A latent subset of human hematopoietic stem cells resists regenerative stress to preserve stemness

Kerstin B. Kaufmann, Andy G. X. Zeng, Etienne-Marie Coyaud, Laura Garcia-Prat, Efthymia Papalexi, Alex Murison, Estelle Laurent, Michelle Chan-Seng-Yue, Olga I. Gan, Kristele Pan, et al.

Figure #	Figure title One sentence only	Filename This should be the name the file is saved as when it is uploaded to our system. Please include the file extension. i.e.: *Smith_ED_Fig1.jpg*	Figure Legend If you are citing a reference for the first time in these legends, please include all new references in the main text Methods References section, and carry on the numbering from the main References section of the paper. If your paper does not have a Methods section, include all new references at the end of the main Reference list.
Extended Data Fig. 1	INKA1 and H4K16ac are mutually exclusive.	Kaufmann_Extended_Data_Fig1.tif	(a) Representative sorting scheme to obtain LT-, ST-, IT90+- IT90- HSCs or CD34+CD38- and CD34+CD38+ cells from lineage depleted CB. (b) Confocal images of 2 individual ST-HSC to visualize heterogeneity and representative for n=3 independent CB pools analyzed. *Scale bar = 2 μm*. (c) Confocal images of 4 individual, uncultured LT-HSC next to thereof derived images for INKA1 and H4K16ac channels after thresholding and binary conversion using ImageJ to visualize intracellular invers staining pattern in double dim cells of the heterogenous cell pool and representative for n=3 independent CB. *Scale bar = 4 μm*. (d) Percentage INKA1^{+} cells in HSC subsets according to INKA1^{+} cell gate in Fig. 1d (immunostaining, Ø50 cells/population and CB, n=3 independent CB pools). *Mean±SD.* (e) Confocal images and quantification (MFI) of prospectively isolated LT- and ST-HSC stained for INKA1 and CDK6 and representative for n=3 independent CB as summarized in right panel. Gating as used for Figure 1F is indicated. *Scale bar = 2 μm*.
Extended Data Fig. 2	Mobilized PB single cell RNA pseudotime analysis.	Kaufmann_Extended_Data_Fig2.tif	(a) Cell cycle analysis (Ki-67, Hoechst, CDK6) of prospectively isolated HSC subsets (LT-, IT90^{+}-, IT90^{-}, ST-HSC) from mPB (n=3) by flow cytometry. CB HSPC (CD34^{+}CD38$^{–}$) and Progenitors (Prog, CD34^{+}CD38^{+}) served as gating controls. *Mean±SD; two-sided paired t-test (**p=0.002).*

			(b-e) Single cell RNA-seq data from $CD34^{+}CD38^{-}$ $CD45RA^{-}$ index-sorted mPB HSC (n=3, also stained and indexed for CD49f, CD90) visualized by diffusion mapping overlaid with (b) cell cycle score (Tirosh et al., 2016) and indicated transcript expression; (c) donor origin, (d) *ITGA* (CD49f) transcript expression, (e) CD49f and CD90 indexed protein surface expression and (f) retrospectively gated HSC subsets (LT-, $IT90^{+}$, $IT90^{-}$, ST-HSC).
Extended Data Fig. 3	Protein single cell pseudotime analysis.	Kaufmann_Extended_Data_Fig3.tif	(a) Gating strategy for cell cycle phases of 2d with 20 hr EdU pulse cultured CB Lin-. (b,c) CB LT- and ST-HSC analyzed by immunostaining for Ki-67, H4K16ac, PAK4 and DAPI after 72 hrs of culture; plots in (h) show data combined from 3 independent CB pools (Ø50 cells/population and sample). *Scale bar = 4 µm.* (d) Longitudinal confocal imaging of LT-HSC harvested at 5 timepoints of culture in high-cytokine conditions. Two individual LT-HSC are shown per timepoint. *Scale bar = 4 µm*. (e) Individual cell fluorescence intensities (IntDen) of nuclear CDK6, H4K16ac, PAK4 and DAPI across pseudotime for HSC subsets (n=2913). Curve fit was performed with nonlinear regression excluding 1-2% of outliers (fifth order polynomial). (f) Pseudotime was split in early, mid and late fractions to visualize distribution of HSC subsets and Progenitors and marker levels across these fractions for all 3442 cells (histograms) or per population and CB (n=2-3, bar plots). Single cell distribution for each population across pseudotime is also shown in horizontal violin plot. *Mean±SD, two sided t-test (compared to "early"; *p<0.05; **p<0.01; ***p<0.001).*

Extended Data Fig. 4	Pharmacological PAK4 inhibition.	Kaufmann_Extended_Data_Fig4.tif	(a) Brightfield images of HSPC ($CD34^+CD38^-$) and Progenitors ($CD34^+CD38^+$) after 5 d treatment with either DMSO or PF-3758309, reproducible across all independent repetitions (n>9). *scale bar = 100 µm.* (b-e) CB Lin^- cells were treated after 2d of prestimulation with either DMSO or PF-3758309 (PF). Concentration as indicated. After 5d of treatment cells were counted (c, n=2), cell cycle (b, d) combined with CD34 analysis was performed (pool of n=2). (f-g) CB Lin^- cells treated after 2 days prestimulation for 5 d with 0.25 uM PF-3758309 or DMSO equivalent were subjected to AnnexinV and CD34 staining (n=3-4). *Mean±SD; two-sided ratio paired t-test.* (h-j) CB HSPC and Progenitors treated for 4 and 6 d with 0.25 uM PF-3758309 or DMSO equivalent were subjected to cell cycle analysis (h), counting (i: HSPC; j: Progenitors). *Mean±SD, two-sided paired t-test (*p<0.05; **p<0.01; ***p<0.001).* (k) Cell counts of treated HSPC on day of injection (d5 treatment, d7 culture). *Mean±SD.* (l) Secondary LDA (8 wk) in NSG and NSG-SGM3 mice of human $CD45^+$ cells from CB Lin^- cells transplanted NSG grafts (24 wk).
Extended Data Fig. 5	*INKA1*-OE mediated effects *in vitro*.	Kaufmann_Extended_Data_Fig5.tif	(a-c) CTRL and *INKA1*-OE HSPC were sorted (BFP+) d6 post transduction and subjected to immunostaining (n=2) and INKA1 levels were determined (a); *two-sided t-test*. Colocalization was quantified in (b) by Coloc2/ImageJ and Spearman's rank correlation values for single cells with Costes' P-value = 1.00 (***p=0.0041*); and in (c) by proximity ligation assay (PLA*)* with background set according to negative controls (IgG1: anti-PAK4 + rabbit IgG; IgG2 anti-INKA1 + mouse IgG); *two-sided Mann-Whitney-test (b,c; ***p<0.0001).*

			(d) Cell viability of transduced HSPC was determined via dye exclusion (PI, 7AAD; n=9). (e) CTRL and *INKA1*-OE LT- and ST-HSC (n=4-5) were sorted (BFP^+) d6 post transduction and subjected to cell cycle analysis by flow cytometry. (f) Distribution of PAK4 protein levels in nucleus vs. cytoplasm according to confocal analysis of CTRL and *INKA1*-OE LT- and ST-HSC; *two-sided Mann-Whitney-test (***p=0.0005).* (g) Differentially expressed genes with adjusted p-value<*0.01 (according to DESeq2, **p=0.003, ***p-value range: <10^{-4}-10^{-26})* in either LT- or ST-HSC transduced with *INKA1*-OE vs CTRL (log_2FC).
Extended Data Fig. 6	*INKA1*-OE mediated effects in *in vitro* differentiation assays.	Kaufmann_Extended_Data_Fig6.tif	(a) CB LT- and ST-HSC were cultured, transduced (OE) and on day 3 ptd BFP^+ cells were sorted into methylcellulose for colony formation, scored and harvested 12 days later and replated (right) for 12 additional days (n=4). Colony type composition of primary and secondary colony formation assays (CFU-C) is shown in the lower panels; *Mean±SD, two-sided paired t-test.* (b) Total single cell colony output (d15) from stromal coculture assays per CB (n=3 independent pools, *Mean±SD*) and subfraction of CTRL and *INKA1*-OE transduced HSC subsets as indicated. *My= myeloid ($CD33^+$), E= erythroid ($CD45^-GlyA^+/CD71^+$), Mk = megakaryocytic ($CD41^+$), NK = $CD56^+$cells.* Upper right panel shows proliferative potential and lower panel total colony output per assay, population and condition as assessed by flow cytometry; *two-sided paired t-test.*
Extended Data Fig. 7	*INKA1*-OE, shINKA1 and shPAK4 in xenograft assays.	Kaufmann_Extended_Data_Fig7.tif	(a) Gating strategy for BM analysis from xenografts. (b) Human CD45 chimerism in injected femur (right, RF) and non-injected bone marrow of CB HSPC transduced

			with CTRL or *INKA1*-OE at various timepoints post-transplant (5 independent CB pools). Each cell represents individual mouse. (c) Relative engraftment in non-injected BM of transduced cells represented as log_2 ratios (%BFP^+ of $CD45^+$ *in vivo* output /%BFP^+ *in vitro* input) of 5 independent experiments, *two-sided t-test with Welch's correction (**p=0.006 and p=0.008).* (d) Lineage composition within BFP^+ xenografts at 20 wk post xenotransplantation of CTRL and *INKA1*-OE HSPC (5 experiments with 4-9 mice/group each). (e) Flow cytometric analysis of 20 wk BM grafts gated on $CD45^+BFP^+CD34^+CD19^-CD38^-$ population; *Mean±SD.* (f) Individual data point representation related to Figure 5e; *mice per group:10/9(4w) - 5/5/5/5 (12w)- 9/7/9/8 (20w).* (g) Human $CD45^+$ cell counts from injected femurs from NSG mice transplanted with CTRL or *INKA1*-OE HSPC at indicated timepoints and after single dose PBS or 5-FU treatment at 4 wk; *Mean±SEM, two-sided Mann-Whitney-test.* (h) Fold change (FC) of $CD45^+$ counts in injected femur at 12 wk relative to 4wk of CTRL and *INKA1*-OE HSPC transplanted and PBS/5FU treated mice; *Mean±SD, two-sided t-test (*p=0.043).* (i) Injected femurs (20w) were analyzed for %$CD34^+$ (left) and total pooled bone marrow was subjected to hierarchy analysis at 20w of CTRL and INKA1-OE HSPC after treatment with PBS or 5-FU at 4w. Pie chart represents CD34 compartment and radius is scaled to percentage CD34. *Mean±SD, two-sided t-test.* (j) Four shRNAs embedded in lentiviral vectors were tested each, for *INKA1* and *PAK4* knockdown, along with

			shRenilla as control in MOLM13 cells. Transduced cells were sorted (BFP^+) and subjected to qRT-PCR for the targeted transcript. Vector copy number (VCN) was estimated by initial transduction efficiency assuming Poisson distribution to correct for VCN introduced bias in knockdown-efficiency. Highlighted in color are the shRNAs chosen for functional assays; *Mean±SD, 3 technical replicates.* (k) Summary of secondary LDA analysis in NSG-SGM3 mice at 8w of shRNA transduced primary grafts (20w). Mice were considered engrafted at human CD45 chimerism >0.1. (l) Lineage composition of BFP^+ NSG grafts from shRNA transduced HSPC as indicated (20 wk). *Mean±SD.*
Extended Data Fig. 8	INKA1 and PAK4 protein interactions identify SIRT1 as INKA1 interactor.	Kaufmann_Extended_Data_Fig8.tif	(a) RNA expression (GSE125345) of PAK4/INKA1 interactome as identified by BioID across human CB HSPC and Progenitors. (b) Pathway (GO processes) enrichment map (FDR<0.05, selected) of BioID interactome via Cytoscape. (c) Proximity ligation assay of HSPC and Progenitors including staining controls (90 cells, 6 variants as in scheme; 3 independent CB pools). Background was set according to controls. Four PLA signal positive cells are shown enlarged. (d) Shown are confocal images of two separate immunostainings of extended chromatin fibers generated from CB HSPC ($CD34^+CD38^-$) that were either stained for H4K16ac and SIRT1 (left) or SIRT1 and INKA1 (right). *Results were reproducible in 2 independent repetitions. Scale bar =2 µm.* (e) Cultures of transduced (OE) CB HSPC ($CD34^+CD38^-$; n=3) were treated with DMSO or 25 µM EX-527 starting 3 days after transduction. EdU was added on d4 and cells

			were analyzed 16 hrs later by flow cytometry for %EdU^{+} cells and DNA synthesis rate by Edu MFI within transduced/BFP^{+} cells; *Mean±SD, two-sided paired t-test (*p<0.05; **p=0.004).*
Extended Data Fig. 9	INKA1 and PAK4 protein interactions implicate CD112 in HSPC biology.	Kaufmann_Extended_Data_Fig9.tif	(a,b) Absolute and relative *Nectin2* and *AFDN* expression from GSE125345(a) and AFDN expression in mPB scRNAseq dataset across diffusion pseudotime (b). (c) CD112 levels in CB mononuclear cells (MNC) gated on CD34^{+}CD19^{+} and CD34^{+}CD19^{-} subpopulations and CD112 vs CD19 expression in CD34^{+} cells from 20w xenografts. (d,e) Flow cytometric analysis of CD34 and CD112 surface expression on CB Lin- cells and CB MNC. *FMO+IgG = fluorescence minus one staining without anti-CD112-APC but IgG-APC; two-sided t-test (*p<0.05, **p<0.01, ***p<0.001).* (f-g) Single cell index-sorting (indexed intensities: CD49f, CD90 and CD112) of 2800 CB CD34^{+}CD38^{-} CD45RA^{-} cells onto MS5 stromal layers followed by 18 days *in vitro* differentiation and analysis via high-throughput flow cytometry (n=4). LT- and ST-HSCs were retrospectively gated according to initially indexed CD49f and CD90 levels and respective colony output (f) and colony output per individual CB pool is shown in (g). (h) Cultured and transduced CB CD34+CD38- cells were analyzed for CD112 levels 5-6 d post-transduction of subgated cLT-(CD34^{+}CD45RA^{-}CD90^{+}), cST-(CD34^{+}CD45RA^{-}CD90^{-}) HSC, cProgenitors (CD34^{+}CD45RA^{+}) or total cHSPC (CD34^{+}). *c= cultured, two-sided paired t-test (*p<0.05, **p=0.004).* (i) Confocal images of BFP+ sorted CB LT- and ST-HSC 6d after transduction followed by immunostaining (n=3; Ø45 cells/ sample); *two-sided Mann-Whitney-test.*

			(j) Confocal images of immunostaining of CB LT- and ST-HSC (total 74 cells from 3 CB) and thereof derived Spearman's rank correlation values (calculated by Coloc2/ImageJ) for the indicated parameter pairs, only non-nuclear cell area was considered.
Extended Data Fig. 10	CD112 separates primed from non-primed LT-HSC.	Kaufmann_Extended_Data_Fig10.tif	(a,b) Post-sort analysis of LT- and ST-HSC sorted for $CD112^{lo}$ and $CD112^{hi}$ either directly after sorting on the same instrument (a) or in parallel with CellRox staining after 30 min cytokine free culture(b; see g,h) on a FACS analyzer. (c) Representative plots of flow cytometric analysis of LT- and ST-HSC sorted for $CD112^{lo}$ and $CD112^{high}$ directly or after 42 hrs culture for CDK6 and Ki-67. (d) Quantification of flow cytometric analysis of LT- and ST-HSC sorted for $CD112^{lo}$ and $CD112^{hi}$ directly or after 42 hrs culture for CDK6 and Ki-67 (n=3; *paired t-test*). (e-g) LT- and ST-HSC sorted for $CD112^{lo}$ and $CD112^{hi}$ (n=4-7) were analyzed directly via immunostaining for INKA1 (e; total of 1587 single cells from 7 independent CB pools, *whiskers indicate 10-90 percentile, Kruskal-Wallis test*) or for CellROX labeling after 30 min cytokine-free culture and assessed by flow cytometry (f) and fluorescence microscopy (g); *box-and-whisker plots with median as centre and either 10-90% percentile (e) or minimal to maximal values are represented by whiskers (f). Two-sided t-test (f,g; *p<0.05, **p<0.01, ***p<0.001)* (h,i) LT-HSC sorted for $CD112^{lo}$ (n=2) and $CD112^{hi}$ (n=3) were subjected to ATAC-seq. The number of 500bp windows overlapping called peaks exclusively in two $CD112^{hi}$ or $CD112^{lo}$ replicates but not in the other is shown in (h) and (i) is a heatmap of -log P-Value of enrichment for known motifs enriched in either $CD112^{hi}$

			or CD112^{lo}, relative to a background of all called peaks. CD112^{hi} versus CD112^{lo} LT-HSC and INKA1-OE vs CTRL LT-HSC. (k) GSEA results for CD112^{hi} versus CD112^{lo} RNAseq data (3472 GO biological processes; gene set size 25-500; FDR<0.1). *nES = normalized enrichment score; *FDR<0.05; **FDR<0.01*. (l) Expression of leading-edge genes from mTORC1 and GOBP gene sets (2-3 independent CB pools). (m) Lineage composition of transplanted CD112^{lo} and CD112^{hi} LT-HSC. Each column represents an individual mouse. (n) Summary of secondary LDA from 12 wk primary grafts into NSG-SGM3 for additional 8 wk for 3 independent CB pools. Same cell numbers per dose were transplanted from initially transplanted CD112^{lo} and CD112^{hi} LT-HSC per CB but varied for individual CB cohorts.

1

2

Item	Present?	Filename This should be the name the file is saved as when it is uploaded to our system, and should include the file extension. The extension must be .pdf	A brief, numerical description of file contents. i.e.: *Supplementary Figures 1-4, Supplementary Discussion, and Supplementary Tables 1-4.*
Supplementary Information	No		
Reporting Summary	Yes	NI-A30897C_reporting_summary.pdf	

3

4

Type	Number	Filename	Legend or Descriptive Caption

	If there are multiple files of the same type this should be the numerical indicator. i.e. "1" for Video 1, "2" for Video 2, etc.	This should be the name the file is saved as when it is uploaded to our system, and should include the file extension. i.e.: *Smith_ Supplementary_Video_1.mov*	Describe the contents of the file
Supplementary Table	1	Kaufmann_Supplementary_Tables.xlsx	Supplementary Table 1 - High Confidence interactors of INKA1 and PAK4 according to proximity-dependent Biotin Identification (BioID) Assay Supplementary Table 2 - Differential expressed genes (289) between $CD112^{lo}$ and $CD112^{hi}$LT-HSC (statistics calculated by DESeq2). Supplementary Table 3 - Detailed information of commercially available reagents and other materials/resources used.

