Donor	Gender	Ethnicity	Age	Cause of death	Donation	Comorbidities	Smoking (>20 PY)	Last P/F	Analysis
018	female	White	61	Anoxia	DCD	HTN, CAD	no	263	FC, CFU
019	male	Hispanic	57	ICH	DBD	CAD	no	107	FC, CFU, scSeq
020	male	Hispanic	46	Anoxia	DBD	HTN/HLD, GERD	yes	257	FC, CFU, scSeq
021	female	White	56	ICH	DBD	HTN/HLD, RA, COPD	yes	388	FC, CFU, scSeq
022	female	Asian	31	ICH	DBD	no	no	288	FC, CFU, Xeno
023	male	White	57	Head trauma	DCD	Alcohol Abuse	no	382	FC, CFU, scSeq
024	male	White	45	ICH	DBD	CAD	no	237	FC, CFU, scSeq
025	female	White	47	ICH	DBD	unknown	no	245	FC, CFU, scSeq
026	male	Hispanic	27	Head trauma	DBD	no	no	297	FC, CFU, scSeq, Xeno
027	male	White	30	Head trauma	DBD	no	no	275	scSeq
028	female	Asian	36	Anoxia	DBD	Asthma	no	350	scSeq, Xeno
029	female	White	59	ICH	DBD	HTN, Asthma	no	260	scSeq, Xeno
030	male	White	43	Anoxia	DCD	Asthma	no	231	FC, Xeno
031	female	White	66	Head trauma	DCD	HTN, Asthma	no	450	FC
033	male	AI/AN	35	Anoxia	DBD	CAD	no	195	FC
034	male	White	19	Head trauma	DBD	no	no	275	FC, scSeq
035	male	B/AF	20	Anoxia	DCD	ARVC	no	196	FC, scSeq, SpOmics
036	male	White	32	Anoxia	DBD	Asthma	no	208	SpOmics
037	male	White	39	Head trauma	DBD	no	no	240	FC
038	male	White	53	Anoxia	DBD	no	yes	307	SpOmics
039	male	White	37	Anoxia	DCD	no	no	165	FC
042	female	White	50	Stroke	DBD	no	no	182	SpOmics

Abbreviations:
AI/AN, American Indian or Alaska Native; ARVC, Arrhythmogenic right ventricular cardiomyopathy; B/AF, Black or African American; CAD, Coronary artery disease; COPD, Chronic obstructive lung disease, DBD, Donor after brain death; DCD, Donor after cardiac death; GERD, Gastroesophageal reflux disease; HLD, Hyperlipidemia; HTN, Hypertension; ICH, Intracerebral hemorrhage; P/F, pO_2/FiO_2 ratio; PY, pack years; RA, Rheumatoid arthritis.

FC, Flow Cytometry (Immunphenotyping); CFU, Colony-Forming Unit Assay; scSeq, single-cell RNA Sequencing; SpOmics, spatial Transcriptomics; Xeno, Xenotransplantation.

Extended Data Table 1: Basic demographics and clinical profiles of deceased organ donors.

Progenitor cell subset	Markers[9-10]
HSC/MPP Pool	$Lin^-CD34^+CD38^-CD45RA^-$
HPC Pool	$Lin^-CD34^+CD38^+$
Hematopoietic Stem Cell (HSC)	$Lin^-CD34^+CD38^-CD45RA^-CD90^+(CD49f^+)$
Multipotent Progenitor (MPP)	$Lin^-CD34^+CD38^-CD45RA^-CD90^-(CD49f^-)$
Multipotent Lymphoid Progenitor (CLP)	$Lin^-CD34^+CD38^-CD45RA^+CD90^-$
Common Myeloid Progenitor (CMP)	$Lin^-CD34^+CD38^+CD45RA^-Flt3^+$
Megakaryocyte-Erythroid Progenitor (MEP)	$Lin^-CD34^+CD38^{lo/+}CD45RA^-Flt3^-$
Colony-Forming Unit-Megakaryocyte (CFU-MK)	$Lin^-CD34^+CD45RA^-Flt3^-CD41^+$
Granulocyte-Macrophage Progenitor (GMP)	$Lin^-CD34^+CD38^+CD45RA^+Flt3^+$

Extended Data Table 2: Haematopoietic progenitor subsets and surface marker expression.

AT1 cell	AGER
	CAV1
	RTKN2
AT2 cell	SFTPC
	SFTPD
	LAMP3
Fibroblast	CLU
	FBLN1
	SERPINF1
	PDGFRA
	CFD
	GPC3
	COL1A2
Pericyte	PDGFRB
	PTN
	HIGD1B
Lymphatic	CCL21
	TFF3
	MMRN1
Smooth Muscle	MYH11
	ACTA2
	ACTG2
Capillary	VWF
	PECAM1
	TEK
	(CD34)
Artery	GJA5
	HEY1
	DKK2
Vein	ACKR1
	CPE
Mesothelial	HP
	CALB2

Leukocyte	PTPRC
Macrophage	MARCO
	C1QB
	FABP4
Monocyte	VCAN
	CD14
	FCGR3A
Neutrophil	CSF3R
	G0S2
	FCGR3B
T cell	CD3D
	IL7R
	CD4 or CD8A
B cell	BANK1
	MS4A1
	CD79A*
DC	IRF7
	LILRA4
	PLD4
NK cell	PRF1
	NKG7
	GZMB
MK/ Platelet	ITGA2B
	PPBP
	PF4
Mast cell	HPGDS
	CPA3
Niche factors	CXCR4
	CXCL12
	KITLG
	THPO

