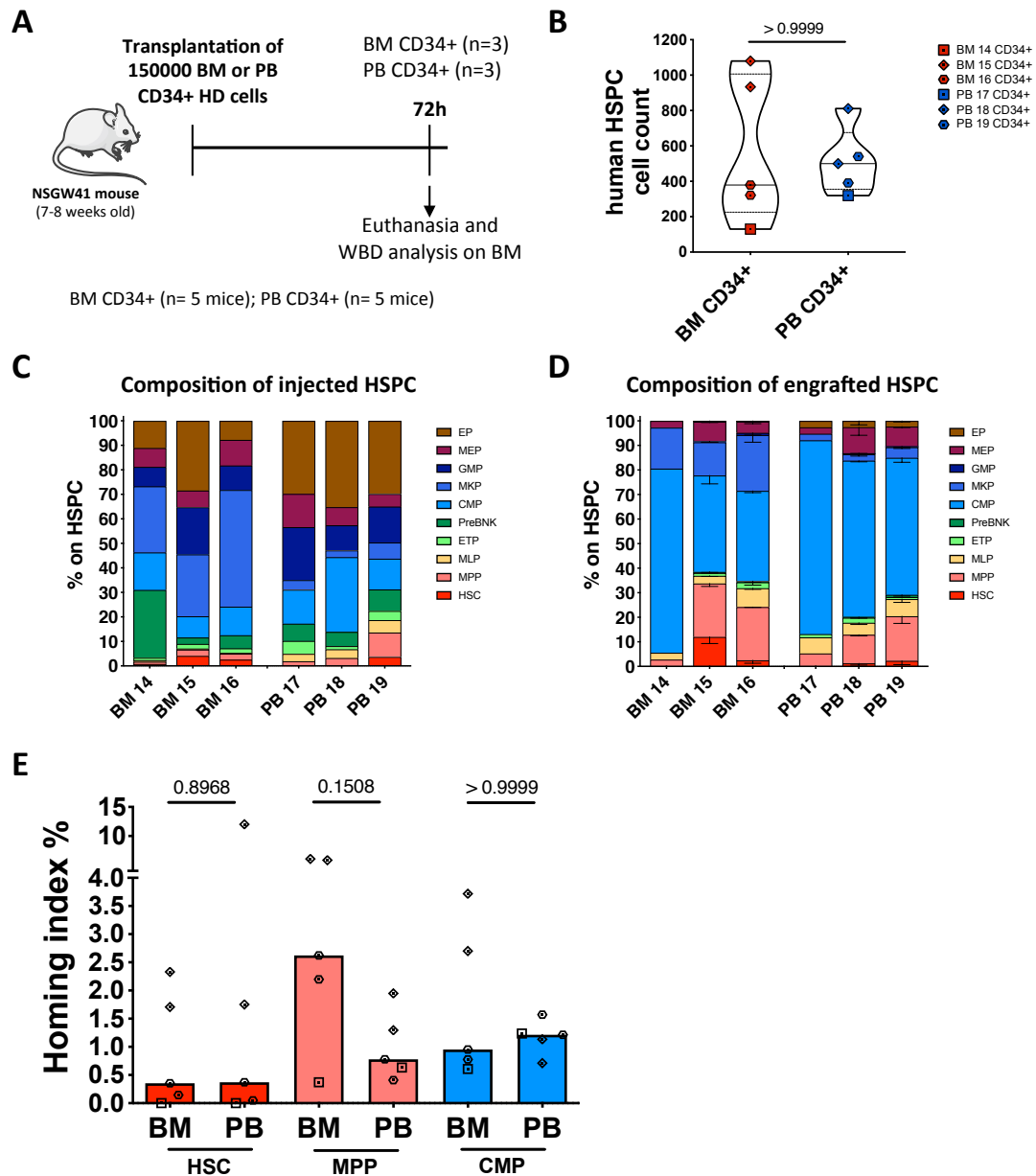


Figure 20. Comparison of in vivo BM homing potential of BM and PB CD34⁺ cells. (A) Experimental scheme for testing in vivo BM homing capability of BM and PB CD34⁺ cells. 150'000 CD34⁺ cells from BM (n=3) or PB (n=3) sources were transplanted in 7-8 weeks-old NSGW41 mice. A total of n=5 and n=5 mice were transplanted with BM and PB CD34⁺ cells, respectively. Mice were euthanized at 72 hours after transplantation and human BM cell content was analyzed by WBD. (B) Violin plot showing the amount of human HSPC detected at sacrifice (72 hours) in the BM of mice transplanted with BM (red) or PB (blue) CD34⁺ cells. Each dot shows one animal. Data are shown as median with interquartile range. (C-D) Stacked bar graph showing the phenotypic composition of HSPC derived from BM and PB sources before (C) and after (D) transplantation into NSGW41. Data are shown as mean with SEM. (E) Histograms displaying BM homing indexes of HSC, MPP and CMP subpopulations after transplantation with BM or PB CD34⁺ cells. Each dot shows one animal and columns represent the median values. For (B, E) Mann-Whitney statistical test was applied for groups' comparison and single p values are reported within the graphs.

In conclusion, our data show that circulating and BM resident HSPC have similar BM homing capability and multilineage output up to 4 weeks upon transplantation in NSGW41 mice (**Figure 19-20**). We also found that the reduced HSPC content and long-term human cell engraftment detected in cHSPC-transplanted mice (**Figure 16** and **Figure 17A-B**) can be only partially explained by the lower amount of infused primitive HSC in adult cHSPC (**Figure 17D-E**), thus suggesting that additional factors, including the biological cellular state of HSPC derived from the two sources, might justify the *in vivo* behavior observed.

3.6 cHSPC show enriched lymphoid and erythroid transcriptional signatures

To investigate the intrinsic molecular properties of HSPC subsets derived from the two sources, we analyzed the transcriptional profiles of BM- and PB-derived HSPC at the single-cell level.

We exploited a high throughput integrated approach, based on the combination of multiparametric immunophenotyping and Cellular Indexing of Transcriptomes and Epitopes by Sequencing (CITE-seq), which allows unbiased single-cell transcriptome and immunophenotypic analysis exploiting drop-seq technology (Stoeckius *et al*, 2017) (**Figure 21**). In brief, before any processing, an aliquot of each adult HD-derived BM and PB sample was analyzed by FACS in order to assess the HSPC composition. Then, we isolated primitive HSPC subsets ($LIN^- CD34^+ CD38^-$) and/or bulk HSPC ($LIN^- CD34^+$) or committed progenitors ($LIN^- CD34^+ CD38^+$) to perform scRNAseq. Moreover, to retrieve the phenotypic composition of sequenced cells and investigate the relationship between transcriptional and phenotypic profiling, samples were stained with TotalSeq-A antibodies, each conjugated to a unique antibody derived tag (ADT), to detect surface markers used for HSPC subset identification (**Table 11**) (please refer to the material and methods section for further technical details).

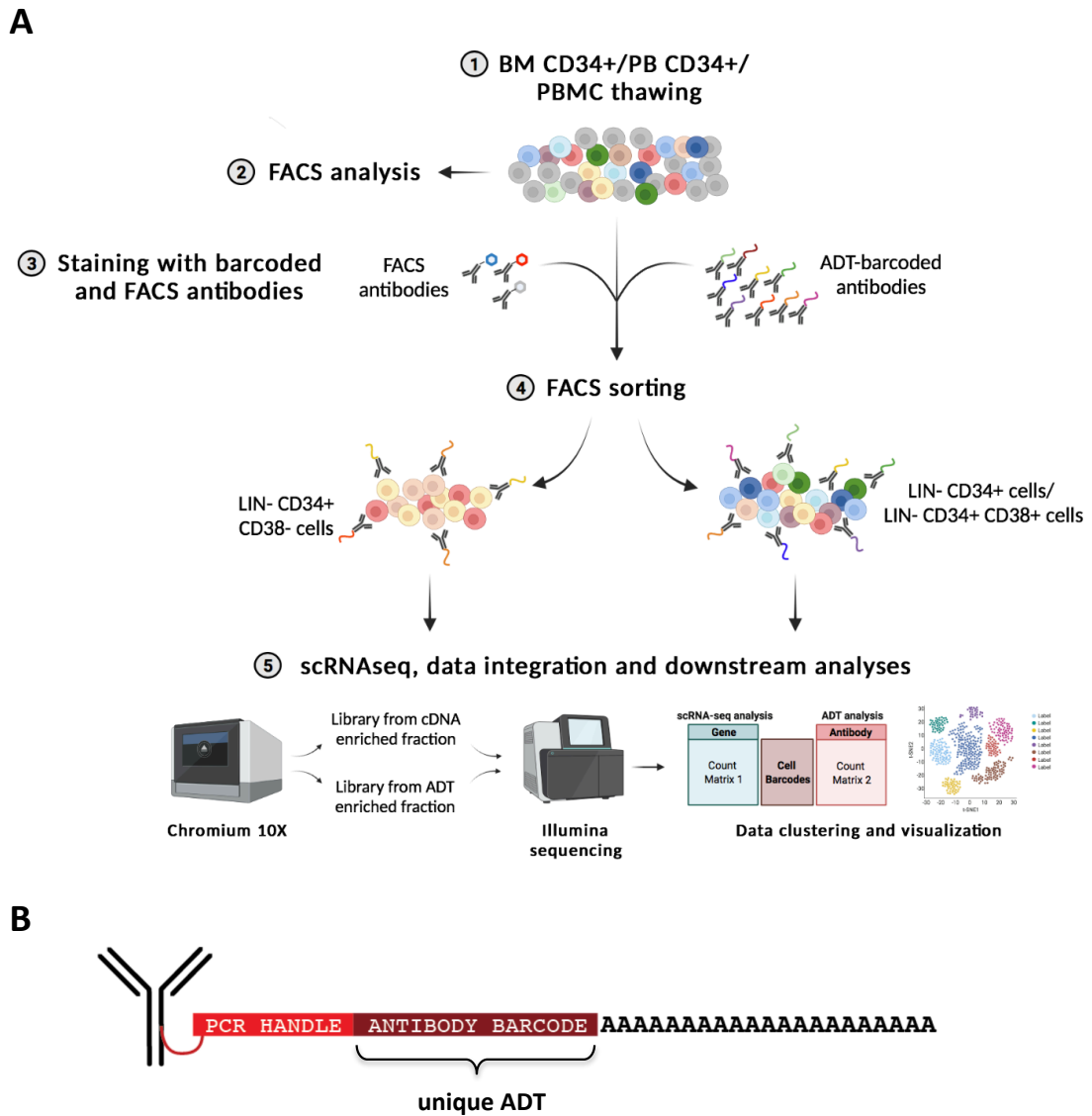


Figure 21. Experimental workflow of our integrated approach based on FACS immunophenotyping and sorting, scRNAseq transcriptome analysis and multiparametric immunophenotyping through ADT protein barcoding (CITE-seq technology). (A) 1) Thawing of human BM CD34⁺ (n=5) or PB CD34⁺ cells (n=4) or total PBMC (n=1). 2) Upon thawing, an aliquot of each sample was analyzed by FACS to assess HSPC phenotypic composition. 3) Samples were stained with a mix of FACS antibodies and ADT-barcoded antibodies directed toward surface antigens used for the phenotypic classification of human HSPC. 4) Sorting of primitive (Lin⁻ CD34⁺ CD38⁻) and/or bulk (Lin⁻ CD34⁺) or committed (Lin⁻ CD34⁺ CD38⁺) HSPC fractions, depending on the sample. 5) Sorted fractions were processed according to Chromium 10x protocol, generating for each sample separated sequencing libraries from cDNA or ADT enriched fractions. Libraries were subsequently sequenced and resulting data were analyzed and integrated by bioinformatic approaches. See the material and methods section for more technical details. (B) Schematic representation of a TotalSeq-A antibody. ADT (antibody derived tag).

Our final dataset comprised a total of 28315 BM HSPC (n=5) and 34461 PB HSPC (n=5) (Table 14). We applied the uniform manifold approximation and projection (UMAP) algorithm for dimensionality reduction (Becht *et al*, 2019) to visualize the analyzed cells. We checked the distribution of BM- and PB-derived cells in the full dataset and we observed that most regions of the UMAP were enriched for one of the two sources (Figure 22A-B). Furthermore, by analyzing cells belonging to PB or BM single donors, we found similar segregation according to the cell source, confirming that the origin of HSPC, rather than the donors, represented the main variable driving UMAP single cell distribution (Figure 22C).

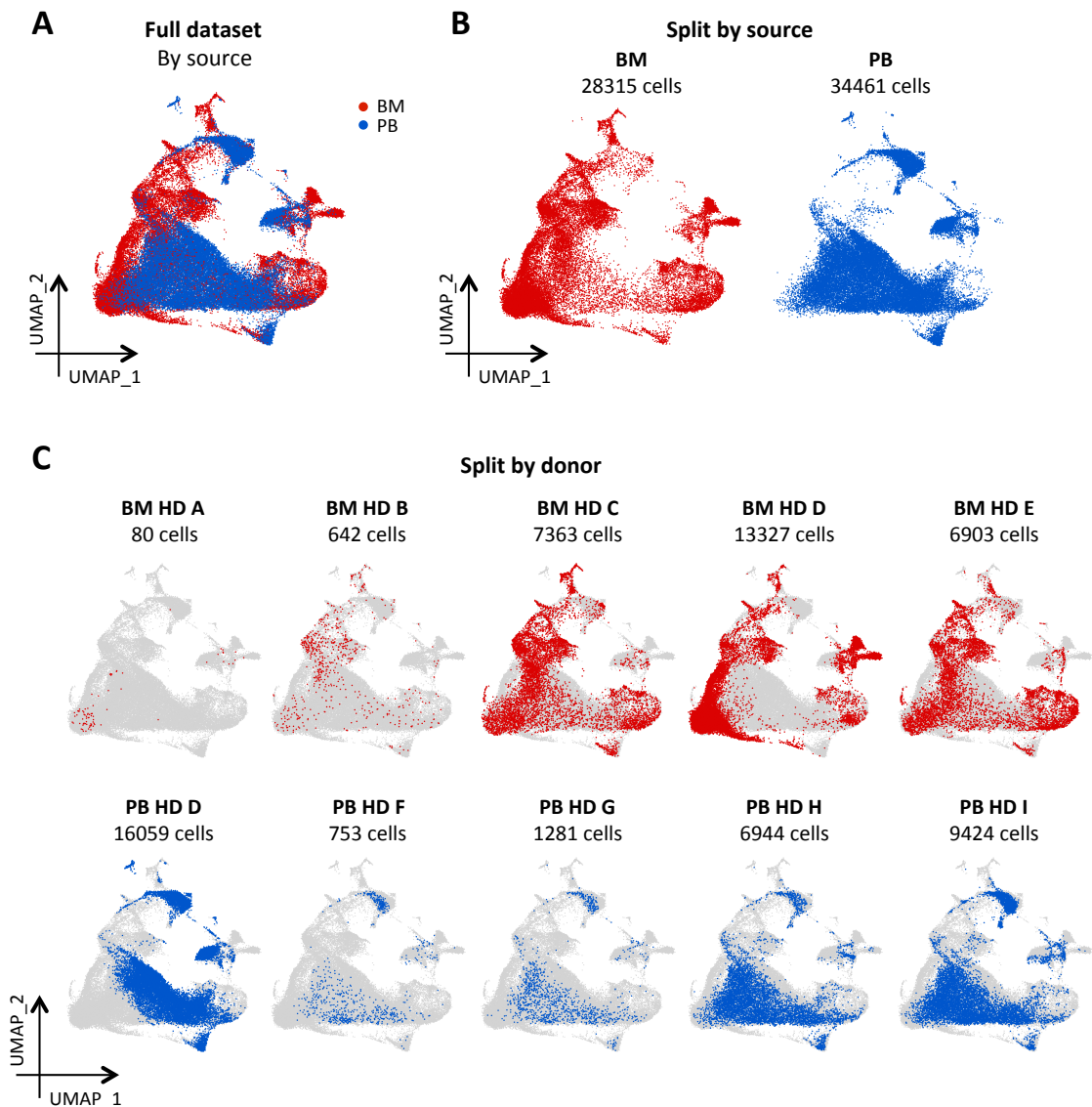


Figure 22. Distribution of scRNAseq-processed cells by source and donor. (A) UMAP representing the distribution of BM- (red) and PB- (blue) derived cells in the full dataset (B)

Distribution of cells in the UMAP according to the source. (C) Distribution of donors' single cells in the full dataset. The numbers of single cells retrieved and analyzed for each donor are reported.

The unsupervised clustering at a resolution of 0.6 partitioned the full dataset in 18 diverse clusters (**Figure 23A-B**). By comparing the cluster transcriptional profiles with published reference datasets and according to differentially expressed HSPC marker genes, we assigned to each cluster an HSPC subset label. Through this approach, we were able to transcriptionally identify in our dataset all the diverse HSPC subpopulations, from the most immature (clusters 1, 11, 12, 0, 9) to myeloid (clusters 5, 7, 10), lymphoid (clusters 4, 15, 17), megakaryocytic (cluster 16) and erythroid (clusters 2, 3, 6, 8, 13, 14) committed ones (**Figure 23C**). Interestingly, we identified 3 diverse clusters associated to primitive HSC signature: cluster 1, 11 and 12; the latter two in close proximity to cells expressing erythroid signatures, which may represent erythroid-biased HSC. Moreover, we identified transcriptional signatures of both immature (clusters 2, 3, 8, 14) and more mature (cluster 6, 13) erythroid progenitors.

Analyzing the distribution of BM and PB HSPC-derived cells, we observed that cHSPC were enriched in clusters with multipotent, erythroid and multi-lymphoid progenitor transcriptional signatures, while BM counterpart displayed a more mixed distribution among all clusters (**Figure 23D**). This result was consistent with our phenotypic characterization that showed a higher proportion of phenotypically defined MLP and MEP/EP in PB than BM HSPC (**Figure 10B**). Moreover, in line with the segregation by source observed in **Figure 22**, the vast majority of the clusters were enriched for PB or BM derived-cells (**Figure 23E and Table 3**).

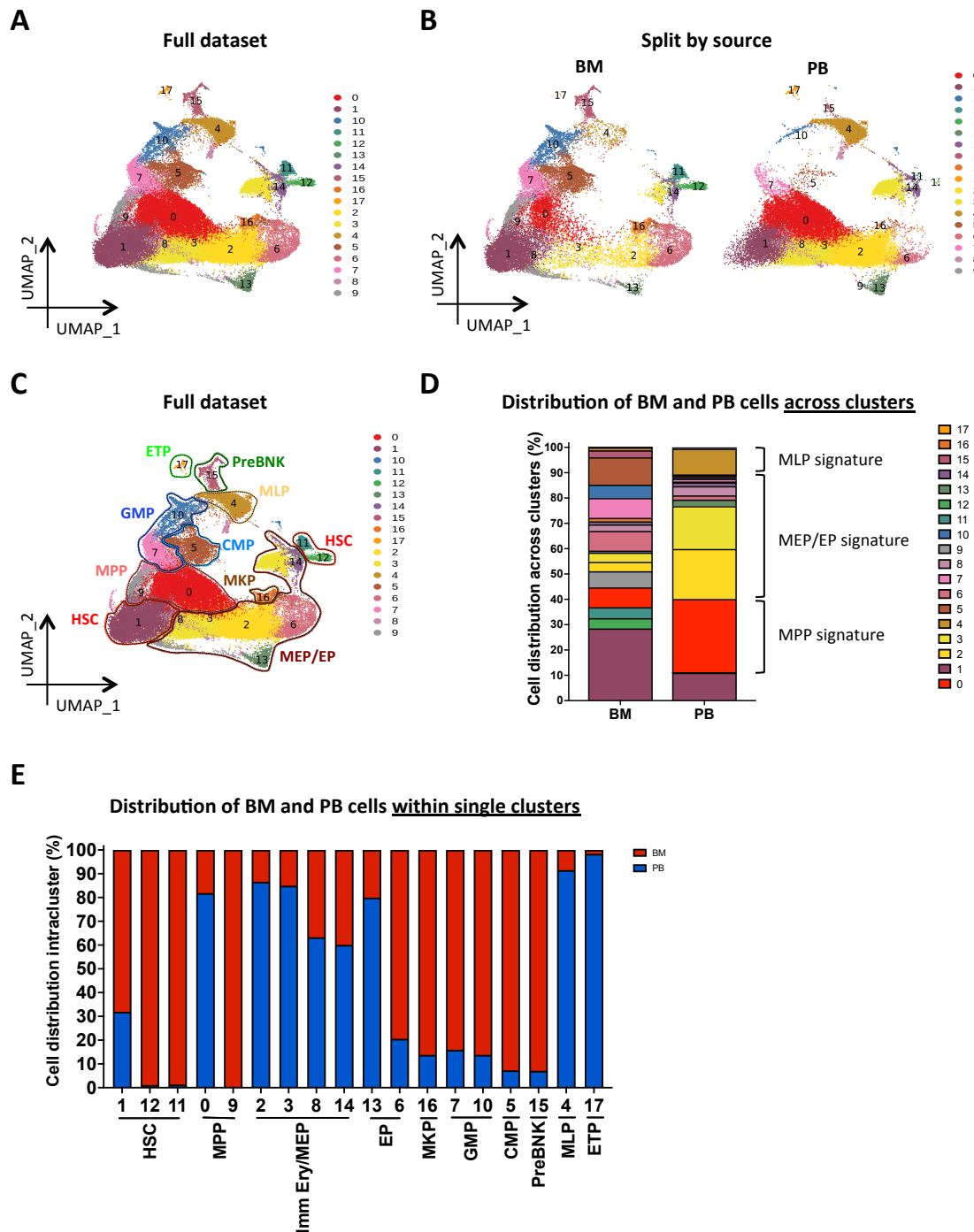


Figure 23. Unsupervised clustering and cluster annotation of HSPC subpopulations in scRNAseq dataset. (A) Unsupervised clustering of the full scRNAseq dataset at resolution 0.6. 18 distinct clusters were identified. (B) Distribution of cells belonging to the distinct clusters according to the source. (C) Cluster annotation based on the expression of published HSPC subsets-associated marker genes. (D) Stacked bar graph showing the distribution across the diverse transcriptional clusters of BM and PB cells. The dominant transcriptional signatures identified for PB cells are reported. (E) Stacked bar graph representing the distribution of BM and PB cells within each single cluster. X axis shows cluster annotation of HSPC subsets.

