Supplementary Figures, Tables and Materials and Methods

Versatile humanized niche model for studying human normal and malignant haematopoieisis

Ander Abarrategi[1]*, Katie Foster[1]*, Ashley Hamilton[1]*, Syed A. Mian[2], Diana Passaro[1], John Gribben[3], Ghulam Mufti[2] and Dominique Bonnet[1]

* Equal contribution.

1- Haematopoietic Stem Cell Laboratory, The Francis Crick Institute, London, WC2A 3LY, UK
2- King's College London School of Medicine, Department of Haematological Medicine, London, SE5 9NU, UK
3- Department of Haemato-Oncology, Barts Cancer Institute, Queen Mary University of London, London, EC1M 6BQ, United Kingdom

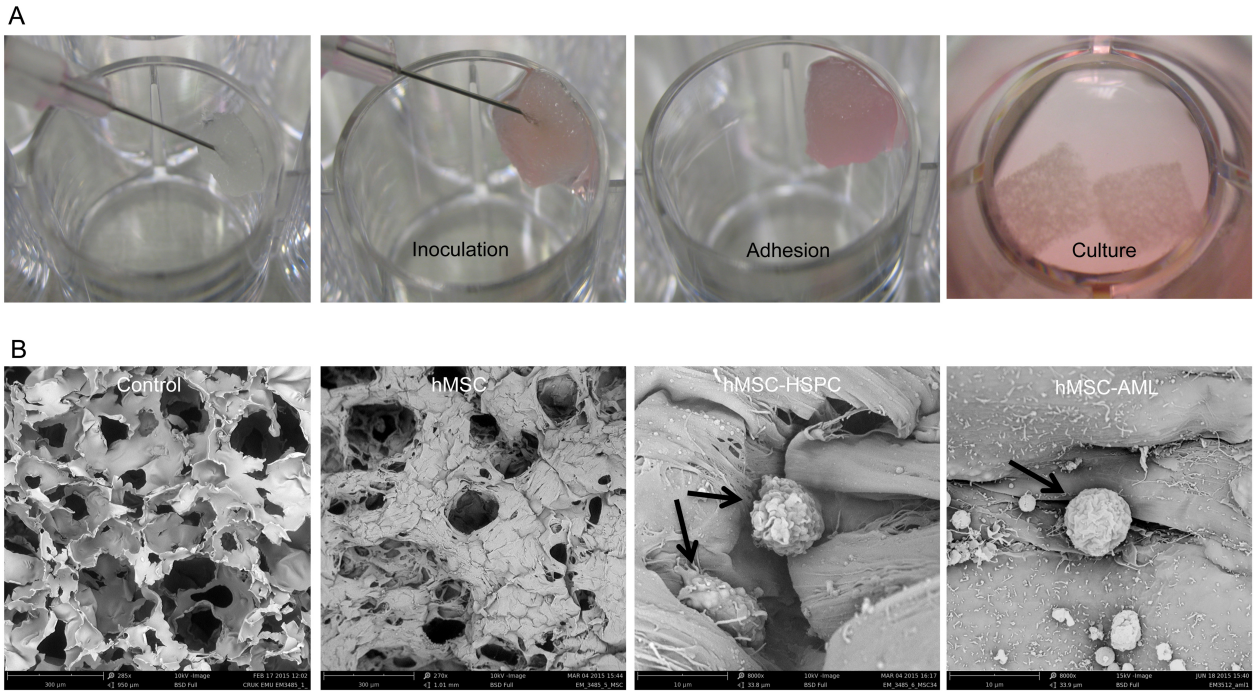
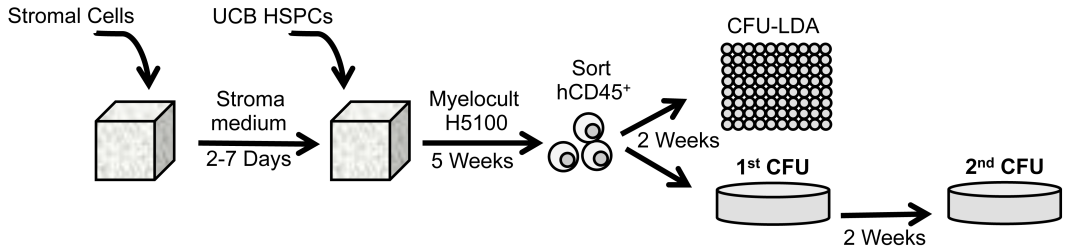
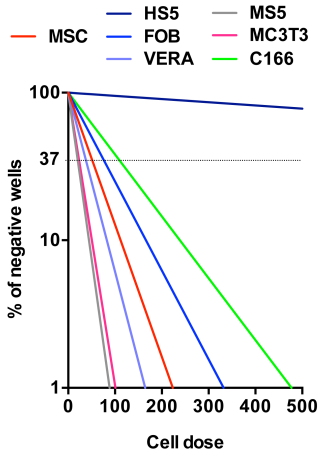


Figure S1. *In vitro* cell culture in 3D scaffolds. A) Cell-inoculation procedure. B) Electron microscopy analysis of scaffolds. Images of an empty scaffold (control), hMSC-coated scaffold, alone or with HSPC or with AML. Arrows are pointing to hematopoietic cells.