Parent Figure or Table	Filename This should be the name the file is saved as when it is uploaded to our system, and should include the file extension. i.e.: *Smith_SourceData_Fig1.xls*, or *Smith_ Unmodified_Gels_Fig1.pdf*	Data description i.e.: Unprocessed Western Blots and/or gels, Statistical Source Data, etc.
Source Data Fig. 1	Kaufmann_SourceData_Fig1.xlsx	Statistical Source Data
Source Data Fig. 2	Kaufmann_SourceData_Fig2.xlsx	Statistical Source Data
Source Data Fig. 3	Kaufmann_SourceData_Fig3.xlsx	Statistical Source Data
Source Data Fig. 4	Kaufmann_SourceData_Fig4.xlsx	Statistical Source Data
Source Data Fig. 5	Kaufmann_SourceData_Fig5.xlsx	Statistical Source Data

Source Data Fig. 6	Kaufmann_SourceData_Fig6.xlsx	Statistical Source Data
Source Data Fig. 7	Kaufmann_SourceData_Fig7.xlsx	Statistical Source Data
Source Data Fig. 8	Kaufmann_SourceData_Fig8.xlsx	Statistical Source Data
Source Data Extended Data Fig. 1	Kaufmann_SourceData_Extended_ Data_Fig1.xlsx	Statistical Source Data
Source Data Extended Data Fig. 2	Kaufmann_SourceData_Extended_ Data_Fig2.xlsx	Statistical Source Data
Source Data Extended Data Fig. 3	Kaufmann_SourceData_Extended_ Data_Fig3.xlsx	Statistical Source Data
Source Data Extended Data Fig. 4	Kaufmann_SourceData_Extended_ Data_Fig4.xlsx	Statistical Source Data
Source Data Extended Data Fig. 5	Kaufmann_SourceData_Extended_ Data_Fig5.xlsx	Statistical Source Data
Source Data Extended Data Fig. 6	Kaufmann_SourceData_Extended_ Data_Fig6.xlsx	Statistical Source Data
Source Data Extended Data Fig. 7	Kaufmann_SourceData_Extended_ Data_Fig7.xlsx	Statistical Source Data
Source Data Extended Data Fig. 8	Kaufmann_SourceData_Extended_ Data_Fig8.xlsx	Statistical Source Data
Source Data Extended Data Fig. 9	Kaufmann_SourceData_Extended_ Data_Fig9.xlsx	Statistical Source Data
Source Data Extended Data Fig. 10	Kaufmann_SourceData_Extended_ Data_Fig10.xlsx	Statistical Source Data

7

8
9

A latent subset of human hematopoietic stem cells resists regenerative stress to preserve stemness

Kerstin B. Kaufmann[1], Andy G. X. Zeng[1,2], Etienne Coyaud[1], Laura Garcia-Prat[1], Efthymia Papalexi[3,4], Alex Murison[1], Estelle M. N. Laurent[1], Michelle Chan-Seng-Yue[1,5], Olga I. Gan[1], Kristele Pan[1], Jessica McLeod[1], Héléna Boutzen[1], Sasan Zandi[1,6], Shin-ichiro Takayanagi[1,7], Rahul Satija[3,4], Brian Raught[1,8], Stephanie Z. Xie[1] and John E. Dick[1,2]*

[1] Princess Margaret Cancer Centre; University Health Network; Toronto, ON, M5G 0A3; Canada

[2] Department of Molecular Genetics; University of Toronto; Toronto, ON, M5S 1A8; Canada

[3] New York Genome Center; New York, NY, 10013; USA

[4] Center for Genomics and Systems Biology; New York University; New York, NY, 10003; USA

[5] PanCuRx Translational Research Initiative; Ontario Institute for Cancer Research; Toronto, ON, M5G 0A3; Canada

[6] Department of Laboratory Medicine and Pathobiology; University of Toronto; Toronto, ON, M5S 1A8; Canada

[7] Cell Therapy Project; R&D Division; Kirin Holdings Company, Limited; Tokyo, 194-8533; Japan

[8] Department of Medical Biophysics; University of Toronto; Toronto, ON, M5S 1A4; Canada

*Corresponding author: john.dick@uhnresearch.ca

ABSTRACT

Continuous supply of immune cells throughout life relies on the delicate balance of the hematopoietic stem cell (HSC) pool between long-term maintenance and meeting the demands of both normal blood production and unexpected stress conditions. Here we identified distinct subsets of human long-term (LT-)HSCs that responded differently to regeneration-mediated stress: an immune checkpoint ligand $CD112^{lo}$ subset that exhibited a transient engraftment restraint (termed latency) before contributing to hematopoietic reconstitution and a primed $CD112^{hi}$ subset that responded rapidly. This functional heterogeneity and CD112 are regulated by INKA1, through direct interaction with PAK4 and SIRT1, inducing epigenetic changes and defining an alternative state of LT-HSC quiescence that serves to preserve self-renewal and regenerative capacity upon regeneration-mediated stress. Collectively, our data uncovered the molecular intricacies underlying HSC heterogeneity and self-renewal regulation and point to latency as an orchestrated physiological response that balances blood cell demands with preserving a stem cell reservoir.

INTRODUCTION

Continuous and life-long blood cell production hinges on the ability of rare multipotent LT-HSCs to balance two fundamental stem cell properties: self-renewal versus capacity to generate differentiated progeny. This balance ensures not only continuous hematopoietic output but also long-term maintenance of the LT-HSC pool. Maintenance occurs because HSC reside predominantly in a quiescent state that can last for weeks to years before intrinsic and extrinsic cues activate some LT-HSC to either self-renew or generate short-term (ST-)HSC, while others remain restrained in quiescence[1-3]. However, serial challenges, such as transplantation or inflammation, can lead to HSC attrition, exhaustion and biases in lineage output manifesting as cytopenias, late graft failure following HSC transplantation and compromised immune function upon aging[4-6]. Understanding *in vivo* HSC dynamics and the mechanisms that ensure HSC longevity on a molecular, cellular and population level is crucial for enabling cell therapies that permit both rapid regeneration and sustained long-term hematopoiesis.

Although murine studies provided considerable insights into cellular and molecular mechanisms of HSC quiescence, poor understanding remains how quiescence control integrates with the fate decisions HSCs must make to either remain quiescent, differentiate and/or self-renew[7,8]. In humans, the longer lifespan and massively higher cellular output make dynamic integration of these fundamental stem cell properties even more acute. We particularly lack understanding of the processes underlying the activation and/or resistance of human HSC to adapt to varying hematopoietic demands during homeostasis or emergency hematopoiesis triggered by inflammation, acute blood loss or transplantation. Although human immunophenotypic $CD90^{+}CD49f^{-}$ LT- and $CD90^{-}CD49f^{-}$ ST-HSCs are both multipotent and possess engraftment capacity, ST-HSCs have markedly reduced self-renewal potential and faster quiescence exit kinetics[3,9]. Despite these functional differences between human LT- and ST-HSC subsets, published studies have not uncovered clear transcriptomic differences at the population level or by single cell (sc)RNA-seq and instead pointed to the continuous nature of early hematopoietic commitment[10,11]. Functional screens offer an alternative approach to identify mechanisms that distinguish LT-HSC from

ST-HSC. As normal HSC and leukemia stem cells (LSC) in acute myeloid leukemia (AML) share many similarities in stemness properties, we recently undertook a functional stemness screen of potential LSC regulators. We revealed *INKA1* (also known as *FAM212A, C3orf54*) as a regulator of LSC self-renewal[12]. This study linked human LSC quiescence and latency to INKA1-mediated inhibition of p21(Cdc42/Rac)-activated kinase 4 (PAK4) and to reduction of H4K16 acetylation (H4K16ac). However, this signaling axis was not studied in normal hematopoiesis. PAK4 plays a role in numerous cellular functions [13] and only a single study in human cell lines and a limited number of studies in frogs, fish and mice established that Inka1 is an endogenous inhibitor of Pak4 and essential for neural crest development[14,15]. H4K16ac is associated with activated transcription of single genes, but this chromatin mark is not strongly correlated with changes in global gene expression as assessed in HeLa cells and murine HSC[16,17]. Here, we uncovered a mechanistic link between PAK4 and H4K16ac, and its role in human hematopoiesis, demonstrating that a molecular network with INKA1 at its center governs the long-term regenerative capacity of human HSC. Moreover, we show that human LT-HSC are not homogeneous but composed of two cellular entities distinguished on the basis of CD112 expression, each with distinct alternative states of quiescence and repopulation kinetics.

RESULTS

INKA1 and H4K16ac delineated HSC heterogeneity

To explore the relevance of INKA1 and PAK4 in normal human hematopoiesis, we first interrogated published RNA expression datasets[18,19]. *INKA1* was highly expressed in HSC and progenitor enriched $CD34^+$ cells compared to mature $CD34^-$ populations, whereas *PAK4* showed low expression across the hematopoietic hierarchy (Fig. 1a). Analysis of protein expression and chromatin acetylation by confocal analysis of four $CD34^+CD38^-CD45RA^-$ HSC subsets sub-fractionated by CD90/CD49f expression from cord blood (CB) showed heterogeneity in INKA1 and H4K16ac abundance at single-cell resolution. The majority of cells (85.5±8.2%) in these four populations were high in H4K16ac with low to background signals for nuclear INKA1 ($INKA1^{lo}$; Fig. 1b-d and Extended Data Fig. 1a-c). Cells with

concomitant faint staining for both markers further revealed intranuclear spatial separation when analyzed after conversion into binary images (Extended Data Fig. 1c) suggesting these cells were in transition to either becoming INKA1^{hi} or H4K16ac^{hi} and that these molecules actively excluded one another. High nuclear signal for INKA1 (INKA1^{hi}) was detected in a minor population (10-20%) with small nuclear diameter and low or negative staining for H4K16ac (Fig. 1e and Extended Data Fig. 1d). To determine if INKA1^{hi} cells were in an alternative state of stem cell quiescence, we used CDK6 as a definitive marker of the activated/poised stem cell state[3] and compared INKA1 versus CDK6 content in LT- and ST-HSCs . Immunofluorescence analysis of LT-HSCs showed that INKA1^{hi} cells had significantly lower CDK6 protein abundance than INKA1^{lo} cells; differences not seen in ST-HSCs (Fig. 1f and Extended Data Fig. 1e). Hence, the inverse nuclear staining patterns of INKA1 and H4K16ac delineated at least two distinct populations within LT-HSCs that might be linked to alternative states of stem cell quiescence.

INKA1 and *CDK6/PAK4* segregated quiescent single cells

To gain more insight into the heterogeneous subpopulations within the LT-HSC pool and determine whether they varied in stem cell activation signatures, we undertook single-cell transcriptomic analysis of HSCs from granulocyte-colony-stimulating factor (G-CSF)-mobilized peripheral blood (mPB). Since mPB cells are activated to exit the bone marrow *in vivo*, we hypothesized that they may result in improved transcriptional resolution of HSC activation dynamics. We confirmed that 95.7±2.9% of CD34$^+$CD38$^-$CD45RA HSCs from mPB were in G_0 (Ki67$^-$, Extended Data Fig. 2a) and that phenotypically defined LT- and ST-HSCs recapitulated the differences in CDK6 protein abundance as previously described for CB (Fig. 2a)[3]. Flow cytometric detection of CDK6 further partitioned Ki67$^-$ G_0 cells into canonically quiescent (CDK6$^-$) and primed (CDK6$^+$) fractions (47.8±14.2%; Fig. 2b).

Single-cell index-sorting (signals indexed: CD34, CD38, CD45RA, CD90, CD49f) followed by full-length transcript scRNA-seq of CD34$^+$CD38$^-$CD45RA$^-$ mPB cells (from 3 donors) generated 1191 single-cell transcriptomes. To identify a trajectory of HSC activation, we applied diffusion pseudotime

analysis while leveraging predictions of future cell states through RNA Velocity[20,21](Fig. 2c and Extended Data Fig. 2b-f). We observed a gradual increase in CDK6 expression as well as cell cycle programs[22] from early to late pseudotime, confirming that we captured the transition from canonically quiescent to activated/poised HSC. Superimposing *INKA1* and *PAK4* expression onto the diffusion map revealed that the highest *INKA1* expression occurred early in pseudotime and thereafter declined, while *PAK4* expression increased late in pseudotime (Fig. 2d and Extended Data Fig. 2b). Despite early enrichment of *ITGA6* transcripts (encoding CD49f), surface expression of CD49f and CD90 was homogenous across pseudotime, implying that all 4 HSC subsets followed the same trajectory of activation (Extended Data Fig. 2d-f). Collectively, these data established that LT-HSCs with high *INKA1* are *CDK6*^{lo} and fall on the non-primed end of the quiescence spectrum.

H4K16ac and PAK4 demarcated HSC cell cycle activation

To examine whether H4K16ac and PAK4 played a role in HSC cell cycle priming, we interrogated cell cycle (Ki67) and H4K16ac status after cytokine stimulation by flow cytometry in lineage-depleted CB cells (CB Lin$^-$) and by immunostaining in LT- and ST-HSCs (also including PAK4). H4K16ac was significantly lower in CB Lin$^-$ cells remaining in G_0 following cytokine exposure as compared to those that progressed into cell cycle. The signal peaked in late S-phase and subsequently declined in G_2/M phase (Fig. 2e,f and Extended Data Fig. 3a). In agreement, higher abundance of H4K16ac and PAK4 were observed in Ki67^{hi} vs. Ki67^{lo}LT- and ST-HSCs (Fig. 2g and Extended Data Fig. 3b,c) suggesting upregulation upon cell cycle entry. To further resolve the longitudinal dynamics of H4K16ac and PAK4 expression during quiescence exit and entry into the first cell cycle of the most primitive HSC subsets, we performed multi-timepoint (6-100 h) single-cell protein imaging analysis under high-cytokine culture conditions. We used CDK6 as a G_0 exit marker[3] and CD34$^+$CD38$^+$ progenitors were used as an activated control. We observed a significant increase in the signal intensities for CDK6, PAK4 and H4K16ac at late timepoints (66 and 100 h) when compared to 6 hrs for both LT- and ST-HSCs. However, LT-HSCs showed an overall lower CDK6 abundance than ST-HSC at 66 hrs in

accordance with the expected delay in upregulation of CDK6[3] (Fig. 2h and Extended Data Fig. 3d). We captured a similar delay of PAK4 and H4K16ac upregulation in LT-HSCs as compared to ST-HSC.

To account for the asynchronous cell cycle priming and progression in the imaging study, we applied diffusion pseudotime analysis[21] to our multidimensional single-cell protein imaging data (6 parameters: nuclear intensity for H4K16ac, CDK6, PAK4, DAPI; total cell intensity for CDK6, PAK4; n= 3454 cells). This method ordered single cells according to their phenotype independently of the actual timepoint harvested. Accordingly, at all timepoints of analysis $CD34^+CD38^+$ progenitors as well as HSCs from all four $CD34^+CD38^-CD45RA^-$ subsets intermixed along the generated pseudotime trajectory that showed gradual upregulation of CDK6, H4K16ac and PAK4 along with DAPI/DNA intensity (Fig. 2i and Extended Data Fig. 3e,f). LT-HSCs were enriched earlier in pseudotime, in accordance with their known delayed cell cycle entry, whereas activated progenitor cells were skewed towards the late end of the trajectory (Extended Data Fig. 3f)[3]. The upregulation of PAK4 protein abundance and H4K16 acetylation preceded the upregulation of CDK6 in this single cell immunofluorescence-based pseudotime analysis (Fig. 2i). Overall, these data suggested that PAK4 and H4K16ac might indicate early cell-cycle priming and progression in HSCs.

PAK4-inhibition induced G_0 maintains regenerative potential

As INKA1 is an endogenous PAK4 inhibitor, we used pharmacological inhibition to obtain independent evidence that cell cycle progression depended on PAK4 activity. Such drug studies are difficult to undertake on highly purified HSC subpopulations so we used combined hematopoietic stem and progenitor cells (HSPC) subsets. Treatment of cytokine-stimulated CB $CD34^+CD38^-$ HSPC and $CD34^+CD38^+$ progenitors for 5 days with the PAK4 inhibitor PF-3758309, known to compete with INKA1[14], resulted in a dose-dependent reduction of cell expansion along with an accumulation of G_0 cells ($Ki67^-$; Fig. 3a,b and Extended Data Fig. 4a-d). Cell viability was only mildly reduced and CD34 surface expression was retained (Extended Data Fig. 4e-g). PF-3758309 treatment without prior cytokine-

stimulation had similar effects on cell counts and cell cycle (Extended Data Fig. 4h-j). *Ex vivo* treatment with PF-3758309 followed by limiting-dilution xenograft assays (LDA) in primary engrafted NSG mice did not demonstrate any change in HSC frequency or lineage output compared to DMSO controls (n=6; Fig. 3c,d and Extended Data Fig. 4k). Subsequently, HSC frequency in 4 of the 6 primary xenograft cohorts was assessed by serial LDA using human cytokine (SCF, GM-CSF, IL3) expressing NSG-SGM3 mice as a more sensitive secondary recipient as demonstrated by direct comparison with NSG mice (Extended Data Fig. 4l). We found a trend towards a 2.0-fold increase in HSC frequency following initial PAK4 inhibitor treatment (*p=0.071*, Fig. 3e,f). These results suggested that PAK4 inhibition transiently stalled cytokine-stimulated HSPC in G_0 thereby maintaining their stem cell regenerative capacity.

INKA1-OE conferred selective resistance to activation

To gain more mechanistic insight into how INKA1 functions as an endogenous PAK4 inhibitor within HSC subsets and influences their priming status, we utilized an *INKA1* overexpressing (*INKA1*-OE) lentiviral vector that co-expressed blue fluorescent protein (BFP) to track transduced cells. By focussing on BFP^+ *INKA1*-OE and control transduced CB $CD34^+CD38^-$ cells, we established that INKA1 protein was elevated and that PAK4 and INKA1 were colocalized (Extended Data Fig. 5a-d). *INKA1*-OE in LT-HSCs, but not in ST-HSCs, reduced CDK6 intensity, H4K16ac and the relative abundance of PAK4 in the nucleus and increased the percentage of cells retained in G_0 (2.6±1.5 fold; n=5; Fig. 4a,b and Extended Data Fig. 5e,f) compared to controls. Gene set enrichment analysis (GSEA) of RNA-seq data further showed that *INKA1*-OE vs. controls preserved established HSC stemness signatures and induced a transcriptionally suppressed cell state in LT-HSCs, as evidenced by negative enrichment of multiple stem cell associated and inflammatory signaling pathways, stress responses, and cell cycle related gene sets (Fig. 4c,d and Extended Data Fig. 5g). To assess whether *INKA1*-OE interfered with HSC function we performed both single cell stroma-based differentiation assays and colony-forming assays (CFU-C). *INKA1*-OE did not change lineage outcome, however, *INKA1*-OE in LT-HSCs resulted in decreased primary CFU-C output without differences in secondary replating assays (Extended Data Fig. 6a,b).