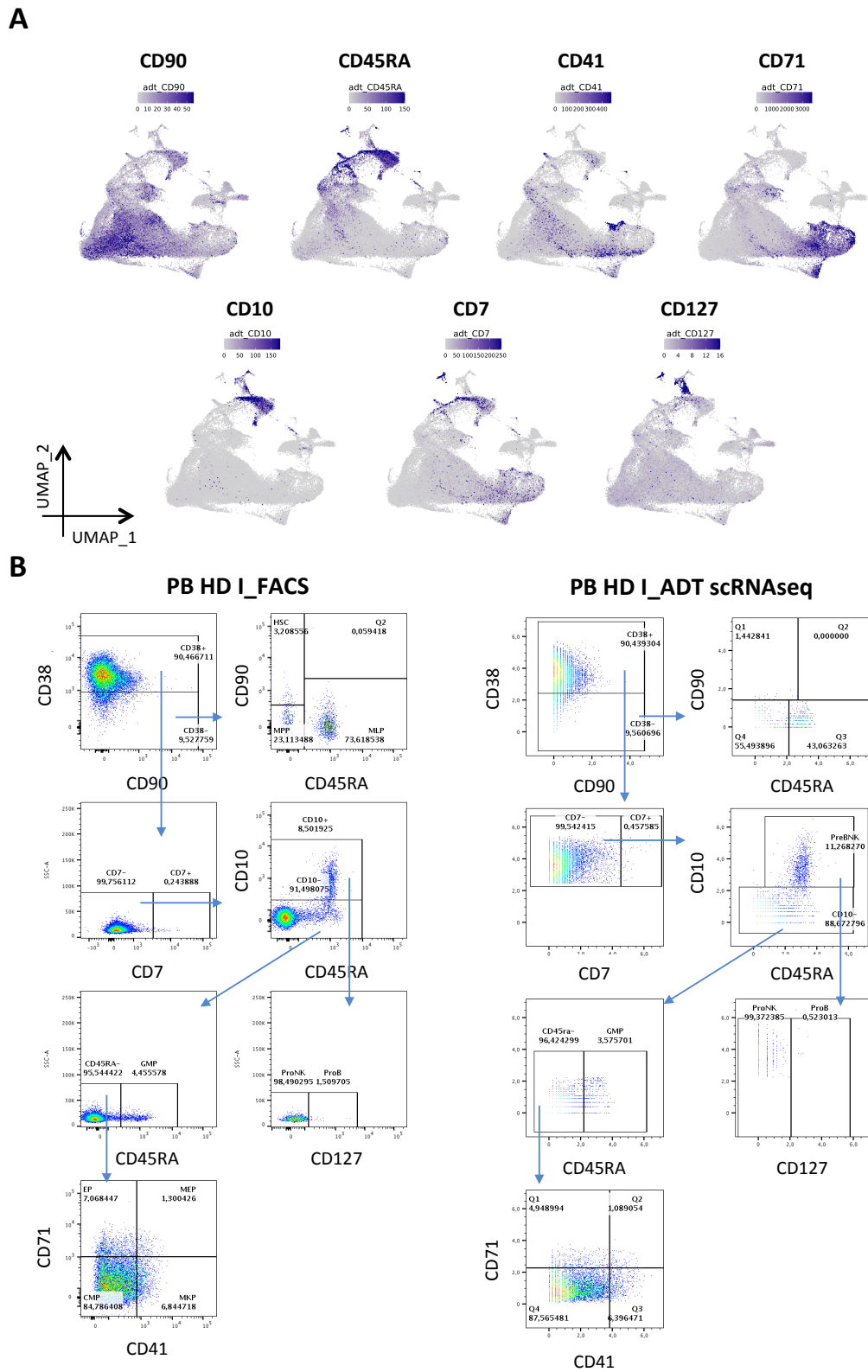
Table 3. Number of BM and PB cells composing transcriptional clusters identified in scRNAseq dataset.

		SOURCE	
		BM	PB
TRANSCRIPTIONAL CLUSTERS	0	2222	10020
	1	8066	3783
	2	1051	6802
	3	1034	5839
	4	330	3565
	5	3072	241
	6	2203	571
	7	2221	421
	8	728	1255
	9	1805	2
	10	1486	238
	11	1218	16
	12	1156	13
	13	220	878
	14	341	513
	15	772	59
	16	387	62
	17	3	183

3.7 Phenotypic circulating CMP and primitive HSPC subsets are transcriptionally and functionally committed toward erythroid lineages

Once identified the transcriptional signatures for all clusters, we analyzed the phenotypic composition of our scRNAseq dataset through TotalSeq-A technology, with the goal of integrating transcriptional and phenotypic profiles of processed single cells.

We were able to detect the expression of the distinct phenotypic markers (**Figure 24A**), with a good matching between the types of antigen expressed and the cluster transcriptional signatures identified (**Figure 23C**). For ADT-based HSPC immunophenotyping, we applied the same gating strategy used for identification of the distinct HSPC subsets by FACS analysis (**Figures 21A, step 2 and 24B**)



By reporting each ADT-phenotypic subset on the UMAP representing the full dataset (**Figure 25A**), we observed that each subpopulation, except CMP, occupied a well-defined UMAP region, with few overlapping areas among each other (**Figure 25C**). We compared ADT- and FACS-phenotyping and we observed that ADT-defined phenotypic composition of BM and PB HSPC (**Figure 25D, left**) was remarkably comparable with the one gained by FACS (**Figure 25D, right**), showing high levels of reproducibility between the two technologies.

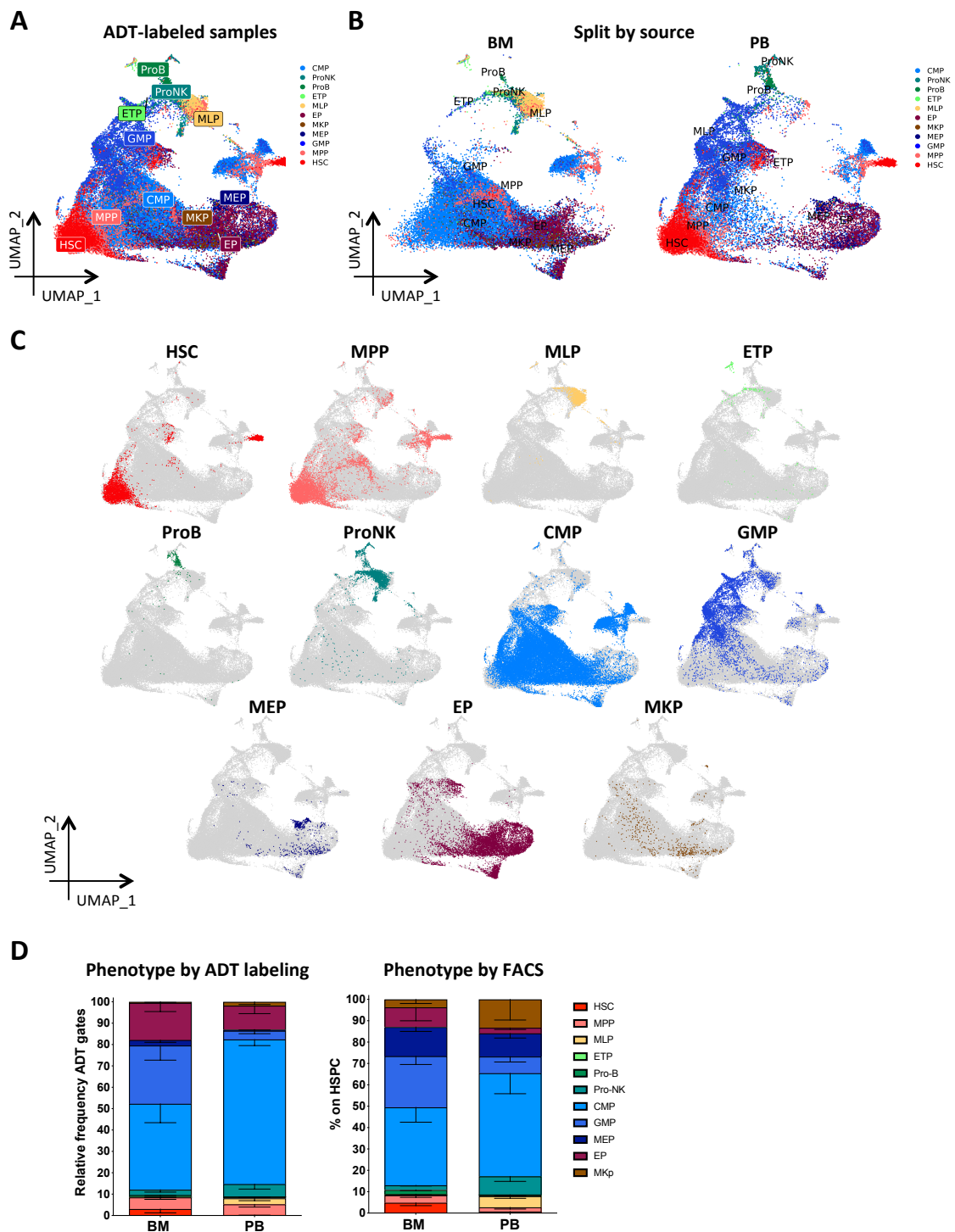


Figure 25. Phenotypic assessment of scRNAseq-processed cells. (A) Phenotypic HSPC subpopulations identified on total scRNAseq dataset by ADT protein barcoding. (B) Distribution of cells belonging to the distinct ADT-identified HSPC subpopulations in the two sources. (C) Distribution of single ADT-identified HSPC subsets in the total dataset. (D) Stacked bar graphs showing phenotypic compositions of total BM ($n=5$) and total PB ($n=5$) cells detected by ADT barcoding (left) and FACS analysis before any processing (right). Data are shown as mean with SEM.

Table 4. Distribution of BM and PB cells among the diverse HSPC subsets identified by ADT barcoding.

		SOURCE	
		BM	PB
ADT-DEFINED HSPC SUBSETS	HSC	3351	55
	MPP	5248	1199
	MLP	9	1267
	CMP	8613	21570
	GMP	5493	1524
	ProB	200	54
	ProNK	505	1875
	ETP	34	203
	MEP	505	154
	EP	3541	4004
	MKP	94	522

Next, we combined phenotypic and transcriptional data by annotating the dominant ADT HSPC phenotype for each cluster-associated transcriptional signature (**Table 5**). We detected a good matching between ADT HSPC phenotype and transcriptional profiles for the majority of clusters, including the most primitive HSC (cluster 1 and 12). Moreover, we observed that clusters 8, 9 and 11, which were not clearly associated to specific HSPC subset transcriptional signatures and showed both a primitive (cluster 11) or an erythroid-primed (cluster 8 and 9) transcriptional bias, were ADT-phenotypically classified as CMP.

Table 5. Transcriptional and ADT phenotypic profiles detected for scRNA-seq single clusters. Table showing the prevalent transcriptional signatures and the ADT-defined phenotypic HSPC subsets identified for each cluster. Question marks identified clusters which do not show univocal transcriptional signatures.

	Transcriptional signature	Phenotype (ADT)
1	HSC	HSC, MPP, CMP
12	HSC (only BM)	HSC
0	MPP	CMP, MPP
4	MLP	MLP
15	PreBNK	ProB, ProNK
17	ETP, T cells	ETP
5	CMP	CMP/GMP
7	GMP	GMP
10	GMP	GMP
2	MEP/Immature Ery	CMP, EP, MEP, MKP
3	MEP	CMP
14	Erythroid/MPP	MPP
6	EP	EP
13	EP	EP
16	MKP	MEP, MKP
11	Primitive (?)	CMP
8	Erythroid (?)	CMP
9	Erythroid (?)	CMP

Looking at the distribution across clusters of ADT-defined CMP from BM and PB, we observed that BM phenotypic CMP had a mixed myeloid, erythroid and primitive signature, while PB phenotypic CMP showed a prominent erythroid/primitive gene expression and a considerably reduced myeloid transcriptional component (**Figure 26A**). Moreover, we found an enriched erythroid transcriptional signature also in PB ADT-phenotypic MPP (**Figure 26B**) and, at a lower extent, HSC (**Figure 26C**). The erythroid transcriptional commitment of PB HSPC was further confirmed by the higher

levels of expression of genes driving erythroid differentiation (*GATA2*, *TALI*, *NFE2*) observed in the majority of PB ADT-phenotypic subsets (**Figure 26D**).

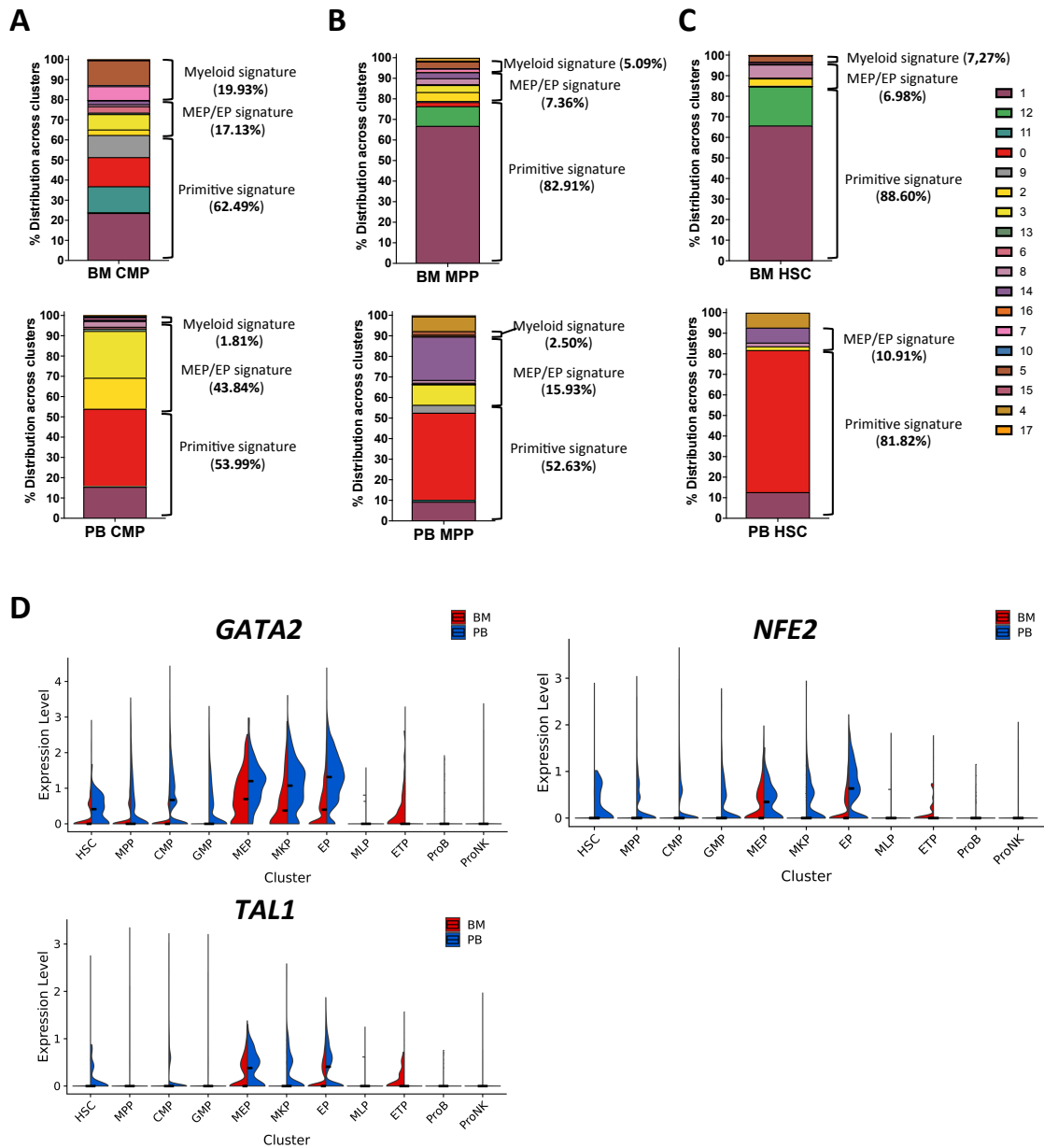


Figure 26. Comparison of transcriptional and ADT-based phenotypic profiles in BM and PB scRNAseq datasets. (A-C) Stacked bar graphs showing the distribution across the diverse transcriptional clusters of BM and PB CMP (A), MPP (B) and HSC (C) identified by ADT barcoding. For each subset, clusters are grouped based on the dominant transcriptional signatures identified. The proportion of cells expressing myeloid, erythroid and primitive transcriptional signatures on total ADT-defined HSPC single subsets are shown. (D) Violin plots showing the expression of genes driving erythroid differentiation (*GATA2*, *TALI*, *NFE2*) in BM (red) and PB (blue) HSPC subsets identified by ADT barcoding. For each plot, thick bars represent median values.

To assess whether this transcriptional erythroid commitment resulted in increased functional erythroid output, we tested the differentiation potential of BM- (n=4) and PB- (n=5) derived HSC, MPP and CMP at single cell level, exploiting our *in vitro* multilineage differentiation protocol. As positive control for erythroid differentiation, we seeded also single MEP and EP derived from either BM or PB sources (**Figure 27**).

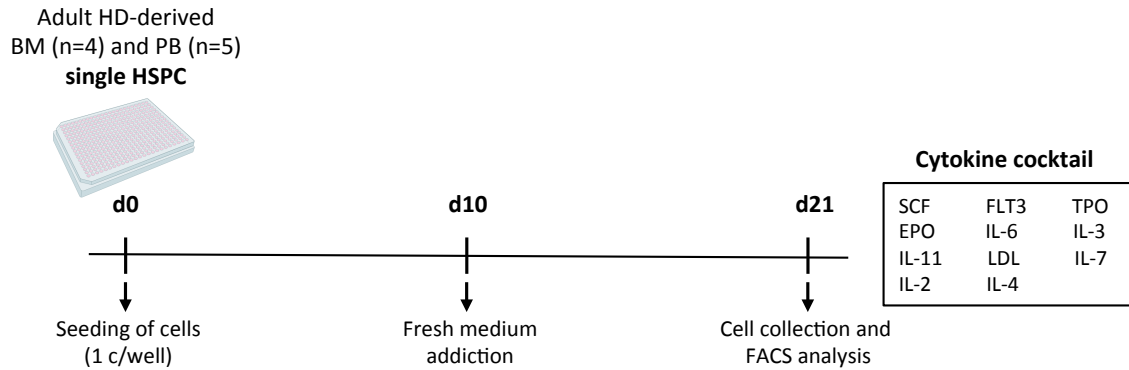


Figure 27. Single cell *in vitro* differentiation assay of BM vs. PB HSPC. Experimental scheme for single-cell *in vitro* differentiation assay of BM and PB HSPC. After sorting, one single HSC/MPP/CMP/MEP/EP per well from BM (n=4) and PB (n=5) sources was cultured in a medium containing a cocktail of cytokines to assess the differentiation into all major hematopoietic lineages (myeloid, lymphoid, erythroid, megakaryocyte). Single cells were cultured for 3 weeks and final cellular outputs were analyzed by FACS. See the material and methods section for more technical details.

First of all, we calculated the differentiation efficiency of each tested population, defined as the number of wells with a cellular output at the end of the culture (*N positive wells*) divided by the total number of sorted single cells (*N seeded wells*):

$$\text{Differentiation efficiency (for each HSPC subset) \%} = \frac{N \text{ positive wells}}{N \text{ seeded wells}} * 100$$

As reported in **Table 6**, we estimated an efficiency of more than 20% for all PB and BM HSPC subpopulations, with PB cells showing a higher differentiation propensity than BM ones, except for EP subset.

Table 6. Differentiation efficiency of BM- and PB-derived HSC, MPP, CMP, MEP, EP subsets at single cell level. Table representing for each BM- or PB-derived HSPC subset 1) the number of wells seeded with one cell, 2) the count of wells showing differentiated progeny after 3 weeks of culture, 3) the total differentiation efficiency starting from one single cell.

		Seeded wells	Wells with output	Efficiency %
HSC	BM	43	12	28
	PB	64	30	47
MPP	BM	43	16	37
	PB	66	29	44
CMP	BM	43	10	23
	PB	66	42	64
MEP	BM	43	23	53
	PB	65	47	72
EP	BM	44	25	57
	PB	64	29	45

Next, we assessed the *in vitro* expansion rate of single cells that showed a cellular output after 21 weeks of culture (positive wells) (**Figure 28**). We detected a statistically significant increased cell output from PB MPP and CMP with respect to the BM counterpart, while no major differences emerged for PB- and BM-derived HSC, MEP and EP subpopulations (**Figure 28**).

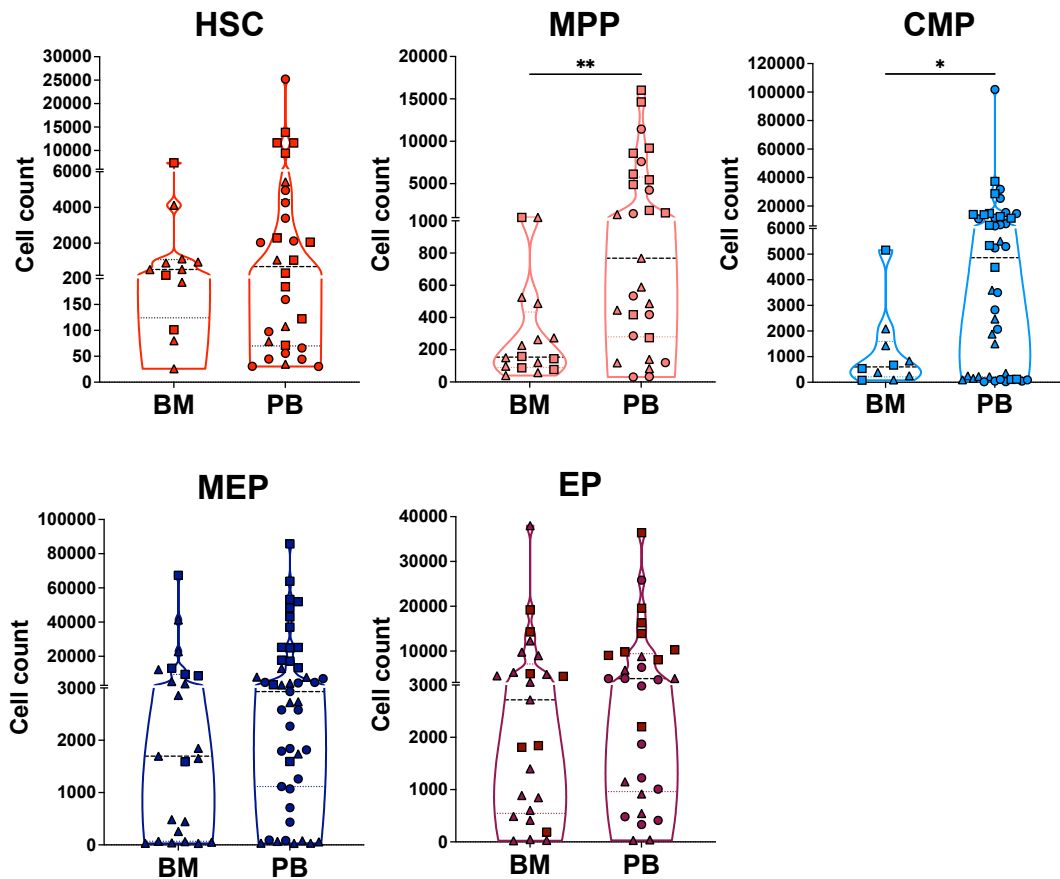


Figure 28. Expansion rate of BM and PB HSPC single-cell cultures. Violin plots showing the total expansion rate of single BM- or PB-derived HSC, MPP, CMP, MEP and EP after 3 weeks of culture. Each point represents the total cell count retrieved from one single cell at the end of the culture and all point referring to a unique HSPC subset, from either BM or PB sources, are pooled together. Points shape refers to experiments run in distinct days. Data are shown as median with interquartile ranges.