A

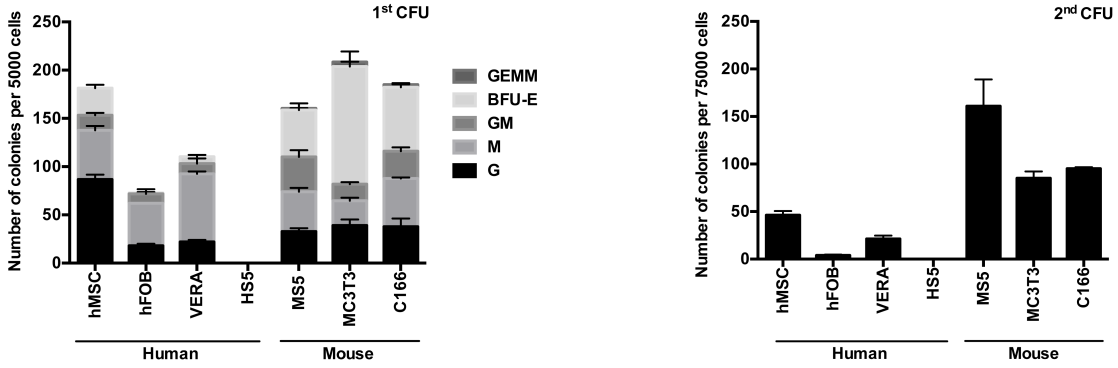


B

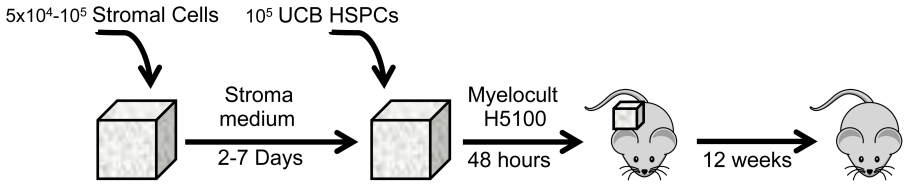


Group	LDA			Total CD45+ cells recovered from scaffolds	Estimated Absolute LTC-IC number
	Lower	Estimate	Upper		
MSC	77.5	48.6	30.4	3.6 x 10e5	7.4 x 10e4
FOB	113.8	72.1	45.7	0.2 x 10e5	0.3 x 10e4
VERA	57.7	35.8	22.2	0.1 x 10e5	0.3 x 10e4
HS5	3345	1996.6	1191	3.85 x 10e5	0.2 x 10e4
MS5	30.8	19.2	12.0	11.7 x 10 e5	60.9 x 10e4
MC3T3	34.8	21.9	13.7	2.7 x 10e5	12.3 x 10e4
C166	161.8	103.6	66.3	12 x 10e5	11.5 x 10e4

C



D



E

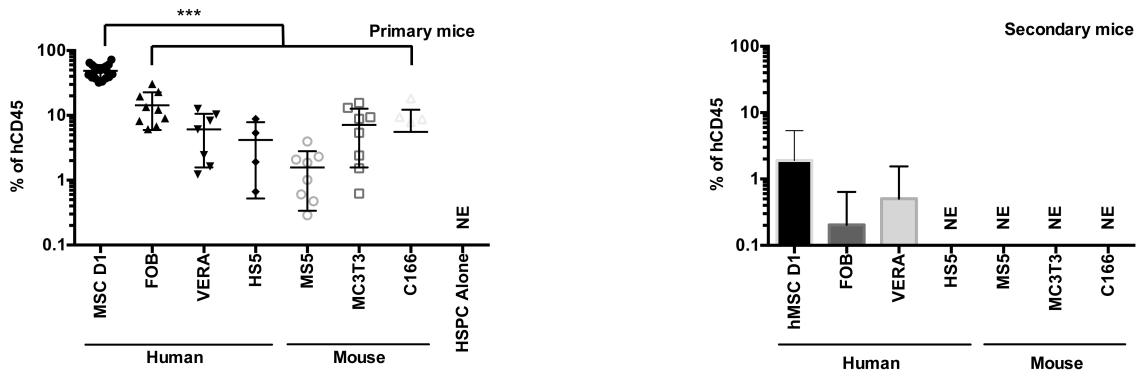


Figure S2. *In vitro* and *in vivo* evaluation of different niche components to support human HSPCs. A) Schematic of the *in vitro* testing procedure. Scaffolds were seeded with different human or mouse stromal cells and then seeded with UCB-derived HSPCs for long-term culture (LTC). In order to estimate the functionality of hematopoietic cells after long-term culture, LTC derived hematopoietic cells were recovered and assayed for primary and secondary Colony Forming Units (CFU) or tested for CFU frequency using limiting dilution assay (LDA). **B)** Data obtained from CFU-LDA assay, indicating the estimated LTC-IC frequency left and right estimated absolute number. **C)** Primary and secondary CFU derived from LTC. **D)** Schematic of scaffold-seeding and further *in vivo* implantation procedure. Samples were implanted in non-irradiated, NSG mice, recovered at 12 weeks and tested for hematopoietic cell engraftment. **E) Left panel**: Percentage of hCD45 engraftment using scaffolds pre-seeded with different human or mouse stroma cells and the same UCB. Each dot represents one mouse implanted with 6 scaffolds. For each stroma type tested, 10 mice were used. "HSPC alone" indicates that the HSPCs were implanted in the scaffold without stromal support. NE, no engraftment; **Right panel**: Cells recovered from primary scaffolds were intra bone marrow injected in sub-lethally irradiated mice (2 to 4 secondary mice were injected per stromal cell type tested). Graphic shows the percentage of $hCD45^+$ cells in secondary recipient mice.