Collectively, these data demonstrated that *INKA1*-OE functioned selectively in LT-HSCs by suppressing stem cell activation programs and inducing a transient restraint in colony formation without affecting their potential to generate multilineage hematopoietic progeny.

INKA1-OE and _PAK4_-KD transiently restrained repopulation

To directly test the functional impact that *INKA1*-OE might have on the *in vivo* properties of HSCs we utilized the xenograft assay. We previously observed that *INKA1*-OE in CB $CD34^+CD38^-$ HSPCs induced delayed repopulation, indicative of a transient proliferative disadvantage of the transduced (BFP^+) fraction at the 4 wk timepoint that is overcome by 20 wk[12]. To explore the effects of INKA1 on stemness maintenance and activation, we carried out a more detailed analysis of CB HSPCs transduced with *INKA1*-OE or control vectors using serial xenotransplantation assays. We observed no differences in relative engraftment in the injected femurs at 2 wk (Fig. 5a and Extended Data Fig. 7a-c), whereas *INKA1*-OE resulted in a relative reduction of 4-fold (log_2BFPratio=2) of gene-marked (BFP^+) cells at 4 wk, a timepoint when the majority of LT-HSCs are normally activated and cycling[3]. Examination of primary xenografts at 20 wk showed that the relative engraftment was comparable in both groups with no overt lineage skewing (Extended Data Fig. 7d). Secondary transplantation into NSG-SGM3 recipients was used to model successive rounds of stem cell activation and forced stress hematopoiesis due to the increased human pro-inflammatory cytokine exposure in NSG-SGM3 mice. When 2 wk primary samples were transplanted, only *INKA1*-OE cells engrafted (Fig. 5b). Similarly, *INKA1*-OE cells isolated from 4 wk primary xenografts generated grafts in all secondary recipients, indicative of self-renewal, while only half of control recipients were engrafted (Fig. 5b). Analysis of 20 wk primary xenografts revealed a dissociation between HSC function and immunophenotype, the *INKA1*-OE group had a 3.6-fold reduction of phenotypic LT-HSCs but a 4.5-fold increase in the frequency of functional HSCs as measured by serial LDA, as compared to controls (Fig. 5b-e and Extended Data Fig. 7e). Collectively, these data indicated that the transient restraint induced by *INKA1*-OE at early timepoints

allowed them to retain their self-renewal and regenerative capacity at later timepoints despite their exposure to regeneration-mediated stress.

As *INKA1*-OE cells exhibited transient restraint during regeneration, we investigated whether the restrained HSCs were cycling more slowly *in vivo*. Xenotransplanted NSG mice were treated after 4 weeks of repopulation with 5-Fluorouracil (5-FU), a treatment known to deplete proliferating cells while inducing quiescent HSCs to become activated[23]. A single dose of 5-FU resulted in higher total $CD45^+$ and BFP^+ cells counts in the *INKA1*-OE group at the 12-wk timepoint and a slight decline at 20 wk compared to 5-FU treated controls (Fig. 5e and Extended Data Fig. 7g,h). There was an expansion of the phenotypic HSC compartment of 5-FU treated, *INKA1*-OE mice, but depletion in control groups (Fig. 5f and Extended Data Fig. 7i). Serial LDA analysis showed higher HSC frequency within 5-FU versus PBS-treated *INKA1*-OE grafts, while repopulation in 5-FU treated control groups was abolished (Fig. 5g,h). These data established that the transient restraint enforced by INKA1 was linked to reduced cycling *in vivo*. In addition to *INKA1*-OE protection from 5-FU-mediated stem cell depletion, exposure of *INKA1*-OE cells to a second challenge (5-FU) during the period of restrained regeneration resulted in activation of HSCs and stem cell expansion without exhaustion of self-renewal potential.

To evaluate whether HSPC engraftment capacity depended on INKA1 and was mediated via PAK4 inhibition, we transduced CB $CD34^+CD38^-$ HSPCs either with control, two *INKA1*-knockdown (KD) and two *PAK4*-KD vectors and performed *in vivo* assays in analogy to *INKA1*-OE studies. *INKA1*-KD resulted in 2-fold reduction in the relative proportion of BFP^+ cells at 4 wk and 20 wk in primary recipients (Fig. 5i and Extended Data Fig. 7j). Upon secondary transplantation, the residual cells harvested at 20 wk and injected at the same cell doses performed as poorly as controls (Fig. 5j). By contrast, both of the *PAK4*-KD groups recapitulated the *INKA1*-OE phenotype, showing delayed repopulation in primary xenograft assays and increased HSC frequency upon secondary transplantation, with the shPAK4-3 cohort showing a 7.6-fold increase (Fig. 5i,j and Extended Data Fig. 7k). *PAK4*-KD

resulted in a higher abundance of lymphoid cells at the expense of myeloid and erythroid cells at 20 wk (Extended Data Fig. 7l). This result contrasts with the balanced output generated upon *INKA1*-OE, further implicating an additional INKA1 function beyond simply inhibiting PAK4. Collectively, *INKA1*-OE or *PAK4*-KD induced a reversible restraint of hematopoietic output that was coincident with enhancement of their long-term self-renewal and regenerative potential.

Interactome linked INKA1 via SIRT1 to H4K16 deacetylation

To gain further mechanistic insight into how INKA1 and PAK4 function and how they are linked to H4K16 acetylation, we studied their broader interactome via proximity-dependent biotin labelling (BioID)[24]. We generated inducible N- and C-terminal BirA* (BirA R118G)-fusions for both proteins. 87 high confidence interactors were identified that were either biotinylated by INKA1 fusions only (34), PAK4 fusions only (30), or both (23)(Fig. 6a, Supplementary Table 1). The transcripts of the majority of these retrieved proteins were also found highly expressed in CB LT-/ST-HSCs (GSE125345; Fig. 6a and Extended Data Fig. 8a).

Pathway enrichment analysis of the interactome confirmed enrichment for GTPase signaling, cell polarity, protein localization, signal transduction, cell cycle and regulation of lysine acetylation (Extended Data Fig. 8b). One of the INKA1 interactors was the NAD^+-dependent deacetylase SIRT1, a known regulator of many biological processes that functions through targeting H4K16ac[25] and other histone and non-histone proteins for deacetylation. We confirmed nuclear and cytoplasmic INKA1 interaction with SIRT1 in CB $CD34^+CD38^-$ cells (Fig. 6b and Extended Data Fig. 8c). Staining of chromatin fibers from $CD34^+CD38^-$ cells further confirmed chromatin association of SIRT1 and INKA1 that was inverse to H4K16ac (Extended Data Fig. 8d). *INKA1*-OE in $CD34^+CD38^-$ cells showed increased relative redistribution of SIRT1 into the nucleus, while pharmacological SIRT1 inhibition with EX-527 accelerated DNA synthesis and partly rescued reduced EdU incorporation upon *INKA1*-OE (Fig. 6c and Extended Data Fig. 8e). Thus, the interaction of INKA1 with SIRT1 provided an explanation both for the

inverse staining patterns of INKA1 and H4K16ac in HSC subsets and the suppression of H4K16ac upon *INKA1*-OE. Broadly, the regulation of SIRT1 localization and activity by INKA1 reveals a novel function beyond PAK4 inhibition.

INKA1/PAK4 interactome identified CD112 as an HSPC marker

Among other high confidence INKA1/PAK4 interactors we found an adhesion and immune regulatory molecule, CD112 (PVRL2, Nectin-2; Fig. 6a) and a protein connecting Nectins to the actin cytoskeleton, MLLT4 (Afadin). *Nectin2* (transcript for CD112) and *Afadin* showed a very distinct HSC/HSPC-specific transcript expression pattern (Extended Data Fig. 9a,b). CD112 and CD34 exhibited similar surface expression on CB mononuclear cells (MNC) although CD112 did not capture lymphoid-primed $CD19^{+}CD34^{+}$ cells (Fig. 6d,e and Extended Data Fig. 9c,d). Across the $CD34^{+}$ hierarchy, CD112 and CD34 expression was consistently higher in the more primitive fractions, with the exception of lymphoid-primed progenitors (MLP), which was $CD112^{lo}CD34^{hi}$ (Extended Data Fig. 9e). In stromal differentiation assays we observed no lineage restriction for single cells ($CD34^{+}CD38^{-}CD45RA^{-}$) with either low or high CD112 expression according to initial index-sorting (Extended Data Fig. 9f,g).

To validate the BioID results, we tested the impact of *INKA1*-OE on surface expression of CD112. *INKA1*-OE reduced CD112 surface expression on culture-derived LT-HSCs (cLT-HSCs)[19] and cST-HSCs, although total protein abundance was not altered (Fig. 6f and Extended Data Fig. 9h,i). We captured individual cST-HSCs with extremely high intracellular CD112 that were almost exclusive to *INKA1*-OE suggesting induction of receptor internalization (Extended Data Fig. 9i). Using the late endosome/lysosome marker LAMP1, we confirmed cytoplasmic colocalization with CD112 and INKA1 in LT- and ST-HSCs (Extended Data Fig. 9j). Extracellular CD112 immunostaining indicated that cells with high expression of CD112 on the surface are $INKA1^{lo}$ and vice versa (Fig. 6g). Finally, our scRNA-seq pseudotime analysis provided independent support for this inverse relationship; *INKA1* expression was enriched early and CD112/*Nectin2* was enriched late in pseudotime (Fig. 6h). Together, these data

support a direct functional connection between INKA1 and CD112 and predict that they may mark alternative states of stem cell quiescence, with primed LT-HSCs being *INKA1*^{lo} and CD112/*Nectin2*^{hi}.

CD112^{hi}/CD112^{lo} marked alternative states of LT-HSCs

To determine if the relative abundance of CD112 can discriminate primed (high in CDK6 and H4K16ac) from non-primed (INKA1^{hi}) cells, we sorted CD112^{hi} and CD112^{lo} cell fractions from purified LT-HSC and ST-HSC populations, that are >90% in G_0[3] (Fig. 7a and Extended Data Fig. 10a,b). Cell cycle analysis revealed that the CD112^{hi}LT-HSC fraction had less non-primed (%CDK6^{lo}Ki67^{-}) cells and less cells in G_0 vs. CD112^{lo} LT-HSCs (91.1±5.3% vs. 95.9±2.2% Ki67^{-}; Fig. 7b and Extended Data Fig. 10c). These differences were still evident upon short-term culture (42 hrs; Extended Data Fig. 10c,d). In accordance, CD112^{hi}LT-HSCs exhibited the highest acetylation of H4K16 and the lowest proportion of INKA1^{hi} cells compared to the other 3 populations, whereas CD112^{lo}LT-HSCs showed the highest INKA1 abundance (Fig. 7c,d and Extended Data Fig. 10e). No significant differences were seen between CD112^{hi} and CD112^{lo}ST-HSCs for all parameters assessed (Fig. 7b-e and Extended Data Fig. 10e-g). These data suggest that CD112 enables prospective enrichment of LT-HSCs in alternative states of quiescence in a context specific manner reminiscent of the LT-HSC specific phenotypes induced by *INKA1*-OE.

Quiescent LT-HSCs generate energy preferentially by glycolysis, in contrast to activated HSPC that rely on oxidative phosphorylation that yields reactive oxygen species (ROS)[5]. CD112^{hi}LT-HSC and both ST-HSC-derived populations showed similar ROS signals (CellROX; Fig. 7e and Extended Data 10f,g). By contrast CD112^{lo}LT-HSCs produced less ROS indicating that CD112^{lo}LT-HSCs are also metabolically less primed for activation. We have recently established chromatin accessibility signatures of quiescent (LT/HSPC signature) vs. activated states (Act/HSPC signature) of human HSPCs[26]. We generated chromatin accessibility signatures of CD112^{hi} and CD112^{lo}LT-HSCs by ATAC-seq and projected them onto our published single cell ATAC-seq data from CB HSPCs[26]. The CD112^{lo} signature

was highly concordant with the LT/HSPC signature and inversely correlated with the Act/HSPC signature (Fig. 7f,g and Extended Data Figure 10h,i). By contrast, the more primed, CD112^{hi} population generated an epigenetic signature that overlapped with both established signatures. Hence, CD112 surface expression also delineated alternative chromatin states within quiescent LT-HSCs. These distinct states were also revealed from GSEA of RNA-seq data comparing CD112^{hi} and CD112^{lo}LT-HSC. CD112^{hi}LT-HSC upregulated several prototypical stem cell response pathways that were downregulated in LT-HSC upon *INKA1*-OE (Extended Data Fig. 10j) or included mTORC1 signaling (Fig. 7h)[27]. Overall, we retrieved 289 differentially expressed genes between these two distinctly quiescent LT-HSC subsets (Fig. 7i and Supplementary Table 2). Further GSEA analysis revealed that CD112^{lo} cells were enriched for the highly interconnected biological processes of cell adhesion, cytoskeleton regulation and GTPase signaling (Fig. 7j and Extended Data Fig. 10k,l). In summary, fractionation of LT-HSCs on the basis of CD112 expression identified pre-existing, molecularly-distinct subsets that fall on opposite ends of the quiescence spectrum, thereby revealing previously undescribed heterogeneity within the LT-HSC subset.

CD112^{lo} captured LT-HSC with delayed repopulation kinetics

To determine whether the quiescence-associated phenotype of CD112^{lo}LT-HSCs is functional and associated with delayed activation and post-transplant stem cell latency similar to what was observed with *INKA1*-OE, we assessed the repopulation kinetics of CD112^{hi} and CD112^{lo}LT-HSCs from 3 CB pools. Human CD45 chimerism in mouse cohorts transplanted with CD112^{lo} LT-HSCs tended to be lower than that of their CD112^{hi}LT-HSC-transplanted counterparts at 4 wk (*p=0.102*; Fig. 8a,b) and reaching significance at 12 wk (*p=0.041*). The percentage of CD45^{+} cells for both LT-HSC subsets increased from 4 to 12 wk resulting in high median engraftment (>60% for 2 of 3 CB). Median engraftment for CD112^{lo}LT-HSC remained stable thereafter to 20 wk, whereas it significantly decreased from 12 to 20 wk for CD112^{hi}LT-HSC. Similar to *INKA1*-OE experiments, no overt mature lineage shift was detected (Extended Data Fig. 10m). In secondary LDA, CD112^{lo}LT-HSCs from all 3 CB pools had a combined 3.6-fold higher HSC frequency compared to their CD112^{hi}LT-HSC counterparts (Fig. 7c,d and Extended

Data Fig. 10n). These data suggested that CD112^{hi}LT-HSCs were primed for activation and able to generate hematopoietic output more rapidly; however, they subsequently showed signs of exhaustion in the primary recipients and possessed limited secondary transplantation capacity. By contrast, the delayed repopulation (latency) of low-primed (CD112^{lo}) LT-HSCs was associated with superior retention of functional HSC, recapitulating the *in vivo* phenotype induced by both *INKA1*-OE or *PAK4*-KD. Thus, these data established the human LT-HSC pool was composed of at least two distinct cellular subsets marked by alternative states of quiescence. In the face of regeneration-mediated stress, these subsets showed differing physiological responses, one subset met acute demand while the other preserved stemness.

DISCUSSION

Our study provides key insights into the molecular underpinnings orchestrating the complex *in vivo* dynamics that human HSCs undergo during transplantation to permit rapid hematopoietic repopulation while also ensuring preservation of a stem cell reservoir. We identified functionally, transcriptionally and epigenetically distinct LT-HSC subsets that are positioned along a spectrum of varying depths of G_0 quiescence. The distinct G_0 states are defined by dichotomy between INKA1/SIRT1 and H4K16ac and along a spectrum of PAK4, CDK6 and CD112 protein expression; a molecular network directly linking cell surface to chromatin state. The G_0 state that a LT-HSC resides in determines its long-term fate upon regeneration-mediated challenge. Activated LT-HSCs contribute rapidly to repopulation but lose self-renewal capacity and exhaust, while the most quiescent LT-HSCs exhibit latency in their repopulation kinetics but retain self-renewal and long-term regenerative capacity. The latency phenotype we uncovered illustrates the complexity of human HSC dynamics *in vivo*, and argues that latency, in addition to self-renewal, quiescence and differentiation, is a fundamental stem cell hallmark that preserves HSC pool size during regeneration-mediated demand.

Our discovery of functionally-distinct LT-HSC subsets is consistent with HSC clonal contribution dynamics from post-transplant gene therapy patients: Initial reconstitution occurs from transient clones, while long-term hematopoiesis originates from clones that only become activated after a 1-2-year latency phase[2,28]. Our genetic models, combining *ex vivo* lentiviral transduction with xenografting, are analogous to gene therapy protocols and to previous clonal tracking studies of mice and xenografts[8,29] that pointed to functional heterogeneity manifesting in two engraftment waves. Our study establishes that these distinct fates are neither stochastic nor induced by *ex vivo* culture but wired into distinct molecular states of human LT-HSCs. These can be prospectively isolated into latent ($CD112^{lo}$) and primed ($CD112^{hi}$) LT-HSC subsets even before exposure to regeneration-mediated stress. The resistance of latent LT-HSCs ($CD112^{lo}$ or *INKA1*-OE) to activation early during reconstitution highlights that the regeneration-mediated stress imposed by transplantation is in itself not a sufficient stressor to activate latent LT-HSC. However, certain stimuli including longer times of repopulation, secondary transplantation or 5-FU treatment lift this restraint and trigger self-renewal of latent LT-HSC. All these stimuli increase demand on ST-HSCs to generate continuous hematopoietic output leading to their exhaustion[9,23]. Thus, we speculate that this loss of immediate progeny either triggers direct activation of latent LT-HSC or lifts their restraint; a concept predicted by murine *in vivo* dynamics modeling[30]. The transient lineage output restraint we observed agrees with reported properties of murine low-output HSCs[8] but contrasts with transient myeloid-restricted "latent-HSCs" from aged mice[31]. Overall, these studies suggest that stimuli stronger than those experienced during primary transplantation might be needed to read out full stem cell potential.