To better quantify the differentiation skewing of BM and PB HSPC subpopulations toward the diverse blood lineages, for each analyzed HSPC subset we constructed a heatmap, representing the relative frequencies of myeloid, lymphoid, erythroid, MK and immature CD34⁺ cells retrieved at the end of the culture (**Figure 29**). We detected an increased proportion of erythroid progeny from PB-derived HSC (**Figure 29A**), MPP (**Figure 29B**) and CMP (**Figure 29C**) than BM counterparts, which in turn displayed more balanced differentiation. By contrast, wells seeded with single erythroid committed progenitors (MEP and EP) isolated from the two sources showed quite comparable differentiation outcomes, displaying almost unilineage erythroid specification (**Figure 29D-E**).

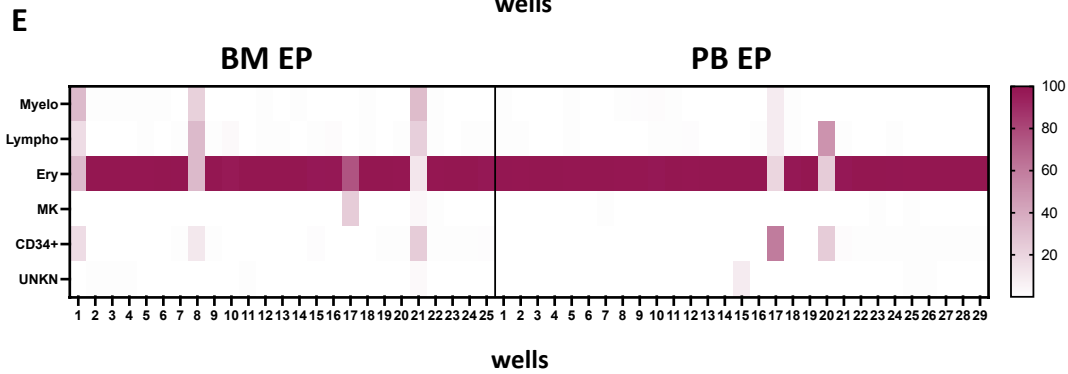
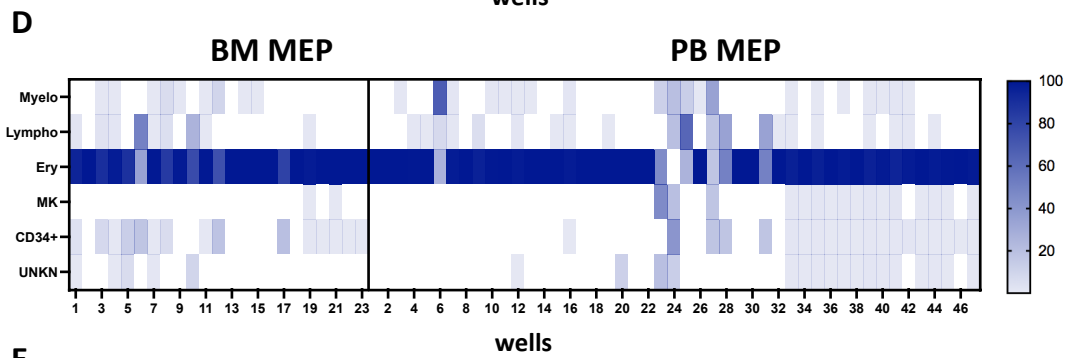
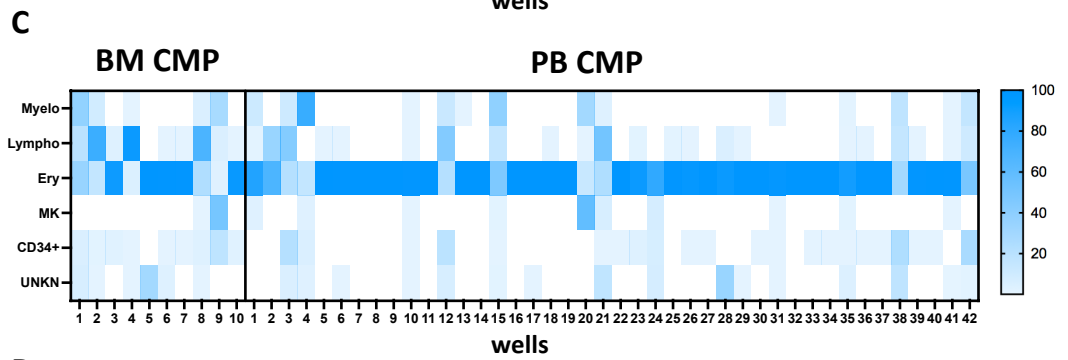
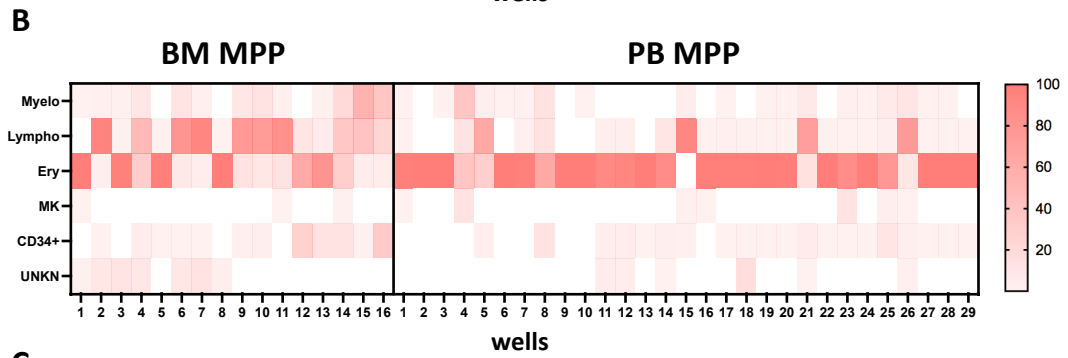
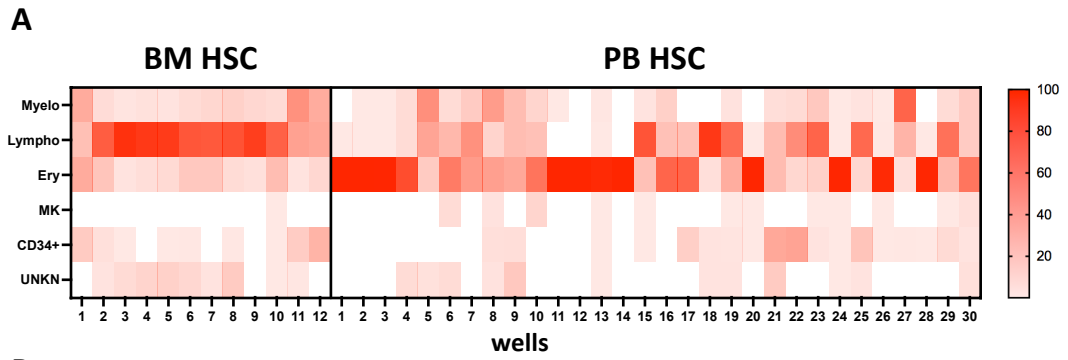


Figure 29. In vitro differentiation propensity of single BM and PB HSPC. (A-E) Heatmaps showing for each well with positive output (x axis) the relative frequencies of myeloid, lymphoid, erythroid, MK, immature CD34⁺ and unclassified (unknow) cells retrieved after 3 weeks of culture (y axis) from single HSC (A), MPP (B), CMP (C), MEP (D), EP (E) isolated from PB or BM sources. Color intensity is proportional to the relative frequencies of the distinct differentiation progenies.

In summary, our integrated analyses showed an overall good matching between phenotype and transcriptome of the distinct HSPC subsets derived from PB and BM, with PB phenotypic CMP, HSC and MPP subsets displaying an increased transcriptional and functional commitment toward erythroid differentiation as compared to BM HSPC subpopulations.

3.8 cHSPC are poised for differentiation, adhesion, and immune activation with respect to BM counterpart

To gain insight into the biological properties of HSPC derived from the two sources, we analyzed the cell cycle activity of BM and PB HSPC through scRNA profiling. According to the expression of marker genes of G1, S and G2M cell cycle phases, we classified clusters as low cycling (G1 cells >75% of cluster cells), intermediate cycling (G1 cells >50% and <70% of cluster cells) and cycling (G1 cells <35% cluster cells) (**Figure 30A-B**). Low cycling clusters were mainly composed of cells with primitive and immature signatures, while higher cell cycle activity was associated with transcriptional profiles of committed progenitors. Furthermore, clusters number 1 and 12, transcriptionally defined as HSC, displayed distinct cell cycle activities, with cluster 12 (which was mainly composed of BM cells) being more cycling than cluster 1 (**Figure 30B**). By splitting the entire dataset into BM (**Figure 30C-D**) and PB (**Figure 30E-F**) components, we found a general increased cell cycle activity in resident than circulating cells, with almost all S and G2M cells of low cycling clusters (0, 1, 2, 3, 4, 8, 17) belonging to BM counterpart.

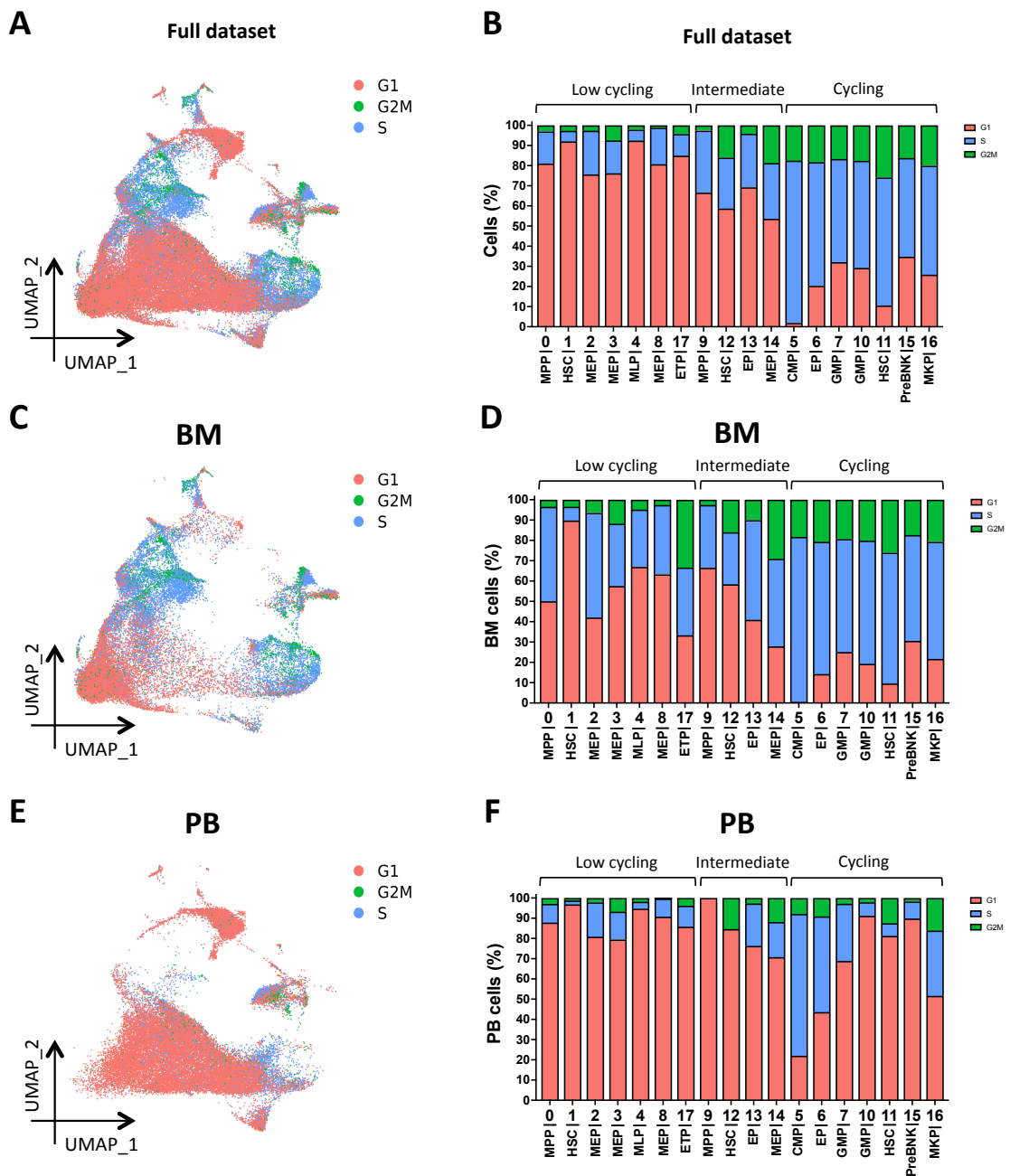


Figure 30. Analysis of cell cycle activity in scRNAseq dataset. (A, C, E) UMAPs displaying the distribution of cells in G1, S or G2M cell cycle phases in total BM+PB (A), only BM (C) and only PB (E) datasets. (B, D, F) Histograms representing the proportion of cells in G1, S and G2M cell cycle phases for clusters of total BM+PB (B), only BM (D) and only PB (F) datasets. X axis shows cluster annotation according to HSPC subset transcriptional signatures. Based on cell cycle activity, clusters of total dataset (B) were classified as low cycling (G1 cells >75% of cluster cells), intermediate cycling (G1 cells >50% and <70% of cluster cells) and cycling (G1 cells <35% cluster cells). The classification defined for the total dataset (B) was then reported on only BM (D) and only PB (F) compartments.

To explore the differential biological role of PB and BM HSPC, we performed intra-cluster comparisons between PB vs. BM cells applying Gene Set Enrichment Analysis (GSEA), focusing on Gene Ontology-Biological Processes (GO-BP) genesets. For each cluster, GO-BP terms were divided into two subgroups, according to the assigned normalized enrichment score (NES): GO-BP terms with negative NES values ($NES < 0$, NESneg subgroup), composed by genesets with an enriched expression in BM vs. PB cells; GO-BP terms with positive NES values ($NES > 0$, NESpos subgroup), encompassing genesets showing an enriched expression in PB vs. BM cells. In order to reduce redundancy within GSEA GO-BP output lists, we performed a semantic reduction for both NESneg (**Figure 31**) and NESpos (**Figure 32**) lists, obtaining a reduced list of significant macro-categories (please refer to the material and method section for further technical details).

We observed that within almost all single clusters, BM cells showed an enriched expression of GO-BP macro-categories related to high transcriptional, metabolic and replicative cell states, as well as cellular responses to stress as compared to PB cells (NESneg subgroup, **Figure 31**). On the other hand, an enriched expression of ontologies associated to differentiation, adhesion and immune response/activation was detected in PB than BM cells within most single clusters (NESpos subgroup, **Figure 32**).

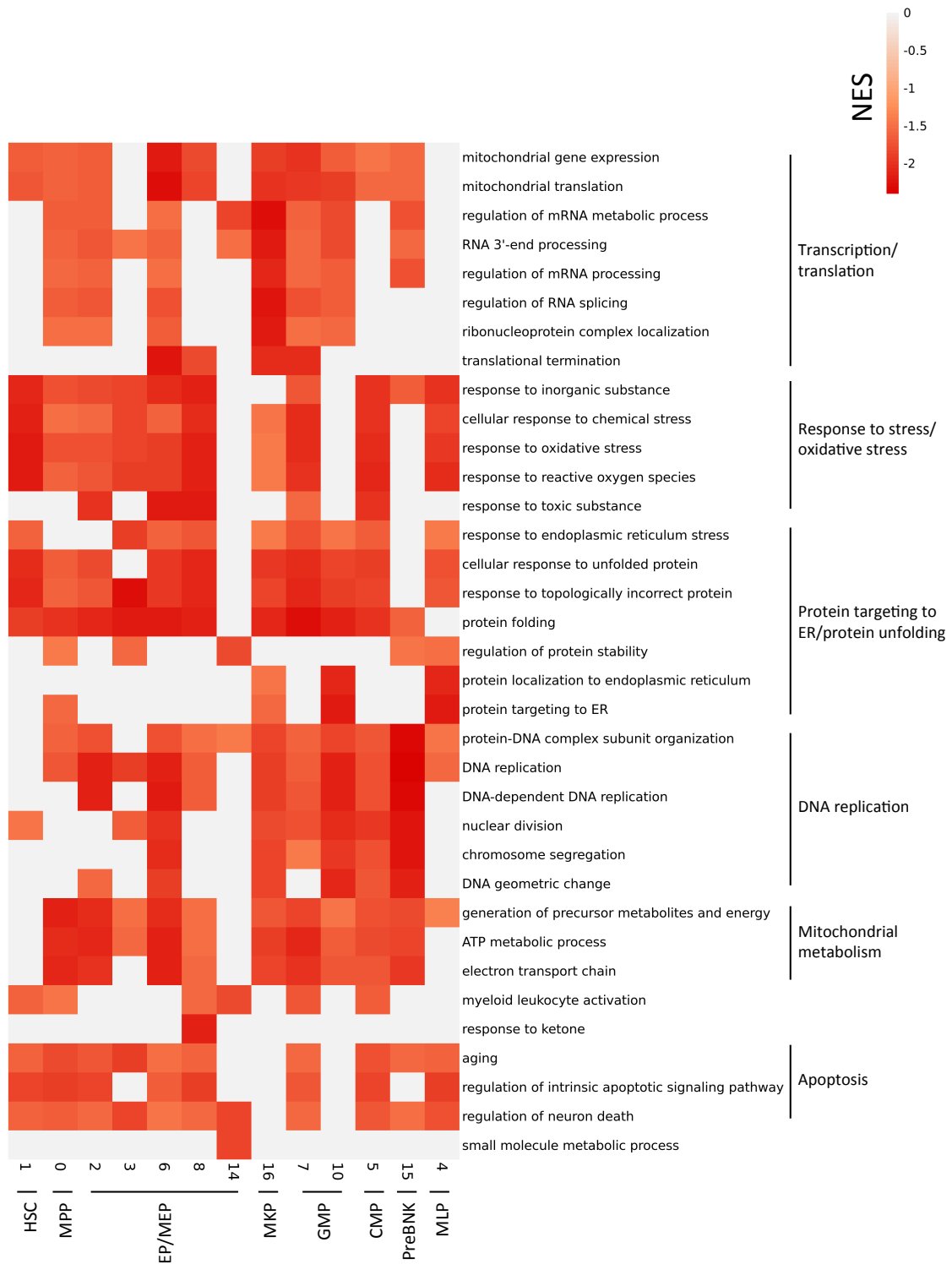


Figure 31. Tile plot of the top 5 Gene Ontology-Biological Processes (GO-BP) macro-categories per cluster expressing NES < 0 values (NESneg subgroup). X axis shows only clusters with statistically significant ($p < 0.05$) GO-PB ontology terms with the annotated transcriptional signatures. For each macro-category, color intensity is proportional to NES absolute values and light grey refers to non-statistically significant comparisons. Macro-categories were classified in diverse groups, according to the associated biological functions.



Figure 32. Tile plot of the top 5 Gene Ontology-Biological Processes (GO-BP) macro-categories per cluster expressing NES>0 values (NESpos subgroup). X axis shows only clusters with statistically significant ($p<0.05$) GO-PB ontology terms with the annotated transcriptional signatures. For each macro-category, color intensity is proportional to NES absolute values and light grey refers to non-statistically significant comparisons. Macro-categories were classified in diverse groups, according to the associated biological functions.

Focusing on marker genes of cluster 1, which shows a transcriptional signature of primitive HSC, we found that BM cells had a higher expression of gene modules related to increased cellular activity, such as transcription/translation, protein folding and cellular response to stress, while PB cells showed an enriched expression of marker genes associated to interferon (IFN) signaling, as well as to low transcriptional/proliferative state (**Figure 33**). These results, combined with the reduced cell cycle activity (**Figure 30**) and the enriched expression of differentiation genesets (**Figure 32**), could suggest that low-cycling trafficking HSPC, including the most primitive HSC subpopulation, are poised for faster activation and differentiation in order to be ready to exert their functions in case of demand.

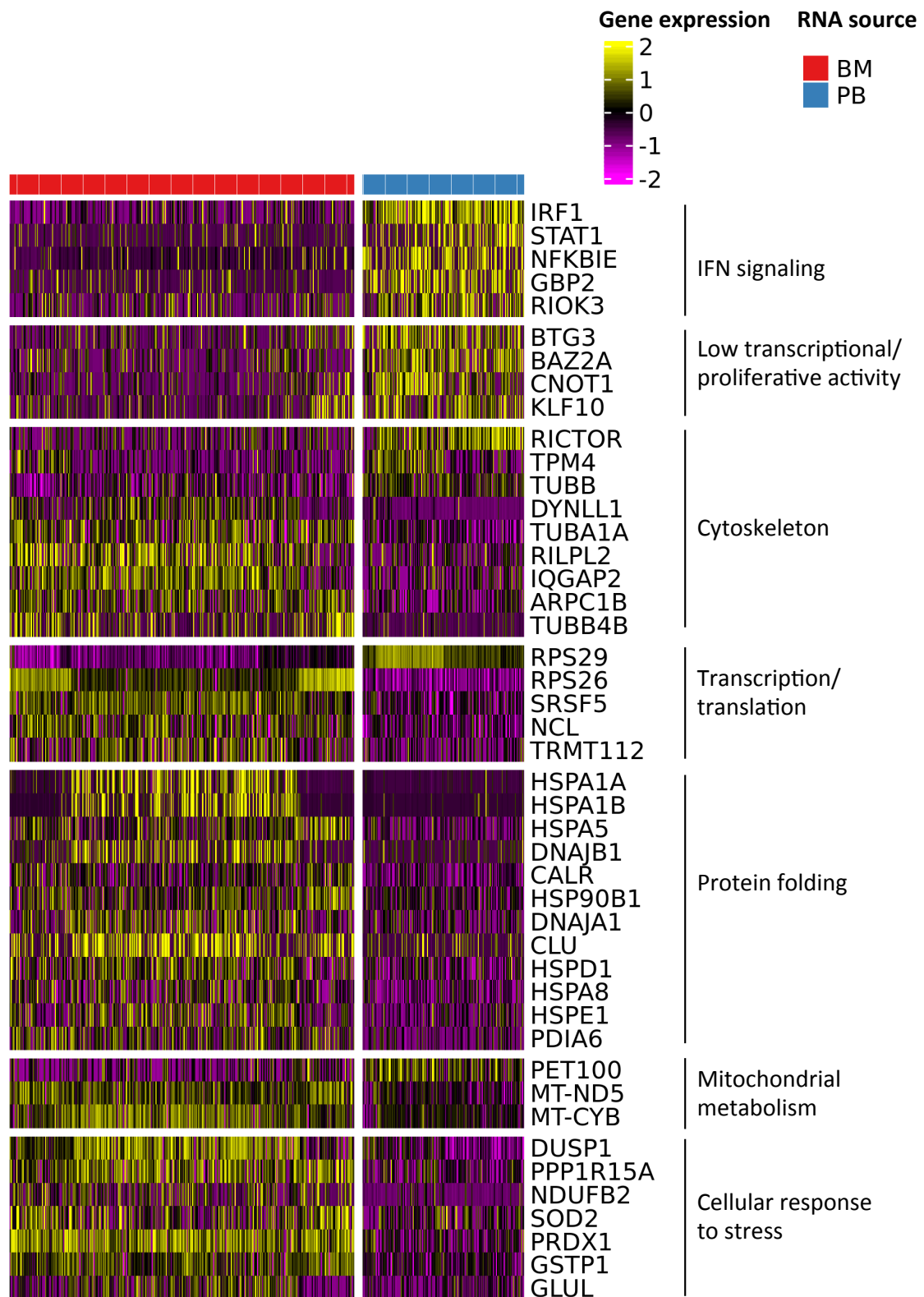


Figure 33. Single cell expression heatmap of selected marker genes of cluster 1, which showed a differential expression between the two sources. Marker genes were classified in diverse groups, according to the associated biological functions. Annotation for BM (red) and PB (blue) cells is reported.