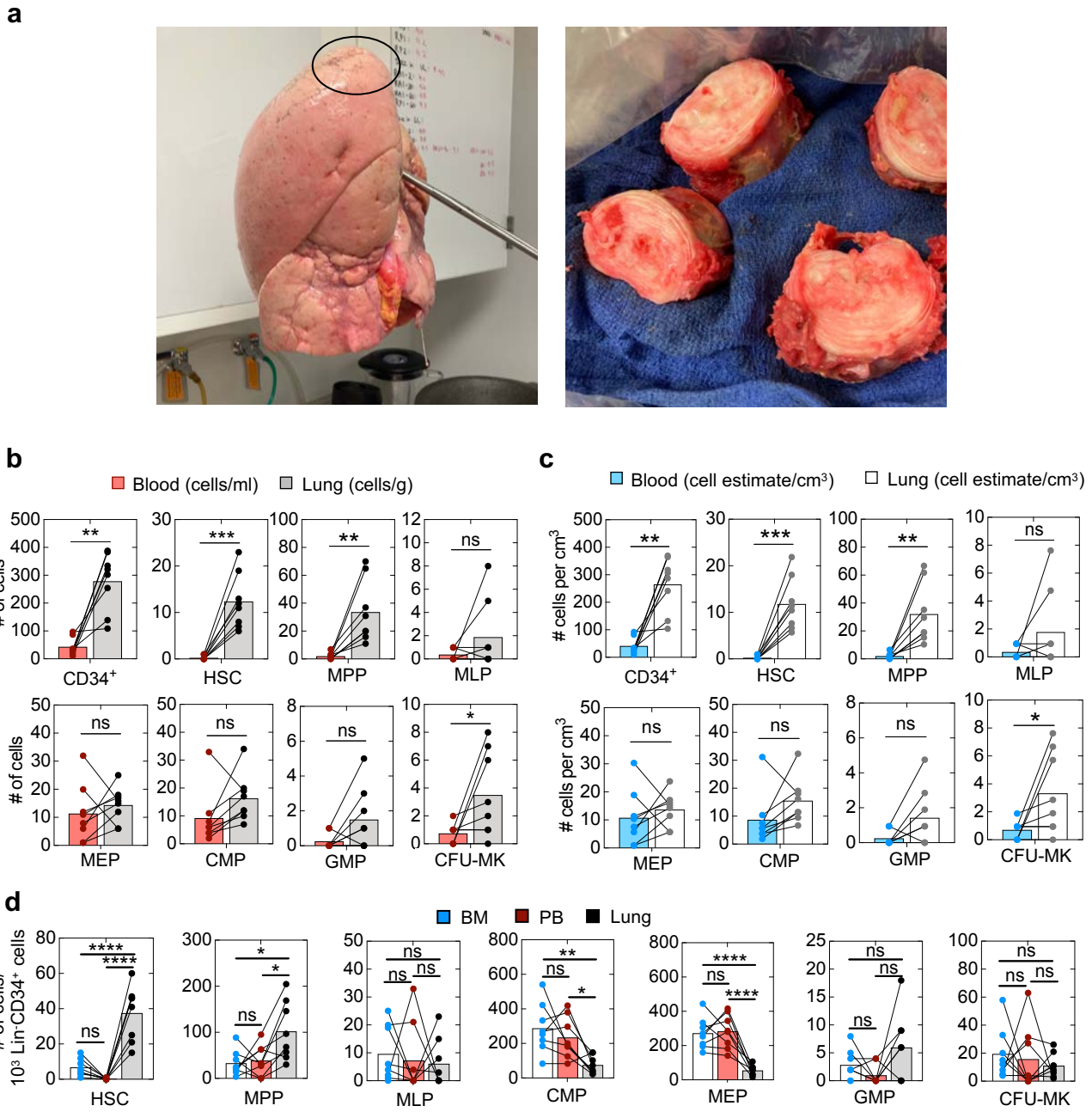
CD34+	CD34
Progenitor	SPINK2
	SELL
	HOPX
	CRHBP
	HBB
	KLF1
	BLVRB
	CA1
	HBD
	FCER1A
	THY1
	CD38
	ITGA6
	MPL
	KIT
	MEG3
	FCER1A
	PCLAF
	AHSA1
	BTF3
	CHCHD2
	TMPO
	ANKRD12
	LAPTM5
	TUBB
	NUCKS1
	TUBA1B
	CENPF
	RORA
	FOSB
	GSTM5
	HMGB2

Extended Data Table 3: Molecular probes to characterize cell types in the human lung in spatial transcriptomics. * Probe excluded due to non-specific staining.