While our earlier study[3] found differences in G_0 exit kinetics between LT- and ST-HSCs, here we uncovered distinct quiescence states prior to LT-HSC activation. The increased purity achieved by CD112 enabled detection of changes in mTORC activity along the trajectory of LT-HSC cell priming that were previously obscured if only "bulk" of LT- versus ST-HSCs were compared[3]. Our findings are reminiscent to the mTorc1 activation properties of G_{Alert} muscle stem cells that are primed to respond rapidly to

stress[27]. Embryonic diapause is a reversible H4K16ac-negative dormant state that requires mTOR signaling and the H4K16 acetyltransferase Mof for activation[32,33]. Thus, the molecular quiescence and priming programs we have uncovered in LT-HSC might be conserved across some species and tissues. The distinct latent and primed molecular signatures we ascribed to distinct CD112 immunophenotypes shed light on RhoGTPase signaling, stress responses and cell adhesion; processes all found relevant for murine HSC function[34,35]. CD112/PVRL2 is a ligand for the immune checkpoint receptors TIGIT and CD226[36] suggesting that cross talk between HSCs and T- and NK-cells might play a role in quiescence induction, HSC recruitment and/or immune surveillance[37]. Furthermore, CD112 upregulation upon cell cycling is similar to MHC-I in tissue stem cells[38] suggesting a role in quality control ensuring HSC pool integrity given their inefficient genomic repair during dormancy[39]. Conversely, INKA1-mediated CD112 downregulation may facilitate immune privilege of quiescent HSCs[40].

Identifying the H4K16 deacetylase SIRT1 as an INKA1 interactor suggests a mechanism explaining the dichotomy between INKA1 and H4K16ac. SIRT1 is a stress sensor that regulates the cellular adaptation to metabolic stress while also exerting anti-inflammatory activity via negative regulation of mTOR[41,42]. H4K16ac, in addition to deacetylation by SIRT1, is also regulated by pH changes, depends on acetyl-CoA and is involved in the regulation of DNA damage response during the cell cycle[17,43,44]. Thus, H4K16ac may indicate the metabolic changes occurring upon HSC activation and cell cycle progression while also ensuring HSC quality control. The molecular interplay that arises from the complex INKA1-interactome we identified, points to additional mechanistic roles for INKA1 beyond PAK4 inhibition and emphasizes the need for studies across all blood cell lineages and other tissue hierarchies.

Collectively, our work has uncovered a potential pivot point for the balance between varying hematopoietic demands and long-term sustainability of the hematopoietic system and identified latency as a fundamental property of stem cells that integrates quiescence control with HSC fate decisions.

REFERENCES

1 Bernitz, J. M., Kim, H. S., MacArthur, B., Sieburg, H. & Moore, K. Hematopoietic Stem Cells Count and Remember Self-Renewal Divisions. *Cell* **167**, 1296-1309 e1210, doi:10.1016/j.cell.2016.10.022 (2016).

2 Scala, S. *et al.* Dynamics of genetically engineered hematopoietic stem and progenitor cells after autologous transplantation in humans. *Nat Med* **24**, 1683-1690, doi:10.1038/s41591-018-0195-3 (2018).

3 Laurenti, E. *et al.* CDK6 levels regulate quiescence exit in human hematopoietic stem cells. *Cell Stem Cell* **16**, 302-313, doi:10.1016/j.stem.2015.01.017 (2015).

4 Takizawa, H. *et al.* Pathogen-Induced TLR4-TRIF Innate Immune Signaling in Hematopoietic Stem Cells Promotes Proliferation but Reduces Competitive Fitness. *Cell Stem Cell* **21**, 225-240 e225, doi:10.1016/j.stem.2017.06.013 (2017).

5 Walter, D. *et al.* Exit from dormancy provokes DNA-damage-induced attrition in haematopoietic stem cells. *Nature* **520**, 549-552, doi:10.1038/nature14131 (2015).

6 Muller, L., Di Benedetto, S. & Pawelec, G. The Immune System and Its Dysregulation with Aging. *Subcell Biochem* **91**, 21-43, doi:10.1007/978-981-13-3681-2_2 (2019).

7 Cabezas-Wallscheid, N. *et al.* Vitamin A-Retinoic Acid Signaling Regulates Hematopoietic Stem Cell Dormancy. *Cell* **169**, 807-823 e819, doi:10.1016/j.cell.2017.04.018 (2017).

8 Rodriguez-Fraticelli, A. E. *et al.* Single-cell lineage tracing unveils a role for TCF15 in haematopoiesis. *Nature* **583**, 585-589, doi:10.1038/s41586-020-2503-6 (2020).

9 Notta, F. *et al.* Isolation of single human hematopoietic stem cells capable of long-term multilineage engraftment. *Science* **333**, 218-221, doi:10.1126/science.1201219 (2011).

10 Velten, L. *et al.* Human haematopoietic stem cell lineage commitment is a continuous process. *Nat Cell Biol* **19**, 271-281, doi:10.1038/ncb3493 (2017).

11 Laurenti, E. *et al.* The transcriptional architecture of early human hematopoiesis identifies multilevel control of lymphoid commitment. *Nat Immunol* **14**, 756-763, doi:10.1038/ni.2615 (2013).

12 Kaufmann, K. B. *et al.* A stemness screen reveals C3orf54/INKA1 as a promoter of human leukemia stem cell latency. *Blood* **133**, 2198-2211, doi:10.1182/blood-2018-10-881441 (2019).

13 Dart, A. E. & Wells, C. M. P21-activated kinase 4--not just one of the PAK. *Eur J Cell Biol* **92**, 129-138, doi:10.1016/j.ejcb.2013.03.002 (2013).

14 Baskaran, Y. *et al.* An in cellulo-derived structure of PAK4 in complex with its inhibitor Inka1. *Nat Commun* **6**, 8681, doi:10.1038/ncomms9681 (2015).

15 Luo, T. *et al.* Inca: a novel p21-activated kinase-associated protein required for cranial neural crest development. *Development* **134**, 1279-1289, doi:10.1242/dev.02813 (2007).

16 Grigoryan, A. *et al.* LaminA/C regulates epigenetic and chromatin architecture changes upon aging of hematopoietic stem cells. *Genome Biol* **19**, 189, doi:10.1186/s13059-018-1557-3 (2018).

17 McBrian, M. A. *et al.* Histone acetylation regulates intracellular pH. *Mol Cell* **49**, 310-321, doi:10.1016/j.molcel.2012.10.025 (2013).

18 Hay, S. B., Ferchen, K., Chetal, K., Grimes, H. L. & Salomonis, N. The Human Cell Atlas bone marrow single-cell interactive web portal. *Exp Hematol* **68**, 51-61, doi:10.1016/j.exphem.2018.09.004 (2018).

19 Xie, S. Z. *et al.* Sphingolipid Modulation Activates Proteostasis Programs to Govern Human Hematopoietic Stem Cell Self-Renewal. *Cell Stem Cell* **25**, 639-653 e637, doi:10.1016/j.stem.2019.09.008 (2019).

20 La Manno, G. *et al.* RNA velocity of single cells. *Nature* **560**, 494-498, doi:10.1038/s41586-018-0414-6 (2018).

21 Haghverdi, L., Buttner, M., Wolf, F. A., Buettner, F. & Theis, F. J. Diffusion pseudotime robustly reconstructs lineage branching. *Nat Methods* **13**, 845-848, doi:10.1038/nmeth.3971 (2016).

22 Tirosh, I. *et al.* Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **352**, 189-196, doi:10.1126/science.aad0501 (2016).

23 Randall, T. D. & Weissman, I. L. Phenotypic and functional changes induced at the clonal level in hematopoietic stem cells after 5-fluorouracil treatment. *Blood* **89**, 3596-3606 (1997).

24 Coyaud, E. *et al.* Global Interactomics Uncovers Extensive Organellar Targeting by Zika Virus. *Mol Cell Proteomics* **17**, 2242-2255, doi:10.1074/mcp.TIR118.000800 (2018).

25 Chen, Y. *et al.* Quantitative acetylome analysis reveals the roles of SIRT1 in regulating diverse substrates and cellular pathways. *Mol Cell Proteomics* **11**, 1048-1062, doi:10.1074/mcp.M112.019547 (2012).

26 Takayama, N. *et al.* The Transition from Quiescent to Activated States in Human Hematopoietic Stem Cells Is Governed by Dynamic 3D Genome Reorganization. *Cell Stem Cell* **28**, 488-501, doi:10.1016/j.stem.2020.11.001 (2021).

27 Rodgers, J. T. *et al.* mTORC1 controls the adaptive transition of quiescent stem cells from G0 to G(Alert). *Nature* **510**, 393-396, doi:10.1038/nature13255 (2014).

28 Biasco, L. *et al.* In Vivo Tracking of Human Hematopoiesis Reveals Patterns of Clonal Dynamics during Early and Steady-State Reconstitution Phases. *Cell Stem Cell* **19**, 107-119, doi:10.1016/j.stem.2016.04.016 (2016).

29 Belderbos, M. E. *et al.* Donor-to-Donor Heterogeneity in the Clonal Dynamics of Transplanted HumanCord Blood Stem Cellsin Murine Xenografts. *Biol Blood Marrow Transplant* **26**, 16-25, doi:10.1016/j.bbmt.2019.08.026 (2020).

30 Klose, M., Florian, M. C., Gerbaulet, A., Geiger, H. & Glauche, I. Hematopoietic Stem Cell Dynamics Are Regulated by Progenitor Demand: Lessons from a Quantitative Modeling Approach. *Stem Cells* **37**, 948-957, doi:10.1002/stem.3005 (2019).

31 Yamamoto, R. *et al.* Large-Scale Clonal Analysis Resolves Aging of the Mouse Hematopoietic Stem Cell Compartment. *Cell Stem Cell* **22**, 600-607 e604, doi:10.1016/j.stem.2018.03.013 (2018).

32 Hussein, A. M. *et al.* Metabolic Control over mTOR-Dependent Diapause-like State. *Dev Cell* **52**, 236-250 e237, doi:10.1016/j.devcel.2019.12.018 (2020).

33 Khoa, L. T. P. *et al.* Histone Acetyltransferase MOF Blocks Acquisition of Quiescence in Ground-State ESCs through Activating Fatty Acid Oxidation. *Cell Stem Cell* **27**, 441-458, doi:10.1016/j.stem.2020.06.005 (2020).

34 Reddy, P. N. *et al.* p21-activated kinase 2 regulates HSPC cytoskeleton, migration, and homing via CDC42 activation and interaction with beta-Pix. *Blood* **127**, 1967-1975, doi:10.1182/blood-2016-01-693572 (2016).

35 Yang, L. *et al.* Rho GTPase Cdc42 coordinates hematopoietic stem cell quiescence and niche interaction in the bone marrow. *Proc Natl Acad Sci U S A* **104**, 5091-5096, doi:10.1073/pnas.0610819104 (2007).

36 Andrews, L. P., Yano, H. & Vignali, D. A. A. Inhibitory receptors and ligands beyond PD-1, PD-L1 and CTLA-4: breakthroughs or backups. *Nat Immunol* **20**, 1425-1434, doi:10.1038/s41590-019-0512-0 (2019).

37 Kalia, V., Penny, L. A., Yuzefpolskiy, Y., Baumann, F. M. & Sarkar, S. Quiescence of Memory CD8(+) T Cells Is Mediated by Regulatory T Cells through Inhibitory Receptor CTLA-4. *Immunity* **42**, 1116-1129, doi:10.1016/j.immuni.2015.05.023 (2015).

38 Agudo, J. *et al.* Quiescent Tissue Stem Cells Evade Immune Surveillance. *Immunity* **48**, 271-285 e275, doi:10.1016/j.immuni.2018.02.001 (2018).

39 Beerman, I., Seita, J., Inlay, M. A., Weissman, I. L. & Rossi, D. J. Quiescent hematopoietic stem cells accumulate DNA damage during aging that is repaired upon entry into cell cycle. *Cell Stem Cell* **15**, 37-50, doi:10.1016/j.stem.2014.04.016 (2014).

40 Hirata, Y. *et al.* CD150(high) Bone Marrow Tregs Maintain Hematopoietic Stem Cell Quiescence and Immune Privilege via Adenosine. *Cell Stem Cell* **22**, 445-453 e445, doi:10.1016/j.stem.2018.01.017 (2018).

41 Ryall, J. G. *et al.* The NAD(+)-dependent SIRT1 deacetylase translates a metabolic switch into regulatory epigenetics in skeletal muscle stem cells. *Cell Stem Cell* **16**, 171-183, doi:10.1016/j.stem.2014.12.004 (2015).

42 Wang, Y. *et al.* Histone Deacetylase SIRT1 Negatively Regulates the Differentiation of Interleukin-9-Producing CD4(+) T Cells. *Immunity* **44**, 1337-1349, doi:10.1016/j.immuni.2016.05.009 (2016).

43 Cluntun, A. A. *et al.* The rate of glycolysis quantitatively mediates specific histone acetylation sites. *Cancer Metab* **3**, 10, doi:10.1186/s40170-015-0135-3 (2015).

44 Sharma, G. G. *et al.* MOF and histone H4 acetylation at lysine 16 are critical for DNA damage response and double-strand break repair. *Mol Cell Biol* **30**, 3582-3595, doi:10.1128/MCB.01476-09 (2010).

Acknowledgments

We thank the obstetrics units of Trillium Health and William Osler Health Partners for CB; The UHN-SickKids Flow cytometry facility for cell sorting; M. Minden and the Leukemia Tissue Bank at Princess Margaret Cancer Centre/University Health Network for providing primary mPB samples. We thank E. Laurenti and all members of the Dick lab, in particular J.C.Y. Wang, for critical feedback. This research was supported by the Deutsche Forschungsgemeinschaft (K.B.K). and is part of the University of Toronto's Medicine by Design initiative which receives funding from the Canada First Research Excellence Fund. Work in J.E.D.'s laboratory is supported by funds from the Princess Margaret Cancer Centre Foundation, the Canadian Institutes of Health Research (Foundation: 154293 (J.E.D), Operating

Grants # 154293 and # 89932 (J.E.D), International Development Research Centre, Canadian Cancer Society (Grant #703212 (J.E.D)), Terry Fox Research Institute Program Project Grant, Ontario Institute for Cancer Research through funding provided by the Government of Ontario, a Canada Research Chair, and the Ontario Ministry of Health and Long Term Care.

Author Contributions

K.B.K. conceived the study, performed research, analyzed data, and wrote the manuscript; A.G.X.Z., A.M. and M.C.S.Y. analyzed data; E.C., E.P. and S.Z.X. performed research and analyzed data; L.G.P., E.M.N.L., O.I.G., K.P., J.M., H.B., S.Z. and S.T. performed research; R.S. and B.R. supervised specific experiments; A.G.X.Z., O.I.G and S.Z.X. edited the manuscript; S.Z.X. provided conceptual input; J.E.D. wrote the manuscript, secured funding and supervised the study.

Competing Interests Statement

J.E. Dick served on the SAB at Trillium Therapeutics, reports receiving a commercial research grant from Celgene, and has ownership interest (including patents) in Trillium Therapeutics Inc.; S.T. is employed by Kirin Holdings Company, Ltd.; all other authors declare no competing interests.

FIGURE LEGENDS

Fig. 1: INKA1 and H4K16ac show inverse abundance and visualize heterogeneity in HSC subsets.

(a) *INKA1/FAM212A* and *PAK4* transcript expression in CB RNA-seq data (upper, GSE125345) and as means of inferred subpopulation from bone marrow (BM) single cell data (lower, Hay *et al* 2018). Corresponding subpopulations (delineated by dotted lines) are listed in middle panel, designation as used in text. *Prog =Progenitors.*

(b) Gating strategy to obtain four HSC subsets (detailed in Extended Data 1a). *IT = intermediate.*

(c) Confocal images of 2 individual $CD90^{+}CD49f^{+}$ LT-HSCs showing the two extremes of the mutually exclusive abundance of INKA1 and H4K16ac representative for n=3 independent CB as summarized in (d). *Scale bar =2 µm.*

(d) Mean fluorescence intensity (MFI) of single cells for nuclear (nucl.) INKA1 and H4K16ac by confocal analysis (Ø50 cells/population and CB, 3 CB pools). *$INKA1^{hi}$ cell gate (dashed) as for Fig. 1e and Extended Data Fig. 1d).*

(e) Nuclear diameter of $INKA1^{hi}$ and $INKA1^{lo}$ (remainder) cells from 3 CB and 4 HSC subsets combined, *two-sided Mann-Whitney-test (***p<0.0001).*

(f) Immunostaining of LT-and ST-HSCs for INKA1 and CDK6 (120 cells/population, 3 CB pools), CDK6 MFI is shown for cells fractionated into $INKA1^{hi}$ (in analogy to (d)) and the remainder ($INKA1^{lo}$); *two-sided t-test (***p=0.00003).*

Fig. 2: Single cell transcriptome and protein analysis validates that *INKA1* vs its target *PAK4* and H4K16ac are on opposing ends of cell cycle priming and progression.

(a) Normalized nuclear CDK6 intensity in LT- and ST-HSCs (CB and mPB; n=4; Ø50 cells/population and sample; immunostaining). *two-sided Mann-Whitney-test (**p=0.0036, ***p<0.0001).*

(b) Percentage CDK6+/- LT- and ST-HSCs within G_0 ($Ki67^{-}$) from mPB (n=3, flow cytometry), *Mean±SD, two-sided paired t-test (*p=0.020, **p=0.004).*

(c) Diffusion mapping of scRNAseq data from $CD34^{+}CD38^{-}CD45RA^{-}$ mPB cells using RNAvelocity. *DC = diffusion component, superimposed scaled color gradient: early (0.0) to late (1.0) pseudotime.*

(d) Indicated transcript expression against diffusion pseudotime, curves fitted by nonlinear regression.

(e,f) CB Lin^{-} cells cultured for 2 d with 20 hr EdU pulse to subgate cell cycle phases (n=5; flow cytometry), *paired t-test*(e), *one-way ANOVA*(f); *minimal to maximal values are represented in box-and-whisker plot with median as centre.*

(g) CB LT- and ST-HSCs analyzed for Ki67, H4K16ac, PAK4 and DAPI after 72 hrs of culture (high cytokine conditions; n=3; Ø50 cells/population/sample; immunostaining), *one-way ANOVA (**p<0.01; ***p<0.001).*

(h) Excerpt of confocal analysis of cultured LT- and ST-HSCs harvested at 5 timepoints (Ø50 cells/population/timepoint/sample; n=3). *One-way ANOVA with Welch's correction; significance level # and * as compared to 6 h timepoint or pair-wise (LT- vs ST-HSCs) for the same timepoint, respectively #p<0.05, ##p<0.001, ###p<0.001; *p<0.05, **p<0.01, ***p<0.001); Scale bar: 4 µm.*

(i) Protein diffusion pseudotime analysis of 3442 cells from 5 timepoints, four $CD45RA^-$ HSC subsets (defined by CD90 and CD49f; n=3) and Progenitors (n=2) by immunostaining. Timepoints or marker levels (gradients) as indicated. Longitudinal marker levels of HSC subsets across pseudotime were inferred by nonlinear regression (upper right).

Fig. 3: PAK4 inhibition induces G_0 arrest *in vitro* maintaining stem cell regenerative capacity.

Sorted CB $CD34^+CD38^-$ HSPC and $CD34^+CD38^+$ progenitors were cultured, after 2 days medium was supplemented with 0.25 μM PF-3758309 and 5 days later cells were analyzed for cell counts, viability and cell cycle or injected into NSG mice.