3.9 Investigating HSPC trafficking *in vivo* in humans through IS analyses

Subjects treated with gene-corrected HSPC represent a unique model to study human hematopoiesis *in vivo* since, upon transduction, each HSPC and its progeny become univocally marked by a distinct vector IS. By combining WBD immunophenotyping and high-throughput IS analysis, we followed HSPC-GT patients treated at SR-TIGET to track cHSPC dynamics overtime and assess their relationship with BM HSPC both during active reconstitution (<1 year post-GT) and at steady-state hematopoiesis (≥ 1 year post-GT).

We phenotypically characterized cHSPC overtime after GT in a cohort of 12 WAS-, 17 MLD- and 8 MPSIH-GT patients and we observed a drop of cHSPC count at 7 days after gene therapy in all patients analyzed, followed by a recovery displaying different kinetics according to the group of treated subjects (**Figure 34**). Indeed, cHSPC count normalized at 30 days after GT in WAS-GT patients, while stabilized starting from 60 days in MLD-GT subjects. MPSIH-GT patients displayed a prominent increase of trafficking HSPC at 30 days after GT, with a 5-fold increase than pre-GT levels, which then normalized in the long term.

By evaluating cHSPC composition before GT, during active reconstitution (30 days post-GT) and after restoration of steady state hematopoiesis (1 year post-GT), we detected no major differences among the patients and the time points analyzed (**Figure 35**).

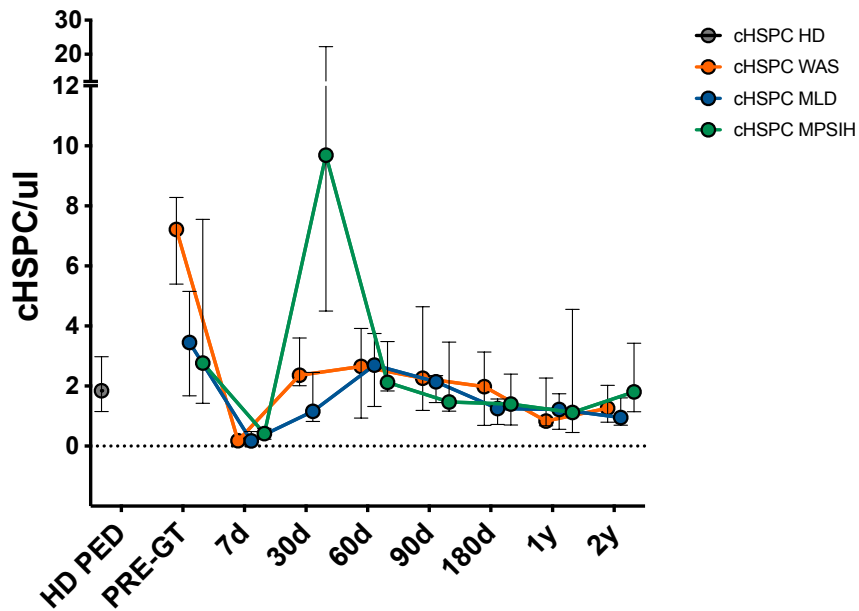


Figure 34. Kinetics of cHSPC count in WAS-, MLD- and MPSIH-GT patients before and after transplantation. Graph showing the amounts of cHSPC detected overtime in WAS- (orange), MLD- (blue)- and MPSIH- (green) GT patients before (PRE-GT) and at different follow-ups after treatment. Numbers of patients analyzed are reported in Table 7. cHSPC count of aged-matched individuals are shown in grey. All data are shown as median with interquartile range.

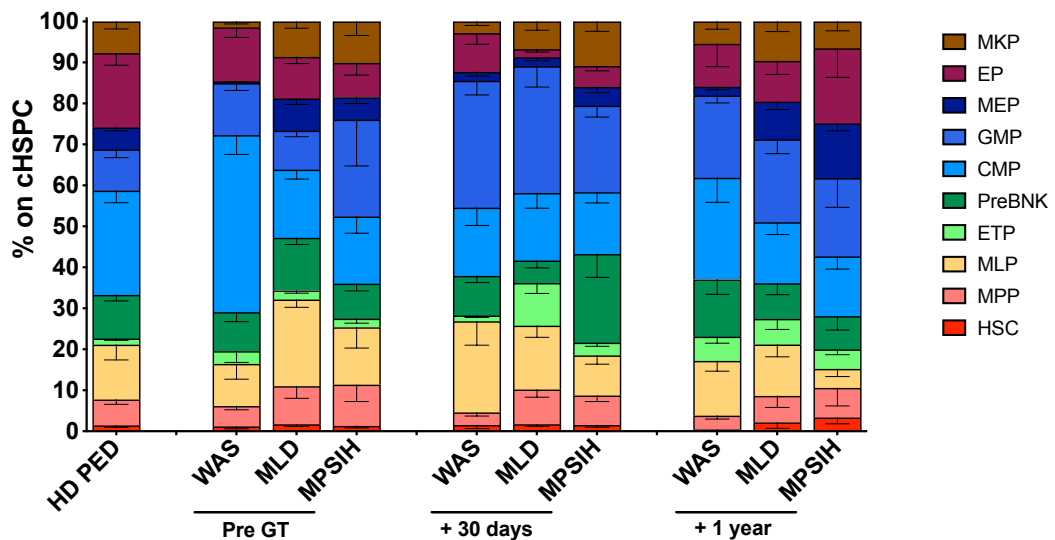


Figure 35. Phenotypic composition of WAS, MLD and MPSIH GT-patients before GT, at 30 days and 1 year after treatment. Reference composition of age matched pediatric HD is reported. Data are shown as mean with Standard Error Mean (SEM).

Thus, to evaluate preferential recirculation of specific HSPC populations, we calculated the CI for each HSPC subset at the same time points indicated above (**Figure 36**). CI analysis showed a higher recirculation of primitive HSPC subsets in MLD- and MPSIH-GT patients at 30 days after treatment, which then normalized at pre-GT levels at 1 year-follow up (**Figure 36B-C**). These data may suggest an early contribution of primitive subsets to HSPC clonal redistribution to diverse BM niches upon transplantation.

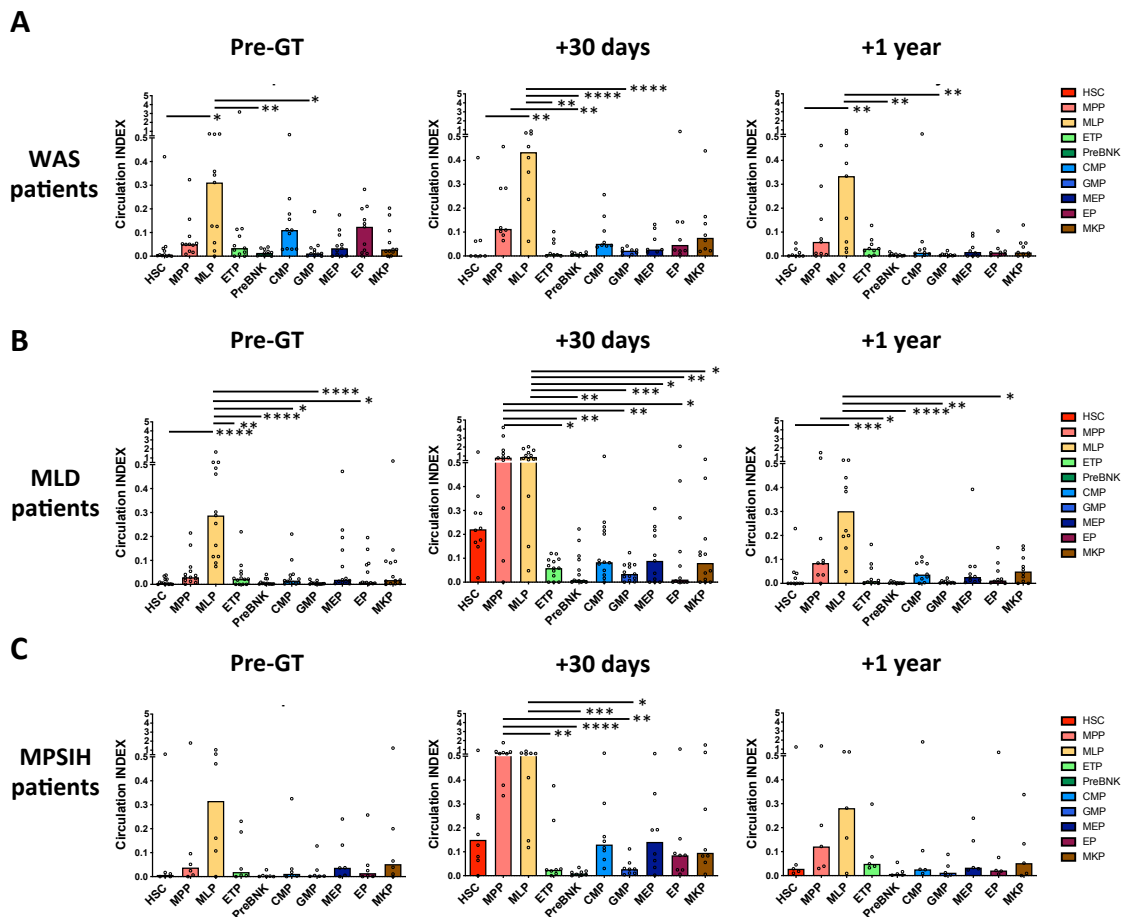


Figure 36. Circulation indexes (CI) in GT-treated patients before and after transplantation. (A-C) CI estimated in all HSPC subpopulations derived from WAS- (A), MLD- (B) and MPSIH- (C) GT patients before GT (pre-GT), at 30 days and at 1 year after transplantation. Statistical tests: Kruskal-Wallis test with Dunn's multiple comparisons test. $*=p<0.05$; $**=p<0.01$; $***=p<0.001$; $****=p<0.0001$. Medians are shown.

To test this hypothesis, we analyzed IS belonging to PB and BM HSPC isolated from 3 WAS GT-patients at early (≤ 90 days) and late (1 year) phases after treatment, already

collected by our group for other research purposes. Importantly, for all the patients, we analyzed HSPC from two different BM sites (right and left sites of iliac crest) at the same time point, both at early and late phases after GT, in order to investigate the IS sharing level of HSPC clones belonging to distant niches during active reconstitution and at steady state (**Figure 37A**).

In line with our hypothesis, we found an almost 2-fold higher level of sharing between PB and BM HSPC at early phases with respect to steady state condition (**Figure 37B**), suggesting recirculation of HSPC clones. This initial trafficking resulted in higher clonal sharing between distant BM niches at steady state. Indeed, we found a 2-fold increase of the IS shared between the left and right BM sites at 1 year after GT with respect to early phases (**Figure 37B**).

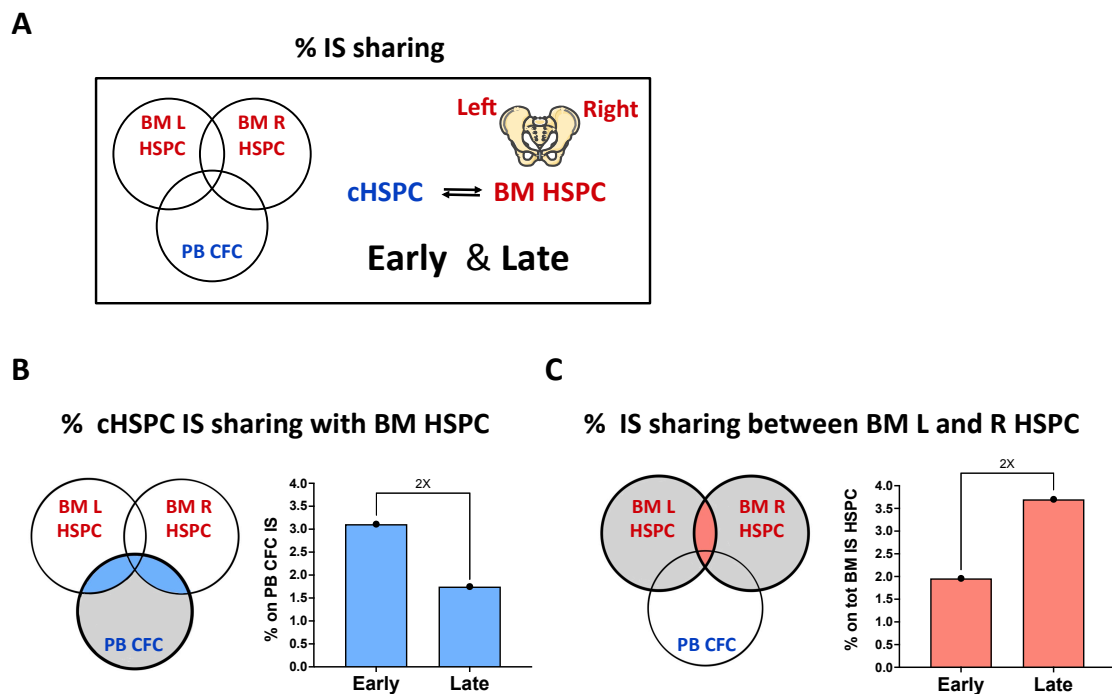


Figure 37. Analyses of IS sharing between PB HSPC and BM HSPC derived from distinct BM sites (left and right) in 3 WAS patients at early and late phases after GT. (A) Schematic representation of analyses based on IS sharing among cHSPC and BM HSPC retrieved from left (BM L) and right (BM R) BM niches of 3 WAS-GT patients. PB HSPC IS were retrieved starting from PB CFC. (B) Histogram showing the frequency of PB CFC IS shared with total BM HSPC at early (≤ 90 days) and late (1 year) phases after GT. (C) Histogram representing the frequency of IS sharing between HSPC derived from left and right BM sites calculated on total BM HSPC at early (≤ 90 days) and late (1 year) phases after GT.

Finally, to have an indication of the hematopoietic output of circulating vs. resident HSPC *in vivo*, we analyzed the IS shared between cHSPC or BM HSPC and PB mature myeloid (Granulocytes+Monocytes) and lymphoid (B, T and NK cells) lineages at 1 year post-GT. Consistently with the enrichment of lymphoid and myeloid transcriptional clusters in PB and BM HSPC, respectively (**Figure 23**), cHSPC shared higher number of IS with the lymphoid compartment, while resident HSPC with the myeloid one (**Figure 38**).

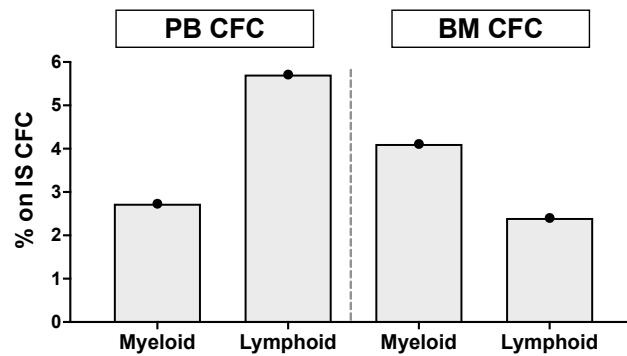


Figure 38. Percentage of IS sharing between PB or BM CFC and PB myeloid and lymphoid mature lineages in 3 WAS patients at 1 year post-GT. The relative frequency is calculated on IS retrieved from PB CFC (left) or total BM CFC (right).

In summary, our preliminary results collected from GT-patients could suggest that during early hematopoietic reconstitution, cHSPC might be in charge of clonal redistribution to recipient's distant BM sites. Moreover, the observation of shared IS between BM and PB HSPC also at late phases could indicate that HSPC trafficking may act as a physiological mechanism to connect multiple BM niches also at steady state in humans. Finally, we have first indication that cHSPC could have a role in seeding lymphoid organs to locally differentiate in lymphoid progeny.

4 DISCUSSION

The existence of cHSPC has been known for a long time. Initially found in mouse, they were shown to repopulate the BM and give rise to functional hematopoiesis (Goodman & Hodgson, 1962; Wright *et al*, 2001; Mazo *et al*, 2011). Several roles have been proposed for these cells in the murine setting, both in physiological and under stress conditions (Wright *et al*, 2001; Massberg *et al*, 2007; Burberry *et al*, 2014). However, cHSPC have been poorly characterized in humans, partly due to intrinsic limitations, such as the low number of cells found in circulation. Few works, mainly with descriptive purposes, have been published so far on this rare human population (Cohen *et al*, 2013; Tsaganos *et al*, 2006; Skirecki *et al*, 2019; Napolitano *et al*, 2016; Abdellatif, 2018; Pizarro *et al*, 2014; Wu *et al*, 2014; Santoro *et al*, 2020), and little information is currently available on their physiological functional properties in relationship with BM resident HSPC.

In the present study, we performed a comprehensive characterization of the biological and functional properties of cHSPC in humans, providing an exhaustive picture of the phenotypic composition, transcriptional profiling and functional features of trafficking HSPC in relationship with their BM counterpart. The rarity of human cHSPC introduced a scientific challenge in studying this population. To face this issue, we exploited a combination of multi-parametric flow-cytometry, scRNAseq and *ad hoc* designed *in vitro* and *in vivo* assays suitable for studying the biology of limited cell numbers isolated from PB of healthy donors, even starting from few milliliters derived from young individuals. Furthermore, to study trafficking and resident HSPC hierarchical relationships and differentiation potential *in vivo* in humans, we exploited IS clonal tracking of cHSPC, BM HSPC and mature PB lineages isolated from a cohort of GT-treated patients.

Despite it was described that cHSPC count negatively correlates with age (Cohen *et al*, 2013), no study has dissected so far the kinetics and changes in composition of human physiologically trafficking HSPC across aging, from newborns to aged subjects. In our study we collected and applied our detailed phenotypic characterization on 114 PB and 48 BM samples derived from HD of various ranges of age, generating a large

dataset of reference for studying cHSPC subsets during hematopoietic system development. Our data confirmed a consistent enrichment of cHSPC count in young individuals (**Figure 8a** and **Figure 9**), and a distinct composition between circulating and resident BM HSPC, indicating different functions (**Figure 10**).

The existence of human trafficking lymphoid-committed progenitors was observed years ago by Six and colleagues, who identified a population of Lin⁻ CD34⁺ CD10⁺ CD24⁻ cells showing B, NK and T cell, but not myelo/erythroid differentiation potential, not only in CB and BM, but also in PB and thymus of human HD (Six *et al*, 2007). Moreover, it was recently described that donor-derived HSPC present in intestinal allografts are capable of seeding recipient thymus after organ allotransplantation, where they undergo selection to generate *de novo* recipient-tolerant T cells (Fu *et al*, 2019). Our phenotypic analyses show that MLP fraction is higher in PB than BM HSPC (**Figure 10**), displaying an increased recirculation propensity among all HSPC subsets and across the different groups of age analyzed (**Figure 11**). These results were in line with the enriched lymphoid transcriptional signature found in PB vs. BM HSPC (**Figure 23**), as well as the higher IS sharing of cHSPC with PB lymphoid than myeloid mature lineages with respect to BM resident counterpart (**Figure 38**). Altogether, these findings suggest that steady-state trafficking lymphoid HSPC may have a role in seeding lymphoid organs, where they could differentiate into mature lymphoid cells. Furthermore, the distinct circulation capability, measured through CI (**Figure 11**), of MLP and ETP populations during aging might suggest diverse functional roles of the two lymphoid-committed progenitors: while ETP, characterized by a lower CI in elderly, may be responsible of seeding the thymus, which shows a declining functionality during aging (Gui *et al*, 2012), MLP might be involved in lifelong extra-thymic lymphoid cell maturation.

Although BM is the site hosting homeostatic erythropoiesis, under stress conditions, such as blood loss, infections or chronic diseases, the higher erythropoietic demand could induce supportive extramedullary erythropoiesis, which mainly occurs in the spleen (Paulson *et al*, 2020). In mice, stress-induced erythropoiesis is supported by the active migration into the spleen of BM ST-HSC, which, once exposed to spleen microenvironment, commit to the erythroid fate and are able to expand and sustain extramedullary erythrocyte production during acute anemic insult (Perry *et al*, 2009;

Harandi *et al.*, 2010). Consistently with these results, Mende and colleagues recently described that steady-state human splenic HSPC are transcriptionally primed toward erythroid differentiation, and spleen HSC/MPP subsets are able to produce a higher number of erythroid colonies than their BM counterparts (Mende *et al.*, 2020). Human extramedullary erythropoiesis becomes relevant especially in case of severe and chronic disorders, however poor information is currently available about the mechanisms driving this phenomenon (Mende & Laurenti, 2021). In the present study, we found that human cHSPC are transcriptionally and functionally committed toward erythroid differentiation at steady state. In particular, we detected a higher production of erythroblasts in mice transplanted with circulating than BM-resident CD34⁺ cells (**Figures 16 and 19**), as well as an increased expression of transcription factors driving erythroid specification (**Figure 26D**) in PB than BM HSPC in our scRNAseq data. Moreover, we found that the proportion of phenotypic CMP, MPP and HSC with myelo/erythroid transcriptional commitment consistently changes between the two sources, with an increased erythroid transcriptional signature (**Figure 26A-C**) and differentiation bias (**Figure 29**) observed in PB than BM cells. In light of these findings, we speculate that erythroid-primed human cHSPC could seed the spleen or other extramedullary organs to guarantee a source of stress-responsive erythroid progenitors for sustaining rapid extramedullary erythropoiesis in case of emergency.

Indeed, our scRNAseq analyses pointed out that all cHSPC subsets, including the most primitive ones: a) are predominantly in G1 cell cycle phase (**Figure 30**); b) show a reduced transcriptional and metabolic cellular state (**Figures 31 and 32**); c) are transcriptionally primed for differentiation and, only for more committed progenitors, for adhesion and immune response in comparison to BM subpopulations, which in turn are more cycling and transcriptionally activated (**Figures 31 and 32**). Altogether, these data suggest a different role of PB and BM HSPC at steady state: while the transcriptional profiles enriched in BM-resident HSPC could be the result of multiple coordinated niche signals aimed at supporting hematopoietic turnover and homeostasis, the poised state of low-cycling PB trafficking HSPC may suggest their role in patrolling peripheral tissues for sustaining rapid activation and local hematopoietic differentiation in case of need. Moreover, the low replicative state of cHSPC could also act as a

protective mechanism of resistance to multiple stimuli they can encounter while circulating.