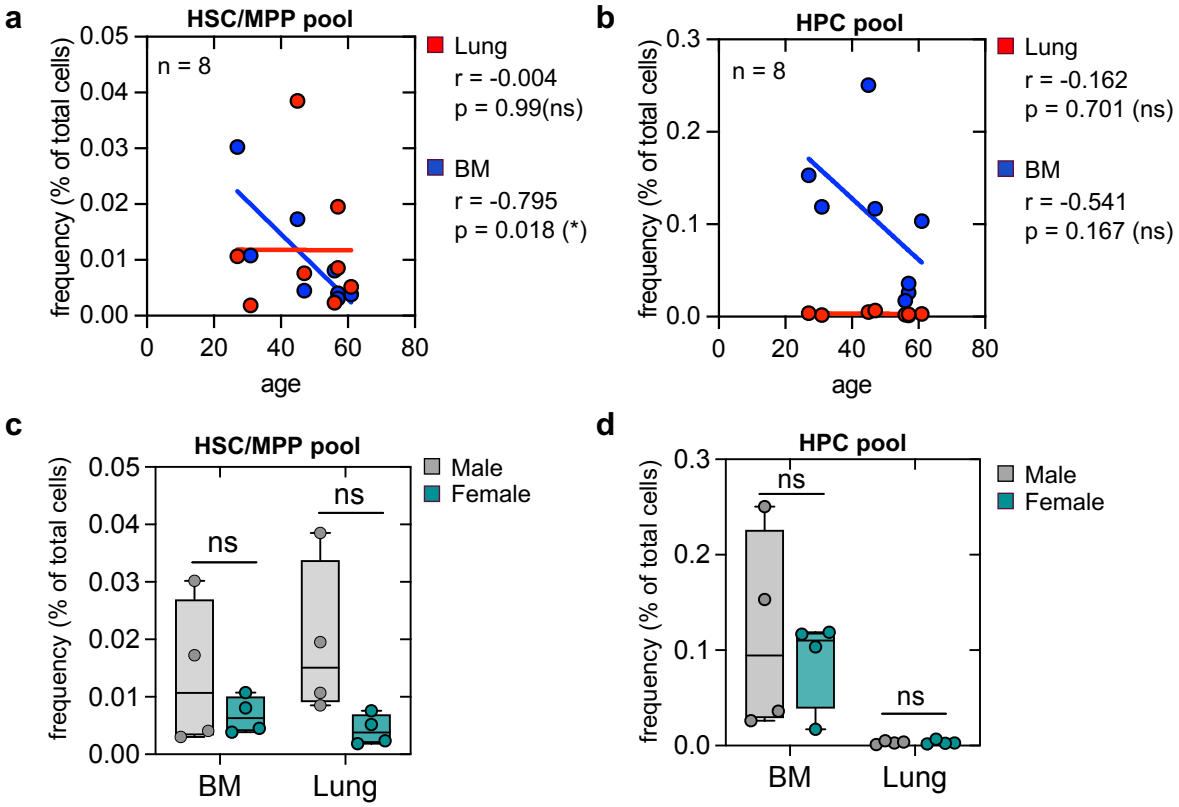
Flow Cytometry		SOURCE	IDENTIFIER
Antibodies	**Host**		
anti-human CD2 Biotin (clone RPA-2.10)	mouse	StemCell	Cat #60007BT
anti-human CD11b Biotin (clone ICRF44)	mouse	StemCell	Cat #60040BT
anti-human CD3 Biotin (clone UCHT1)	mouse	StemCell	Cat #60011BT
anti-human CD66b Biotin (clone G10F5)	mouse	BioLegend	Cat #305120
anti-human CD16 Biotin (clone 3G8)	mouse	StemCell	Cat #60041BT
anti-human CD11c Biotin (clone 3.9)	mouse	Invitrogen	Cat #11-0116-82,
anti-human CD14 Biotin (clone 6103)	mouse	Invitrogen	Cat #13-0149-82
anti-human CD19 Biotin (clone HIB19)	mouse	StemCell	Cat #60005BT
anti-human CD235 Biotin (clone HIR2)	mouse	BioLegend	Cat #306618
anti-human CD56 Biotin (clone HCD56)	mouse	StemCell	Cat #60021BT
anti-human CD24 Biotin (eBioSN3)	mouse	Invitrogen	Cat #13-0247-82
anti-human CD31 Biotin (WM-59)	mouse	Invitrogen	Cat #13-0319-82
anti-human CD326 Biotin (1B7)	mouse	Invitrogen	Cat #13-9326-82
anti-human CD140a Biotin (16A1)	mouse	Biolegend	Cat #323503
anti-human CD34 FITC (clone 581)	mouse	StemCell	Cat #60013FI
anti-human CD38 APC (clone HIT2)	mouse	Invitrogen	Cat #17-0389-42
anti-human CD41/61 PerCP-Cy5.5 (clone A2A9/6)	mouse	BioLegend	Cat #359813
anti-human CD45RA APC-Cy7 (clone HI100)	mouse	BD Pharmingen	Cat #560674
anti-human CD49f Pacific Blue (clone GoH3)	rat	BD Pharmingen	Cat #60037PB.1
anti-human CD90 PE (clone 5E10)	mouse	StemCell	Cat #60045PE.1
anti-human CD135 BV711 (clone 4G8)	mouse	BD Horizon	Cat #563908
anti-human CD14 PE (clone 61D3)	mouse	Invitrogen	Cat #12-0149-41
anti-human CD15 APC (clone 4G8)	mouse	Invitrogen	Cat #17-0158-41
anti-human CD45 APC-Cy7 (clone 2D1)	mouse	Biolegend	Cat #368515
anti-human CD45 APC-eFluor 780 (clone 30-F11)	mouse	Invitrogen	Cat # 47-0451-82
anti-human CD45 AF647 (clone HI30)	mouse	Biolegend	Cat # 304056
anti-human GlyA Pacific Blue (clone HI264)	mouse	Invitrogen	Cat #349107
anti-human GlyA PE (clone HIR2)	mouse	Invitrogen	Cat #12-9987-80
anti-human CD41a AF488 (clone HIP8)	mouse	BioLegend	Cat #303723
anti-human CD41a FITC (clone HIP8)	mouse	BD Pharmingen	Cat # 555466
anti-human Ki-67 PerCP-Cy5.5	mouse	BioLegend	Cat #350519
anti-human CD33 PerCP (clone WM2)	mouse	Invitrogen	Cat # 50-113-7770
anti-human CD33 PE (Clone 6C5/2)	mouse	R&D Systems	Cat # FAB1137P
anti-human CD71 PerCP-Cy5.5 (clone CY1G4)	mouse	BioLegend	Cat # 334113
anti-human CD3 BB700 (clone SK7)	mouse	BD Horizon	Cat # 566575
anti-human CD19 BB700 (clone SJ25C1)	mouse	BD Horizon	Cat # 566396
anti-human CD19 PB (clone LT19)	mouse	BioRad	Cat # MCA1940PBT
anti-human CD19 PE-Cy7 (clone HIB19)	mouse	NovusBio	Cat # NBP1-42967
anti-mouse CD45 PE-Cy7 (clone 30-F11)	rat	BD Biosciences	Cat # 552828
anti-mouse CD45 APC (clone 30-F11)	rat	BioLegend	Cat # 103112
anti-mouse CD45 APC-Cy7 (clone 30-F11)	rat	BioLegend	Cat # 557659
Streptavidin-BV605		BioLegend	Cat #405229
Streptavidin-AF647		BioLegend	Cat # 405237
LIVE/DEAD Fixable Yellow Stain		ThermoFisher	Cat #L34959
Human TruStain FcX™ (Fc Blocker)		BioLegend	Cat #422302
TruStain FcX™ (anti-mouse CD16/32) Antibody		BioLegend	Cat # 101319

Immunhistochemistry		SOURCE	IDENTIFIER
Antibodies	**Host**		
anti-human CD45 (clone D9M8I)	rabbit	CellSignalling	Cat # 13917S
anti-human GlyA (clone EPR8200)	rabbit	Abcam	Cat #ab129024
anti-rabbit IgG (HRP)	donkey	Abcam	Cat #ab98493

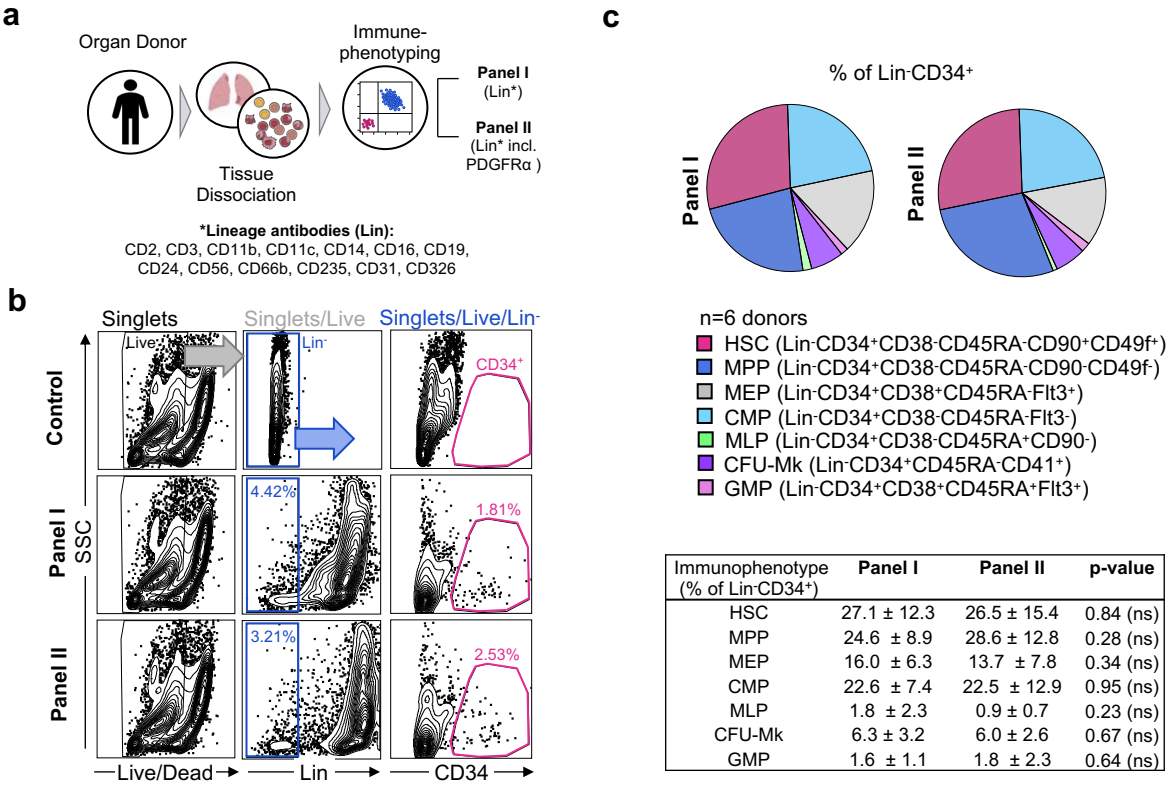
Extended Data Table 4: Flow cytometry and immunohistochemistry antibodies.