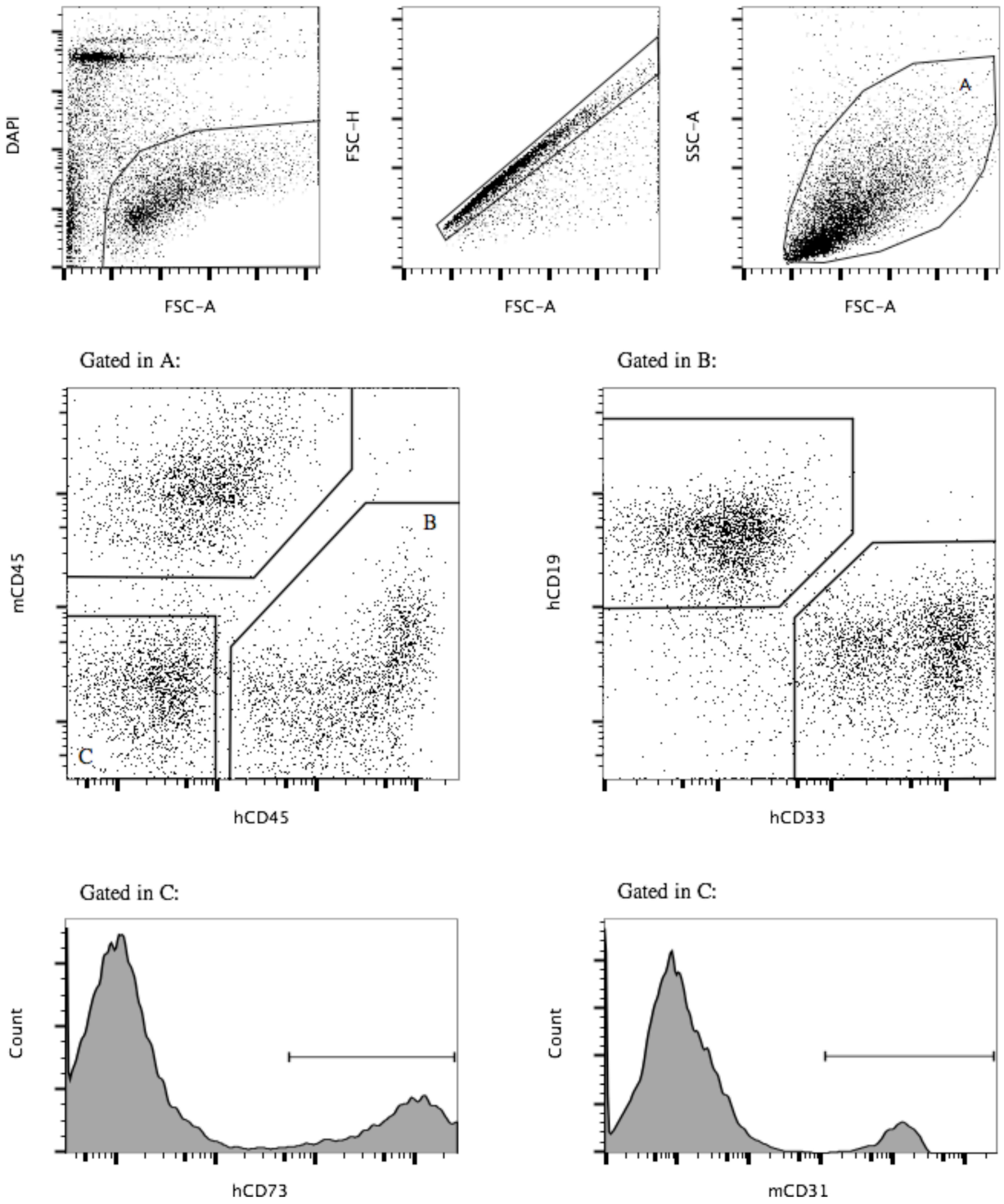


Figure S3. Gating strategy for cells retrieved from *in vivo* implanted scaffolds. After extraction of the scaffold and collagenase treatment, retrieved cells were analyzed by flow cytometry. An example of the analysis is presented here related to scaffolds seeded with hMSC and cord blood derived hHSPCs recovered 12 weeks post-implantation. The three top plots show the procedure to remove dead cells and select single cells. The two plots in the middle (gate A; live cells) show the gating strategy to evaluate the percentage of human $CD45^+$ cells (left panel), gating out mouse $mCD45^+$ hematopoietic cells and human and mouse CD45 negative cells. Right panel, gated on human $CD45^+$ cells (gate B), show the proportion of human B ($CD19^+$) and myeloid ($CD33^+$) cells. Human T ($CD3^+$) cells were also measured inside gate B, in a different channel (plot not shown) .The two plots below, gated on $mCD45^-hCD45^-$ double negative population, show the presence of h $CD73^+$ hMSCs or mouse $mCD31^+$ endothelial cells.