Some evidence in mouse suggested that cHSPC might have a role in connecting different BM niches, indicating continuous trafficking of HSPC between PB and BM. Our preliminary data on HSPC-GT patients showed an enrichment of cHSPC amount (**Figure 34**), with an increased recirculation of most primitive subsets (**Figure 36**), during active hematopoietic reconstitution after GT. Moreover, we observed an increased IS sharing between circulating and resident HSPC at early phases post-GT with respect to steady state condition, resulting in higher clonal sharing between distant BM sites at 1 year after treatment (**Figure 37**). Altogether, these results may suggest that trafficking HSPC may play a pivotal role in connecting multiple BM niches to allow clonal redistribution during active hematopoietic recovery after transplantation. Furthermore, finding shared IS between BM and PB HSPC also at steady state (**Figure 37B**) may suggest that, together with patrolling peripheral tissues for supporting local differentiation, the constant trafficking of human cHPSC could also act as a homeostatic mechanism for connecting multiple distant BM niches.

Given the clinical exploitation of HSPC for transplantation or gene therapy purposes, we also assessed the homing and repopulating potential of trafficking HSPC. Over the last years, few studies showed that human allograft-passenger HSPC can contribute to *in vivo* long-lasting hematopoietic chimerism in the recipient after organ allotransplantation (Alexander *et al*, 2008; Wang *et al*, 2012; Fu *et al*, 2019). Furthermore, the *in vitro* clonogenic potential and the *in vivo* ST-engraftment capability of adult HD-derived human cHSPC was recently proved (Brunet de la Grange *et al*, 2013; Bourdieu *et al*, 2018). However, an extensive evaluation of the functional properties of human cHSPC in relationship with their BM counterpart was still missing. Here, we assessed with unprecedented details that, similarly to their BM counterpart, human adult cHSPC are endowed with multilineage differentiation potential *in vitro* (**Figure 14**), and are able to home and stably engraft into murine BM upon transplantation, supporting a multilineage output *in vivo*, although at a lower extent than BM HSPC at 20 weeks (**Figures 16, 17, 19, 20**). Multiple factors could explain the differential long-term *in vivo* behaviors of HSPC derived from the two sources. Since it is known that hematopoietic repopulation is supported by primitive HSC in the long

term after transplantation (Scala *et al*, 2018), the reduced LT-engraftment potential could be the effect of the lower amount of infused primitive HSC in PB cells, as suggested by the positive correlation found between LT-human CD45⁺ cell content in the murine BM and the injected HSC number (**Figure 17D**). However, when we normalized the amount of human CD45⁺ cells detected in murine BM at long-term phases by the number of infused primitive HSC, the differences between mice transplanted with the two sources did not abrogate (**Figure 17E**), thus suggesting diverse intrinsic biological states between BM and PB HSPC. Indeed, the results of our scRNAseq transcriptional profiling showed that steady-state PB HSC have lower replicative and cellular activity, but higher basal expression of differentiation and IFN-pathway gene modules than BM HSC (**Figure 33**). Of interest, the activation of the IFN pathway in HSPC was described to be associated to increased proliferation, terminal differentiation and, if chronically stimulated, loss of self-renewal and exhaustion (Demerdash *et al*, 2021; Essers *et al*, 2009; King *et al*, 2015; Yang *et al*, 2005). Therefore, our results suggest that the low quantitative amount, together with the peculiar basal pre-activated state of PB HSC, could be at the basis of their faster exhaustion after transplantation, resulting in reduced LT-engraftment as compared to the BM counterpart. To functionally prove this hypothesis, future *in vitro* and *in vivo* studies based on the functional comparison of equal number of BM and PB HSC would be required. Moreover, further investigations will have to validate at protein level the IFN signaling activation in steady-state cHSPC, and to explore the biological reasons driving the basal increased expression of genes associated to this molecular pathway.

From a clinical perspective, since the wider application of newborn screening is expected to result in increased number of early-diagnosed patients within the first years of life, the higher number of circulating primitive HSPC observed in young individuals (**Figures 8 and 9**) could suggest the possible use of cHSPC as alternative stem cell source for this cohort of subjects. Indeed, both HSPC mobilization and collection through leukapheresis, as well as HSPC harvest through BM aspirates are still considered difficult clinical procedures for very young pediatric individuals, that can be performed only in experienced centers. Thus, targeted mobilization protocol based on new combination of drugs or shorter drug exposure could reduce collection times and make collection feasible even in very young infants, who have a higher propensity of

HSPC recirculation (**Figure 11**). In addition, collected HSPC could be subjected to *ex vivo* expansion to further increase the amount of cHSPC available for autologous GT applications. Although requiring functional validations, our results on adult HD-derived cHSPC support the idea that also pediatric cells could be endowed with both engraftment and multilineage differentiation potential. However, given the transcriptional pre-activated state of PB HSC isolated from adult HD (**Figure 33**) and the reduced *in vivo* LT-engraftment potential observed for adult-derived PB CD34⁺ cells (**Figures 16 and 17**), future investigations will have to address the transcriptional profile of pediatric cHSPC, as well as their *in vitro* and *in vivo* functional properties. These analyses will be essential to fully elucidate the repopulation and engraftment potential of pediatric cHSPC for therapeutic purposes.

Although providing novel findings on the biological role of human cHSPC at steady state, our study have left some pending questions that remain to be addressed. First, in our analyses we did not elucidate the molecular mechanisms driving HSPC egress from the BM and trafficking. Thus, future investigations will be focused on the dissection of the molecular factors tuning physiological HSPC recirculation, which could also provide novel targets for HSPC mobilization. Moreover, although previous works already described low cell cycle (Mende *et al*, 2020) and metabolic (Bourdieu *et al*, 2018) activity of extramedullary HSPC, additional studies will have to functionally validate the poised state detected at transcriptional level in cHSPC, as well as to shed light on their preferential routes of migration toward peripheral tissues capable of hosting lymphoid- or erythroid-primed differentiation, both at steady state and in stress conditions. Finally, we observed that the early raise in cHSPC count showed different kinetics among the diverse GT trials analyzed (**Figures 34**). This differential behavior could be partially justified by the diverse regimen conditionings, types of *ex vivo* HSPC culture and transduction protocols applied (Aiuti *et al*, 2013; Biffi *et al*, 2013; Gentner *et al*, 2021), as summarized in **Table 7**. For this reason we are currently collecting IS from resident and circulating HSPC and PB mature lineages of additional GT-patients overtime after treatment, in order to confirm and expand on an enlarged number of patients our preliminary findings on the *in vivo* relationship and differentiation output of BM and PB HSPC.

In conclusion, our results provided substantial novel findings over previous reports on human cHSPC properties. First, circulating HSPC show a progressive quantitative reduction during physiological aging and a diverse phenotypic composition than BM-resident counterpart, with a consistent enrichment of MPP, MLP, CMP and erythroid-committed progenitors. Second, cHSPC are endowed with multilineage differentiation potential both *in vitro* and *in vivo* after transplantation, displaying comparable BM homing capability but reduced long-term human cell engraftment as compared to transplanted BM-derived HSPC. Third, we identified a unique transcriptional profile of cHSPC, including the primitive subpopulation, characterized by lower replicative and cellular activity, but increased differentiation- and, only for more committed progenitors, adhesion- and immune response-priming than BM counterpart. Finally, our preliminary data on a cohort of HSPC-GT patients suggest that cHSPC may sustain clonal redistribution to distant BM sites both during active hematopoietic reconstitution and, at a lower extent, during steady-state conditions. Altogether, our findings indicate PB trafficking HSPC as a peculiar steady-state reservoir of low-cycling, pre-activated hematopoietic progenitors, which continuously recirculate among multiple BM sites and are poised for promptly sustaining activation and *in situ* local hematopoietic differentiation in case of demand.

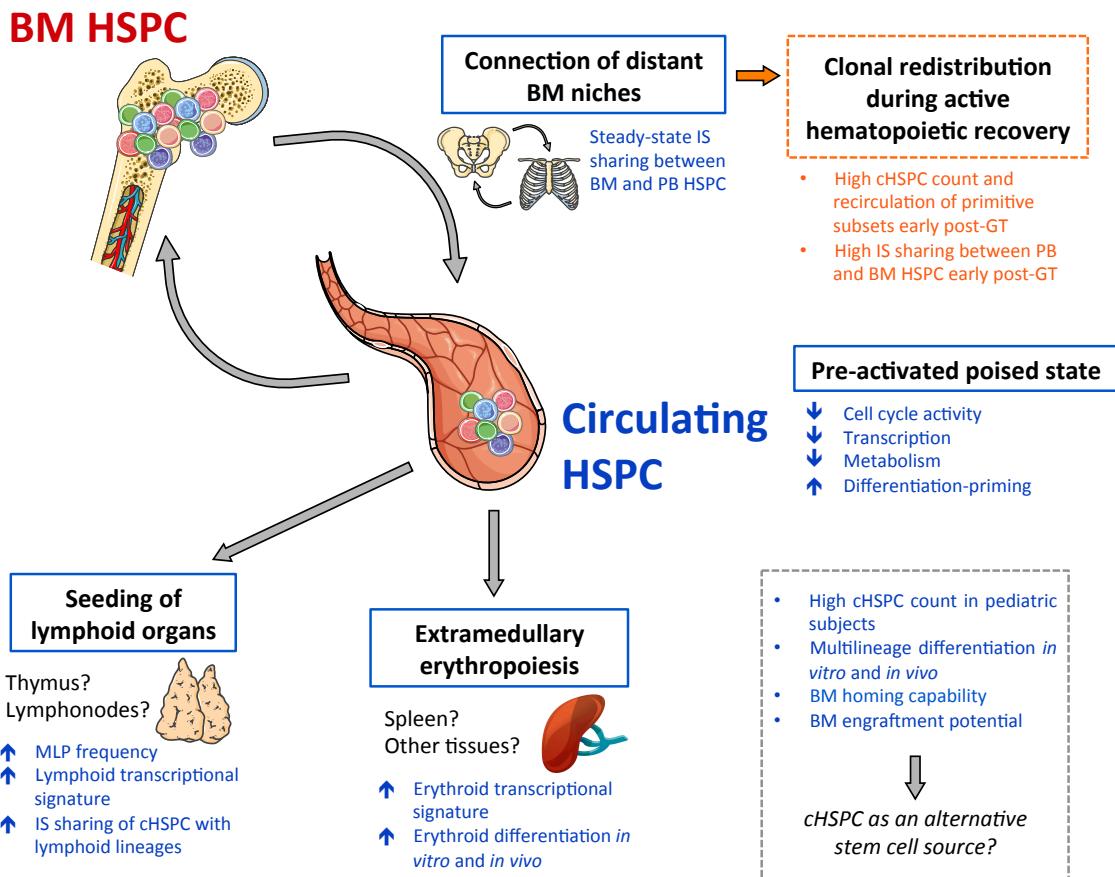


Figure 39. Proposed functions of human cHSPC at steady state and during hematopoietic reconstitution. Main findings from this study supporting the proposed cHSPC biological roles at steady state are reported in blue. Up and down blue arrows describe cHSPC properties with respect to BM counterpart. Main preliminary results supporting cHSPC functional role during hematopoietic reconstitution after HSPC-GT are reported in orange.

5 MATERIALS AND METHODS

5.1 Characteristics of healthy donors and gene therapy patients involved in the study

We have complied with all the ethical regulations for retrieving biological materials from healthy donors and patients. The sample size was determined by the number of individuals for whom excess material was available from procedures performed for clinical reasons, and after signing informed consent for research protocols approved by the San Raffaele Scientific Institute's Ethics Committee (TIGET06 and TIGET09). Informed consent for pediatric individuals was signed by their parents.

For cHSPC studies, we collected PB samples from 26 healthy subjects with 0-1 day of age, 7 subjects with 0-1 years of age, 13 subjects with 1-6 years of age (median age: 3.5 years), 15 subjects with 6-12 years of age (median age: 7.9 years), 11 subjects with 12-18 years of age (median age: 14.8 years), 27 subjects with 18-65 years of age (median age: 38.7 years) and 15 subjects with >70 years of age (median age: 75.1 years) (**Table 1**). The group of adult HD for BM analyses was composed of 9 pediatric subjects with 3-17 years of age (median age: 9.8 years), 20 adult individuals with 18-65 years of age (median age: 41.4 years) and 19 subjects with >70 years of age (median age: 75.6 years) (**Table 2**).

Gene therapy treated patients were enrolled in open-label, non-randomized, phase 1/2 clinical studies at the Pediatric Clinical Research Unit and Pediatric Immunohematology and Bone Marrow Transplantation Unit of the San Raffaele Scientific Institute (Milan, Italy) (Ferrua *et al*, 2019; Gentner *et al*, 2021; Fumagalli *et al*, in press). WAS patients were treated under early access program (compassionate use program or hospital exemption) or were enrolled in clinical trials registered with ClinicalTrials.gov (number NCT01515462 and NCT03837483) and EudraCT (number 2009-017346-32 and 2018-003842-18). MLD patients were enrolled in a clinical trial registered with ClinicalTrials.gov (number NCT01560182) and EudraCT (number 2009-017349-77). MPSIH patients were enrolled in a clinical trial registered with ClinicalTrials.gov (number NCT03488394) and EudraCT number (2017-002430-23).

To study cHSPC dynamics during hematopoietic reconstitution after gene therapy, we collected PB samples from 12 WAS-, 17 MLD- and 8 MPSIH-GT patients before and at different time points after treatment (7 days, 30 days, 60 days, 90 days, 180 days, 1 year and 2 years). Additionally, from the same patients we retrieved BM samples before GT and at 30 days-, 90 days- and 1 year-follow up. Detailed patients' characteristics, including the age at treatment, the conditioning regimen and the type of transduction protocol applied on autologous HSPC, are reported in **Table 7**. BM and PB samples were collected at Ospedale San Raffaele in Milan, Italy, in conjunction with diagnostic procedures or as specified by the GT protocol, with approval of the San Raffaele Scientific Institute's Ethics Committee and informed consent from patients' parents (TIGET06 and TIGET09).

The present work is a research-based study, which is not intended to report on the outcome of these clinical trials and early access programs.

Table 7. Cohort of GT-patients included in the study. *Bu (Busulfan); Flu (Fludarabin); PGE2: prostaglandin E2.*

	Number	Age at treatment (years)	Conditioning	Transduction protocol
WAS	12	3-10	Reduced intensity (Bu, Flu, Rituximab)	2 hits
MLD	17	1-4	Sub/Full Myeloablative (Bu)	2 hits
MPSIH	8	1-3	Sub/Full Myeloablative (Bu, Flu, Rituximab)	1 hit

5.2 Flow cytometry analyses by WBD on healthy donors' and patients' samples

Both BM and PB samples collected from healthy donors, GT-patients and transplanted mice were analyzed using our newly developed multi-parametric flow-cytometry assay (Whole Blood Dissection) (Basso-Ricci *et al*, 2017). In brief, after RBC lysis with ACK (STEMCELL Technologies #07850), samples were labeled with 100ul of fluorescent antibody mix (**Table 8**). Titration assays were performed to assess the best antibody concentration. Only for murine BM or PB samples, cells were incubated with a mouse FcR blocking reagent (BD #6148596, dilution 1:100) before staining with antibody cocktail. After surface marking, cells were incubated with propidium iodide (PI) (Biolegend #421301) to stain dead cells. Absolute cell quantification was performed by adding precision count beads (Biolegend #424902) to BM or PB samples before WBD procedure. Human hematopoietic subsets identified by WBD protocol are reported in **Table 9**.

All stained samples were acquired through BD Symphony A5 (BD Bioscience) cytofluorimeter after Rainbow beads (Spherotech #RCP-30-5A) calibration and raw data were collected through DIVA software (BD Biosciences). Data were subsequently analyzed with FlowJo software Version 10.5.3 (BD Biosciences) and the graphical output was automatically generated through Prism 9.0.0 (GraphPad software). Technically validated results were always included in the analyses, and we did not apply any exclusion criteria for outliers.

Table 8. List of fluorescent antibodies for WBD phenotyping.

Antibody	Source	Cat. Number
Mouse anti-human CD3 BV605	Biolegend	317322
Mouse anti-human CD56 PC5	Biolegend	362516
Mouse anti-human CD14 BV510	Biolegend	301842
Mouse anti-human CD33 BB515	BD Biosciences	564588
Mouse anti-human CD41/CD61 PC7	Biolegend	359812
Mouse anti-human CD66b BB515	BD Biosciences	564679
Mouse anti-human CD7 BB700	BD Biosciences	566488
Mouse anti-human CD45 BUV395	BD Biosciences	563792
Mouse anti-human CD38 BUV737	BD Biosciences	612824
Mouse anti-human CD90 APC	BD Biosciences	559869
Mouse anti-human CD184 (CXCR4) PE	Biolegend	306506
Mouse anti-human CD11c BV650	BD Biosciences	563404
Mouse anti-human CD10 BV786	BD Biosciences	564960
Mouse anti-human CD34 BV421	Biolegend	343610
Mouse anti-human CD45RA APCH7	Biolegend	304128
Mouse anti-human CD71 BV711	BD Biosciences	563767
Mouse anti-human CD19 APCR700	BD Biosciences	659121

Table 9. Hematopoietic populations identified by WBD analysis.

Hematopoietic output	Markers
iPMN	CD45+ CD33+ CD66b+ SShigh CD10- and/or CD11c-
PMN	CD45+ CD33+ CD66b+ SShigh CD10+ CD11c+
Monocyte	CD45+ CD33+ CD14+
DC	CD45+ CD33+ CD14- CD11c+
Myeloblast	CD45+ CD33+ CD14- CD11c- CD34-
T cells	CD45+ CD33- CD66b- CD3+ CD56-

NKt Cell	CD45+ CD33- CD66b- CD3+ CD56+
NK Cell	CD45+ CD33- CD66b- CD3- CD19- CD56+
B cell	CD45+ CD33- CD66b- CD3- CD19+ CD10- CD34-
Pre-B cell	CD45+ CD33- CD66b- CD3- CD19+ CD10+ CD34-
Pro-B cell	CD45+ CD33- CD66b- CD3- CD19+ CD10+ CD34+
Pro-lymphocyte	CD45+ CD33+ CD66b- CD3- CD19- CD56- CD34- CD71- CD41/61- CD7+ or CD10+
Pro-erythroblast	CD45+ CD33+ CD66b- CD3- CD19- CD56- CD34- CD71+
Erythroblast	CD45- CD71+
HSC	CD45+ CD14- CD11c- CD3- CD19- CD56- CD34+ CD38- CD90+ CD45RA-
MPP	CD45+ CD14- CD11c- CD3- CD19- CD56- CD34+ CD38- CD90- CD45RA-
MLP	CD45+ CD14- CD11c- CD3- CD19- CD56- CD34+ CD38- CD90- CD45RA+
ETP	CD45+ CD14- CD11c- CD3- CD19- CD56- CD34+ CD38+ CD7+
PreBNK	CD45+ CD14- CD11c- CD3- CD19- CD56- CD34+ CD38+ CD7- CD10+
GMP	CD45+ CD14- CD11c- CD3- CD19- CD56- CD34+ CD38+ CD7- CD10- CD45RA+
CMP	CD45+ CD14- CD11c- CD3- CD19- CD56- CD34+ CD38+ CD7- CD10- CD45RA- CD71- CD41/61-
MEP	CD45+ CD14- CD11c- CD3- CD19- CD56- CD34+ CD38+ CD7- CD10- CD45RA- CD71+ CD41/61+
EP	CD45+ CD14- CD11c- CD3- CD19- CD56- CD34+ CD38+ CD7- CD10- CD45RA- CD71+ CD41/61-
MKP	CD45+ CD14- CD11c- CD3- CD19- CD56- CD34+ CD38+ CD7- CD10- CD45RA- CD71- CD41/61+

5.3 Isolation of human mononuclear cells and CD34⁺ cells from BM and PB samples

PB and BM mononuclear cells (MC) from healthy donors were isolated using Lympholyte human cell separation medium (Euroclone #CL5026). CD34⁺ cells were purified from PBMC or BM MC using positive selection with immunomagnetic beads (average purity, 94.6%) according to the manufacturer's specifications (Miltenyi Biotec #130-046-702). Both mononuclear cells and purified PB and BM CD34⁺ cell fractions were frozen in liquid nitrogen until use.

5.4 Single-cell RNA sequencing and multiparametric single-cell protein barcoding

5.4.1 *Sample preparation*

Upon thawing, HD BM- or PB-derived purified CD34⁺ or PBMC samples were labeled with 100ul of a mix of fluorescent antibodies (**Table 10**) and TotalSeq-A antibodies (**Table 11**) for ADT barcoding. Briefly, TotalSeq-A antibodies are conjugated to short DNA oligonucleotides containing unique ADT, which can be captured and sequenced together with cellular mRNAs on a scRNAseq platform, thus allowing to detect the expression of selected surface markers on sequenced cells. Staining with TotalSeq-A antibodies and reagents was performed according to the manufacturer's specifications, and TotalSeq-A antibodies were used at concentrations defined by titration of corresponding fluorescent antibodies. After surface marking, cells were incubated with PI to label dead cells. We applied three different strategies for sample preparation prior to scRNAseq processing:

- 1) isolation by FACS sorting of both bulk HSPC (live LIN⁻ CD34⁺) and primitive HSPC (live LIN⁻ CD34⁺ CD38⁻) fractions starting from purified CD34⁺ cells;
- 2) isolation by FACS sorting of both primitive (live LIN⁻ CD34⁺ CD38⁻) and progenitor (live LIN⁻ CD34⁺ CD38⁺) HSPC fractions starting from purified CD34⁺ cells or PBMC;

3) isolation by FACS sorting of bulk HSPC (live LIN⁻ CD34⁺) starting from purified CD34⁺ cells.

Moreover, to compare the initial phenotypic composition of human HSPC at thawing with phenotypic data obtained by ADT barcoding, an aliquot of each sample was processed by WBD protocol and/or *ad hoc* designed FACS staining (“HSPC staining”, **Table 12**), aimed at identifying the same HSPC subsets detected by TotalSeq-A surface marking.

A detailed list of FACS stainings applied upon thawing, cellular fractions obtained after sorting, and TotalSeq-A antibody mix used for each sample is reported in **Table 13**.

Table 10. List of fluorescent antibodies used for cell sorting. * Antibody not used for BM_C and PB_I samples.

Antibody	Source	Cat. Number
Mouse anti-human Lineage cocktail (anti-CD3/CD14/ CD16/CD19/CD20/CD56) BV510	Biolegend	348807
Mouse anti-human CD15 BV510	Biolegend	323028
Mouse anti-human CD38 PC5*	Biolegend	303508
Mouse anti-human CD34 BV421	Biolegend	343610

Table 11. List of TotalSeq-A barcoded antibodies. * TotalSeq-A antibody used only for BM_C and PB_I samples.

TotalSeq-A Antibody	Source	Cat. Number
CD90	Biolegend	328135
CD45RA	Biolegend	304157
CD7	Biolegend	343123
CD71	Biolegend	334123

CD41	Biolegend	303737
CD10	Biolegend	312231
CD127	Biolegend	351352
CD38*	Biolegend	303541

Table 12. List of fluorescent antibodies used for HSPC staining of scRNAseq-processed samples upon thawing, before sorting and scRNAseq library preparation.