Extended Data Figure 1: Measured and predicted numbers of haematopoietic progenitor cells in the lung, PB and BM. (a) Representative donor lung and matching vertebral bodies. Patients with no significant lung pathologies or haematological disorders of any age, gender and ethnicity were included in our study. Lungs were inspected for visible injury and tissue was only collected from normal-appearing regions (black circle). **(b)** The absolute numbers of immunophenotypic haematopoietic progenitors were quantified in 1 mL of peripheral blood (red) and 1 gram of lung tissue (grey) by flow cytometry. **(c)** To facilitate direct comparison, cell numbers per cm^3 for each tissue were calculated based on published densities (https://www.aqua-calc.com/calculate/weight-to-volume) for blood (blue, Ref ID 362, 1.0565 g/cm^3) and for lung (white, Ref ID 1762, 1.050 g/cm^3). **(d)** Numbers of haematopoietic progenitor cell subsets in the BM (blue), PB (red) and lung (grey) per 10^3 Lin^-CD34^+ cells. n=8 donors. Student's t-test *p< 0.05; **p<0.01; ***p<0.001; ****p<0.0001.

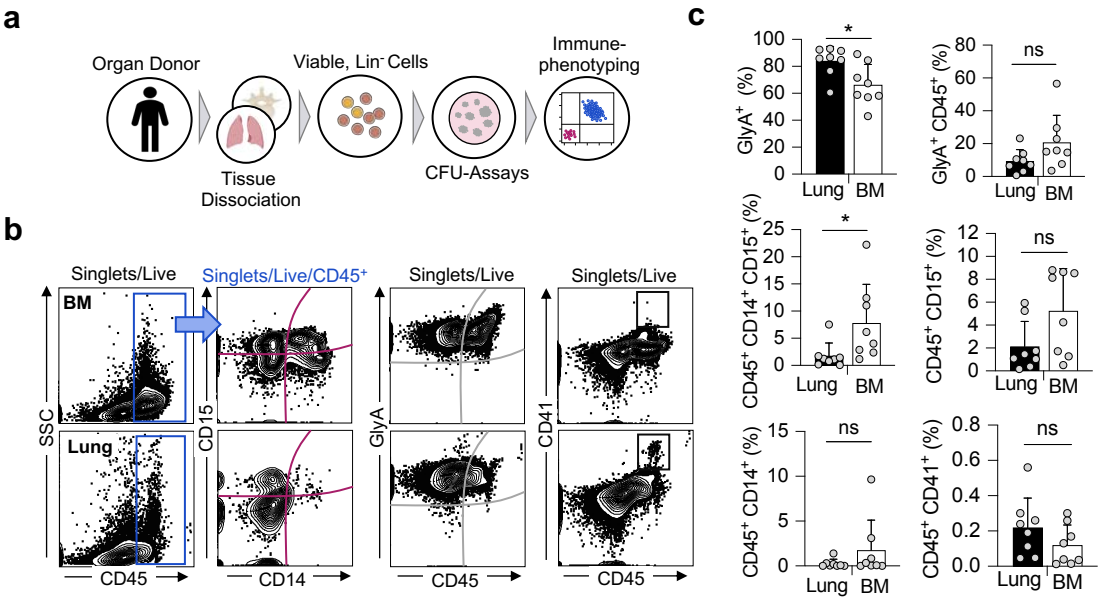


Extended Data Figure 2: Association of HSPC frequency with age and gender. (a) Scatterplot illustrating the correlation between age and HSC/MPP frequency in the BM (blue) and lung (red). The regression line visualizes the association of the variable, Pearson's correlation coefficient (r) and associated p-values (p) are indicated. **(b)** Correlation between age and HPC frequency in the BM and lung, respectively. **(c, d)** Box and whisker plot with individual values showing the HSC/MPP **(c)** and HPC (d) frequencies separated by donor sex (male, grey; female, green). ANOVA followed by Sidak's multiple comparison test. ns, not significant.