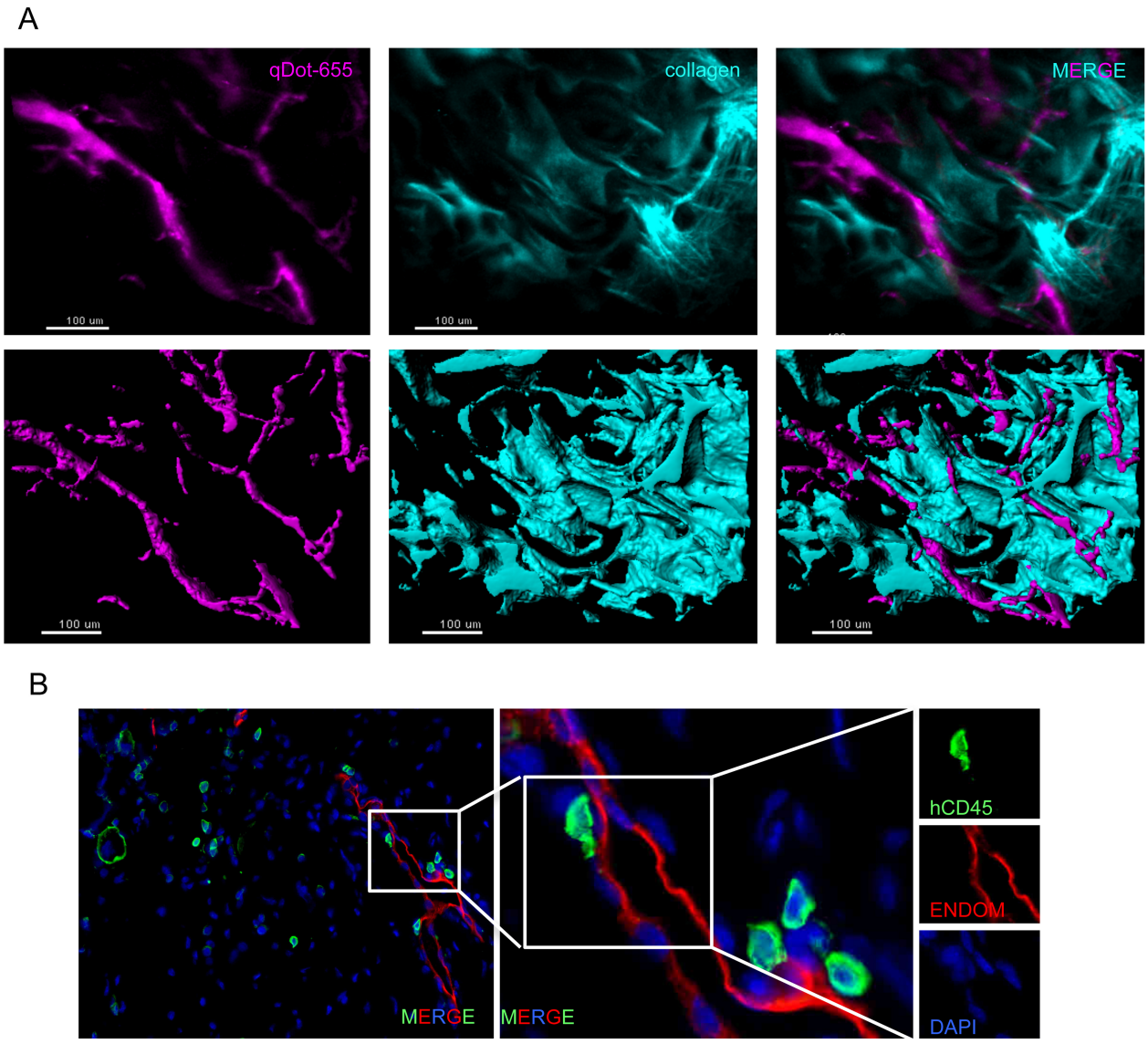


Figure S4. Characterization of *in vivo* implanted hHSPC-hMSC scaffolds. Example of scaffolds seeded with hMSCs and UCB-derived HSPCs harvested 12 weeks post-implantation and processed independently for, A) 2-photon confocal study. Images show vasculature in purple (signal from qDots-655) and collagen-based scaffold in blue (Second Harmonic Generation signal). Fluorescence images are provided on top and 3D reconstruction of the signal on the bottom. B) Histology and immunofluorescence. Human hCD45^{+} cells (in Green) were observed thorough the scaffold structure and also in close contact with scaffold-colonized mouse vasculature structure (in Red: Endomucin^{+}). Blue: dapi nuclear staining.

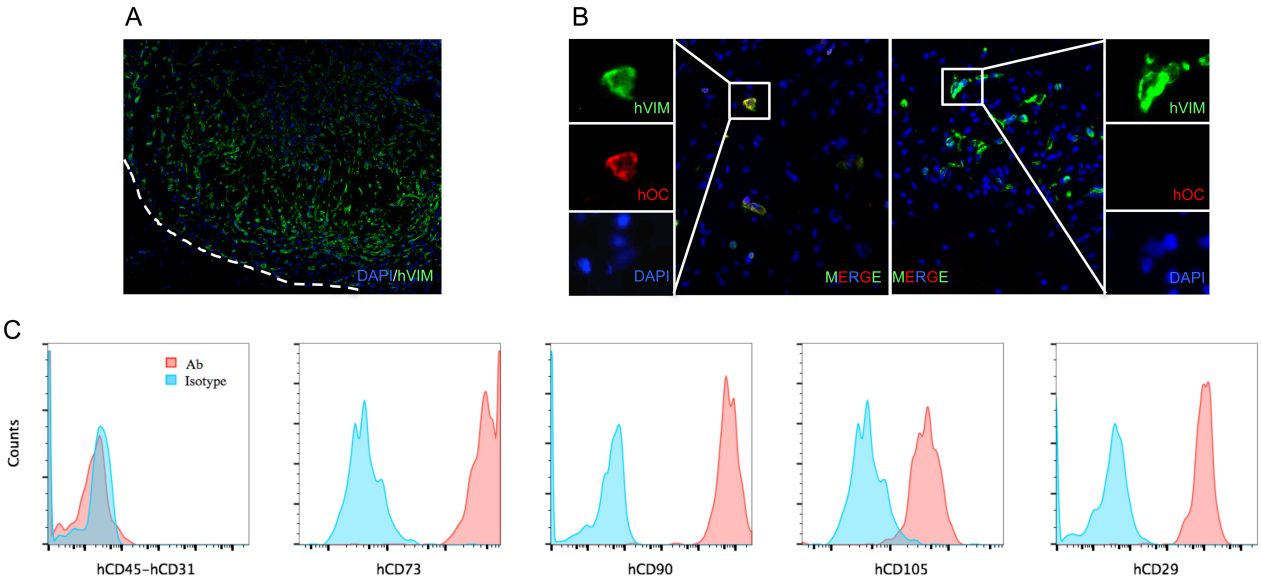


Figure S5 – Characterization of hMSCs in *in vivo* implanted scaffolds. Example of scaffolds seeded with hMSCs and UCB-derived HSCs, and harvested 12 weeks post-implantation. A) Immunofluorescence showing h-Vimentin^{+} MSCs in green, B) Left panel: human Vimentin^{+} (green) co-expressing osteocalcin^{+} (red) and Right panel: human Vimentin^{+} (green) and negative for osteocalcin C) Flow cytometry. mCD45^{-}hCD45^{-}mCD31^{-} cells analysed for the expression of hMSC markers: hCD73, hCD90, hCD105 and hCD29. (Red: specific antibody; Blue: isotype control) (Note same isotype control was used for hCD73 and hCD105).

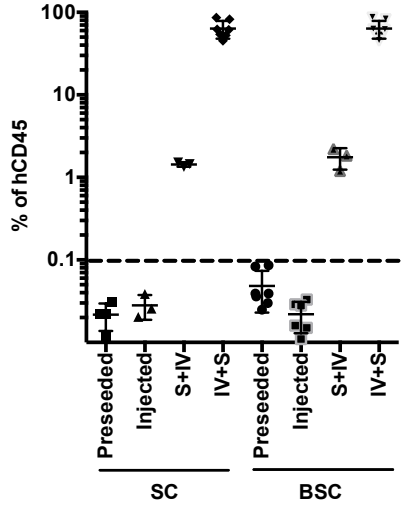


Figure S6. Analysis of hCD45$^+$ engraftment level in the bone marrow of mice implanted with scaffold shown in figure 2B. Samples were considered engrafted when the number of detectable hCD45$^+$ cells was higher than 0.1%. Each point represents one mouse. NT, non-tested.