Antibody	Source	Cat. Number
Mouse anti-human CD34 BV421	Biolegend	343610
Mouse anti-human CD38 BUV737	BD Biosciences	612824
Mouse anti-human Lineage cocktail (anti-CD3/CD14/ CD16/CD19/CD20/CD56) BV510	Biolegend	348807
Mouse anti-human CD15 BV510	Biolegend	323028
Mouse anti-human CD45 BUV395	BD Biosciences	563792
Mouse anti-human CD90 APC	BD Biosciences	559869
Mouse anti-human CD7 BB700	BD Biosciences	566488
Mouse anti-human CD45RA APCH7	Biolegend	304128
Mouse anti-human CD71 BV711	BD Biosciences	563767
Mouse anti-human CD41a FITC	Biolegend	303704
Mouse anti-human CD10 BV786	BD Biosciences	564960
Mouse anti-human CD127 PC7	Biolegend	351304

Table 13. Detailed list of donor ID, cell source, type of FACS staining applied before any processing, cellular fractions obtained after sorting, and TotalSeq-A antibody mix used for each scRNAseq-processed sample. BM_D and PB_D (bold) were isolated from the same donor.

Donor ID	Source	FACS stainings applied at thawing	Sorted populations		TotalSeq-A antibody mix
			Populations	Cell number	
BM_A	BM CD34+	WBD	Lin- CD34+	20000	NO
			Lin- CD34+ CD38-	22600	NO
BM_B	BM CD34+	WBD	Lin- CD34+	50000	NO
			Lin- CD34+ CD38-	40200	NO
BM_C	BM CD34+	WBD HSPC staining	Lin- CD34+ CD38-	57300	YES
			Lin- CD34+ CD38+	314000	YES
BM_D	BM CD34+	WBD HSPC staining	Lin- CD34+ CD38-	35220	YES
			Lin- CD34+ CD38+	186000	YES
BM_E	BM CD34+	WBD HSPC staining	Lin- CD34+	236000	YES
PB_F	PB CD34+	WBD	Lin- CD34+	20000	NO
			Lin- CD34+ CD38-	16483	NO
PB_G	PB CD34+	WBD	Lin- CD34+	20000	NO
			Lin- CD34+ CD38-	2683	NO
PB_D	PBMC	WBD HSPC staining	Lin- CD34+ CD38-	8820	YES
			Lin- CD34+ CD38+	50523	YES
PB_D	PBMC	WBD HSPC staining	Lin- CD34+ CD38-	5100	YES
			Lin- CD34+ CD38+	23800	YES
PB_H	PB CD34+	WBD HSPC staining	Lin- CD34+ CD38-	11400	YES
			Lin- CD34+ CD38+	98400	YES
PB_I	PB CD34+	WBD HSPC staining	Lin- CD34+	200000	YES

5.4.2 Library preparation

Sorted BM samples were resuspended in known volume and counted with trypan blue for vitality when FACS cell recovery was >10,000 cells. Then cells were resuspended at the appropriate concentration for loading into the Chromium 10X Single Cell 3' Gene Expression v2 or v3 chemistry. Samples were processed according to the Chromium 10x 3' Gene Expression protocol v2 or v3 chemistry and TotalSeq™-A Antibodies and Cell Hashing with 10x Single Cell 3' Reagent Kit v3 Protocol. Details on cell recovery and sequencing depth for gene expression (GEX) and ADT libraries are reported in **Table 14** and **Table 15**, respectively.

Table 14. scRNAseq cell recovery and sequencing depth (median unique molecular identifier (UMI)/cell and median genes/cell) for Chromium 10x 3' GEX libraries. BM_D and PB_D (bold) were isolated from the same donor.

	Chromium 10X 3' GEX libraries				
Donor ID	Cell type	Chromium Chemistry	Sequenced cells	Median UMI/cell	Median genes/cell
BM_A	Lin- CD34+	v2	Failed quality control		
	Lin- CD34+ CD38-	v2	80	10472	1908
BM_B	Lin- CD34+ CD38-	v2	642	15589	3131
	Lin- CD34+ CD38+	v2	Failed quality control		
BM_C	Lin- CD34+ CD38-	v3	1605	14080	3200
	Lin- CD34+ CD38+	v3	5758	13373	3104
BM_D	Lin- CD34+ CD38-	v3	6562	11331	2797
	Lin- CD34+ CD38+	v3	6765	13133	3171
BM_E	Lin- CD34+	v3	6903	14621	3124
PB_F	Lin- CD34+	v2	422	10725	2095
	Lin- CD34+ CD38-	v2	331	6089	1541

PB_G	Lin- CD34+	v2	1116	11663	2303
	Lin- CD34+ CD38-	v2	165	7644	1588
PB_D	Lin- CD34+ CD38-	v3	157	11561	3374
	Lin- CD34+ CD38+	v3	7912	12916	3102
PB_D	Lin- CD34+ CD38-	v3	1423	11392	3480
	Lin- CD34+ CD38+	v3	6567	9727	2674
PB_H	Lin- CD34+ CD38-	v3	132	926	362
	Lin- CD34+ CD38+	v3	6812	12840	2822
PB_I	Lin- CD34+	v3	9424	11414	2566

Table 15. Median and interquartile values of UMI counts/cell for each ADT surface marker detected in Chromium 10x 3' ADT libraries. BM_D and PB_D (bold) were isolated from the same donor.

Chromium 10X 3' ADT libraries					
ADT surface marker	Donor ID	Cell type	UMI/cell		
			Median	Q1	Q3
CD90	BM_C	Lin- CD34+ CD38-	0	0	1
		Lin- CD34+ CD38+	0	0	0
	BM_D	Lin- CD34+ CD38-	7	5	10
		Lin- CD34+ CD38+	6	4	8
	BM_E	Lin- CD34+	5	2	11
	PB_D	Lin- CD34+ CD38-	2	0	4
		Lin- CD34+ CD38+	4	2	7
	PB_D	Lin- CD34+ CD38-	6	2	13
		Lin- CD34+ CD38+	2	1	6
	PB_H	Lin- CD34+ CD38-	2	1	6
		Lin- CD34+ CD38+	2	0	4

	PB_I	Lin- CD34+	2	0	4
CD45RA	BM_C	Lin- CD34+ CD38-	0	0	1
		Lin- CD34+ CD38+	2	1	9
	BM_D	Lin- CD34+ CD38-	1	0	1
		Lin- CD34+ CD38+	1	0	2
	BM_E	Lin- CD34+	10	5	28.5
	PB_D	Lin- CD34+ CD38-	2	0	7
		Lin- CD34+ CD38+	1	0	2
	PB_D	Lin- CD34+ CD38-	56	3	92
		Lin- CD34+ CD38+	1	0	3
	PB_H	Lin- CD34+ CD38-	8	2	1.25
Lin- CD34+ CD38+		2	1	3	
PB_I	Lin- CD34+	2	1	6	
CD7	BM_C	Lin- CD34+ CD38-	0	0	1
		Lin- CD34+ CD38+	2	0	4
	BM_D	Lin- CD34+ CD38-	1	0	2
		Lin- CD34+ CD38+	1	0	2
	BM_E	Lin- CD34+	1	0	4
	PB_D	Lin- CD34+ CD38-	1	0	4
		Lin- CD34+ CD38+	3	1	5
	PB_D	Lin- CD34+ CD38-	2	1	6
		Lin- CD34+ CD38+	3	1	8
	PB_H	Lin- CD34+ CD38-	3	1	8
Lin- CD34+ CD38+		2	0	4	
PB_I	Lin- CD34+	3	1	7	
CD71	BM_C	Lin- CD34+ CD38-	12	5	30
		Lin- CD34+ CD38+	106	52	276
	BM_D	Lin- CD34+ CD38-	10	3	28
		Lin- CD34+ CD38+	32	9	96
	BM_E	Lin- CD34+	281	111	741
	PB_D	Lin- CD34+ CD38-	2	0	5

		Lin- CD34+ CD38+	15	4	56
	PB_D	Lin- CD34+ CD38-	54	29	106
		Lin- CD34+ CD38+	239	110	798
	PB_H	Lin- CD34+ CD38-	26.5	18.75	35
		Lin- CD34+ CD38+	116	61	282
	PB_I	Lin- CD34+	87	40	180
CD41	BM_C	Lin- CD34+ CD38-	0	0	0
		Lin- CD34+ CD38+	4	2	12
	BM_D	Lin- CD34+ CD38-	0	0	1
		Lin- CD34+ CD38+	0	0	1
	BM_E	Lin- CD34+	5	2	16
	PB_D	Lin- CD34+ CD38-	0	0	0
		Lin- CD34+ CD38+	1	0	2
	PB_D	Lin- CD34+ CD38-	0	0	1
		Lin- CD34+ CD38+	1	0	4
	PB_H	Lin- CD34+ CD38-	3	1	5
		Lin- CD34+ CD38+	7	2	23
	PB_I	Lin- CD34+	16	4	47
CD10	BM_C	Lin- CD34+ CD38-	0	0	0
		Lin- CD34+ CD38+	0	0	1
	BM_D	Lin- CD34+ CD38-	1	0	2
		Lin- CD34+ CD38+	1	0	2
	BM_E	Lin- CD34+	0	0	1
	PB_D	Lin- CD34+ CD38-	0	0	2
		Lin- CD34+ CD38+	0	0	1
	PB_D	Lin- CD34+ CD38-	13	2	41
		Lin- CD34+ CD38+	0	0	1
	PB_H	Lin- CD34+ CD38-	13.5	3	32
		Lin- CD34+ CD38+	0	0	1
	PB_I	Lin- CD34+	2	1	3
CD127	BM_C	Lin- CD34+ CD38-	0	0	0

	Lin- CD34+ CD38+	0	0	0
BM_D	Lin- CD34+ CD38-	1	0	1
	Lin- CD34+ CD38+	0	0	1
BM_E	Lin- CD34+	0	0	0
PB_D	Lin- CD34+ CD38-	0	0	0
	Lin- CD34+ CD38+	0	0	1
PB_D	Lin- CD34+ CD38-	0	0	1
	Lin- CD34+ CD38+	0	0	1
PB_H	Lin- CD34+ CD38-	1	0	2
	Lin- CD34+ CD38+	1	0	1
PB_I	Lin- CD34+	0	0	0

5.4.3 Sequencing

Chromium 10x 3' GEX libraries were sequenced on Illumina NovaSeq S1 with the following paired-end, single indexing protocol: Read 1 = 28, Read 2 = 91 and Index 1 = 8 cycles. ADT libraries were sequenced together with GEX libraries or independently on an Illumina NextSeq sequencer with the following paired-end, single indexing protocol: Read 1 = 28, Read 2 = 25 and Index 1 = 8 cycles. Libraries were diluted and re-quantified by Qubit spectrophotometer, denatured with NaOH at final concentration of 0.2M. Loading concentrations were 2.4pM for NextSeq and 1nM for NovaSeq. Output was ~400M clusters for NextSeq and ~1.5G clusters for NovaSeq S1. Target sequencing depth was of 30,000 reads/cell for GEX libraries and 5,000 reads/cell for ADT libraries.

5.4.4 Sample demultiplexing, barcode processing and UMI count

We processed and demultiplexed raw data base call files (.bcl) to FASTQ files by the 10x Genomics pipeline Cell Ranger (v.4.0.0) by mkfastq function. We generated a genes-barcode matrix of UMI counts for each of the samples separately and aligned the reads to the GRCh38 human reference genome using STAR v2.5.1b. We used the

“cellranger count” function to generate a UMI count matrix for both GEX and ADT libraries.

5.4.5 *Gene expression data analysis*

We set up a standardized analysis using Seurat R package v.3.0.1 (Butler *et al*, 2018; Stuart *et al*, 2019). After importing UMI counts matrix, gene and cell barcode lists, we generated a minimal Seurat object including only cells with a minimum of 50 expressed genes.

Samples were merged together into a single Seurat object. Cell quality control by filtering of low-quality cells, which could negatively impact the analysis, was conducted by filtering cells with a percentage of mitochondrial genes to nuclear ones higher than 20 or cells with less than 200 or more than 8000 expressed genes.

After removing poor quality cells, counts were log-normalized by the default parameters in Seurat v.3.0.1. Successive analyses were performed on the most variable genes within the dataset (n = 7321), defined as the top 20% by mean expression and dispersion of the total genes present in the sample dataset.

In order to mitigate cell cycle status driven effects on gene expression, we first inferred the cell cycle phase of each cell by the expression levels of specific cell cycle marker genes (Regev list) with the “CellCycleScore” function in Seurat. We did not want to completely regress out the effect of cell cycle because cell cycle status may be interlinked with cardinal properties of hematopoietic stem cells. Thus, we applied a regression that would maintain the distinction between quiescent/non cycling (G1) cells and those that are actively cycling (G2M/S) (https://satijalab.org/seurat/v3.1/cell_cycle_vignette.html) (Nestorowa *et al*, 2016). Each single cell UMI counts were scaled and regressed for cell cycle difference, number of UMIs and percentage of mitochondrial genes.

The dataset was harmonized in order to reduce donor-driven variability. This was achieved by calling on the Seurat object the RunHarmony() from the Harmony R package (Korsunsky *et al*, 2019) with the following parameters: group.by.vars = c(“RNA_Donor”).

Dimensionality reduction was carried out by principal component analysis using only the top variable genes. The number of significant principal components (PCs) (n = 80)

was chosen after manual inspection of both elbow plot and jackstraw plot as suggested by the developers. Unsupervised clustering was performed on the most significant PCs selected by using FindNeighbors and FindClusters built-in functions with standard parameters. Clustering at a resolution of 0.6 was used for downstream analysis.

Dimensionality reduction for visualization purposes was performed by UMAP algorithm with a seed parameter of 123 for reproducibility purposes.

Differentially expressed genes across clusters were identified by the FindAllMarkers function running *wilcox* test with the following parameters:

logFC threshold = 0.25 – minimum logFC across groups

min.pct = 0.1 – genes evaluated must be expressed in at least 10 percent of cells within each group

pval.t = 1e-06 – p-value threshold for significant genes

min.cells.group = 10 – consider groups of at least 10 cells

5.4.6 Cell type annotation

In order to classify cells according to their cell type, we set up two different approaches, the first based on expression of specific lineage marker genes and published HSPC gene signatures (Velten *et al*, 2017; Fares *et al*, 2017; Zheng *et al*, 2018; Laurenti *et al*, 2013; Chen *et al*, 2014; Doulatov *et al*, 2013), the second based on SingleR (Aran *et al*, 2019), a computational method for unbiased cell type annotation in scRNAseq through comparisons with published reference datasets. We manually classified clusters by inspecting the expression of curated markers genes (FeaturePlot function in Seurat) and published gene signatures (AddModuleScore function in Seurat).

Annotation of single cells based on their transcriptional profile was then compared to that obtained by ADT surface marker staining. For each donor a custom gating strategy was defined by first converting sequencing data into “.fcs” files to be explored manually with FlowJo software Version 10.5.3 (BD Biosciences) through the *flowcore* R package. Gates were manually set based on comparison with conventional flow cytometry analysis performed on the same sample prior to single cell library preparation. ADT cell gating was then stored as meta.data within the Seurat object for

downstream analysis. For samples enriched for primitive CD38⁺ HSPC fraction, the proportions of ADT-detected HSPC subsets were normalized according to FACS immunophenotyping performed prior to single-cell library preparation.

5.4.7 *Downstream analysis*

We performed intra-cluster comparisons between PB vs. BM cells using the *FindMarkers* function, setting `test.use = wilcox`, a `logFC` threshold of 0 and `return.thresh` parameter equal to 1 in order to obtain unfiltered gene lists for all tested conditions. Only intra-cluster comparisons between groups with at least 5 cells were considered for differential gene expression analysis. Downstream analyses were carried out on the unfiltered output marker gene list, ranked by decreasing `logFC`, with positive values indicating higher expression in PB cells. We performed Gene Set Enrichment Analysis using the *GSEA* function of *ClusterProfiler* R package (Yu *et al*, 2012) (v3.16) focusing mainly on GO-BP genesets. For each cluster, all statistically significant (adjusted $p < 0.05$) GO-BP terms were divided into $NES > 0$ (NESpos) and $NES < 0$ (NESneg) subgroups. In order to reduce redundancy within GSEA GO-BP output lists, we ran a separate semantic reduction algorithm for NESpos and NESneg lists, through *REVIGO* web interface with the following parameters: J&C algorithm and “Tiny” filter (Supek *et al*, 2011), obtaining a reduced list of significant macro-categories. For each macro-category we defined a NES value based on the mean of NES values from all the grouped GO-BP single terms. We next generated a tile plot visualizing the NES value for each of the top 5 macro-categories for each cluster with the *ggplot2* R package, ordering the terms (rows) by biological functions.

For analysis focused on cluster 1, we generated a heatmap of selected significant (adjusted $p < 0.05$) marker genes showing an increased expression in PB ($\logFC \geq 0.4$) or BM ($\logFC \leq -0.4$) sources.

5.5 Colony forming units (CFU) assay

Cells deriving from healthy donors' and patients' PB or BM were cultured in MethoCultGF M3434 (STEMCELL Technologies #04434), a methylcellulose semisolid medium that allows the differentiation of human stem progenitors, giving rise to CFU. For CFU assay on whole peripheral blood cells, 250 ul of PB were lysed with 2.5 ml ACK at room temperature. Cell pellets were resuspended in 2 ml of MethoCult and then plated in two petri dishes. For CFU assay on patients' BM cells, 5,000 purified CD34⁺ cells were seeded in two petri dishes. After 14 days, CFU were counted and evaluated for their morphology. BM and PB patient-derived CFU at post-GT follow-ups were collected in bulk and pellets of cells were stored at -80°C.

5.6 *In vitro* multi-lineage differentiation assays

BM CD34⁺ cells isolated from adult healthy subjects were purchased from STEMCELL Technologies (#70002), while PB CD34⁺ cells were isolated from PB samples of adult HD collected at Ospedale San Raffaele in Milan, Italy, with approval of the San Raffaele Scientific Institute's Ethics Committee and informed consent (TIGET06 and TIGET09).

In vitro differentiation assay was applied on either BM- (n=8) and PB- (n=5) derived bulk CD34⁺ or BM- (n=4) and PB- (n=5) derived HSC, MPP, CMP, MEP and EP at single cell level. For bulk CD34⁺, the assay was performed in no-tissue culture treated 96-well flat bottom plate (Falcon #351172), while for single-cell differentiation assay not-treated 384 well plates (Thermo Scientific #265202) were used. 2 hours before cell seeding, plates were coated with StemSpan differentiation coating material (STEMCELL Technologies #09925) according to manufacturer specifications.

For bulk CD34⁺ cells differentiation assay, CD34⁺ cells were thawed in a medium containing PBS and 20% Fetal Bovine Serum (Euroclone #ECS0180L). Upon thawing, 500 cells were seeded into 100 ul of SFEM II medium (STEMCELL Technologies #09655) complemented with 1% Penicillin-Streptomycin (Euroclone #ECB3001D) and cytokines (**Table 16**). Changes with fresh medium were performed every 3-4 days.

To isolate single HSPC, CD34⁺ cells were labeled with the conjugated antibodies specified in **Table 17**, and stained samples were FACS purified with the BD FACSAria™ Fusion Cell Sorter (BD Biosciences), achieving purity ranging between 92% and 99%. 22 single cells for each subset were sorted directly in 384 well plates into 50 ul of SFEM II medium complemented with cytokines. For each plate, two wells with only medium and differentiation coating material were plated as negative controls. During sorting, single-cell index sorting was registered. 30 ul of fresh medium/well was added to each single-cell culture after 2 weeks.

After 3 weeks of culture, cells were harvested and labeled with anti-human conjugated antibodies listed in **Table 18**, in order to identify hematopoietic outputs reported in **Table 19**. All stained samples were acquired through BD FACS Symphony A5 (BD Biosciences) cytofluorimeter after Rainbow bead calibration. Raw FACS data were collected through DIVA software (BD Biosciences) and subsequently analyzed with FlowJo software Version 10.5.3, and the graphical output was generated through Prism 9.0.0 (GraphPad). The threshold for positive wells was determined based on negative controls. Technically validated results were always included in the analyses, and we did not apply any exclusion criteria for outliers.

Table 16. Cytokine cocktail added to SFEM II medium for in vitro multilineage differentiation assay.

Human recombinant cytokine	Source	Cat. Number	Concentration
Human stem cell factor (hSCF)	Peprotech	300-07	100 ng/ml
Human FLT3-L	Peprotech	300-19	10 ng/ml
Human thrombopoietin (hTPO)	Peprotech	300-18	75 ng/ml
Human Interleukin-3 (hIL-3)	Peprotech	200-03	10 ng/ml
Human Interleukin-7 (hIL-7)	Peprotech	200-07	100 ng/ml
Human Interleukin-2 (hIL-2)	Novartis	027131010	10 ng/ml
Human Interleukin-6 (hIL-6)	Peprotech	200-06	10 ng/ml
Human Interleukin-11 (hIL-11)	Peprotech	200-11	40 ng/ml
Human Erythropoietin (hEPO)	Peprotech	100-64	0.1 U/ml

Human Interleukin-4 (hIL-4)	Miltenyi Biotec	130-093-917	10 ng/ml
hLDL	Stem Cell technologies	02698	4 ug/ml

Table 17. List of fluorescent antibodies used for HSPC sorting for single-cell multilineage differentiation assay.

Antibody	Source	Cat. Number
Mouse anti-human Lineage cocktail (anti-CD3/CD14/ CD16/CD19/CD20/CD56) BV510	Biolegend	348807
Mouse anti-human CD34 BV421	Biolegend	343610
Mouse anti-human CD38 PC5	Biolegend	303508
Mouse anti-human CD90 APC	BD Biosciences	559869
Mouse anti-human CD45RA APCH7	Biolegend	304128
Mouse anti-human CD7 APCR700	BD Biosciences	659124
Mouse anti-human CD10 PC7	Biolegend	312214
Mouse anti-human CD71 PE	BD Biosciences	55537
Mouse anti-human CD41a FITC	Biolegend	303704

Table 18. List of fluorescent antibodies used for phenotypic characterization of cellular outputs at the end of multilineage differentiation assay.

Antibody	Source	Cat. Number
Mouse anti-human CD235a PE	BD Biosciences	561051
Mouse anti-human CD10 BV510	Biolegend	312219
Mouse anti-human CD3 BV605	Biolegend	317322
Mouse anti-human CD56 PC5	Biolegend	362516
Mouse anti-human CD33 BB515	BD Biosciences	564588
Mouse anti-human CD41 PC7	Biolegend	303718

Mouse anti-human CD15 APC fire750	Biolegend	323041
Mouse anti-human CD7 BB700	BD Biosciences	566488
Mouse anti-human CD45 BUV395	BD Biosciences	563792
Mouse anti-human CD1a APC	BD Biosciences	561755
Mouse anti-human CD5 BUV737	BD Biosciences	612842
Mouse anti-human CD34 BV421	Biolegend	343610
Mouse anti-human CD11c BV650	BD Biosciences	563404
Mouse anti-human CD42b BV786	BD Biosciences	740976
Mouse anti-human CD71 BV711	BD Biosciences	563767
Mouse anti-human CD19 APCR700	BD Biosciences	659121

Table 19. Hematopoietic outputs identified by FACS analysis at the end of multilineage differentiation culture. Each subset is identified through the combination of multiple phenotypic markers of early differentiation.