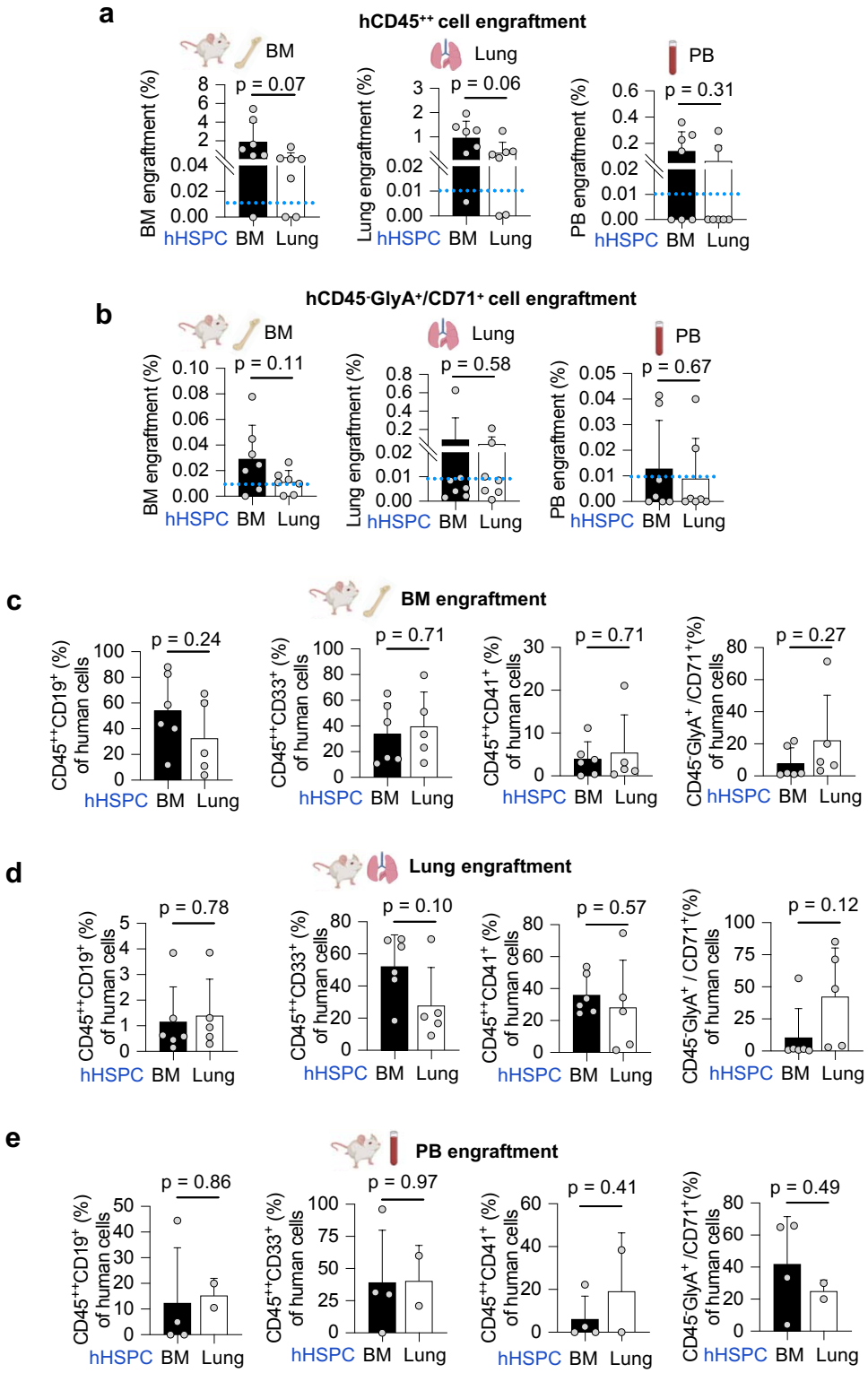


Immunophenotype (% of Lin-CD34+)	Panel I	Panel II	p-value
HSC	27.1 ± 12.3	26.5 ± 15.4	0.84 (ns)
MPP	24.6 ± 8.9	28.6 ± 12.8	0.28 (ns)
MEP	16.0 ± 6.3	13.7 ± 7.8	0.34 (ns)
CMP	22.6 ± 7.4	22.5 ± 12.9	0.95 (ns)
MLP	1.8 ± 2.3	0.9 ± 0.7	0.23 (ns)
CFU-Mk	6.3 ± 3.2	6.0 ± 2.6	0.67 (ns)
GMP	1.6 ± 1.1	1.8 ± 2.3	0.64 (ns)

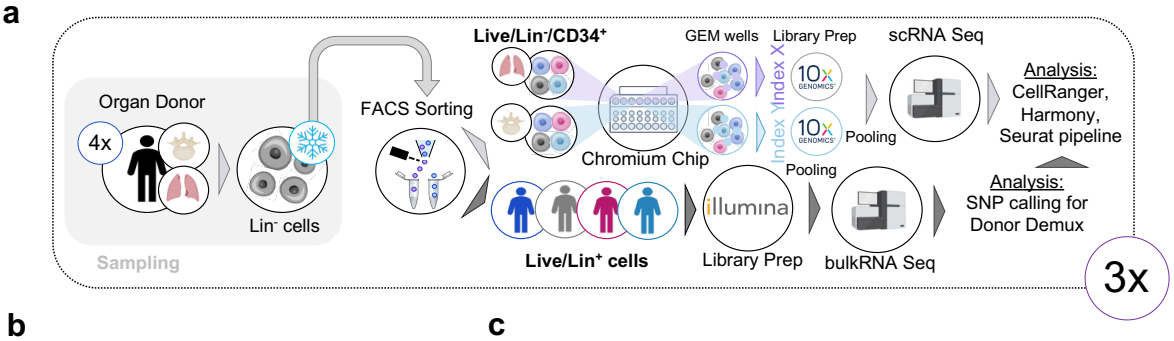
Extended Data Figure 3: Lineage panel modification to reduce fibroblast capture. (a) Published datasets[13] as well as our own scRNA-seq data suggest that platelet-derived growth factor receptor alpha (PDGFRα) could mark most fibroblasts in the lung. To test this, an antibody against PDGFRα was added to the lineage panel (Panel II) and the results were compared to the lineage panel without PDGFRα (Panel I). **(b)** Representative flow cytometry plots of lung cells show the impact of the PDGFRα antibody on the Live/Lin$^-$/CD34$^+$ cell population. 'Control' cells were stained with the viability dye only to determine autofluorescence. **(c)** Frequencies of immunophenotypes as percentage of Live/Lin$^-$/CD34$^+$ cells with Panel I versus II; no significant differences were detected across 6 donors (ANOVA followed by Sidak's multiple comparison test.)



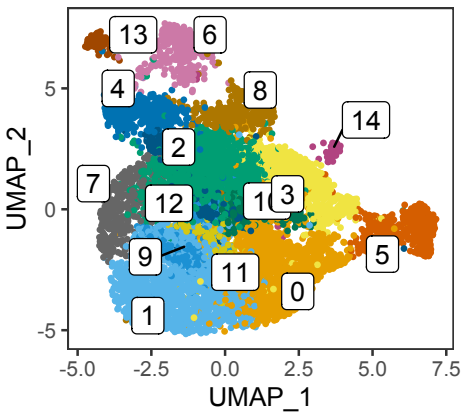
Extended Data Figure 4: Confirmation of hematopoietic surface marker expression on colonies produced by BM or lung progenitor cells using flow cytometry. (a) Following visual colony identification at 14 days of culture (Figure 2e, MethoCult™ media was dissolved to generate a single cell suspension for flow cytometric analysis. **(b)** Representative flow plots of cellular lineage marker expression on BM- and lung-derived colonies (GlyA, erythroid; CD45, leukocytic; CD14, monocytic; CD15, neutrophilic; CD41, megakaryocytic). **(c)** Cellular composition of colonies across 8 matched donors, marker expression given as percentage of single, live cells (%). Student's ttest * $p<0.03$; ns, not significant.



Extended Data Figure 5. **Engraftment efficiency of human cells in the BM, lung and PB of recipient mice after xenotransplantation of HSPCs from BM or lung. (a)** Bar graphs representing the percentage of hCD45^{++} cell engraftment in the BM, lung and PB of recipient mice after transplantation of HSPCs from human BM (black) or lung (white). Mean ± SD; individual data points for each animal are plotted as gray dots. Blue dotted line indicates threshold for positive engraftment. **(b)** Bar graphs representing the percentage of human erythroid engraftment (CD45^{-}GlyA^{+}CD71^{+}) in BM, lung, and PB. **(c-e)** Lineage expansion of human lymphoid (CD45^{++}CD19^{+}), human myeloid (hCD45^{++}CD19^{+}), human megakaryocytic (CD45^{++}CD41^{+}) and human erythroid (CD45^{-}GlyA^{+}CD71^{+}) cells as percentage of all human cells in the BM (**c**), lung (**d**), and PB (**e**) of recipient mice. Mean ± SD; individual data points for each animal are plotted as gray dots. Student's t-test.