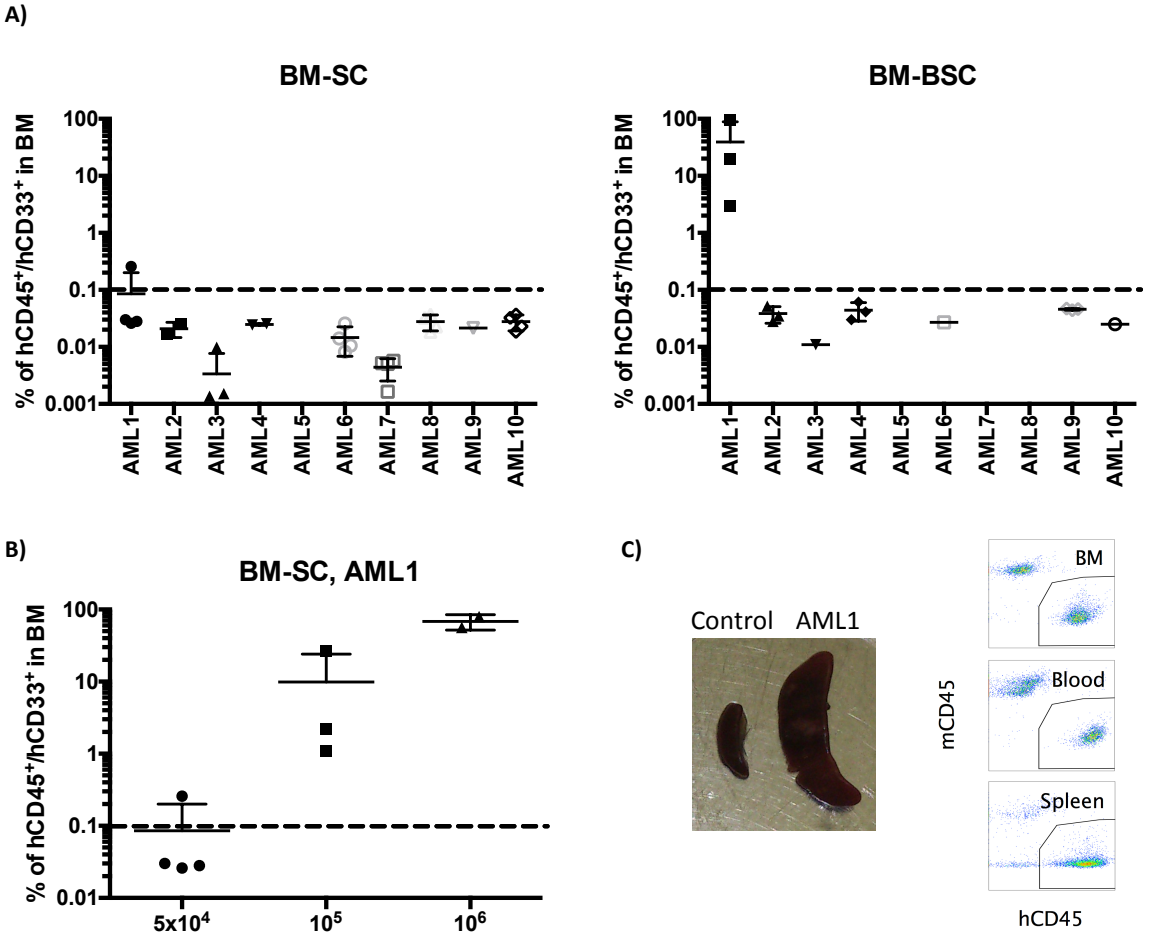


Figure S7. Study of bone marrow engraftment of AML in scaffold-implanted mice. BM-CS, Bone marrow of scaffold-implanted mice, BM-BSC Bone marrow of scaffold-implanted mice. A) Percentage of AML cells in bone marrow of mice shown in figure 3. Each dot represent a single mouse. Sample was considered engrafted when the number of target cells was higher than 0.1%. The dotted line in the graphics shows this engraftment limit. B) Spreading of AML1 patient-cells from scaffolds to bone marrow (BM) in scaffolds seeded with different amount of AML1 patient cells. C) Detail of the study of samples obtained in mice implanted with scaffolds carrying 10^6 AML1 cells.. Image shows the size of spleens coming from control mouse or AML1 implanted mouse. FASC plot show human leukemia cells in mouse bone marrow (BM), blood and spleen.

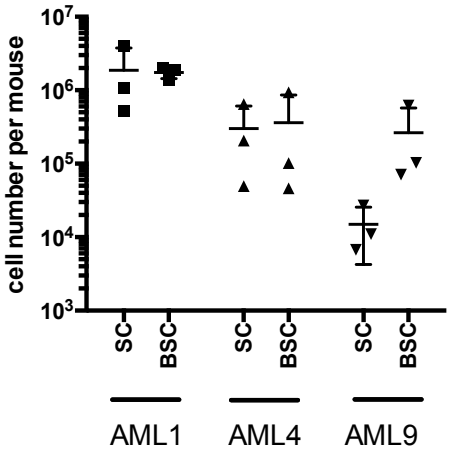


Figure S8. Absolute number of cells recovered per mouse and used to seed secondary scaffolds shown in figure 3C.

Supplementary table 1. Clinical data of the patient samples used in this study.

Patient ID	AGE	GENDER	DIAGNOSIS	FAB CLASSIFICATION	CYTOGENETICS (Karyotype)
AML-1	72	F	AML	M2	46XX
AML-2	43	M	AML	M5	t(9;11)
AML-3	58	M	AML	M5	46XY
AML-4	47	F	AML	M4	t(3;7)
AML-5	35	M	AML	M4	+3, +10
AML-6	63	F	AML	M1	46XX
AML-7	28	F	AML	M4	46XX
AML-8	65	M	AML	M4	Inv(16)
AML-9	28	M	AML	M1	m1 i(21)(q10)
AML-10	59	F	AML	M4	Inv(16)
AML-11	66	F	AML	M1	46XX
AML-12	76	M	AML	MIXED PHENOTYPE	trisomy 8,21
AML-13	Unknown	Unknown	AML	M6	Complex
AML-14	41	M	AML	M5b	t(1;12)
AML-15	52	M	AML	M2	t(6;9)

Supplementary table 2. Mutation data of the hCD33+ pre-transplanted cells and hCD45+CD33 xenografted cells. Targeted mutation screen was performed on the xenografted cells for mutant genes for respective patients. IV- intravenous, SC- scaffold, BSC- bone-forming scaffold, MAB- Mutant allele burden, NA- Sample not available, NE-Sample not engrafted. * represents mutations detected at background level <3% MAB.