Hematopoietic output	Markers
CD19+ cell	CD45+ CD19+ CD3-
Precursor of T cell (Pre T)	Sum of: <ul style="list-style-type: none"> • CD45+ CD19- CD3+ • CD45+ CD19- CD3- CD15- CD235- CD71- CD42b- CD41- CD56- CD11c- CD7+ CD5- • CD45+ CD19- CD3- CD15- CD235- CD71- CD42b- CD41- CD56- CD11c- CD7+CD5+
CD10+ cell	CD45+ CD19- CD3- CD15- CD235- CD71- CD42b- CD41- CD56- CD11c- CD7- CD5- CD33- CD34- CD10+
CD56+ CD33- cell	CD45+ CD19- CD3- CD15- CD235- CD71- CD42b- CD41- CD56+ CD33-
CD56+ CD33+ cell	CD45+ CD19- CD3- CD15- CD235- CD71- CD42b- CD41- CD56+ CD33+
CD11c+ CD1a+ cell	CD45+ CD19- CD3- CD15- CD235- CD71- CD42b- CD41- CD56- CD11c+ CD1a+
CD11c+ CD1a- cell	CD45+ CD19- CD3- CD15- CD235- CD71- CD42b-

	CD41- CD56- CD11c+ CD1a-
CD15+ cell	CD45+ CD19- CD3- CD15+
CD33+ cell	CD45+ CD19- CD3- CD15- CD235- CD71- CD42b- CD41- CD56- CD11c- CD7- CD5- CD33+
Immature erythroid cell	CD45+ CD19- CD3- CD15- CD235- CD71+
Intermediate erythroid cell	Sum of: <ul style="list-style-type: none"> • CD45+ CD19- CD3- CD15- CD235+ CD71+ • CD45+ CD19- CD3- CD15- CD235+ CD71- • CD45- CD71+ CD235- • CD45- CD71+ CD235+
Mature erythroid cell	CD45- CD71- CD235+
MK CD45+ cell	Sum of: <ul style="list-style-type: none"> • CD45+ CD19- CD3- CD15- CD235- CD71- CD42b- CD41+ • CD45+ CD19- CD3- CD15- CD235- CD71- CD42b+ CD41+
MK CD45- cell	Sum of: <ul style="list-style-type: none"> • CD45- CD71+ CD235- CD42b- CD41+ • CD45- CD71+ CD235- CD42b+ CD41+
Unknown	CD45+ CD19- CD3- CD15- CD235- CD71- CD42b- CD41- CD56- CD11c- CD7- CD5- CD33- CD34- CD10-

5.7 *In vivo* models

Mouse studies were conducted according to protocols approved by the San Raffaele Scientific Institute and Institutional Animal Care and Use Committee (IACUC, #1183), adhering to the Italian Ministry of Health guidelines for the use and the care of experimental animals. All efforts were made to minimize the mice's number and the pain or distress during and after experimental procedures. NOD.Cg-Kit^{W-41J} Prkdc^{scid} Il2rgtm1Wjl/^{WaskJ} (NSGW41, stock #026497) mice were purchased from the Jackson Laboratory. Mice were maintained in specific pathogen-free conditions at San Raffaele Scientific Institute SPF Animal Facility.

5.8 *In vivo* transplantation assays

Human CD34⁺ cells from either BM or PB source were resuspended in 200 µl PBS and transplanted by tail vein injection into 7-8 week-old female NSGW41 mice.

For long-term transplantation experiments, NSGW41 mice were transplanted with 90,000 CD34⁺ cells from BM (n=3) or PB (n=3), with a total of n=17 and n=16 mice infused with BM and PB CD34⁺ cells, respectively. PB bleedings were performed at 7, 12 and 20 weeks after transplantation and mice were euthanized after 20 weeks.

For short-term transplantation experiments, NSGW41 mice were transplanted with 90,000 CD34⁺ cells from BM (n=3) or PB (n=4) sources, with a total of n=20 and n=24 mice infused with BM and PB CD34⁺ cells, respectively. Mice were divided into two groups (A and B), which were euthanized at 4 (group A) and 12 (group B) weeks after transplantation, while PB bleedings were performed at 4 (group A and B) and 12 (group B) weeks post-transplant.

For homing experiments, NSGW41 mice were transplanted with 150,000 CD34⁺ cells from BM (n=3) or PB (n=3) sources. A total of n=5 and n=5 mice were infused with BM and PB CD34⁺ cells, respectively. All animals were euthanized at 72 hours after transplantation to collect respective BM.

Human cell engraftment in PB was monitored by periodic tail vein blood sampling. At sacrifice, bone marrow cells were harvested by flushing of femurs and tibiae and filtered through a 40-µm cell strainer. For all experimental settings, human cell content in murine BM and PB was assessed by applying WBD protocol (see section 5.2).

5.9 IS retrieval and analysis

5.9.1 *Isolation of PB mature lineages, HSPC subpopulations and PB/BM CFC from GT-patients*

IS retrieval and analysis were performed on a selected cohort of 3 WAS GT-patients at early (≤ 90 days) and late (1 year) phases after GT (**Table 20**). Patients' PB mature lineages and BM HSPC subsets were purified as previously reported (Scala *et al*, 2018).

In brief, PB mononuclear cells were isolated through Ficoll-Hypaque gradient separation (Lymphoprep, Fresenius). Then, whole CD3⁺, CD4⁺, CD8⁺, CD14⁺, CD15⁺, CD19⁺ and CD56⁺ cells were purified using positive selection with immunomagnetic beads (average purity, 94.6%) according to the manufacturer's specifications (Miltenyi Biotec). To isolate patients' BM HSPC subpopulations, purified CD34⁺ cells were labeled with the following conjugated antibodies: anti-human LIN cocktail (anti-CD3/CD14/CD16/CD19/CD20/CD56), anti-CD15, anti-CD34, anti-CD38, anti-CD45RA, anti-CD10, anti-CD135 (Biolegend), anti-CD90 and anti-CD7 (BD Biosciences). The stained samples were FACS purified with the MoFlo XDP cell sorter (Beckman Coulter), achieving purity ranging between 92% and 99%. At each follow-up, BM aspirates were performed from two distinct anatomical sites (left and right iliac crests) to collect sufficient material for clinical tests.

Moreover, PB and BM CFC were plated starting respectively from patients' whole PB and purified BM CD34⁺ cells (derived from left and right sites), as previously described (section 5.5).

5.9.2 DNA extraction and IS retrieval

Genomic DNA was extracted from patients' PB CFC, BM CFC, PB mature lineages, and BM HSPC subpopulations using the QIAamp DNA Blood Mini (QIAGEN, #51304) or Micro Kit (QIAGEN, #56304), according to manufacturer's instructions. Whole-genome amplification was performed with the Repli-g Mini Kit (QIAGEN, #150025) on DNA from FACS-sorted HSPC subpopulations as previously described (Biasco *et al*, 2015; Scala *et al*, 2018). IS were retrieved through LAM-PCR (Schmidt *et al*, 2007) (patient 1's samples), or SLIM-PCR (Benedicenti F. *et al* manuscript in preparation) (patient 2 and 3's samples), a novel technology that combines fragmentation of genomic DNA by sonication and tagging of the IS-containing genomic fragments with random barcodes prior to PCR amplification. Briefly, extracted DNA undergoes mechanical fragmentation through sonication followed by a reaction of ligation with a barcoded linker cassette. Through this approach, each sample is univocally labeled with specific identifier. Barcoded linker cassettes also include a small region with random nucleotides that allow tagging differently each single

fragment of a sample. After ligation, the samples are purified and a round of exponential amplification coupled with annealing with Illumina fusion primers is performed. High-throughput sequencing with Illumina platform (Hiseq) and validated bioinformatic pipeline (Spinozzi *et al*, 2017) allow the identification and quantification of IS present in each sample.

5.9.3 IS mapping and filtering

All retrieved IS were mapped with the software VISPA2 (Spinozzi *et al*, 2017) and univocally assigned to a single patient to minimize any potential contamination. We extended our previous approach, described in (Aiuti *et al*, 2013; Biffi *et al*, 2013), introducing the attribution of each single IS to a patient based on the date of sample processing: whenever an IS is shared between two patients, we firstly consider whether the same IS was previously assigned to one of the two subjects, then we applied the 10-fold rule presented in (Aiuti *et al*, 2013; Biffi *et al*, 2013), that assigns the IS to the patient with higher sequence count. Finally, we removed all the IS with a sum of sequencing count below 3 in the entire dataset.

All IS derived from the 3 distinct WAS GT-patients were pooled together (**Table 20**). We then analyzed the level of IS sharing between the following groups:

- PB CFC early and total BM HSPC early;
- PB CFC late and total BM HSPC late;
- BM left HSPC early and BM right HSPC early;
- BM left HSPC late and BM right HSPC late;
- PB CFC late and PB mature myeloid lineages (Granulocytes+Monocytes) late;
- PB CFC late and PB mature lymphoid lineages (B, T and NK cells) late;
- BM CFC late and PB mature myeloid lineages (Granulocytes+Monocytes) late;
- BM CFC late and PB mature lymphoid lineages (B, T and NK cells) late.

Graphical output of percentages of IS sharing was automatically generated through Prism 9.0.0 (GraphPad software).

Table 20. Number of IS retrieved from cHSPC, BM HSPC and PB mature lineages from GT-patients. Table reporting the number of IS retrieved from PB CFC, BM CFC (only late phases), BM HSPC left, BM HSPC right, total BM HSPC (left+right), PB myeloid mature lineages (Granulocytes+Monocytes), and PB lymphoid mature lineages (B, T, NK cells) from 3 distinct WAS HSPC-GT patients at early (≤ 90 days) and late (1 year) phases after GT. The total IS number for each cell type analyzed is reported. For patient WAS Pt1, we were able to retrieve late HSPC IS from total BM but not from the distinct left and right sites. For patient WAS Pt3, we were not able to retrieve IS from late BM CFC.

Cell type	Time point	Patient	IS number	Total IS number
PB CFC	EARLY	WAS Pt1	847	1287
		WAS Pt2	313	
		WAS Pt3	127	
	LATE	WAS Pt1	853	1541
		WAS Pt2	475	
		WAS Pt3	213	
BM CFC	LATE	WAS Pt1	731	1291
		WAS Pt2	560	
		WAS Pt3		
BM HSPC LEFT	EARLY	WAS Pt1	1576	3366
		WAS Pt2	1610	
		WAS Pt3	180	
	LATE	WAS Pt1		945
		WAS Pt2	481	
		WAS Pt3	464	
BM HSPC RIGHT	EARLY	WAS Pt1	2340	3583
		WAS Pt2	1176	
		WAS Pt3	67	

	LATE	WAS Pt1		1301
		WAS Pt2	779	
		WAS Pt3	522	
TOTAL BM HSPC	EARLY	WAS Pt1	3897	6813
		WAS Pt2	2688	
		WAS Pt3	228	
	LATE	WAS Pt1	2288	4451
		WAS Pt2	1182	
		WAS Pt3	981	
PB MYELOID	EARLY	WAS Pt1	17803	42150
		WAS Pt2	13923	
		WAS Pt3	10424	
	LATE	WAS Pt1	8205	12414
		WAS Pt2	3544	
		WAS Pt3	665	
PB LYMPHOID	EARLY	WAS Pt1	8103	28029
		WAS Pt2	15654	
		WAS Pt3	4272	
	LATE	WAS Pt1	6381	16201
		WAS Pt2	3318	
		WAS Pt3	6502	

5.10 Statistical tests

Statistical tests were performed using Prism v9.1.0 software (GraphPad). Analytical tests for statistical significance between two groups used the Mann–Whitney test (p values are specified in each figure legend or within figure graphs). If more than two groups were compared, analytical tests for statistical significance among groups used the Dunn’s multiple comparisons test (p values are specified in each figure legend or within figure graphs). Before performing Dunn’s tests, we always analyzed intragroup variance and validated statistical significance of mean and median differences among groups through the nonparametric Kruskal–Wallis statistic.

Correlations between variables were assessed through Spearman r test. All the exact p values and r values are specified within each figure graph.

REFERENCES

- Abbas AK & Lichtman AH (2013) *Le basi dell'immunologia: fisiopatologia del sistema immunitario* Elsevier-Masson
- Abdellatif H (2018) Circulating CD34 + hematopoietic stem / progenitor cells paralleled with level of viremia in patients chronically infected with hepatitis B virus. *Regen Med Res* 6: 1–8
- Aiuti A, Biasco L, Scaramuzza S, Ferrua F, Cicalese MP, Baricordi C, Dionisio F, Calabria A, Giannelli S, Castiello MC, *et al* (2013) Lentiviral Hematopoietic Stem Cell Gene Therapy in Patients with Wiskott-Aldrich Syndrome. *Science (80-)* 341: 1233151–1233151
- Aiuti A, Roncarolo MG & Naldini L (2017) Gene therapy for ADA-SCID, the first marketing approval of an ex vivo gene therapy in Europe: paving the road for the next generation of advanced therapy medicinal products. *EMBO Mol Med* 9: 737–740
- Aiuti A, Slavin S, Aker M, Ficara F, Deola S, Mortellaro A, Morecki S, Andolfi G, Tabucchi A, Carlucci F, *et al* (2002) Correction of ADA-SCID by stem cell gene therapy combined with nonmyeloablative conditioning. *Science (80-)* 296: 2410–2413
- Aiuti BA, Webb IJ, Bleul C & Springer T (1997) New Mechanism to Explain the Mobilization of CD34+ Progenitors to Peripheral Blood. *J Exp Med* 185: 111–120
- Albakri M, Tashkandi H & Zhou L (2020) A Review of Advances in Hematopoietic Stem Cell Mobilization and the Potential Role of Notch2 Blockade. *Cell Transplant* 29: 1–13
- Alexander SI, Smith N, Hu M, Verran D, Shun A, Dorney S, Smith A, Webster B, Shaw PJ, Lammi A, *et al* (2008) Chimerism and Tolerance in a Recipient of a Deceased-Donor Liver Transplant. *N Engl J Med* 358: 369–374
- Aran D, Looney AP, Liu L, Wu E, Fong V, Hsu A, Chak S, Naikawadi RP, Wolters PJ, Abate AR, *et al* (2019) Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat Immunol* 20: 163–172
- Avigdor A, Goichberg P, Shvitiel S, Dar A, Peled A, Samira S, Kollet O, Hershkoviz R,

- Alon R, Hardan I, *et al* (2004) CD44 and hyaluronic acid cooperate with SDF-1 in the trafficking of human CD34+ stem / progenitor cells to bone marrow. *Blood* 103: 2981–2989
- Basso-Ricci L, Scala S, Milani R, Migliavacca M, Rovelli A, Bernardo ME, Ciceri F, Aiuti A & Biasco L (2017) Multiparametric Whole Blood Dissection: A one-shot comprehensive picture of the human hematopoietic system. *Cytom Part A* 91: 952–965
- Becht E, McInnes L, Healy J, Dutertre CA, Kwok IWH, Ng LG, Ginhoux F & Newell EW (2019) Dimensionality reduction for visualizing single-cell data using UMAP. *Nat Biotechnol* 37: 38–47
- Bernardo ME, Gentner B, Tucci F, Fumagalli F, Ciotti F, Sarzana M, Pontesilli S, Baldoli C, Darin S, De Bellis D, *et al* (2020) First-in-human phase I/II clinical trial of hematopoietic stem cell-gene therapy for Mucopolysaccharidosis Type I, Hurler (MPS-IH): preliminary evidence of extensive metabolic correction. *EHA25 virtual*; 295101; S281.
- Biasco L, Pellin D, Scala S, Dionisio F, Basso-Ricci L, Leonardelli L, Scaramuzza S, Baricordi C, Ferrua F, Cicalese MP, *et al* (2016) In Vivo Tracking of Human Hematopoiesis Reveals Patterns of Clonal Dynamics during Early and Steady-State Reconstitution Phases. *Cell Stem Cell* 19: 107–119
- Biasco L, Scala S, Basso Ricci L, Dionisio F, Baricordi C, Calabria A, Giannelli S, Cieri N, Barzaghi F, Pajno R, *et al* (2015) In vivo tracking of T cells in humans unveils decade-long survival and activity of genetically modified T memory stem cells. *Sci Transl Med* 7: 273ra13
- Biffi A, Montini E, Lorioli L, Cesani M, Fumagalli F, Plati T, Baldoli C, Martino S, Calabria A, Canale S, *et al* (2013) Lentiviral hematopoietic stem cell gene therapy benefits metachromatic leukodystrophy. *Science* (80-) 341
- Birbrair A & Frenette PS (2016) Niche heterogeneity in the bone marrow. *Ann N Y Acad Sci* 1370: 82–96
- Bourdieu A, Avalon M, Lapostolle V, Ismail S, Mombled M, Debeissat C, Guérinet M, Duchez P, Chevaleyre J, Vlaski-Lafarge M, *et al* (2018) Steady state peripheral blood provides cells with functional and metabolic characteristics of real hematopoietic stem cells. *J Cell Physiol* 233: 338–349

- Bozdag SC, Tekgunduz E & Altuntas F (2015) The Current Status in Hematopoietic Stem Cell Mobilization. *J Clin Apher* 30: 273–280
- Bradley TR & Metcalf D (1966) The growth of mouse bone marrow cells in vitro. *Aust J Exp Biol Med Sci* 44: 287–299
- Braun CJ, Boztug K, Paruzynski A, Witzel M, Schwarzer A, Rothe M, Modlich U, Beier R, Göhring G, Steinemann D, *et al* (2014) Gene therapy for Wiskott-Aldrich syndrome - Long-Term efficacy and genotoxicity. *Sci Transl Med* 6: 1–15
- Brehm MA, Shultz LD, Luban J & Greiner DL (2013) Overcoming current limitations in humanized mouse research. *J Infect Dis* 208 Suppl: 125–130
- Brunet de la Grange P, Vlaski M, Duchez P, Chevaleyre J, Lapostolle V, Boiron JM, Praloran V & Ivanovic Z (2013) Long-term repopulating hematopoietic stem cells and ‘side population’ in human steady state peripheral blood. *Stem Cell Res* 11: 625–633
- Buenrostro JD, Corces MR, Lareau CA, Wu B, Schep AN, Aryee MJ, Majeti R, Chang HY & Greenleaf WJ (2018) Integrated Single-Cell Analysis Maps the Continuous Regulatory Landscape of Human Hematopoietic Differentiation. *Cell* 173: 1535–1548
- Bueren JA, Quintana-Bustamante O, Almarza E, Navarro S, Río P, Segovia JC & Guenechea G (2020) Advances in the gene therapy of monogenic blood cell diseases. *Clin Genet* 97: 89–102
- Buffone A, Anderson NR, Hammer DA, Buffone A, Anderson NR & Hammer DA (2018) Migration against the direction of flow is LFA-1- dependent in human hematopoietic stem and progenitor cells. *J Cell Sci* 205575
- Burberry A, Zeng MY, Ding L, Wicks I, Inohara N, Morrison SJ & Núñez G (2014) Infection mobilizes hematopoietic stem cells through cooperative NOD-like receptor and toll-like receptor signaling. *Cell Host Microbe* 15: 779–791
- Butler A, Hoffman P, Smibert P, Papalexi E & Satija R (2018) Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol* 36: 411–420
- Calvi LM, Adams GB, Weibrecht KW, Weber JM, Olson DP, Knight MC, Martin RP, Schipani E, Divieti P, Bringhurst FR, *et al* (2003) Osteoblastic cells regulate the haematopoietic stem cell niche. *Nature* 425: 841–846

- Cardier JE & Barberá-Guillem E (1997) Extramedullary hematopoiesis in the adult mouse liver is associated with specific hepatic sinusoidal endothelial cells. *Hepatology* 26: 165–175
- Cartier N, Haccin-Bey-Abina S, Bartholomae CC, Veres G, Schmidt M, Kutschera I, Vidaud M, Abel U, Dal-Cortivo L, Caccavelli L, *et al* (2009) Hematopoietic stem cell gene therapy with a lentiviral vector in X-linked adrenoleukodystrophy. *Science* (80-) 326: 818–823
- Ceradini DJ, Kulkarni AR, Callaghan MJ, Tepper OM, Bastidas N, Kleinman ME, Capla JM, Galiano RD, Levine JP & Gurtner GC (2004) Progenitor cell trafficking is regulated by hypoxic gradients through HIF-1 induction of SDF-1. *Nat Med* 10: 858–864
- Chen L, Kostadima M, Martens JHA, Canu G, Garcia SP, Turro E, Downes K, Macaulay IC, Bielczyk-Maczynska E, Coe S, *et al* (2014) Transcriptional diversity during lineage commitment of human blood progenitors. *Science* 345: 1251033
- Cicalese MP, Ferrua F, Castagnaro L, Pajno R, Barzaghi F, Giannelli S, Dionisio F, Brigida I, Bonopane M, Casiraghi M, *et al* (2016) Update on the safety and efficacy of retroviral gene therapy for immunodeficiency due to adenosine deaminase deficiency. *Blood* 128: 45–54
- Cohen KS, Cheng S, Larson MG, Cupples LA, McCabe EL, Wang YA, Ngwa JS, Martin RP, Klein RJ, Hashmi B, *et al* (2013) Circulating CD34+ progenitor cell frequency is associated with clinical and genetic factors. *Blood* 121: 50–57
- Cosgun KN, Rahmig S, Mende N, Reinke S, Hauber I, Schäfer C, Petzold A, Weisbach H, Heidkamp G, Purbojo A, *et al* (2014) Kit regulates HSC engraftment across the human-mouse species barrier. *Cell Stem Cell* 15: 227–238
- Crippa S & Bernardo ME (2018) Mesenchymal Stromal Cells: Role in the BM Niche and in the Support of Hematopoietic Stem Cell Transplantation. *HemaSphere* 2
- Day M & Schultz R (2010) An Overview of the Immune System. In *Veterinary Immunology* pp 9–18.
- Demerdash Y, Kain B, Essers MAG & King KY (2021) Yin and Yang: The dual effects of interferons on hematopoiesis. *Exp Hematol* 96: 1–12
- Deutsch VR & Tomer A (2006) Megakaryocyte development and platelet production. *Br J Haematol* 134: 453–466