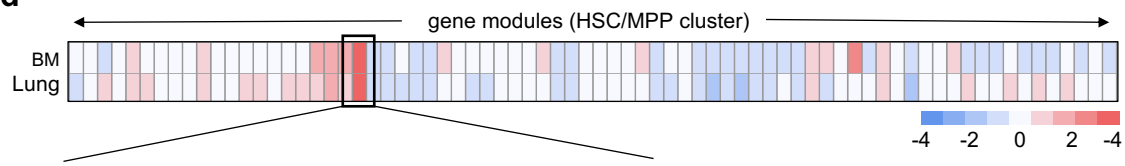
b



c

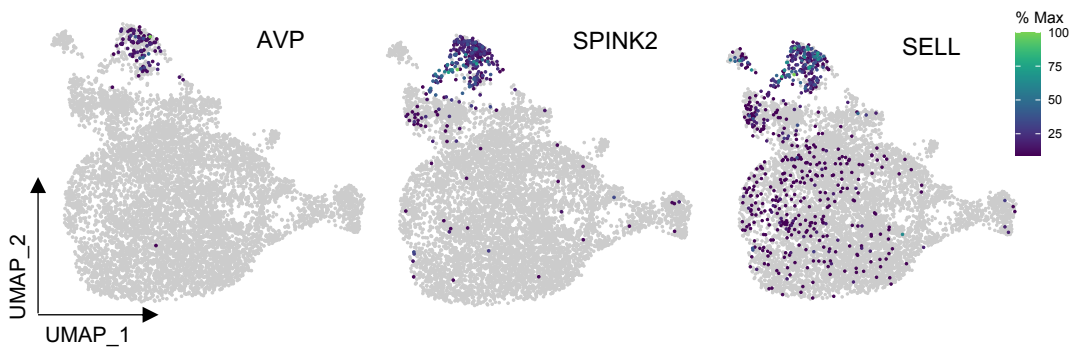
Cluster	Progenitor subset	Marker genes
0	Erythroid	HBB, GLRX5, SOD1
1	lateERY	HBD, CA1, HBB, HMGB2, STMN1, DUT
2	MPP/ multiLin	LEPROT, CELF2, PNRC1, YPEL3, CCNI, DUSP1
3	Stromal/prog mix	APOC1, CTSH, VIM, APOE, PDLXDC2
4	BaEoMa	TPSB2, CD69, CPA3, HPGDS, GATA2
5	Stromal	APOD, DCN, GPX3, CFD, COL6A2
6	HSC/MPP	AVP, CD74, SPINK2, SELL, MLLT3, HOPX, CD34
7	EMP	PLEK, GP1BB, HBD, ITGA2B, CD38, HEMGN
8	MPP	SOX4, KLF2, JUN, JUNB, FOSB, BTG2
9	Cycling cells	TUBA1A, TUBB, STMN1, BIRC5, TOP2A
10	Not specific	CCNL1, FOSB, NEAT1, NKTR
11	earlyERY	CA1, CA2, HBD, FAM178B, KLF1, BLVRB
12	Not specific	no markers
13	My	LYZ, TYROBP, FCER1G, S100A8, CD14
14	ND	no markers