Patient ID	Patient MNC Day0 Molecular Marker(s)			IV MAB (%)			SC MAB (%)			BSC MAB (%)		
	Gene	Amino Acid Change	MAB (%)	CD33+ M1	CD33+ M2	CD33+ M3	CD33+ M1	CD33+ M2	CD33+ M3	CD33+ M1	CD33+ M2	CD33+ M3
AML-1	NPM	W/C288fsX12	45	42	NA	NA	45	40	NA	38	NA	NA
	FLT3	P606_R607insYDLKWEFP	10	27	NA	NA	19	11	NA	18	NA	NA
AML-2	FLT3	D/N835	41	50	50	45	49	46	46	53	NA	NA
AML-3	NPM1	W/C288fsX12	45	41	51	NA	42	NA	NA	19	NA	NA
	FLT3	L601_K602insREYEYDL	50	40	47	NA	43	NA	NA	52	NA	NA
	RUNX1	L/S56	51	41	45	NA	17	NA	NA	0	NA	NA
AML-4	NRAS	G/D12	8	3	26	NA	0	0	0	0	0	NA
	KRAS	QL61	30	0	0	NA	0	0	0	0	0	NA
	NPM1	W/C288fsX12	52	20	40	NA	0	35	20	3	44	NA
AML-5	DNMT3A	R/H882	51	NA	NA	NA	NA	NA	NA	NT	NT	NT
	RUNX1	S/P222fsX15	50	NA	NA	NA	NA	NA	NA	NT	NT	NT
	RUNX1	D62GfsX77	48	NA	NA	NA	NA	NA	NA	NT	NT	NT
AML-6	IDH1	R/C132	51	40	36	42	28	44	0	NE	NE	NE
	NPM	W/C288fsX12	48	24	22	32	27	24	0	NE	NE	NE
AML-7	DNMT3A	R/H882	47	19	NA	NA	82	53	NA	NA	NA	NA
	IDH2	R/Q140	50	32	NA	NA	16	53	NA	NA	NA	NA
AML-8	KIT	D/V816	51	30	30	16	0	0	NA	NA	NA	NA
AML-9	CEBPA	L317_T317insGlnQ	26	0	NA	NA	3	41	46	NA	NA	NA
AML-10	None Detected		-	-	-	-	-	-	-	-	-	-
AML-11	NPM	W/C288fsX12	44	NE	NE	NE	30	38	28	NT	NT	NT
	FLT3	R/G845	5	NE	NE	NE	0*	0*	6	NT	NT	NT
AML-12	DNMT3A	R/H882	52	NE	NE	NE	15	30	NA	52	NA	NA
	SRSF2	P/H95	53	NE	NE	NE	NA	30	NA	39	NA	NA
	RUNX1	SS_donor 508+1G>A	60	NE	NE	NE	NA	NA	NA	NA	NA	NA
AML-13	SF3B1	K/N666	45	NE	NE	NE	48	52	49	29	NA	NA
	JAK2	V/F617	46	NE	NE	NE	56	39	43	21	NA	NA
AML-14	KRAS	G/S12	71	NE	NE	NE	0*	NA	NA	0	NA	NA
	ZRSR2	G438_S439InsSR	83	NE	NE	NE	70	NA	NA	0	NA	NA
AML-15	None Detected		-	-	-	-	-	-	-	-	-	-

Supplementary Materials and Methods

Cell lines and human MSC

The cell lines used were the mouse endothelial cell line; C166, mouse pre-osteoblastic cell line; MC3T3, murine stromal cell line; MS5, human stromal cell line; HS5, human foetal osteoblast cell line; hFOB and human venous endothelial cell line; VeraVec. C166, MC3T3, MS5, HS5 and hFOB were purchased from LGC standards (Middlesex, UK). VeraVec was purchased from Agiocrine Bioscience Inc, (NY, USA). All were grown as suggested by the supplier. Primary human mesenchymal stroma cells (hMSCs) were purchased (Lonza) or kindly provided by Dr. Dosquet (Universite Paris Diderot, Paris) from human bone marrow obtained during orthopaedic surgery under ethical approval 10-038 from IRB00006477. hMSCs were grown in αMEM and hMSC-specific FBS (Gibco) and used at low passages (below 5^{th} in all cases). Phenotype of the hMSCs cells was determined before implantation in scaffold as being: $hCD45^-$, $hCD31^-$, $hCD90^+$, $hCD73^+$ and $hCD105^+$. All media was supplemented with 10% FBS and 1x Penicillin-Streptomycin and all reagents were from Gibco®-Life Technologies (Paisley, UK).

Cell purification from human primary samples

Umbilical Cord Blood (UCB) samples were obtained from the Royal London Hospital (London, UK) after informed consent. Mononuclear cells (MNCs) were isolated by centrifugation using Ficoll-Paque TM PLUS (GE Healthcare Life Sciences, Buckinghamshire, UK). Cells were enriched for $CD34^+$, using an EasySep Human CD34 Positive Selection kit (StemCell Technologies, Vancouver, Canada) according to the manufacturer's instructions, with a purity of 85% to 99%. Those cells are named as HSPCs thorough the text. AML samples were collected at diagnosis and obtained after informed consent from St Bartholomew's Hospital (London, UK). Both UCB and AML protocol were approved by the East London Ethical Committee and carried out in accordance with the Declaration of Helsinki (See "study approval" paragraph in the article). Details of patient samples are listed in **Supplementary Table 1**. T cells were depleted from all AML samples using the anti-CD3 mAb OKT-3 (West Lebanon, USA) before injection.

Cell seeding in scaffolds

All the experimental procedures were performed in sterile conditions. Gelfoam gelatin sponges (2 cm x 6 cm x 7mm) (Pfizer, Kalamazoo, MI, USA) were sectioned into 24 pieces, washed with ethanol 70% and rehydrated in sterile PBS. After that, each sample was sectioned again in half and leftover media removed before proceed with cell seeding. Stroma cells were diluted in relevant culture media at $1x10^6$ cells/ml and 100μl ($1x10^5$ cells) were carefully inoculated in each scaffold using a syringe. Cell-seeded scaffolds were transferred to polystyrene ultra-low attachment 24 well plates (Corning) and maintained in cell culture conditions for 3-5 hours. Then culture media was added (specific media for each stroma cell) and scaffolds were cultured for 3 to 7 days. Human hematopoietic cells ($5x10^4$ $CD34^+$ or Lin-depleted cells) were seeded in these stroma-coated scaffolds following a similar process and media was replaced to Myelocult H5100 (StemCell Technologies). C166, MS5 and MC3T3 cells were always irradiated before hematopoietic cell seeding to stop over proliferation that would block the pores of the scaffold. Myelocult H5100 (StemCell Technologies) was supplemented with cytokines (20ng/ml G-CSF, 20ng/ml IL-3 and 20ng/ml TPO from PeproTech, London, UK) when AML patient samples were used, or with endothelial cell growth factor (ECGF, Millipore) when VeraVec cells were used as stromal cells. Stroma-seeded scaffolds, with or without pre-seeded hHSPCs, were transplanted in mice following the in vivo surgical procedure described below.