(a) Fold expansion according to input and cell count on day 7 of culture (3 CB pools; HSPC: $CD34^+CD38^-$; Prog: $CD34^+CD38^+$), *Mean±SD, two-sided t-test (*p=0.048, ***p=0.00045).*

(b) Cell cycle analysis of 5 d treated cultures (n=4), *Mean±SD, two-sided paired t-test (**p<0.01).*

(c) HSC frequencies as estimated by primary LDA from 5 d treated cultures transplanted into a total of 105 NSG (16-20 wk, 6 CB donors/pools, each dot representative of 6-10 injected mice, hCD45 engraftment > 0.1%). *Initial cell culture input (resulting progeny injected/mouse): 160-500 (low), 800-2500 (mid), 4000-12500 (high dose). Mean frequency indicated by fraction numbers. Frequency was calculated either according to the initial cell doses cultured irrespective of the actual cell number*

transplanted or according to the actual number of progeny after treatment transplanted. TX= transplantation.

(d) Lineage composition (human $CD45^+$) of engrafted primary NSG mice (n=9-12). *Mean±SD.*

(e,f) Individual assay (e) and combined (f) HSC frequency estimates according to secondary LDA (8 wk) in NSG-SGM3 mice upon transplantation of sorted $hCD45^+$ cells from NSG grafts generated from cultured and treated HSPC. *∇: in 1 of 4 assays no mouse engrafted in DMSO cohort resulting in a frequency of 1/∞. $CD45^+$ cell dose transplanted into secondary recipients (n=86): 180K-540K (high), 60K-180K (mid), 20K-60K (low).*

Fig. 4: *INKA1*-OE LT-HSCs resist activation *in vitro.*

(a,b) Immunostaining of prospectively isolated, transduced (OE) and BFP^+-sorted CB LT- and ST-HSCs (n=3, day 6 post-transduction), *one-sided Kruskal-Wallis-test*; *minimal to maximal values are represented in box-and-whisker plot with median as centre (*p<0.005, ***p<0.001).*

(c,d) Gene sets enriched upon *INKA1*-OE vs CTRL in LT-HSCs or ST-HSCs with FDR q-value <0.1 at least in one dataset (RNA-seq, n=3, d6 ptd). *Kolmogorov-Smirnov statistic *FDR < 0.10; **FDR <0.05; ***FDR<0.01.*

Fig. 5: *INKA1*-OE and *PAK4*-KD preserve long-term stem cell regenerative potential under regeneration-mediated stress.

Progeny of 10,000 CB $Lin^-CD34^+CD38^-$ cells were transduced with OE or knock-down (KD) vectors and transplanted into NSG mice 1 day later. *Input: mean of transduction efficiency at day 3 and 6.*

(a) Relative engraftment of transduced cells (log_2 ratios: $\%BFP^+$ of $CD45^+$ *in vivo* output injected femur/$\%BFP^+$ *in vitro* input) of 1 to 5 independent experiments, *Mean±SD, one-way ANOVA with Tukey's post hoc test (**p=0.0078 and p=0.0027, ***p<0.0001).*

(b,c) Scored engraftment (>0.1% hCD45$^+$) of secondary transplants from bone marrow (BM) grafts harvested and CD45$^+$BFP$^+$ sorted ((injected cell dose indicated) at multiple timepoints (2w, 4w, 20w) ; 20w grafts were assessed in secondary LDA (n=3; c).

(d) Flow cytometric analysis of pooled BM from 20 w primary xenografts (n=4) gated on CD45$^+$BFP$^+$CD34$^+$CD19$^-$CD38$^-$ population, *Mean±SD, two-sided paired t-test*.

(e,f) NSG mice (n=80) were transplanted with CTRL or *INKA1*-OE CB CD34$^+$CD38$^-$ cells. At 4 w 150 mg/kg 5-Fluorouracil (5-FU) or PBS equivalent was injected intraperitoneally. Multiplicity of infection was adjusted to result at comparable %BFP$^+$ at 4 w for CTRL and *INKA1*-OE; untreated xenografts at 4 w served as baseline.

(e) BFP$^+$ cell counts of input and in BM of injected femurs as inferred by total cell count/femur x %BFP$^+$. *Mean±SEM of 5-10 mice/timepoint and condition, two-sided Mann-Whitney-test.*

(f) Total BM (20w) of 5-FU or PBS treated mice was subjected to hierarchy analysis.

(g,h) Secondary LDA analysis of CD45$^+$BFP$^+$ cells from 20 w primary grafts (+8wk; >0.1% hCD45$^+$).

(i,j) Relative %BFP$^+$ engraftment (primary xenografts; i, *Mean±SD, two-sided t-test, *p<0.05; **p=0.0010*) and HSC frequency by secondary LDA (8 wk) for KD vectors (j).

Fig. 6: INKA1/PAK4 interactome identifies SIRT1 as link between INKA1 and H4K16ac and CD112 as a HSPC marker.

(a) INKA1-PAK4 protein interaction network by BioID. *Edge width indicates SpecSum/abundance of biotinylated peptides; node color resembles differential RNA expression in CB LT- vs. ST-HSCs (log_2; GSE125345). *previously reported PAK4 interactors.*

(b) CB CD34$^+$CD38$^-$ HSPC and CD38$^+$CD38$^+$ progenitors were subjected to proximity ligation assay (PLA) for INKA1 and SIRT1 (n=3, Ø100 cells/ sample). *Scale bar =6 µm. Mean±SD, two-sided t-test (**p=0.005).*

(c) CTRL and INKA1-OE CB CD34$^+$CD38$^-$ HSPC were BFP$^+$ sorted 6d after transduction followed by immunostaining for SIRT1 and INKA1 and confocal microscopy. Relative distribution of SIRT1 in

nucleus was calculated dividing nuclear intensity by total cell intensity (n=3; ∅40 cells/ sample). *Scale bar =5 µm; two-sided t-test with Welch's correction (**p=0.008).*

(d, e) Flow cytometric analysis of CB Lin^- cells (n=11, d) and CB mononuclear cells (MNC, n=4, e) for CD112 MFI across subsets. *Mean±S, two-sided t-test (***p<0.0001).*

(f) Cultured and transduced HSPC were analyzed 6 days later for CD112 levels of subgated cLT-($CD34^+CD45RA^-CD90^+$), cST-($CD34^+CD45RA^-CD90^-$) HSCs, cProgenitors ($CD34^+CD45RA^+$) or total cHSPCs ($CD34^+$) *c= cultured.*

(g) LT-HSCs preincubated with anti-CD112 before fixation to only stain for cell surface expression were subjected to immunostaining and quantified intensity for total INKA1 and surface CD112 were plotted against each other (n=68 from 3 CB). *Scale bar =2 µm.*

(h) *Nectin2*/CD112 RNA expression levels against scRNAseq diffusion pseudotime.

Fig. 7: CD112 surface expression delineates alternatively quiescent states of LT-HSCs.

(a) Gating strategy for sorting $CD112^{hi}$ and $CD112^{lo}$ CB $CD90^+CD49f^+$ LT- and $CD90^-CD49f^-$ ST-HSCs.

(b) LT- and ST-HSCs sorted for $CD112^{hi}$ and $CD112^{lo}$ and subsequently analyzed by intracellular flow cytometry for CDK6 and Ki67; *two-sided paired t-test.*

(c,d) Immunostaining of $CD112^{hi}$ and $CD112^{lo}$ sorted LT- and ST-HSCs (n=5-7; ∅50 cells/population and sample), *one-way ANOVA and unpaired t-test; minimal to maximal values are represented in box-and-whisker plot with median as centre (c; ***p<0.0001)and two-sided Mann-Whitney test (d; **p=0.0067).*

(e) Flow cytometric analysis of CellROX labelling.

(f,g) LT-HSCs sorted for $CD112^{lo}$(n=2) and $CD112^{hi}$(n=3) were subjected to ATAC-seq. (f) shows UMAP 2D projection of the enrichment of all 10 NMF signatures previously identified[26] in CD34+CD38- or CD34+CD38+ CB single cells (combined shown*), colours indicate the enrichment (Z-Score) of the $CD112^{lo}$ or $CD112^{hi}$ LT-HSC specific regions. Red indicates enrichment, blue depletion.* Pearson's correlation between the single-cell enrichment of the $CD112^{lo}$ or $CD112^{hi}$ LT-HSC specific regions and the LT/HSPC (yellow) and Act/HSPC (green) signatures[26] is shown in (g).

(h) Hallmark GSEA of CD112^{hi} vs. CD112^{lo} LT-HSCs and *INKA1* expression (bulk RNA-seq). *Kolmogorov-Smirnov statistic.*

(i) Differential gene expression for CD112^{hi} vs. CD112^{lo} LT-HSCs (RNA-seq, 2-3 independent CB pools).

(j) Enrichment map visualization of GSEA results for CD112^{hi} versus CD112^{lo} RNA-seq data (3472 GO biological processes; gene set size 25-500; FDR<0.1).

Fig. 8: CD112^{hi} LT-HSCs are primed whereas CD112^{lo} demarcates latent LT-HSCs.

Prospectively isolated CD112^{hi} and CD112^{lo}LT-HSCs from 3 independent CB pools were transplanted separately at equal cell doses (as indicated) into NSG mice (n=90).

(a, b) Engraftment kinetics according to human CD45^{+} chimerism in injected femurs, two-sided *paired t-test; minimal to maximal values are represented in box-and-whisker plot with median as centre (*p<0.05).*

(c) HSC frequency of individual CB pools estimated by secondary LDA of sorted hCD45^{+} cells from 12 w primary xenografts transplanted into NSG-SGM3 mice and analyzed after 8 w, *two-sided ratio paired t-test (*p=0.0476).*

(d) Combined secondary LDA analysis from 3 CB pools. *Chisquare likelihood ratio test statistic.*

METHODS

Materials and Reagents

Detailed information of commercially available reagents and other materials used in this study is summarized in Supplementary Table 3.

Human Specimen Collection

Human CB samples were obtained with informed consent from William Osler, Trillium and Credit Valley Hospitals according to procedures approved by the University Health Network (UHN) Research Ethics Board. Mononuclear cells (MNC) from single CB units or independent pools of male and female CB units (~4-15 units) were obtained by centrifugation on Lymphoprep medium, and after ammonium chloride lysis MNC were depleted of lineage positive cells (lineage depletion) by negative selection with the StemSep Human Progenitor Cell Enrichment Kit according to the manufacturer's protocol. Resulting lineage negative (Lin-) CB cells were stored viably at −80°C or −150°C. Human G-CSF mobilized peripheral blood samples (mPB, female and male, age: 40-63 years) were obtained with informed consent from the Leukemia Tissue Bank at Princess Margaret Cancer Centre/University Health Network according to procedures approved by the UHN Research Ethics Board. Following thawing and MNC isolation, CD34+ cells from mPB were enriched by positive selection with the CD34 Microbead kit (Miltenyi) and LS column purification with MACS magnet technology (Miltenyi).

Mice

Animal experiments were done in accordance with institutional guidelines approved by the University Health Network (UHN) Animal care and we complied with all relevant ethical regulations for animal testing and research. NOD.Cg PrkdcscidIl2rgtm1Wjl /SzJ (NSG) and NOD.Cg-*Prkdc*^{scid}*Il2rg*^{tm1Wjl}Tg(CMV-IL3,CSF2,KITLG)1Eav/MloySzJ (NSG-SGM3) mice were housed in a controlled environment with 12:12 light-dark cycle with a 30 min transition, room temperature of 21–23 °C and humidity of 30-60%, and had ad libitum access to dry laboratory food and water .at the animal facility (ARC) at Princess Margaret Cancer Centre, They were kept in a room designated only for immunocompromised mice with individually ventilated racks equipped with complete sterile micro-isolator caging (IVC), on corn-cob bedding and supplied with environmental enrichment in the form of a red house/tube and a cotton nestlet.

Cages are changed every <7 days under a biological safety cabinet. Health status is monitored using a combination of soiled bedding sentinels and environmental monitoring.

Flow Cytometric Analysis and Sorting

Primary human cells were thawed by dropwise addition of X-VIVO 10 + 50% fetal calf serum supplemented with DNase (100 µg/mL final concentration) and resuspended at a density of 5×10^6 cells/mL. Cells were then incubated (30 min, 4°C) with the following antibodies: FITC– or BV421-anti-CD45RA (1/50), PE– or APC-anti-CD90 (1/100), PECy5–anti-CD49f (1/50), PECy7–anti-CD38 (1/100), APCCy7–anti-CD34 (1/200) and if applicable APC-anti-CD112(1/100) for sorting and/or flow cytometric analysis or anti-Nectin2 (goat) for subsequent surface immunostaining via confocal analysis. After washing cells were resuspended at 0.5-10x10^6 cells/mL in PBS, 5% FBS supplemented with 0.1 µg/L propidium iodide (PI) and sorted on a BD FACS Aria III instrument.

For flow cytometric analysis of CB MNC in addition to SYTOX blue the following antibodies were used: V500-anti-CD45 (1/100), APC-Cy7-anti-CD34 (1/200), PE-Dazzle-CD38 (1/100), APC-anti-CD112 (1/100), BV786-anti-CD33 (1/100), PE-anti-CD19(1/100), PE-anti-GlyA (1/100), FITC-anti-CD3 (1/100). Cell preparation for intracellular staining of sorted cells was performed with BD Cytofix/Cytoperm as described before [3] and incubated with antibodies as applicable (FITC or PE-anti-Ki67 (1/30), AF647-anti-CDK6 (1/30), AF488-anti-H4K16ac (1/100)) in PermWash solution (BD) overnight. If applicable, Edu Click-IT reaction was performed according to manufacturer's instructions and prior to flow cytometry analysis cells were stained with Hoechst 33342 (1:2000; Thermo Fisher) or DAPI (1 µg/mL; Sigma). Other flow cytometry-based assays were performed according to the manufacturer's protocol: AnnexinV-FITC or -APC Kit (BD) and CellROX staining (ThermoFisher) in X-Vivo 10, 1% BSA, L-Glutamine, Pen/Strep. Cells were sorted on BD FACS Aria III or analyzed on BD Canto II, BD Fortessa or BD Celesta instruments. Positive gates were set according to unstained or isotype controls. Data was analyzed by FlowJo.

Immunostaining

Sorted cells were spun onto Poly-L-Lysine (Sigma)-coated slides (Ibidi, 200 xg, 10 min), fixed with 4% formaldehyde (Sigma) and permeabilized with 0.5% Triton X-100 (Sigma) before blocking (PBS, 10% FBS, 5% BSA; 30 min, RT). Slides were incubated with primary antibodies (Anti-INKA1(1/100), Anti-PAK4 (1/50), AF647-Anti-CDK6 (1/50), AF488-Anti-H4K16ac(1/100), Anti-SIRT1(1/100), Anti-Nectin2 (1/100), AF647-Anti-Ki67 (1/100)) in blocking solution O/N at 4°C. If necessary Anti-FAM212A and Anti-PAK4 antibodies were directly labeled using the Zip Alexa Fluor 555 Rapid Antibody Labeling Kit (Thermo Fisher) at a 0.1x antibody concentration but otherwise according to manufacturer's instructions to allow co-staining with other directly labeled rabbit or mouse antibodies, respectively. As applicable, secondary antibodies (1/400) were added (PBS, 0.025% Tween 20, Sigma, 1.5 h, RT). After washing, nuclei were stained with 1 µg/mL DAPI (Sigma) and slides were mounted (FluoromountG). Single cell images were captured by a Zeiss LSM700 Confocal Microscope (oil, 63x/1.4NA, Zen 2012) and processed and analyzed with ImageJ/Fiji and FlowJo10 or Python. In case of high inter-experimental variability between independent stainings data were normalized by z-score transformation. Pixel Intensity Spatial Correlation Analysis of colocalization was carried out using the Coloc2/Fiji plugin for ImageJ Software to obtain Spearman's rank correlation values for a region of interest incorporating either an individual cell, its nucleus or its non-nuclear area. The Coloc2 plugin was run with a PSF of 4.1-4.7 depending on channel combination and with 50 Costes' randomizations. For pseudotime analysis using single cell immunofluorescence data, fluorescence values were log transformed and scaled (CDK6, nuclear CDK6, DAPI, H4K16ac, PAK4, and nuclear PAK4). Diffusion mapping embedding was subsequently calculated as well as diffusion pseudotime [21] within scanpy . For the CellROX assay sorted cells were washed, resuspended in cytokine-free medium (X-Vivo 10, 1% BSA, L-Glutamine, Pen/Strep) supplemented with 5 µM CellROX Green and then incubated for 30 min at 37°C in Poly-L-Lysine coated chambered glass coverslips/ µ-slides (Ibidi). Subsequently, slides were spun (200 xg, 10 min), medium was removed, and cells were fixed, washed, permeabilized and blocking was performed as before. Slides were then incubated with AF647-anti-CDK6 (1/50) antibody in blocking

solution for 1 hr at RT. After washing, nuclei were stained with 1 µg/mL DAPI, chambers were washed, covered with FluoromountG and images were captured by a Zeiss LSM700 Confocal Microscope (20x/0.8 NA, Zen 2012).

Proximity Ligation Assay

Proximity ligation assay (PLA) was performed after sample preparation for immunostaining as described followed by using the Duolink® In Situ Orange Starter Kit Mouse/Rabbit (Millipore Sigma) according to manufacturer's instructions with few adaptations: Before primary antibody incubation overnight in PBS, 10% FBS, 5% BSA, slides were blocked with PBS, 10% FBS, 5% BSA for 5 hrs at 4°C. After two washes with Wash Buffer A (Kit), slides were incubated with the Duolink® Blocking solution for 3 hrs at 37°C. For all subsequent steps manufacturer's instructions were followed. Images were captured by a Zeiss LSM700 Confocal Microscope (20x/0.8 NA, Zen 2012) and processed and analyzed with ImageJ/Fiji. Background signal was determined according to several control HSPC/Progenitor specimens incubated with either no primary antibody (AB), no primary or secondary AB, anti-SIRT1(1/100) as only primary AB, anti-FAM212A (1/100) as only primary AB, anti-SIRT191/100) or anti-PAK4 (1/50) in combination with rabbit IgG (anti-V5; 1/100), anti-FAM212A in combination with mouse IgG (anti-BrdU; (1/100 or 1/50; respectively)).