d

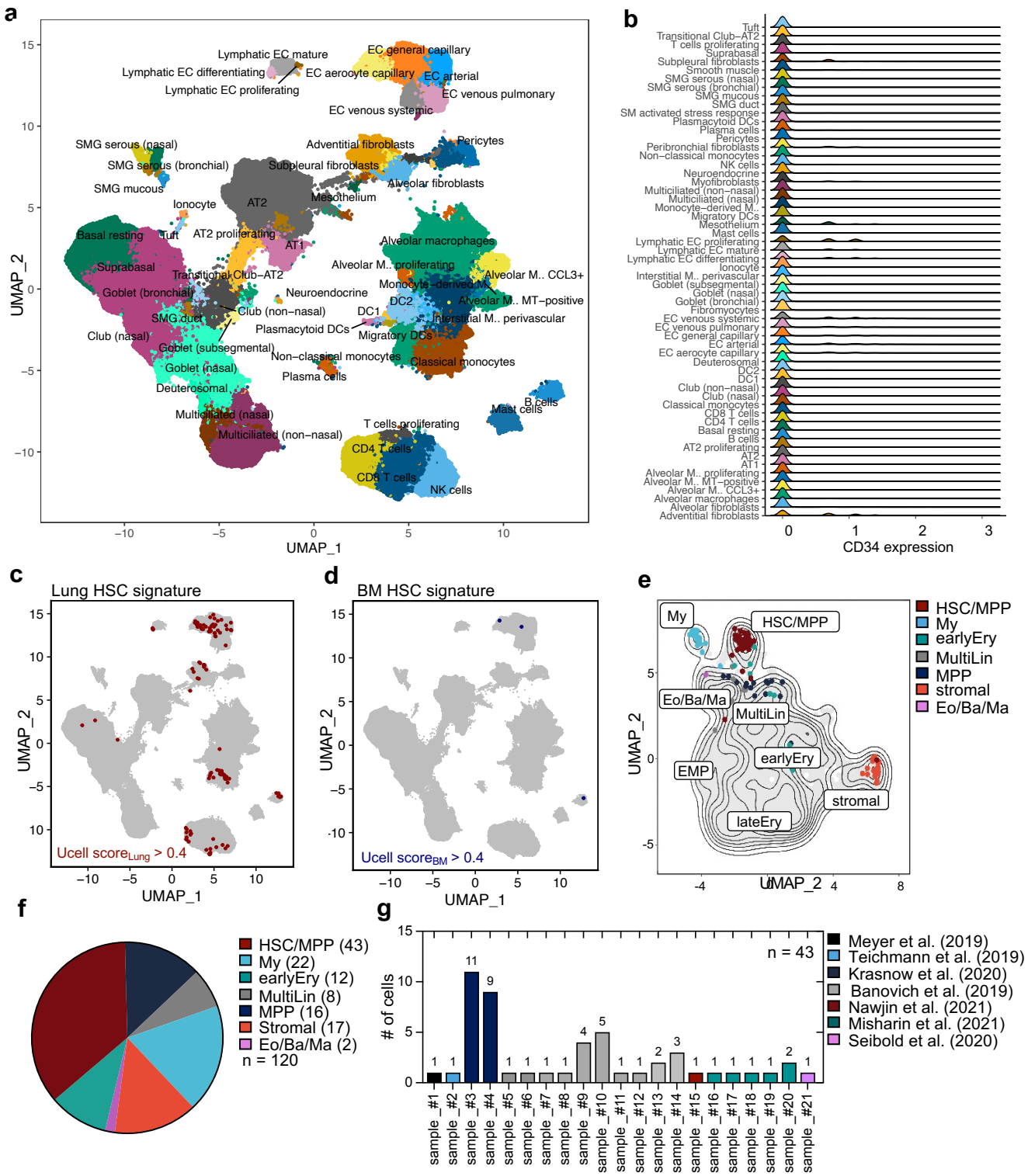


Module 54: *SPINK2, HOPX, TPT1, SELL, AREG, AVP, AJ009632.2, HLA-E, SOCS2, HEMGN, C1QTNF4, ADGRG6, BST2, CSF3R, TXNIP, MLLT3, RFLNB, GLIPR1, TCEAL2, CXCL2, FLT3, RCSD1, NAAA, BAALC, ATP8B4, TFEC, LPCAT2, INPP4B, LAIR1, ICAM3, ATP1B1, GUCY1A1, MMP7, PIK3IP1, TCTEX1D1, PDZK1IP1, PREX2, CALN1, PCDH9, TMEM163, BEX2, ELMO1, SLC2A5, BEX1, SORL1, GIMAP1, PTGER4, ABCB1, SMIM3, GSTM5, LYSMD2, MIR155HG, SULT1C4, RIC3, CDADC1, SSBP2, AGPS, LINC00891, CBX2, IL18, NAALADL1, PTAFR, STARD3, MSRB3, A1BG, CYSLTR1, SOCS2-AS1, NPM2, MBOAT7, SCN9A, CPLX1, GSTM2, ADAM28, CLDN15, AC004130.2, BCAS4, ITGA6, HLF, ACOT11, ROBO4, GIMAP5, TNFSF4, PRKCH, BEX5, GPAT3, SH3D21, F2RL1, PTP4A3, PGM5, KRT18, KRT8, KIAA1257, LPIN1, RHPN1, HOXA6, MMP28, COL24A1, TSPAN2, C9orf43, ALS2, PLAG1, AP000547.3, AC007336.1, PANX2, MYO5C, DENND6B, FAM47E, CYB561D2.1, WAKMAR2, LINC00893, OXT, CACNA1F, STX17-AS1, LY6G5C, PRX, BEND7, AL359220.1, UBOX5, COL9A2, GRAP, FCMR, TNFRSF10D*