Scanning Electron Microscopy

Cell-seeded scaffolds were fixed (2.5% (v/v) glutaraldehyde / 4% (v/v) formaldehyde in 0.1M phosphate buffer) for 30 min. Samples were then incubated in 1% (v/v) reduced osmium, dehydrated through a graded series of ethanol, transferred to acetone and critical point dried (Leica EM CPD300). Samples were mounted on aluminium stubs, coated with 9nm platinum (Quorum QR 150 coater) and imaged at 10kV with a Phenom ProX scanning electron microscope.

Flow cytometry analysis and cell sorting

All scaffold samples were digested in collagenase (Sigma), 2mg/ml in PBS, before processing for FACS analysis. Flow cytometry analysis was performed using a LSRII flow cytometer (BD Biosciences, Oxford, UK). Cell sorting was performed using a FACS Aria SORP (BD Biosciences, Oxford, UK). Dead cells and debris were excluded from the analysis using 4,6, diamidino-2-phenylindole (DAPI) staining. Isotype controls used were mouse immunoglobulin-G conjugated to FITC (BD Biosciences, 555909), PE (BD Biosciences, 555749), APC (BD Biosciences, 550854), PercP (BD Biosciences, 552991), PC5 (Beckman Coulter, IM2663U), or PC7 (Beckman Coulter, 737662). CD34 expression of primary hematopoietic samples was assessed using human CD34-FITC and human CD38-APC antibodies (BD Pharmagen, 555821 and 555462). Human grafts in mice were assessed using mouseCD45-PerCP and humanCD45-PC7 (eBioscience, 45-0451-82 and 25-0459-42), humanCD19-APC, humanCD33-PE and humanCD3-FITC antibodies (BD Pharmagen, 555415, 555450 and 555339). Human mesenchymal cells recovered from scaffolds were assessed using mouseCD45-PerCP (eBioscience, 45-0451-82), humanCD45-FITC, humanCD31-FITC, humanCD73-PE and humanCD29-PC5 (BD Phramagen, 555482, 555445, 550257 and 559882), humanCD90-APC (eBioscience, 17-0909-42), and humanCD105-PE (Serotec, MCA1557PE).

In vitro long term culture initiating cell assays (LTC-IC)

Scaffolds were seeded and cultured as mentioned in “Cell seeding in scaffolds” section. 48 scaffolds were prepared per stroma cell type ($1x10^5$ stromal cells/scaffold, $5x10^4$ UCB-derived HSPCs/scaffold). Following a 5-week *in vitro* co-culture, scaffolds were digested (collagenase 2mg/ml in PBS, 1 hour at 37C), cells were recovered and human $CD45^+$ cells were sorted for further colony forming unit (CFU) or limiting dilution analysis (LDA) assays. To perform CFU assays, $CD45^+$ cells were plated in the methylcellulose media, MethoCult H4434 (StemCell Technologies, Vancouver, Canada), before colonies were scored according to their morphology at 14 days using an inverted microscope. Colonies were harvested by washing with PBS and re-plated in secondary assays for an additional 14 days. For estimating the frequency of LTC-IC, LDA was performed. In this case, $CD45^+$ cells were plated in 96 well plates in methylcellulose

media, MethoCult H4434 (StemCell Technologies, Vancouver, Canada). 8 different dilutions were performed, from 5000 to 10 cells per well, and 10 wells were plated per each dilution. 14 days after plating, wells with and without colonies were counted and LTC-IC frequency was calculated as previously reported (1) using the ELDA online web tool (http://bioinf.wehi.edu.au/software/elda/index.html).

In vivo studies: Surgical Implantation of Scaffolds

NSG mice were a kind gift from Dr Leonard Shultz (The Jackson Laboratory, Bar Harbor, ME). All animal experiments were performed in accordance to Home Office and Crick guidelines. Scaffolds were seeded and cultured as mentioned in "Cell seeding in scaffolds" section. 2 hours before surgical procedure Caprofren (Rimadyl, Zoetis) anti-inflammatory and pain-killer drug was administrated to each animal, both subcutaneously and in the drinking water. Anesthesia was induced with 2.5% isoflurane and O_2 at 2-4%. A wide section of fur from the back was shaved. Then skin was sterilized twice with surgiscrub, and washed twice with PBS. For each scaffold implantation, 0.5 cm incision was made 1cm away from the spine on each side of the animal. With forceps, a pocket under the skin was made in the incision, down the side of the animal. A scaffold was inserted, making sure it was placed deep within the pocket, and then incisions were glued (3M surgical glue, Vetbond, St Paul, MN, USA). The procedure was repeated in order to implant successive scaffolds in the same mouse, and up to six scaffolds were implanted per mice. Buprenorphine (Vetergesic, Alstoe) post-operative analgesia was administrated under the skin. Animals were placed on its side in a pre-warmed cage and left to recover. After surgery, animals were checked frequently for their well-being. Rimadyl in the drinking water was removed 48 hours after surgery. At endpoint, all scaffolds per each animal were harvested and pooled, and then digested in collagenase (Sigma), 2mg/ml in PBS, before processing for FACS analysis as described above. Bone samples (femurs and tibias) for each mouse were also harvested, bone marrow was flushed and treated with ammonium chloride red cell lysis buffer before processing for FACS analysis.

Additional methods for human hematopoietic cell-incorporation

In specified assays, mice were implanted with hMSC-seeded scaffolds as described above and then hHSPCs were injected intravenously. In other assays, hHSPCs were injected intra-scaffold in previously implanted scaffolds. For this procedure, mice were anesthetized with 2.5% isoflurane and O_2 at 2-4% before intra-scaffold injection of hHSPCs through the skin. In other assays, mice were sublethally irradiated (3.75 Gy) and hHSPCs were injected intravenously. Four to 6 weeks later, hMSC-seeded scaffolds were implanted as described above.