Chromatin fiber assay

For visualizing epigenetic information and proteins bound to chromatin extended chromatin fibers were prepared as described elsewhere[45,46] with few modifications: CB HSPC $CD34^{+}CD38^{-}$ were resuspended in PBS at $2.5x10^{5}$ cells/mL, 8 µl were spotted as a line on a glass slide and air-dried before 150 µl lysis buffer (500 mM NaCl, 500 mM Urea, 25 mM Tris, 1% Triton X-100, pH 7.5) was added. After 15 min at RT slides were placed in a 25° angle to extend fibers by gravity. For fixation slides were treated for 10 min with 4% PFA and then washed for 30 min in PBS. After 1 hr of blocking (PBS, 10% FBS, 5% BSA) primary antibody incubation followed (90 min, RT; 1/50 in blocking solution). Subsequently slides were

washed in PBS, 0.025% Tween (10 min, RT) and incubated with secondary antibodies (1/400 in PBS, 0.025% Tween; 90 min, RT). After 1x washing in PBS, 0.025% Tween (10 min), slides were washed 2x with PBS and incubated with 5 µg/mL Hoechst 33342 in PBS (30 min, RT). Slides were then washed (PBS), mounted with FluoromountG and analyzed with a Zeiss LSM700 Confocal Microscope (oil, 63x/1.4NA, Zen 2012).

Cell Culture

For the time course experiment ~1000 sorted cells were cultured in Terasaki plates in 25 μl. To align G_0 exit kinetics with previous data[3] cells were cultured under high cytokine conditions: StemPro-34 SFM media (Thermo Fisher) with the supplied supplement, 1x L-Glutamine, 1x Pen/Strep, 0.02% Human LDL and the following cytokines: FLT3L (20 ng/mL), GM-CSF (20 ng/mL), SCF (100 ng/mL), TPO (100 ng/mL), EPO (3 U/mL), IL-3 (10 ng/mL), IL-6 (50 ng/mL). Cells were harvested at indicated timepoints and subjected to immunostaining. For all other primary cell cultures cells were cultured at a density of 0.4-5x10^5 cell/mL in 96-well round-bottom plates with low-cytokine condition media: X-VIVO 10 media, 1 % BSA supplemented with 1x Pen/Strep, L-Glutamine and the following cytokines: SCF (100 ng/ml), Flt3L (100 ng/ml), TPO (50 ng/ml) and IL7 (10 ng/ml). For experiments with the PAK4 inhibitor PF-3758309 (solvent: DMSO), the SIRT1 inhibitor EX-527 (final concentration 25 μM; solvent: DMSO) or EdU (final concentration 10 μM) 1 volume of complete medium with 2x reagent or equivalent DMSO concentration as indicated was added to 1 volume of cell suspension. For PF-3758309 treated cells after 3 days a half medium exchange was performed with complete medium supplemented with 1x inhibitor concentration. For LDA experiments initial cell input was: 160-500 (low dose), 800-2500 (mid dose), 4000-12500 (high dose).

Lentiviral Vector Production

VSV-G pseudotyped lentiviral vector particles were produced, and titers were determined as previously described[12]. The negative control (CTRL) vector for overexpression encodes *gp91^{phox}P415H* (catalytic

inactive gp91^{phox}/*CYBB*) and for knockdown a shRNA directed against Renilla luciferase (shCTRL)[12]. Additional shRNA were generated as described before[12].

Lentiviral Transduction

If for a particular assay not stated otherwise, after 16-20 hours of pre-stimulation in low cytokine condition medium cells were transduced with OE or KD and control vectors at matching multiplicity of infection[12] aiming at mid-range (20-40%) transduction efficiencies but without lentiviral preparation exceeding 20% of total culture volume. Transduction efficiency (%BFP$^+$) was determined at day 3 and day 6 post-transduction by flow cytometry on a BD Celesta and for xenotransplantation assays the mean of both timepoints served as initial input estimate.

Quantitative RT-PCR for *PAK4* and *INKA1*

RNA from FACS-sorted fractions was isolated with TRIzol (Invitrogen) or for low cell numbers (<20,000 cells) with PicoPure RNA Isolation Kit (Thermo Fisher) according to manufacturer's protocol. Quantitative RT-PCR analysis was performed using the QuantiFast SYBR Green PCR Kit (Qiagen) on the LightCycler 480 Instrument II (Roche). All signals were quantified using the ΔCt method and were normalized to the levels of GAPDH.

Methylcellulose Colony Forming Unit Assay

LT-HSC and ST-HSC were sorted and cultured in low cytokine medium for 4 hours and then transduced with CTRL or INKA1-OE lentiviral vectors. At day 3 post-transduction, 200 BFP$^+$ cells were sorted directly into 2 mL methylcellulose (H4034, Stem Cell Technologies), supplemented with FLT3 Ligand (10 ng/mL) and IL6 (10 ng/mL) and plated onto 2x35 mm dishes as duplicates. Colonies were allowed to differentiate for 12-14 days and then morphologically assessed in a blind fashion by a second investigator. Subsequently, colonies from replicate plates were pooled, resuspended in PBS/5% FBS, washed and 0.5%

of progeny from LT-HSC and 1% of progeny of ST-HSC was added to fresh methylcellulose as above, replated and scored after 12-13 days.

Single Cell Stromal Assays

Single cell *in vitro* assays were set up as with low passage murine MS-5 stroma cells[47] seeded at a density of 1500 cells per 96-well and grown for 2-4 days in H5100 media (StemCell Technologies, 05150). One-day prior to co-culture initiation, the H5100 media was removed and replaced with 100 µl erythro-myeloid differentiation media: StemPro-34 SFM media (Thermo Fisher) with the supplied supplement, 1x L-Glutamine, 1x Pen/Strep, 0.02% Human LDL and the following cytokines: FLT3L (20 ng/mL), GM-CSF (20 ng/mL), SCF (100 ng/mL), TPO (100 ng/mL), EPO (3 U/mL), IL-2 (10 ng/mL), IL-3 (10 ng/mL), IL-6 (50 ng/mL), IL-7 (20 ng/mL), and IL-11 (50 ng/mL).

Single cell stromal assays: *INKA1*-OE

LT-, IT90+, IT90- and ST-HSC were sorted from CB Lin- and cultured in low cytokine medium for 4 hours before transduction with CTRL or *INKA1*-OE lentiviral vectors. On day 3 post-transduction, cells were washed, resuspended in PBS, 2.5% FBS, 0.1 µg/L PI and $PI^{-}BFP^{+}$ single cells were deposited onto seeded MS5 stromal layers in 96 well plates using a BD FACSAria III. Single cells were co-cultured to proliferate and differentiate for 15 days (18 days post-transduction) with an addition of 100 µl erythro-myeloid differentiation media per well on day 8. Wells with hematopoietic cell content were marked 1 day prior and the total number of wells with colonies was used to estimate single cell cloning efficiency. On the day of final analysis cultures were trypsinized, blocked with FBS, filtered (8027, Pall), stained and flow cytometric analyzed on a BD Celesta using the high-throughput sampler (HTS) unit. The following antibodies were used: V500-anti-CD45 (1/100), BV786-anti-CD33 (1/100), FITC-anti-CD56 (1/100), PE-anti-GlyA (1/100), PECy5-anti-CD14 (1/100), APC-anti-CD71 (1/100) and APC-Cy7-anti-CD41 (1/200). Erythroid cells were defined as CD45 negative and positive for CD71 and/or GlyA.

CD112 Index-sorted Single Cell Stromal Assay

On the day of the experiment, CB Lin$^-$ cells were stained as above to sort and deposit Sytox Blue$^-$ CD34$^+$CD38$^-$CD45RA$^-$ single cells onto seeded MS5 stromal layers in 96 well plates while single cell fluorescence intensities for PECy5-anti-CD49f, PE-anti-CD90 and APC-anti-CD112 were indexed using a BD FACSAria III. Single cells were cultured for 18 days with an addition of 100 µl erythro-myeloid differentiation media per well at day 8. Wells with hematopoietic cell content were marked 1 day prior and the total number of wells with colonies was used to estimate single cell cloning efficiency. On day 18 of co-culture cells hematopoietic cells were resuspended, filtered (8027, Pall), stained and flow cytometric analyzed on a BD Celesta using the HTS unit. The following antibodies were used: V500-anti-CD45(1/100), BV786-anti-CD33 (1/100), BV605-anti-CD56 (1/100), PE-anti-GlyA (1/100), PECy5-anti-CD11b (1/200), FITC-anti-CD71 (1/100), APC-Cy7-anti-CD41(1/200) and APC-anti-CD34(1/100). Clonal lineage output was scored and aligned with indexed single cell fluorescence intensity for CD112 at initial single cell deposition generating an .fcs file for visualization in FlowJo10.

Xenotransplantation

Aged and gender matched 8-16 wk old male and female recipient NOD.Cg-PrkdcscidIl2rgtm1Wjl/SzJ (NSG) or NOD.Cg-PrkdcscidIl2rgtm1WjlTg(CMVIL3,CSF2,KITLG)1Eav/MloySzJ (NSG-SGM3) mice were sublethally irradiated (224-225 cGy) 24 hours prior to intra-femoral injection. Cells were either injected directly after sorting, the progeny of 10^4 CD34$^+$CD38$^-$ transduced cells one day after transduction or the progeny of initially plated CD34$^+$CD38$^-$ cells treated in culture with PF-3758309 or DMSO. For analysis, mice were euthanized, injected femur and other long bones (non-injected femur, tibiae) were flushed separately in Iscove's modified Dulbecco's medium (IMDM) and 5% of cells were analyzed by flow cytometry (BD Celesta). For primary transplant analysis of initially *ex vivo* cultured cells, mice were euthanized at indicated timepoints (2, 4, 12, 20 weeks) after transplantation and human chimerism was assessed along with BFP and the following antibodies in addition to staining with Fixable Viability Dye eFluor 506 (1/500): PECy5–anti-CD45(1/100), APCCy7–anti-CD34(1/200), BV786-anti-CD33(1/100),

PE–anti-CD19(1/100), PE–anti-GlyA(1/100), FITC-anti-CD71(1/100). For experiments with sorted and injected LT-HSC $CD112^{high}$ and $CD112^{low}$ fractions the antibody panel was: V500–anti-CD45(1/100), FITC-anti-CD34(1/50), BV786-anti-CD33(1/100), PE–anti-CD19(1/100), PE–anti-GlyA(1/100), APCCy7-anti-CD41(1/200) and APC-anti-CD112(1/100);in addition to SYTOX Blue. Sick and mis-injected mice were excluded from analysis. For purification of human cells from xenotransplanted mice, fresh or thawed BM from individual or from pools of 2-5 mice were either directly analyzed by flow cytometry and/or flow sorted for serial transplantation, or after mouse cell depletion (Mouse Cell Depletion Kit, Miltenyi) or $CD34^+$ enrichment (human CD34 Microbead kit, Miltenyi) according to manufacturer's protocol. $CD34^+$ enriched cells were analyzed by flow cytometry with the following antibodies: FITC–anti-CD45RA (1/50), APC–anti-CD90 (1/100) or PE-anti-CD90(1/100), PECy5–anti-CD49f (1/50), PECy7–anti-CD38(1/200), APCCy7–anti-CD34(1/200), PE-anti-CD19(1/100) or BV711-anti-CD19(1/100), V500-anti-CD45(1/100) and if applicable with APC-anti-CD112(1/100). For secondary transplantation human double $CD45^+$ (APC-anti-CD45(1/100); FITC-anti-CD45(1/100); and if applicable BFP^+) cells were sorted from each pooled sample and injected at indicated doses into irradiated (224-225 cGy) NSG-SGM3 mice and engraftment assessed at 8 weeks post-transplant. Injected femur and non-injected femurs were isolated, flushed and analyzed separately with the following reagents: PECy5–anti-CD45(1/100), APC–anti-CD45 (1/100), PE-anti-CD19(1/100), PE–anti-GlyA(1/100), BV786–anti-CD33(1/100), FITC-anti-CD3(1/100), APCCy7–anti-CD34 (1/200) and if applicable SYTOX Blue or BFP. A mouse was considered engrafted if human $\%CD45^{++} > 0.1$. For LDA experiments HSC frequency was estimated using the ELDA software (http://bioinf.wehi.edu.au/software/elda/).

5-Fluorouracil injections

Male NSG mice xenotransplanted with transduced HSPC 4 weeks earlier (as above) were weighted and accordingly i.p. injected with a single dose of 150 mg/kg clinical grade 5-Fluorouracil (Hospira) or an equivalent volume of PBS. Sub-cohort were then analyzed 8 and 16 weeks later (corresponding to 12- and

20-weeks post-transplant, respectively) as described above. Mice in parallel xenotransplanted from the same transduced HSPC pool and euthanized at 4 weeks post-transplant served as baseline control.

Single cell RNAseq and processing

$CD34^+$ enriched mPB samples from three healthy donors (age 58-63; 2x male, 1x female) were thawed and stained (V500-anti-CD45(1/100), FITC-anti-CD45RA(1/50), APC-anti-CD90(1/100), PECy5–anti-CD49f(1/50), PECy7–anti-CD38(1/100), APCCy7–anti-CD34(1/200), PI) to sort and deposit PI^- $CD45^+CD34^+CD38^-CD45RA^-$ single cells into 96 well plates while single cell fluorescence intensities for PECy5-anti-CD49f and PE-anti-CD90 were additionally indexed using a BD FACSAria III. Single cells were immediately lysed when sorted into wells with a mix of 5× Maxima Transcription Buffer, dNTPs, and NXGen RNase inhibitor (Lucigen), custom 3'end reverse transcriptase primers[48] and water (detailed protocol deposited: https://dx.doi.org/10.17504/protocols.io.nkgdctw). Plates were spun and immediately transferred onto dry ice and stored at -80 until further processing. A modified version of the SMARTseq2 protocol[49] was performed with a custom modification in which a 12-base cell barcode was included in the 3′-end reverse transcriptase primer. This enabled us to perform multiplexed pooling before library preparation with the Nextera XT DNA sample prep kit (Illumina).The resulting libraries of 1536 single cell and simultaneously processed 192 $CD34^+CD38^-CD45RA^-$ BM cells were sequenced in a total of 4 lanes in Rapid Run mode on an Illumina HiSeq2500 instrument (Single index, paired end, 26bp read 1, 32bp read 2, 8bp index read with custom barcodes). Reads were aligned with STAR 2.6.0c. Default parameters were used except for the following: sjdbOverhang 30. RNA velocity was run on the bam files using the run_smartseq2 function in velocyto 0.17.8 with default settings [20]. Cells with total read counts or total genes 4 median absolute deviations (MADs) below the median were removed, as were cells with a proportion of mitochondrial genes 4 MADs above the median. Denoising was performed using dca 0.2.2 and batch correction was applied using an mnncorrect implementation within scanpy 1.3.3[50,51]. After preprocessing, mPB data was projected onto a diffusion map embedding and diffusion pseudotime[21] was calculated using default parameters. Imputed gene expression was subsequently compared against

977 pseudotime. To assign cells as LT-HSC, ST-HSC, IT90+ and IT90-, 40th percentile cutoffs were applied
978 for both CD90 and CD49f surface intensities as measured through index sorting.
979
980 **Bulk RNA-seq processing and analysis**
981 CD112^{high} and CD112^{low}LT-HSC, total LT-HSC and ST-HSC were sorted from 3 independent CB pools
982 and either directly resuspended and frozen (-80°C) in PicoPure RNA Isolation Kit Extraction Buffer or
983 after transduction followed by a second sort (PI^{-}BFP^{+}) on day 6 post-transduction (average cell count:
984 10^4; range 0.25-3x10^4). RNA was isolated using the PicoPure RNA Isolation Kit (Thermo Fisher)
985 according to manufacturer's instructions. Samples that passed quality control according to integrity
986 (RIN>7) and concentration as verified on a Bioanalyzer pico chip (Agilent Technologies) were subjected
987 to further processing by the Center for Applied Genomics, Sick Kids Hospital: cDNA conversion was
988 performed using SMART-Seq v4 Ultra Low Input RNA Kit for Sequencing (Takara) and libraries were
989 prepared using Nextera XT DNA Library Preparation Kit (Illumina). Equimolar quantities of libraries
990 were pooled and sequenced with 3-4 cDNA libraries per lane on a High Throughput Run Mode Flowcell
991 with v4 sequencing chemistry on the Illumina HiSeq 2500 following manufacturer's protocol generating
992 paired-end reads of 125-bp in length to reach depth of 60-80 million reads per sample. Reads were then
993 aligned with STAR v2.5.2b against hg38 and annotated with ensembl v90. Default parameters were used
994 except for the following: chimSegmentMin 12; chimJunctionOverhangMin 12; alignSJDBoverhangMin
995 10; alignMatesGapMax 100000; alignIntronMax 100000; chimSegmentReadGapMax parameter 3;
996 alignSJstitchMismatchNmax 5 -1 5 5. Read counts were generated using HTSeq v0.7.2 and general
997 statistics were obtained from picard v2.6.0. Differential gene expression was performed using DESeq2
998 v1.22.2 following recommended practices. Genes were ranked using the formula -sign(logFC) * -
999 log10(pvalue). The rank files from each comparison were used in pathway analysis (GSEA) using 2000
1000 permutations and default parameters-if not specified otherwise, against pathway collections from
1001 databases (MsigDB; HALLMARK, GOBP from http://baderlab.org/GeneSets, version February 2020) or
1002 retrieved from the literature (HSC_120_COMMON[52]). The gene set aHSC/MPP_UP_(vs_dHSC) was

derived from published murine RNAseq data[7] combining genes with logFC > 1 and FDR < 0.01 in either activated HSC (aHSC) or MPP versus dormant HSC (dHSC). For the *INKA1*-OE datasets pathways were excluded from analysis if gene set enrichment was driven by a single gene. EnrichmentMap version 3.1.0 in Cytoscape was used to visualize enriched pathway gene sets at FDR < 0.1 with an overlap coefficient set to 0.5 .

Proximity-dependent Biotin Identification (BioID) Assay

Coding sequences for *INKA1* and *PAK4* were PCR amplified from pENTR223_C3orf54_STOP of pDONR221_PAK4 (primers: ASCI-INKA1_fw, Not1_INKA1_rev, ASCI-PAK4_fw, Not1_PAK4_rev), respectively. Thereby, AscI and NotI restriction sites were introduced and used to clone both genes into pcDNA5/FRT/TO-FLAG-BirA*-[MCS] and pcDNA5/FRT/TO-[MCS]-FLAG-BirA* to generate N- and C-terminal BirA* fusion constructs. Transfection, induction, mass spectrometric analysis and data processing for the BioID assay was performed as described previously in HEK293[24]. Briefly, high confidence interactors were identified fulfilling these criteria: iProphet>0.9, at least two unique peptides in each of 4 runs, and a Bayesian FDR< 1% comparing the two highest peptide counts of each given ID between 20 controls (FlagBirA* only construct) and 4 bait protein sample analysis. Data was loaded into Cytoscape for network visualization and in Cytoscape STRING enrichment (GO processes) was used for pathway analysis to subsequently generate the enrichment map of selected pathways with FDR<0.05.