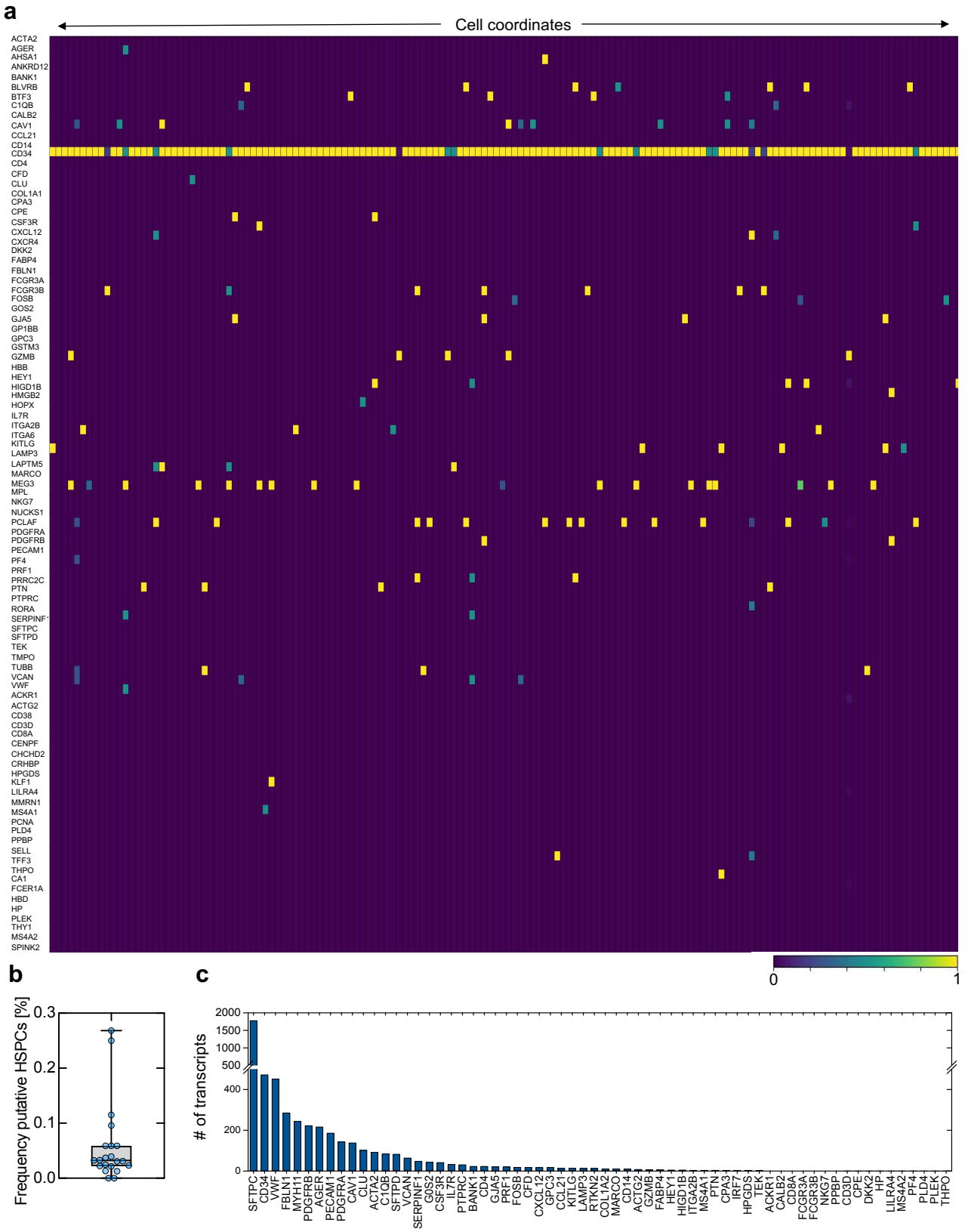
e



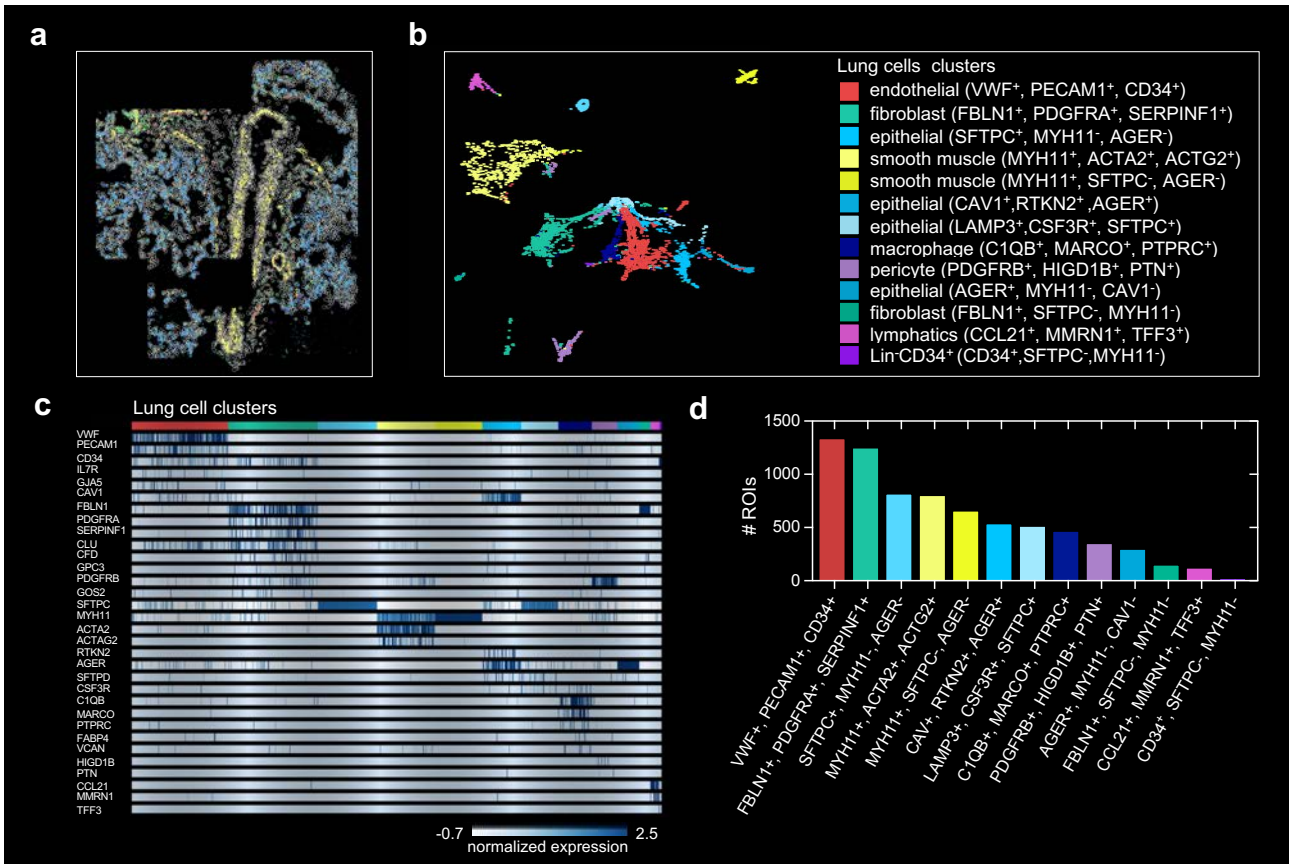
Extended Data Figure 6. Pipeline for multiplexed scRNAseq of Lin$^-$CD34$^+$ cells from matched lung and BM and annotation of hematopoietic progenitor subsets. (a) Single-cell suspensions generated from lung and BM were lineage-depleted (Lin$^-$ cells) and cryopreserved. For each experimental batch (4 donors), Live/Lin$^+$/CD34$^+$ cells were flow sorted and encapsulated into 2 GEMs (Chip, GEM wells). 10x Chromium^{TM} Single Cell 3' v2 libraries were prepared, pooled and sequenced. Live/Lin$^+$ cells were collected for bulk RNA-sequencing and subsequent SNP calling for donor demultiplexing. The experiment was carried out in 3 batches (3x). **(b)** Batch corrected UMAP representation and Louvain clustering (resolution 0.5). **(c)** Overview of marker genes associated with each cluster based on Seurat's 'FindConservedMarkers' function. **(d)** Aggregate module scores generated with Monocle3's 'find_gene_modules' function grouping similar patterns of gene expression within the HSC/MPP cluster. Genes co-regulated in the module highly specific for HSCs (Module 54) are noted below. **(e)** Expression values of selected HSC-associated genes.



Extended Data Figure 7. Identification of HSC signatures in published scRNA-seq datasets of the human lung. Due to their rarity, HSCs might be masked by the noise of other highly abundant cell types in the lung. **(a)** UMAP representation with annotations of the lung reference data set (https://azimuth.hubmapconsortium.org/references/human_lung_v2/) representing 584,944 cells from 9 datasets[13,30-37]. **(b)** CD34 expression levels across all lung cell entities in the reference dataset. **(c, d)** Ucell gene signature scoring to identify putative HSCs based on their gene expression profile in the lung. **(c)** UMAP projection highlighting cells with a Ucell score for lung HSC signatures >0.4. **(d)** UMAP projection highlighting cells with a Ucell score for BM HSC signatures >0.4. **(e)** Using Ucell signature scoring, we identified 120 putative HSCs in the human cell lung atlas V2 (HCLA V2). These cells were projected on the UMAP structure of haematopoietic progenitors from the lung and BM generated in Fig. 3a. Predicted IDs of the putative HSCs are shown in the legend. **(f)** Pie graph showing the proportion of predicted cell identities across the 120 putative HSCs in the HLCA V2 categorized by UCell scoring. Cell counts are indicated in parentheses. **(g)** Representation of cells projecting on the HSC/MPP cluster (n= 43) in individual samples from datasets within the integrated lung reference atlas. Out of 584,944 cells in the HLCA V2, 43 HSC/MPPs were identified representing a frequency of 0.007%.



Extended Data Figure 8. Gene expression of putative HSPCs across all lung tissue sections. (a) Heatmap displaying gene expression normalized per cell of putative HSPCs across all lung tissue sections. **(b)** Frequency of putative HSCs among all lung cells based on QuPath segmentation. **(c)** Number of marker transcripts within a radius of 20µm from putative HSPCs across all anatomic locations.



Extended Data Figure 9. Analytical pipeline to delineate neighboring cells of putative HSPCs. (a) Mapping of target transcripts at subcellular resolution using the Molecular Cartography platform by Resolve. Cell segmentation was performed using DAPI-based cell detection and adapting the QuPath algorithm for lung tissue. The generated cell segmentation ROI sets were used to compute cell type clustering. **(b)** Cells were clustered into distinct populations based on their marker gene expression and annotated through comparison with reference gene sets. UMAP visualization pseudo-colored by annotated lung cell types. **(c)** Heatmap representation of marker gene expression values normalized for each gene. Cell segmentation ROIs have been sorted by assigned cell type cluster, indicated by the color in the top row (see **(b)** for color legend). **(d)** Bar graph illustrating the distribution of annotated cell type clusters in the lung tissue shown in **(a)**.