BMP-2 incorporation to the system

BMP-2 (Noricum, Spain) was incorporated into scaffolds before implantation in a fibrin coating as previously described (2). Briefly, each scaffold was allocated in a 15mL falcon tube (Falcon) and 8 µL of BMP-2 (Noricum, Spain) (reconstituted in acetic acid 50mM at 5µg/µL) were added. Then, 30µL of thrombin from human plasma (Sigma) (reconstituted in 2% $CaCl_2$) and 30µL of fibrinogen from human plasma (Sigma) (water reconstituted) were incorporated. Solidification was allowed during 30 min in cell culture conditions before proceeding with *in vivo* implantation.

microCT studies

Samples were scanned using a SkyScan-1176 µCT scanner (Bruker MicroCT, Kontich, Belgium). The X-ray source was operated at 40kV and 600µA, no filter was used. The scans were made over a trajectory of 180º with a 0.4º step size with a 9.01µm pixel size. The images were reconstructed using nRecon (Bruker MicroCT) and further analysed using CTan (Bruker MicroCT).

2-photon confocal microscopy study.

Twelve weeks after scaffold-implantation in mice Streptavidin conjugated Quantum dots (Qdots)-655 (Thermo Fisher Scientific) were IV injected to label endothelial cells. It was done 30 minutes prior to euthanasia and scaffold collection. Images of scaffolds were obtained on a Zeiss 710 NLO laser scanning multiphoton microscope with a 20x 1.0 NA water immersion lens. The

microscope is equipped with a MaiTai “High Performance” fully automated 1-box mode-locked Ti:Sapphire laser with DeepSee dispersion compensation (Spectra-Physics), tuned to 800nm excitation wavelength. Scaffold-derived collagen signal (Second Harmonic Generation signal) was collected at 380-485nm and qDots-655 at 640-690nm by non-descanned detectors.

Histology and IHC

Harvested samples were fixed overnight in 10% neutral buffered formalin and then decalcified with 17% EDTA (Osteosoft, Millipore) during 7 days. Samples were processed, paraffin embedded and sectioned (5 μm) for histological studies. Hematoxylin/eosin was performed first to assess quality of the sections. For immunofluorescence (IF) studies heat antigen retrieval was performed in all cases. Primary unconjugated antibodies employed were specific for the following proteins: human CD45 (Dako, M0701), human Vimentin (Santa Cruz, sc-6260), Endomucin (Santa Cruz, sc-65495), Osterix (Santa Cruz, sc-22536-R), and human Osteocalcin (Santa Cruz, sc-18319). Secondary fluorescent antibodies were from Invitrogen. Images were obtained in a Nikon Eclipse 90i microscope equipped with Intensilight C-HGFI Fiber Illuminator for Fluorescent microscopy.

Mutation analysis

Gene mutation screening for myeloid-related gene aberrations was performed by using King’s College Hospital gene panel. DNA from patient peripheral blood $CD33^{+}$ cells was used to amplify entire coding regions or known hotspots for 24 genes: ASXL1, CBL, CEBPA, DNMT3A, ETV6, EZH2, FLT3, GATA2, IDH1, IDH2, JAK2, KDM6A, KIT, KRAS, NPM1, NRAS, RUNX1, SF3B1, SRSF2, STAG2, TET2, TP53, U2AF1 and ZRSR2; using the Illumina TruSeq Amplicon panel as described previously (3). Sequencing was performed using the MiSeq Instrument with version 3 sequencing chemistry (Illumina, San Diego, CA, USA), as per the manufacture's protocol.

Sequencing data was processed using the GATK pipeline package (Broad Institute, Boston, MA, USA). Processed VCF and BAM files were visualized using variant studio (Illimina) and integrated genome viewer (IGV), respectively. Candidate variants obtained were passed for validation if the

variants were deemed to cause protein changes and not found in dbSNP142, esp5400 and 6500, and 1000 genomes databases at <0.01 population allele frequency.

Gene variants were deemed to be candidate variants and were included for further analysis only if they had an allele burden of ≥5%, present in COSMIC (Catalogue of Somatic Mutations in Cancer) or had been previously reported. In addition, novel variants were also included if the coverage across the variant was >200 sequencing reads. All the candidate variants were subjected to confirmation via independent targeted PCR followed by transposon-based Illumina Nextera technology sequencing as described in next section. The gene variants present in each patient are summarized in **table 1 and supplementary table 2**.

Targeted Amplicon Sequencing

PCR amplicon libraries were prepared as described previously (4). DNA from hCD33+ cells obtained from pre-xeno transplant (Day 0) and post-xenografted mice was used for PCR library preparation. After the PCR amplification of the targeted gene regions, transposon-based Illumina Nextera technology (Illumina, UK) was used to prepare the Illumina sequencing libraries by following the manufacturer's protocol. All patient indexed amplicon mixes were quantified by using the Quant-iT pico-green dsDNA assay kit (Life Science Technologies) by following manufacturers protocol. Subsequently, patient libraries were pooled together, purified using the Agencourt AMpure XP beads (Beckman Coulter) and then quantitated using the Quant-iT picoGreen dsDNA assay kit (Life Science Technologies) by following manufacturers protocol. Pooled libraries were sequenced on the Illumina MiSeq platform utilizing version 2 chemistry with 150-250 paired-end reads. On average gene amplicons were covered with >1000 sequencing reads, therefore providing a high depth mutation screening.

Statistical analysis

Prism Version 6 software (GraphPad) was used for statistical analysis. Data is presented as the mean ± S.E.M. Statistical analysis was performed using the two-tailed student's t-test for

comparison of two groups or using ordinary one-way ANOVA for multiple comparisons test. A P value less than 0.05 was considered significant (* $p \leq 0.05$; ** $p \leq 0.005$; *** $p \leq 0.0005$).

References to materials and methods section:

1- Hu, Y., & Smyth, G. K. ELDA: extreme limiting dilution analysis for comparing depleted and enriched populations in stem cell and other assays. *Journal of immunological methods*. 2009;*347*(1):70-78.

2- Abarrategi A, et al. In Vivo Ectopic Implantation Model to Assess Human Mesenchymal Progenitor Cell Potential. *Stem Cell Rev. Rep*. 2013;9:833–846.

3- Mohamedali A.M., et al. High concordance of genomic and cytogenetic aberrations between peripheral blood and bone marrow in myelodysplastic syndrome (MDS). *Leukemia*. 2015; 29(9):1928-38.

4- Mian S.A., et al. SF3B1 mutant MDS-initiating cells may arise from the haematopoietic stem cell compartment. *Nature Comm..* 2015; 8:6:10004.