ATAC-seq processing and analysis

Library preparation for ATAC-Seq was performed on 2000 cells with Nextera DNA Sample Preparation kit (Illumina), according to a previously reported protocol[53]. 4 ATAC-seq libraries were sequenced per lane in HiSeq 2500 System (Illumina) to generate paired-end 50-bp reads. Reads were mapped against the hg19 human reference genome using BWA with default parameters. All duplicate reads, and reads mapped to mitochondria, chrY, an ENCODE blacklisted region or an unspecified contig were removed (ENCODE Project Consortium 2012). MACS 2.0.10 was used to call peaks with a uniform extension size

of 147 bp. A set of 500 bp windows over hg19, overlapping by 250bp was generated and for each sample the windows overlapping a called peak were identified. Windows were considered to be present in each condition if they overlapped a peak in 2 replicates. Windows only considered present in one condition were considered specific to that condition. Motif enrichment was performed using HOMER against a catalogue of all peaks called in any hematopoietic population was produced by merging all called peaks which overlapped by at least one base pair using bedtools. As described in *Takayama et al.*[26], cells from three sorted CD34+CD38-CD45RA- and three CD34+CD38+ populations were processed on the 10X Genomics single cell ATAC-seq platform, and 12,414 single cells were retained based on default cellranger QC criterion and chromVAR depth filtering. Cellranger-reanalyze (1.1, 10x Genomics) was subsequently used to map reads over the sites identified in the bulk ATAC-Seq catalogue, and read counts were binarized. chromVAR (with default settings) was used to calculate the enrichment of each of the hematopoietic signatures identified in this paper in each single cell, as well as for the windows considered specific to the $CD112^{low}$ or $CD112^{high}$ LT-HSC subsets.

Statistical Analysis

GraphPad Prism or MS Excel were used for all statistical analyses except RNA-seq, which was analyzed with R and Python. Unless otherwise indicated in Fig. legend, mean ± SD values are reported in the graphs and minimal to maximal values are represented in box-and-whisker plots. Statistical significance ($*p < 0.05$, $**p < 0.01$ and $***p < 0.001$, n.s. represents non-significance) was determined if not otherwise indicated in Figure legends with two-sided Student's t-tests with or without Welch's correction as applicable. If normal-distribution assumptions were not valid, statistical significance was evaluated using the Mann-Whitney test (two-tailed) for single comparisons or Kruskal-Wallis test for multiple comparisons. A normal distribution of the data was tested using the Kolmogorov–Smirnov test if the sample size allowed. One-way ANOVA was used to compare means among three or more independent

groups. Applicable post hoc tests were used to compare all pairs of treatment groups when the overall P value was <0.05.

Data availability

The accession number for RNA-seq data generated for this manuscript is GEO: GSE148885, and GEO: GSE125345 for previously published RNAseq data used in this study. Raw RNA-seq and ATACseq data is available on EGA: EGAD00001006541 (Bulk RNA-Seq: EGAS00001004768; scRNAseq: EGAS00001004769; ATACseq: EGAS00001005121). For GSEA GOBP were retrieved from http://baderlab.org/GeneSets (version February 2020), the Hallmark gene set v.7.0 from ftp.broadinstitute.org://pub/gsea/gene_sets/h.all.v7.0.symbols.gmt and Eppert HSC-R from http://www.gsea-msigdb.org/gsea/msigdb/cards/EPPERT_HSC_R.html.

Code availability

Code generated during this study is available at GitHub https://github.com/andygxzeng/HSC_Pseudotime.

Methods-only References

45 Wooten, M. *et al.* Superresolution imaging of chromatin fibers to visualize epigenetic information on replicative DNA. *Nat Protoc* **15**, 1188-1208, doi:10.1038/s41596-019-0283-y (2020).

46 Sullivan, B. A. Optical mapping of protein-DNA complexes on chromatin fibers. *Methods Mol Biol* **659**, 99-115, doi:10.1007/978-1-60761-789-1_7 (2010).

47 Itoh, K. *et al.* Reproducible establishment of hemopoietic supportive stromal cell lines from murine bone marrow. *Exp Hematol* **17**, 145-153 (1989).

48 Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction of single-cell gene expression data. *Nat Biotechnol* **33**, 495-502, doi:10.1038/nbt.3192 (2015).

49 Picelli, S. *et al.* Full-length RNA-seq from single cells using Smart-seq2. *Nat Protoc* **9**, 171-181, doi:10.1038/nprot.2014.006 (2014).

50 Haghverdi, L., Lun, A. T. L., Morgan, M. D. & Marioni, J. C. Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. *Nat Biotechnol* **36**, 421-427, doi:10.1038/nbt.4091 (2018).

51 Eraslan, G., Simon, L. M., Mircea, M., Mueller, N. S. & Theis, F. J. Single-cell RNA-seq denoising using a deep count autoencoder. *Nat Commun* **10**, 390, doi:10.1038/s41467-018-07931-2 (2019).

52 Fares, I. *et al.* EPCR expression marks UM171-expanded CD34(+) cord blood stem cells. *Blood* **129**, 3344-3351, doi:10.1182/blood-2016-11-750729 (2017).

53 Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods* **10**, 1213-1218, doi:10.1038/nmeth.2688 (2013).

Fig. 1

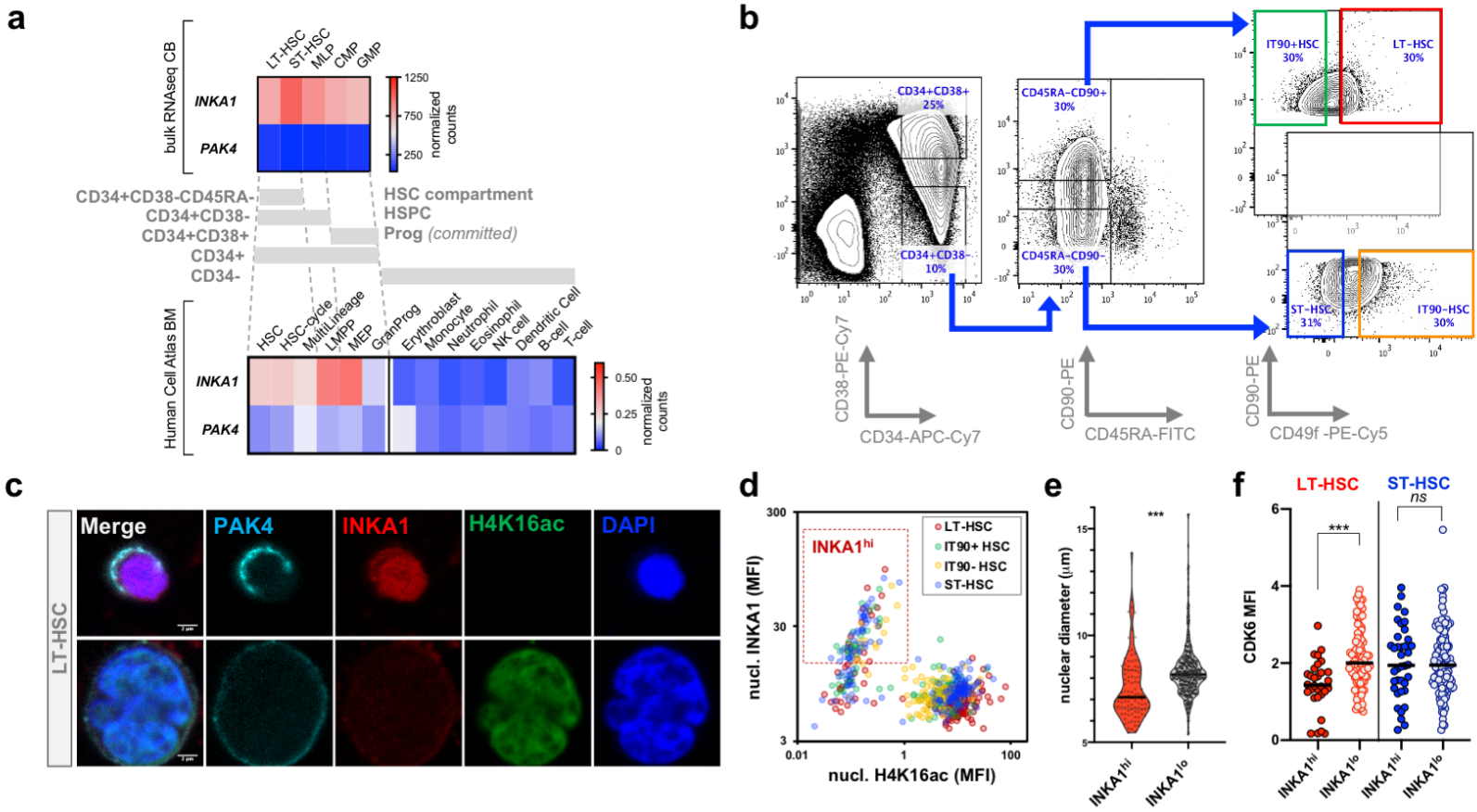

a
bulk RNAseq CB
LT-HSC
ST-HSC
MLP
CMP
GMP
INKA1
PAK4
1250
750
250
normalized counts
CD34+CD38-CD45RA-
CD34+CD38-
CD34+CD38+
CD34+
CD34-
HSC compartment
HSPC
Prog (committed)
Human Cell Atlas BM
HSC
HSC-cycle
Multilineage
LMPP
MEP
GranProg
Erythroblast
Monocyte
Neutrophil
Eosinophil
NK cell
Dendritic Cell
B-cell
T-cell
0.50
0.25
0
b
CD34+CD38+ 25%
CD34+CD38- 10%
CD45RA-CD90+ 30%
CD45RA-CD90- 30%
IT90+HSC 30%
LT-HSC 30%
ST-HSC 31%
IT90-HSC 30%
CD38-PE-Cy7
CD34-APC-Cy7
CD90-PE
CD45RA-FITC
CD90-PE
CD49f -PE-Cy5
c
LT-HSC
Merge
PAK4
INKA1
H4K16ac
DAPI
d
INKA1hi
LT-HSC
IT90+ HSC
IT90- HSC
ST-HSC
nucl. INKA1 (MFI)
nucl. H4K16ac (MFI)
e
***
nuclear diameter (µm)
INKA1hi
INKA1lo
f
LT-HSC
ST-HSC
ns
***
CDK6 MFI
INKA1hi
INKA1lo
INKA1hi
INKA1lo

Fig. 2

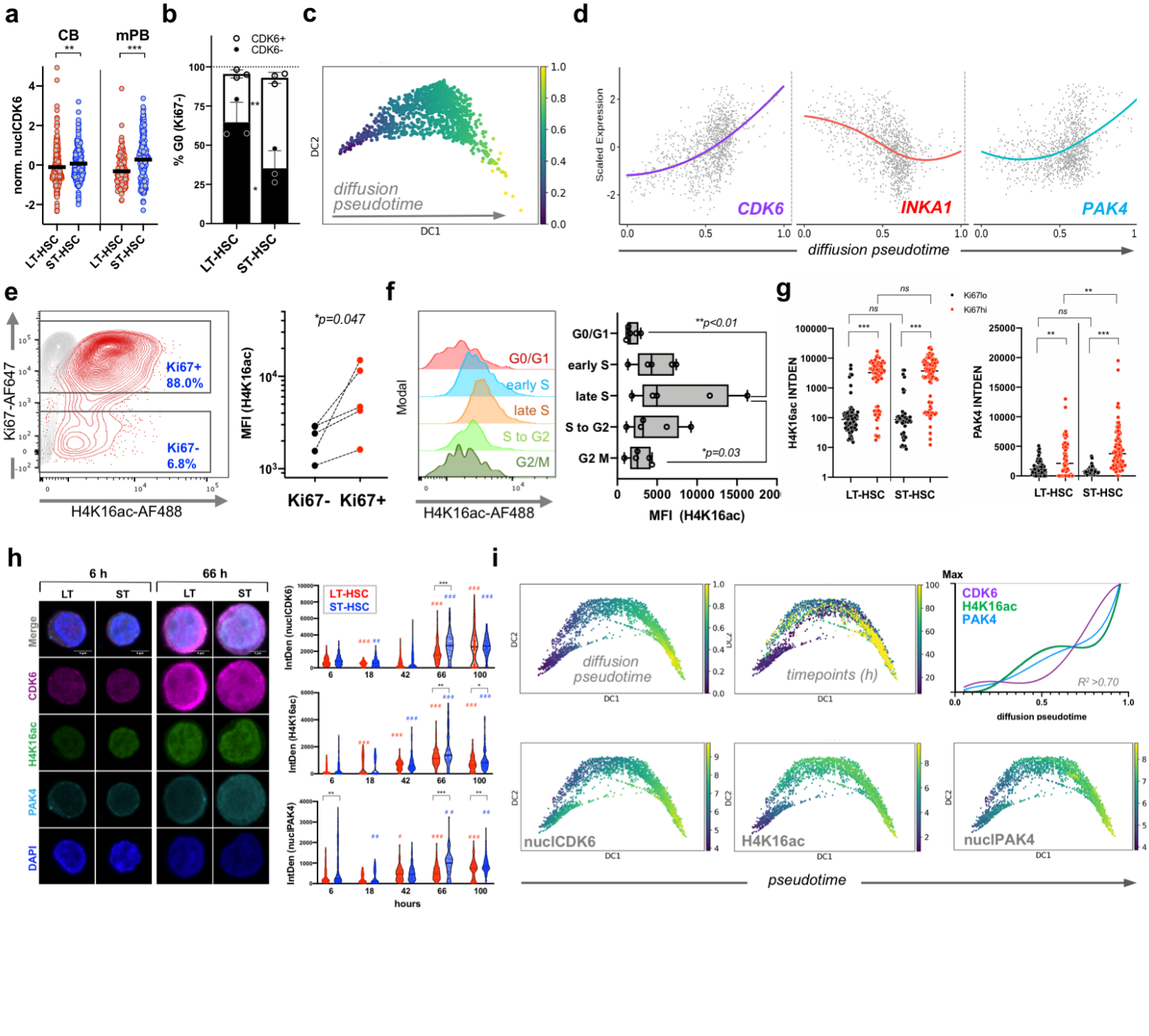

a
CB
mPB
norm. nuclCDK6
LT-HSC
ST-HSC
b
CDK6+
CDK6-
% G0 (Ki67-)
c
DC2
DC1
diffusion pseudotime
d
Scaled Expression
CDK6
INKA1
PAK4
diffiusion pseudotime
e
Ki67-AF647
Ki67+ 88.0%
Ki67- 6.8%
H4K16ac-AF488
*p=0.047
MFI (H4K16ac)
Ki67- Ki67+
f
Modal
G0/G1
early S
late S
S to G2
G2/M
H4K16ac-AF488
**p<0.01
*p=0.03
MFI (H4K16ac)
g
Ki67lo
Ki67hi
ns
H4K16ac INTDEN
PAK4 INTDEN
LT-HSC
ST-HSC
h
6 h
66 h
LT
ST
Merge
CDK6
H4K16ac
PAK4
DAPI
IntDen (nuclCDK6)
IntDen (H4K16ac)
IntDen (nuclPAK4)
LT-HSC
ST-HSC
hours
i
diffusion pseudotime
timepoints (h)
Max
CDK6
H4K16ac
PAK4
R² >0.70
diffusion pseudotime
nuclCDK6
H4K16ac
nuclPAK4
pseudotime

Fig. 3

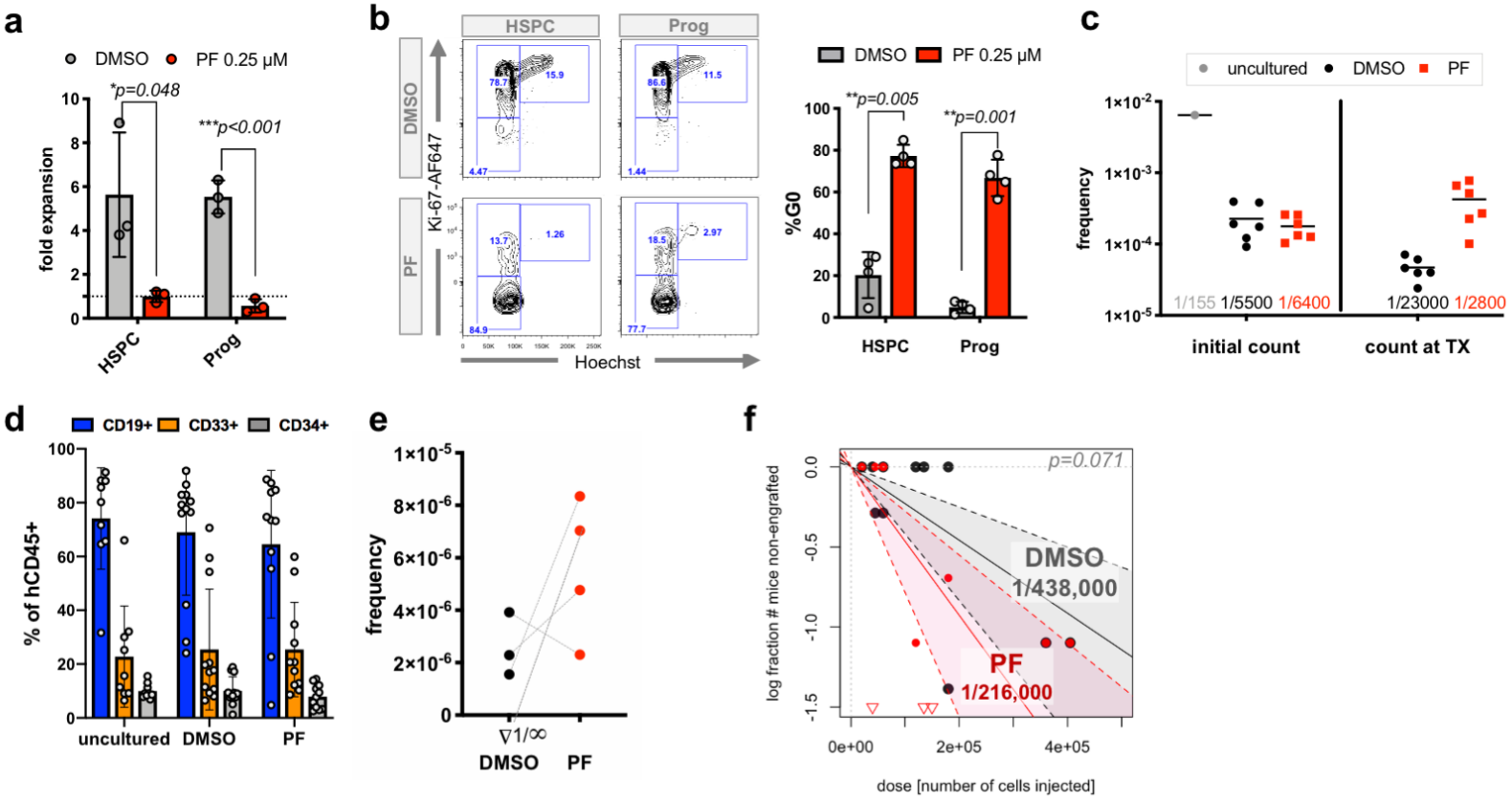

a
DMSO
PF 0.25 µM
*p=0.048
***p<0.001
fold expansion
HSPC
Prog
b
HSPC
Prog
DMSO
PF
Ki-67-AF647
Hoechst
78.7
15.9
4.47
86.6
11.5
1.44
13.7
1.26
84.9
18.5
2.97
77.7
DMSO
PF 0.25 µM
**p=0.005
**p=0.001
%G0
HSPC
Prog
c
uncultured
DMSO
PF
frequency
1/155
1/5500
1/6400
1/23000
1/2800
initial count
count at TX
d
CD19+
CD33+
CD34+
% of hCD45+
uncultured
DMSO
PF
e
frequency
∇1/∞
DMSO
PF
f
p=0.071
DMSO
1/438,000
PF
1/216,000
log fraction # mice non-engrafted
dose [number of cells injected]