- DiPersio JF, Karpova D & Rettig MP (2019) Mobilized peripheral blood: An updated perspective. *F1000Research* 8: 1–14
- Domingues MJ, Nilsson SK & Cao B (2017) New agents in HSC mobilization. *Int J Hematol* 105: 141–152
- Doulatov S, Notta F, Eppert K, Nguyen LT, Ohashi PS & Dick JE (2010) Revised map of the human progenitor hierarchy shows the origin of macrophages and dendritic cells in early lymphoid development. *Nat Immunol* 11: 585–593
- Doulatov S, Notta F, Laurenti E & Dick JE (2012) Hematopoiesis: A human perspective. *Cell Stem Cell* 10: 120–136
- Doulatov S, Vo LT, Chou SS, Kim PG, Arora N, Li H, Hadland BK, Bernstein ID, Collins JJ, Zon LI, *et al* (2013) Induction of multipotential hematopoietic progenitors from human pluripotent stem cells via respecification of lineage-restricted precursors. *Cell Stem Cell* 13: 459–470
- Dzierzak E & Philipsen S (2013) Erythropoiesis: Development and differentiation. *Cold Spring Harb Perspect Med* 3: 1–16
- Eichler F, Duncan C, Musolino PL, Orchard PJ, De Oliveira S, Thrasher AJ, Armant M, Dansereau C, Lund TC, Miller WP, *et al* (2017) Hematopoietic Stem-Cell Gene Therapy for Cerebral Adrenoleukodystrophy. *N Engl J Med* 377: 1630–1638
- Essers MAG, Offner S, Blanco-Bose WE, Waibler Z, Kalinke U, Duchosal MA & Trumpp A (2009) IFN α activates dormant haematopoietic stem cells in vivo. *Nature* 458: 904–908
- Fares I, Chagraoui J, Lehnertz B, MacRae T, Mayotte N, Tomellini E, Aubert L, Roux PP & Sauvageau G (2017) EPCR expression marks UM171-expanded CD34+ cord blood stem cells. *Blood* 129: 3344–3351
- Ferrari G, Thrasher AJ & Aiuti A (2021) Gene therapy using haematopoietic stem and progenitor cells. *Nat Rev Genet* 22: 216–234
- Ferrua F, Cicalese MP, Galimberti S, Giannelli S, Dionisio F, Barzaghi F, Migliavacca M, Bernardo ME, Calbi V, Assanelli AA, *et al* (2019) Lentiviral haematopoietic stem/progenitor cell gene therapy for treatment of Wiskott-Aldrich syndrome: interim results of a non-randomised, open-label, phase 1/2 clinical study. *Lancet Haematol* 6: e239–e253
- Fliedner TM (1998) The role of blood stem cells in hematopoietic cell renewal. *Stem*

Cells 16: 361–374

- Fu J, Zuber J, Martinez M, Shonts B, Obradovic A, Wang H, Lau S ping, Xia A, Waffarn EE, Frangaj K, *et al* (2019) Human Intestinal Allografts Contain Functional Hematopoietic Stem and Progenitor Cells that Are Maintained by a Circulating Pool. *Cell Stem Cell* 24: 227-239.e8
- Fukuda S, Bian H, King AG & Pelus LM (2007) The chemokine GRO β mobilizes early hematopoietic stem cells characterized by enhanced homing and engraftment. *Blood* 110: 860–869
- Fumagalli, Calbi, *et al* (in press), *Lancet*
- Fumagalli F, Zambon AA, Rancoita PMV, Baldoli C, Canale S, Spiga I, Medaglini S, Penati R, Facchini M, Ciotti F, *et al* (2021) Metachromatic leukodystrophy: A single-center longitudinal study of 45 patients. *J Inherit Metab Dis*: 1–14
- Gaspar HB, Parsley KL, Howe S, King D, Gilmour KC, Sinclair J, Brouns G, Schmidt M, Von Kalle C, Barington T, *et al* (2004) Gene therapy of X-linked severe combined immunodeficiency by use of a pseudotyped gammaretroviral vector. *Lancet* 364: 2181–2187
- Gentner B, Tucci F, Galimberti S, Fumagalli F, De Pellegrin M, Silvani P, Camesasca C, Pontesilli S, Darin S, Ciotti F, *et al* (2021) Hematopoietic Stem- and Progenitor-Cell Gene Therapy for Hurler Syndrome. *N Engl J Med* 385: 1929–1940
- Goodman JW & Hodgson GS (1962) Evidence for stem cells in the peripheral blood of mice. *Blood* 19: 702–714
- Gu Y, Kortessmaa J, Tryggvason K, Persson J, Ekblom P, Jacobsen S & Ekblom M (2003) Laminin isoform – specific promotion of adhesion and migration of human bone marrow progenitor cells. *Blood* 101: 877–885
- Gui J, Mustachio LM, Su DM & Craig RW (2012) Thymus size and age-related thymic involution: Early programming, sexual dimorphism, progenitors and stroma. *Aging Dis* 3: 280–290
- Gulbas Z (2018) Haploidentical stem cell transplantation-bone marrow vs peripheral blood. *Transfus Apher Sci* 57: 168–170
- Hacein-Bey-Abina S, Pai S-Y, Gaspar HB, Armant M, Berry CC, Blanche S, Bleesing J, Blondeau J, de Boer H, Buckland KF, *et al* (2014) A Modified γ -Retrovirus Vector for X-Linked Severe Combined Immunodeficiency. *N Engl J Med* 371:

1407–1417

- Harandi OF, Mckeone D & Paulson RF (2010) Murine erythroid short-term radioprotection requires a BMP4- dependent, self-renewing population of stress erythroid progenitors. *Blood* 120: 4507–4519
- Hübel K (2019) Mobilization and collection of HSC. In *The EBMT Handbook: Hematopoietic Stem Cell Transplantation and Cellular Therapies* pp 117–122.
- Itkin T, Gur-Cohen S, Spencer JA, Schajnovitz A, Ramasamy SK, Kusumbe AP, Ledergor G, Jung Y, Milo I, Poulos MG, *et al* (2016) Distinct bone marrow blood vessels differentially regulate haematopoiesis. *Nature* 532: 323–328
- Ito M, Hiramatsu H, Kobayashi K, Suzue K, Kawahata M, Hioki K, Ueyama Y, Koyanagi Y, Sugamura K, Tsuji K, *et al* (2002) NOD/SCID/ γ c null mouse: an excellent recipient mouse model for engraftment of human cells. *Bone* 100: 3175–3182
- Karamitros D, Stoilova B, Aboukhalil Z, Hamey F, Reinisch A, Samitsch M, Quek L, Otto G, Repapi E, Doondea J, *et al* (2018) Single-cell analysis reveals the continuum of human lympho-myeloid progenitor cells article. *Nat Immunol* 19: 85–97
- Karpova D, Link DC, Dipersio JF, Karpova D, Rettig MP, Ritchey J, Cancilla D, Christ S, Gehrs L, Chendamarai E, *et al* (2019) Targeting VLA4 integrin and CXCR2 mobilizes serially repopulating hematopoietic stem cells. *J Clin Invest* 129: 2745–2759
- King KY, Matatall KA, Shen CC, Goodell MA, Swierczek SI & Prchal JT (2015) Comparative long-term effects of interferon α and hydroxyurea on human hematopoietic progenitor cells. *Exp Hematol* 43: 912-918.e2
- Korsunsky I, Millard N, Fan J, Slowikowski K, Zhang F, Wei K, Baglaenko Y, Brenner M, Loh P-R & Raychaudhuri S (2019) Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat Methods* 16: 1289–1296
- de Kruijf EJFM, Fibbe WE & van Pel M (2019) Cytokine-induced hematopoietic stem and progenitor cell mobilization: unraveling interactions between stem cells and their niche. *Ann N Y Acad Sci* 1466: 24–38
- Kunisaki Y, Bruns I, Scheiermann C, Ahmed J, Pinho S, Zhang D, Mizoguchi T, Wei Q, Lucas D, Ito K, *et al* (2013) Arteriolar niches maintain haematopoietic stem cell

- quiescence. *Nature* 502: 637–643
- Laurenti E, Doulatov S, Zandi S, Plumb I, Chen J, April C, Fan JB & Dick JE (2013) The transcriptional architecture of early human hematopoiesis identifies multilevel control of lymphoid commitment. *Nat Immunol* 14: 756–763
- Laurenti E, Frelin C, Xie S, Ferrari R, Dunant CF, Zandi S, Neumann A, Plumb I, Doulatov S, Chen J, *et al* (2015) CDK6 levels regulate quiescence exit in human hematopoietic stem cells. *Cell Stem Cell* 16: 302–313
- Laurenti E & Göttgens B (2018) From haematopoietic stem cells to complex differentiation landscapes. *Nature* 553: 418–426
- Lefrançais E, Ortiz-Muñoz G, Caudrillier A, Mallavia B, Liu F, Sayah DM, Thornton EE, Headley MB, David T, Coughlin SR, *et al* (2017) The lung is a site of platelet biogenesis and a reservoir for haematopoietic progenitors. *Nature* 544: 105–109
- Markt S, Scaramuzza S, Cicalese MP, Giglio F, Galimberti S, Lidonnici MR, Calbi V, Assanelli A, Bernardo ME, Rossi C, *et al* (2019) Intrabone hematopoietic stem cell gene therapy for adult and pediatric patients affected by transfusion-dependent β -thalassemia. *Nat Med* 25: 234–241
- Massberg S, Schaerli P, Knezevic-Maramica I, Köllnberger M, Tubo N, Moseman EA, Huff I V., Junt T, Wagers AJ, Mazo IB, *et al* (2007) Immunosurveillance by Hematopoietic Progenitor Cells Trafficking through Blood, Lymph, and Peripheral Tissues. *Cell* 131: 994–1008
- Mazo IB, Massberg S & von Andrian UH (2011) Hematopoietic stem and progenitor cell trafficking. *Trends Immunol* 32: 493–503
- McKinney-Freeman SL, Jackson KA, Camargo FD, Ferrari G, Mavilio F & Goodell MA (2002) Muscle-derived hematopoietic stem cells are hematopoietic in origin. *Proc Natl Acad Sci U S A* 99: 1341–1346
- Mende N, Bastos HP, Santoro A, Sham K, Mahbubani KT, Curd A, Takizawa H, Wilson NK, Göttgens B, Saeb-Parsy K, *et al* (2020) Quantitative and molecular differences distinguish adult human medullary and extramedullary haematopoietic stem and progenitor cell landscapes. *bioRxiv*: doi.org/10.1101/2020.01.26.919753
- Mende N & Laurenti E (2021) Hematopoietic stem and progenitor cells outside the bone marrow: where, when, and why. *Exp Hematol*: 1–8
- Méndez-Ferrer S, Lucas D, Battista M & Frenette PS (2008) Haematopoietic stem cell

- release is regulated by circadian oscillations. *Nature* 452: 442–447
- Miyawaki K, Jiromaru T, Kusumoto H, Yurino A, Sugio T, Uehara Y, Odawara J, Daitoku S, Kunisaki Y, Mori Y, *et al* (2017) Identification of unipotent megakaryocyte progenitors in human hematopoiesis. *Blood* 129: 3332–3343
- La Motte-Mohs RN, Herer E & Zúñiga-Pflücker JC (2005) Induction of T-cell development from human cord blood hematopoietic stem cells by Delta-like 1 in vitro. *Blood* 105: 1431–1439
- Müller A, Homey B, Soto H, Ge N, Catron D, Buchanan ME, McClanahan T, Murphy E, Yuan W, Wagner SN, *et al* (2001) Involvement of chemokine receptors in breast cancer metastasis. *Nature* 410: 50–56
- Naldini L (2011) Ex vivo gene transfer and correction for cell-based therapies. *Nat Rev Genet* 12: 301–315
- Napolitano M, Gerardi C, Lucia A Di, Accardo PA, Rizzuto L, Ferraro M, Siragusa S & Buscemi F (2016) Hematopoietic peripheral circulating blood stem cells as an independent marker of good transfusion management in patients with β -thalassemia: results from a preliminary study. *Transfusion* 56: 827–830
- Naveiras O, Nardi V, Wenzel PL, Hauschka P V., Fahey F & Daley GQ (2009) Bone-marrow adipocytes as negative regulators of the haematopoietic microenvironment. *Nature* 460: 259–263
- Nestorowa S, Hamey FK, Pijuan Sala B, Diamanti E, Shepherd M, Laurenti E, Wilson NK, Kent DG & Göttgens B (2016) A single-cell resolution map of mouse hematopoietic stem and progenitor cell differentiation. *Blood* 128: e20–e31
- Nilsson AR, Soneji S, Adolfsson S, Bryder D & Pronk CJ (2016) Human and murine hematopoietic stem cell aging is associated with functional impairments and intrinsic megakaryocytic/erythroid bias. *PLoS One* 11: 1–20
- Notta F, Doulatov S, Laurenti E, Poeppl A, Jurisica I & Dick JE (2011) Isolation of Single Human Hematopoietic Stem Cells Capable of Long-Term Multilineage Engraftment. *Science (80-)* 333: 218–221
- Notta F, Zandi S, Takayama N, Dobson S, Gan OI, Wilson G, Kaufmann KB, McLeod J, Laurenti E, Dunant CF, *et al* (2016) Distinct routes of lineage development reshape the human blood hierarchy across ontogeny. *Science (80-)* 351
- Ott MG, Schmidt M, Schwarzwaelder K, Stein S, Siler U, Koehl U, Glimm H, Kühlcke

- K, Schilz A, Kunkel H, *et al* (2006) Correction of X-linked chronic granulomatous disease by gene therapy, augmented by insertional activation of MDS1-EVI1, PRDM16 or SETBP1. *Nat Med* 12: 401–409
- Pablos JL, Amara A, Bouloc A, Santiago B, Caruz A, Galindo M, Delaunay T, Virelizier JL & Arenzana-Seisdedos F (1999) Stromal-cell derived factor is expressed by dendritic cells and endothelium in human skin. *Am J Pathol* 155: 1577–1586
- Panch SR, Szymanski J, Savani BN & Stroncek DF (2017) Sources of Hematopoietic Stem and Progenitor Cells and Methods to Optimize Yields for Clinical Cell Therapy. *Biol Blood Marrow Transplant* 23: 1241–1249
- Parkin J & Cohen B (2001) An overview of the immune system. *Lancet (London, England)* 357: 1777–1789
- Patel B, Pearson H & Zacharoulis S (2015) Mobilisation of Haematopoietic Stem Cells in Paediatric Patients Prior to Autologous Transplantation Following Administration of Plerixafor and G-CSF. *Pediatr Blood Cancer* 62: 1477–1480
- Paulson RF, Ruan B, Hao S & Chen Y (2020) Stress Erythropoiesis is a Key Inflammatory Response. *Cells* 9: 1–14
- Perry JM, Harandi OF, Porayette P, Hegde S, Kannan AK & Paulson RF (2009) Maintenance of the BMP4-dependent stress erythropoiesis pathway in the murine spleen requires hedgehog signaling. *Blood* 113: 911–918
- Pinho S & Frenette PS (2019) Haematopoietic stem cell activity and interactions with the niche. *Nat Rev Mol Cell Biol* 20: 303–320
- Pizarro S, García-Lucio J, Peinado VI, Tura-Ceide O, Díez M, Blanco I, Sitges M, Petriz J, Torralba Y, Marín P, *et al* (2014) Circulating progenitor cells and vascular dysfunction in chronic obstructive pulmonary disease. *PLoS One* 9
- Psaila B, Barkas N, Iskander D, Roy A, Anderson S, Ashley N, Caputo VS, Lichtenberg J, Loaiza S, Bodine DM, *et al* (2016) Single-cell profiling of human megakaryocyte-erythroid progenitors identifies distinct megakaryocyte and erythroid differentiation pathways. *Genome Biol* 17: 1–19
- Qian H, Tryggvason K, Jacobsen SE & Ekblom M (2006) Contribution of $\alpha 6$ integrins to hematopoietic stem and progenitor cell homing to bone marrow and collaboration with $\alpha 4$ integrins. *Blood* 107: 3503–3510

- Rahmig S, Kronstein-Wiedemann R, Fohgrub J, Kronstein N, Nevmerzhitskaya A, Bornhäuser M, Gassmann M, Platz A, Ordemann R, Tonn T, *et al* (2016) Improved Human Erythropoiesis and Platelet Formation in Humanized NSGW41 Mice. *Stem Cell Reports* 7: 591–601
- Ramirez P, Rettig MP, Uy GL, Deych E, Holt MS, Ritchey JK & DiPersio JF (2009) BIO5192, a small molecule inhibitor of VLA-4, mobilizes hematopoietic stem and progenitor cells. *Blood* 114: 1340–1343
- De Ravin SS, Wu X, Moir S, Anaya-O'Brien S, Kwatema N, Littel P, Theobald N, Choi U, Su L, Marquesen M, *et al* (2016) Lentiviral hematopoietic stem cell gene therapy for X-linked severe combined immunodeficiency. *Sci Transl Med* 8: 1–12
- Ribeil J-A, Hacein-Bey-Abina S, Payen E, Magnani A, Semeraro M, Magrin E, Caccavelli L, Neven B, Bourget P, El Nemer W, *et al* (2017) Gene Therapy in a Patient with Sickle Cell Disease. *N Engl J Med* 376: 848–855
- Río P, Navarro S, Wang W, Sánchez-Domínguez R, Pujol RM, Segovia JC, Bogliolo M, Merino E, Wu N, Salgado R, *et al* (2019) Successful engraftment of gene-corrected hematopoietic stem cells in non-conditioned patients with Fanconi anemia. *Nat Med* 25: 1396–1401
- Romero-Moya D, Bueno C, Montes R, Navarro-Montero O, Iborra FJ, López LC, Martín M & Menendez P (2013) Cord blood-derived CD34+ hematopoietic cells with low mitochondrial mass are enriched in hematopoietic repopulating stem cell function. *Haematologica* 98: 1022–1029
- Santoro A, Andrei C, Bryant C, Calderbank E, Wray A, Baxter JE, Godfrey A, Laurenti E & Ringshausen I (2020) Chronic lymphocytic leukemia increases the pool of peripheral blood hematopoietic stem cells and skews differentiation. *Blood Adv* 4: 6310–6314
- Sarma NJ, Takeda A & Yaseen NR (2010) Colony forming cell (CFC) assay for human hematopoietic cells. *J Vis Exp*
- Scala S & Aiuti A (2019) In vivo dynamics of human hematopoietic stem cells: Novel concepts and future directions. *Blood Adv* 3: 1916–1924
- Scala S, Basso-Ricci L, Dionisio F, Pellin D, Giannelli S, Salerio FA, Leonardelli L, Cicalese MP, Ferrua F, Aiuti A, *et al* (2018) Dynamics of genetically engineered hematopoietic stem and progenitor cells after autologous transplantation in

- humans. *Nat Med* 24: 1683–1690
- Schmidt M, Schwarzwaelder K, Bartholomae C, Zaoui K, Ball C, Pilz I, Braun S, Glimm H & von Kalle C (2007) High-resolution insertion-site analysis by linear amplification-mediated PCR (LAM-PCR). *Nat Methods* 4: 1051–1057
- Schmitt TM & Zúñiga-Pflücker JC (2006) T-cell development, doing it in a dish. *Immunol Rev* 209: 95–102
- Schreiber TD, Steinl C, Essl M, Abele H, Geiger K, Müller CA, Aicher WK & Klein G (2009) The integrin $\alpha 9\beta 1$ on hematopoietic stem and progenitor cells : involvement in cell adhesion , proliferation and differentiation. *Haematologica* 94: 1493–1501
- Seita J & Weissman IL (2010) Hematopoietic stem cell: Self-renewal versus differentiation. *Wiley Interdiscip Rev Syst Biol Med* 2: 640–653
- Sessa M, Lorioli L, Fumagalli F, Acquati S, Redaelli D, Baldoli C, Canale S, Lopez ID, Morena F, Calabria A, *et al* (2016) Lentiviral haemopoietic stem-cell gene therapy in early-onset metachromatic leukodystrophy: an ad-hoc analysis of a non-randomised, open-label, phase 1/2 trial. *Lancet* 388: 476–487
- Six E, Guilloux A, Denis A, Lecoules A, Magnani A, Vilette R, Male F, Cagnard N, Delville M, Magrin E, *et al* (2020) Clonal tracking in gene therapy patients reveals a diversity of human hematopoietic differentiation programs. *Blood* 135: 1219–1231
- Six EM, Bonhomme D, Monteiro M, Beldjord K, Jurkowska M, Cordier-Garcia C, Garrigue A, Dal Cortivo L, Rocha B, Fischer A, *et al* (2007) A human postnatal lymphoid progenitor capable of circulating and seeding the thymus. *J Exp Med* 204: 3085–3093
- Skirecki T, Mikaszewska-Sokolewicz M, Godlewska M, Dołęgowska B, Czubak J, Hoser G, Kawiak J & Zielińska-Borkowska U (2019) Mobilization of Stem and Progenitor Cells in Septic Shock Patients. *Sci Rep* 9: 1–10
- Spinozzi G, Calabria A, Brasca S, Beretta S, Merelli I, Milanese L & Montini E (2017) VISPA2: A scalable pipeline for high-throughput identification and annotation of vector integration sites. *BMC Bioinformatics* 18: 1–12
- Stoeckius M, Hafemeister C, Stephenson W, Houck-Loomis B, Chattopadhyay PK, Swerdlow H, Satija R & Smibert P (2017) Simultaneous epitope and transcriptome measurement in single cells. *Nat Methods* 14: 865–868

- Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, Hao Y, Stoeckius M, Smibert P & Satija R (2019) Comprehensive Integration of Single-Cell Data. *Cell* 177: 1888-1902.e21
- Suárez-Álvarez B, López-Vázquez A & López-Larrea C (2012) Mobilization and homing of hematopoietic stem cells. *Adv Exp Med Biol* 741: 152–170
- Supek F, Bošnjak M, Škunca N & Šmuc T (2011) Revigo summarizes and visualizes long lists of gene ontology terms. *PLoS One* 6
- Thompson AA, Walters MC, Kwiatkowski J, Rasko JEJ, Ribeil J-AA, Hongeng S, Magrin E, Schiller GJ, Payen E, Semeraro M, *et al* (2018) Gene therapy in patients with transfusion-dependent β -thalassemia. *N Engl J Med* 378: 1479–1493
- Tsaganos T, Giamarellos-Bourboulis EJ, Kollias S, Zervakis D, Karagianni V, Pelekanou A, Tampaki EC, Kontogiorgi M, Koroneos A, Drakoulis N, *et al* (2006) Kinetics of progenitor hemopoetic stem cells in sepsis: Correlation with patients survival? *BMC Infect Dis* 6
- Tucci F, Galimberti S, Naldini L & Aiuti A Gene therapy with hematopoietic stem and progenitor cell for monogenic disorders : a systematic review and meta-analysis (In press). *Res Sq*: 1-16 DOI: <https://doi.org/10.21203/rs.3>.
- Velten L, Haas SF, Raffel S, Blaszkiewicz S, Islam S, Hennig BP, Hirche C, Lutz C, Buss EC, Nowak D, *et al* (2017) Human haematopoietic stem cell lineage commitment is a continuous process. *Nat Cell Biol* 19: 271–281
- Wang XQ, Lo CM, Chen L, Cheung CKY, Yang ZF, Chen YX, Ng MN, Yu WC, Ming X, Zhang W, *et al* (2012) Hematopoietic chimerism in liver transplantation patients and hematopoietic stem/progenitor cells in adult human liver. *Hepatology* 56: 1557–1566
- Wright DE, Wagers AJ, Gulati AP, Johnson FL & Weissman IL (2001) Physiological Migration of Hematopoietic Stem and Progenitor Cells. *Science (80-)* 294: 1933 LP – 1936
- Wu W-C, Sun H-W, Chen H-T, Liang J, Yu X-J, Wu C, Wang Z & Zheng L (2014) Circulating hematopoietic stem and progenitor cells are myeloid-biased in cancer patients. *Proc Natl Acad Sci* 111: 4221–4226
- Yamazaki S, Ema H, Karlsson G, Yamaguchi T, Miyoshi H, Shioda S, Taketo MM, Karlsson S, Iwama A & Nakauchi H (2011) Nonmyelinating schwann cells