Fig. 4

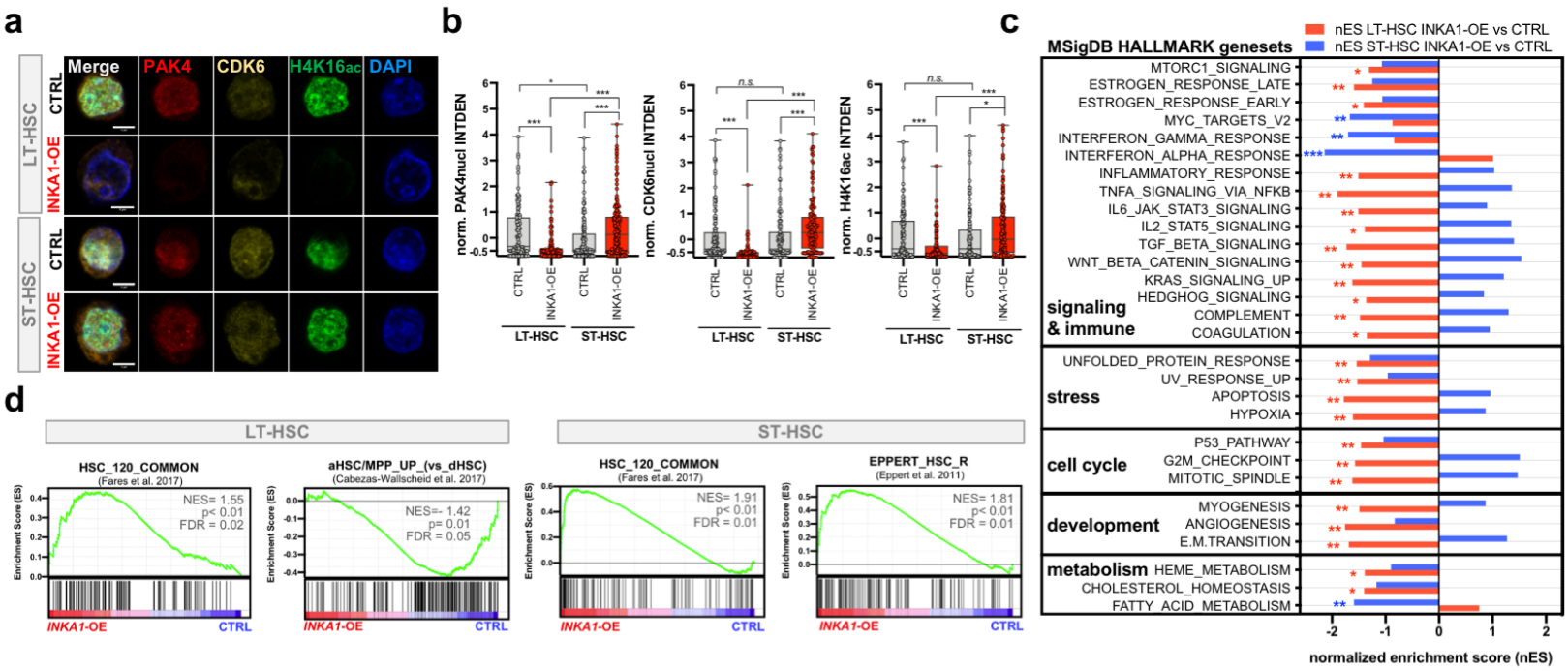

a
Merge
PAK4
CDK6
H4K16ac
DAPI
LT-HSC
ST-HSC
CTRL
INKA1-OE
b
norm. PAK4nucl INTDEN
norm. CDK6nucl INTDEN
norm. H4K16ac INTDEN
n.s.
LT-HSC
ST-HSC
c
MSigDB HALLMARK genesets
nES LT-HSC INKA1-OE vs CTRL
nES ST-HSC INKA1-OE vs CTRL
MTORC1_SIGNALING
ESTROGEN_RESPONSE_LATE
ESTROGEN_RESPONSE_EARLY
MYC_TARGETS_V2
INTERFERON_GAMMA_RESPONSE
INTERFERON_ALPHA_RESPONSE
INFLAMMATORY_RESPONSE
TNFA_SIGNALING_VIA_NFKB
IL6_JAK_STAT3_SIGNALING
IL2_STAT5_SIGNALING
TGF_BETA_SIGNALING
WNT_BETA_CATENIN_SIGNALING
KRAS_SIGNALING_UP
HEDGHOG_SIGNALING
COMPLEMENT
COAGULATION
signaling & immune
UNFOLDED_PROTEIN_RESPONSE
UV_RESPONSE_UP
APOPTOSIS
HYPOXIA
stress
P53_PATHWAY
G2M_CHECKPOINT
MITOTIC_SPINDLE
cell cycle
MYOGENESIS
ANGIOGENESIS
E.M.TRANSITION
development
HEME_METABOLISM
CHOLESTEROL_HOMEOSTASIS
FATTY_ACID_METABOLISM
metabolism
normalized enrichment score (nES)
d
LT-HSC
ST-HSC
HSC_120_COMMON
(Fares et al. 2017)
NES= 1.55
p< 0.01
FDR = 0.02
aHSC/MPP_UP_(vs_dHSC)
(Cabezas-Wallscheid et al. 2017)
NES=- 1.42
p= 0.01
FDR = 0.05
HSC_120_COMMON
(Fares et al. 2017)
NES= 1.91
p< 0.01
FDR = 0.01
EPPERT_HSC_R
(Eppert et al. 2011)
NES= 1.81
p< 0.01
FDR = 0.01
Enrichment Score (ES)
INKA1-OE
CTRL

Fig. 5

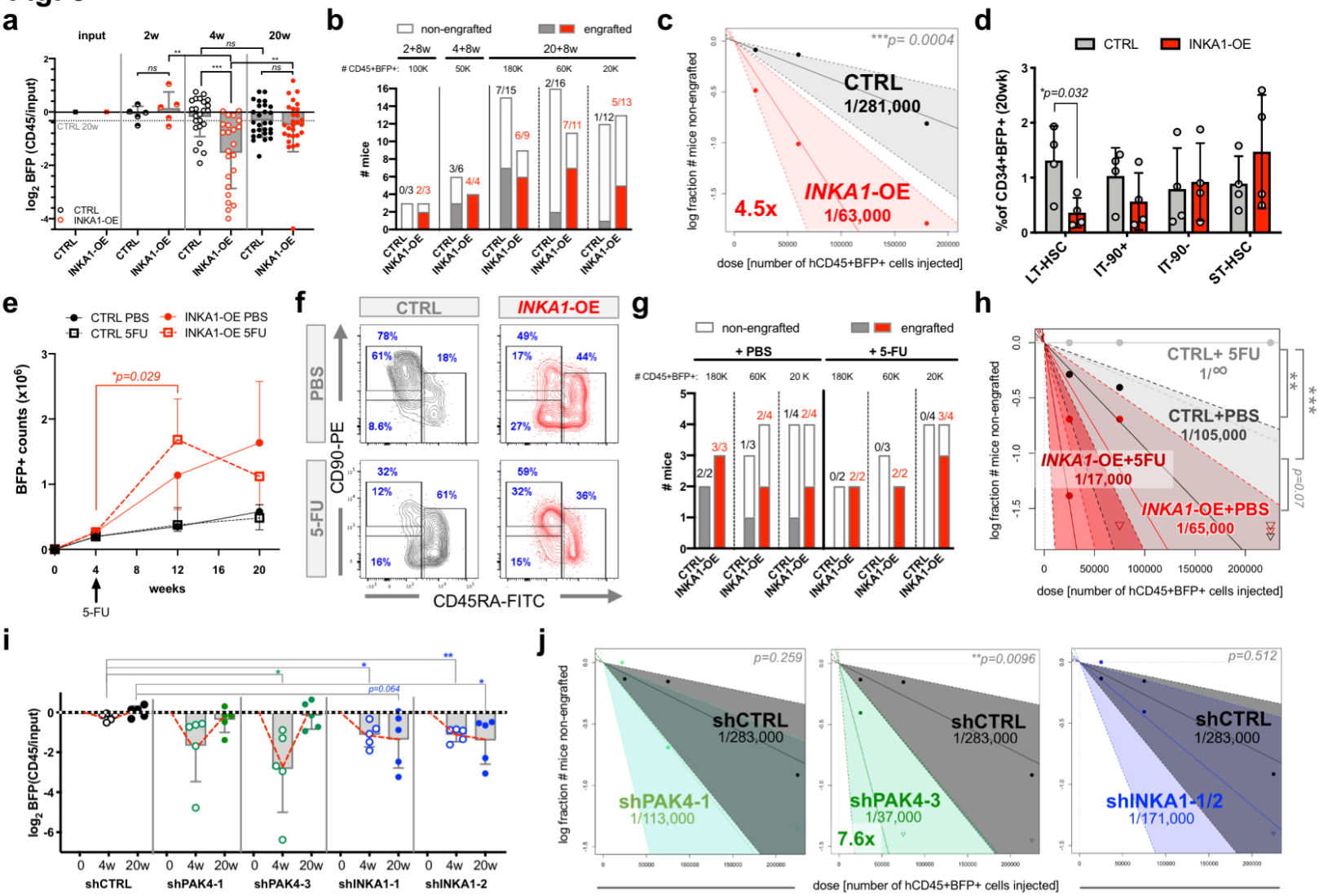

a
input
2w
4w
20w
log2 BFP (CD45/input)
CTRL 20w
CTRL
INKA1-OE
b
non-engrafted
engrafted
2+8w
4+8w
20+8w
CD45+BFP+:
100K
50K
180K
60K
20K
mice
0/3 2/3
3/6 4/4
7/15 6/9
2/16 7/11
1/12 5/13
c
***p= 0.0004
CTRL
1/281,000
INKA1-OE
1/63,000
4.5x
log fraction # mice non-engrafted
dose [number of hCD45+BFP+ cells injected]
d
CTRL
INKA1-OE
*p=0.032
%of CD34+BFP+ (20wk)
LT-HSC
IT-90+
IT-90-
ST-HSC
e
CTRL PBS
INKA1-OE PBS
CTRL 5FU
INKA1-OE 5FU
*p=0.029
BFP+ counts (x10^6)
weeks
5-FU
f
CTRL
INKA1-OE
PBS
5-FU
CD90-PE
CD45RA-FITC
78%
61%
18%
8.6%
49%
17%
44%
27%
32%
12%
61%
16%
59%
32%
36%
15%
g
non-engrafted
engrafted
+ PBS
+ 5-FU
# CD45+BFP+:
180K
60K
20 K
mice
2/2 3/3
1/3 2/4
1/4 2/4
0/2 2/2
0/3 2/2
0/4 3/4
h
CTRL+ 5FU
1/∞
CTRL+PBS
1/105,000
INKA1-OE+5FU
1/17,000
INKA1-OE+PBS
1/65,000
**
***
p=0.07
log fraction # mice non-engrafted
dose [number of hCD45+BFP+ cells injected]
i
p=0.064
log2 BFP(CD45/input)
shCTRL
shPAK4-1
shPAK4-3
shINKA1-1
shINKA1-2
j
p=0.259
shCTRL
1/283,000
shPAK4-1
1/113,000
**p=0.0096
shCTRL
1/283,000
shPAK4-3
1/37,000
7.6x
p=0.512
shCTRL
1/283,000
shINKA1-1/2
1/171,000
log fraction # mice non-engrafted
dose [number of hCD45+BFP+ cells injected]

Fig. 6

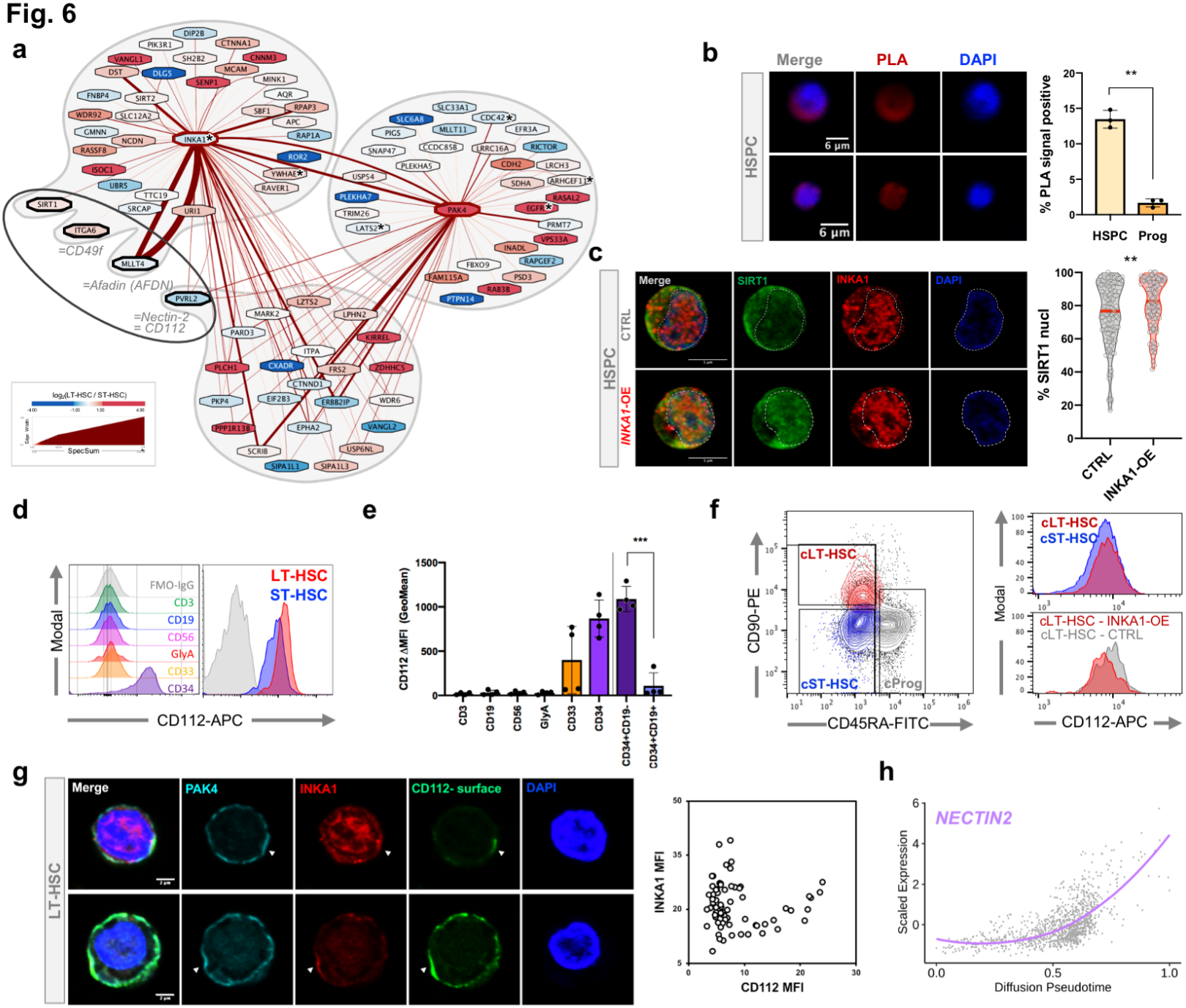

a
DIP2B
PIK3R1
CTNNA1
VANGL1
SH2B2
CNNM3
DST
DLG5
MCAM
SENP1
MINK1
FNBP4
SIRT2
AQR
WDR92
SLC12A2
SBF1
RPAP3
APC
GMNN
NCDN
INKA1*
RAP1A
RASSF8
ROR2
ISOC1
YWHAE*
UBR5
TTC19
RAVER1
SIRT1
SRCAP
URI1
ITGA6
=CD49f
MLLT4
=Afadin (AFDN)
PVRL2
=Nectin-2
= CD112
SLC33A1
SLC6A8
CDC42*
PIGS
MLLT11
EFR3A
SNAP47
CCDC85B
LRRC16A
RICTOR
PLEKHA5
CDH2
LRCH3
USP54
SDHA
ARHGEF11*
PLEKHA7
RASAL2
TRIM26
PAK4
EGFR*
LATS2*
PRMT7
VPS33A
INADL
RAPGEF2
FBXO9
TAM115A
PSD3
RAB3B
PTPN14
LZTS2
MARK2
LPHN2
PARD3
KIRREL
ITPA
PLCH1
CXADR
FRS2
ZDHHC5
CTNND1
PKP4
EIF2B3
ERBB2IP
WDR6
PPP1R13B
EPHA2
VANGL2
SCRIB
USP6NL
SIPA1L1
SIPA1L3
log2(LT-HSC / ST-HSC)
SpecSum
b
Merge
PLA
DAPI
HSPC
6 µm
% PLA signal positive
**
HSPC
Prog
c
HSPC
CTRL
INKA1-OE
Merge
SIRT1
INKA1
DAPI
% SIRT1 nucl
**
CTRL
INKA1-OE
d
Modal
FMO-IgG
CD3
CD19
CD56
GlyA
CD33
CD34
LT-HSC
ST-HSC
CD112-APC
e
CD112 ΔMFI (GeoMean)
***
CD3
CD19
CD56
GlyA
CD33
CD34
CD34+CD19-
CD34+CD19+
f
CD90-PE
cLT-HSC
cST-HSC
cProg
CD45RA-FITC
Modal
cLT-HSC
cST-HSC
cLT-HSC - INKA1-OE
cLT-HSC - CTRL
CD112-APC
g
LT-HSC
Merge
PAK4
INKA1
CD112- surface
DAPI
INKA1 MFI
CD112 MFI
h
NECTIN2
Scaled Expression
Diffusion Pseudotime

Fig. 7

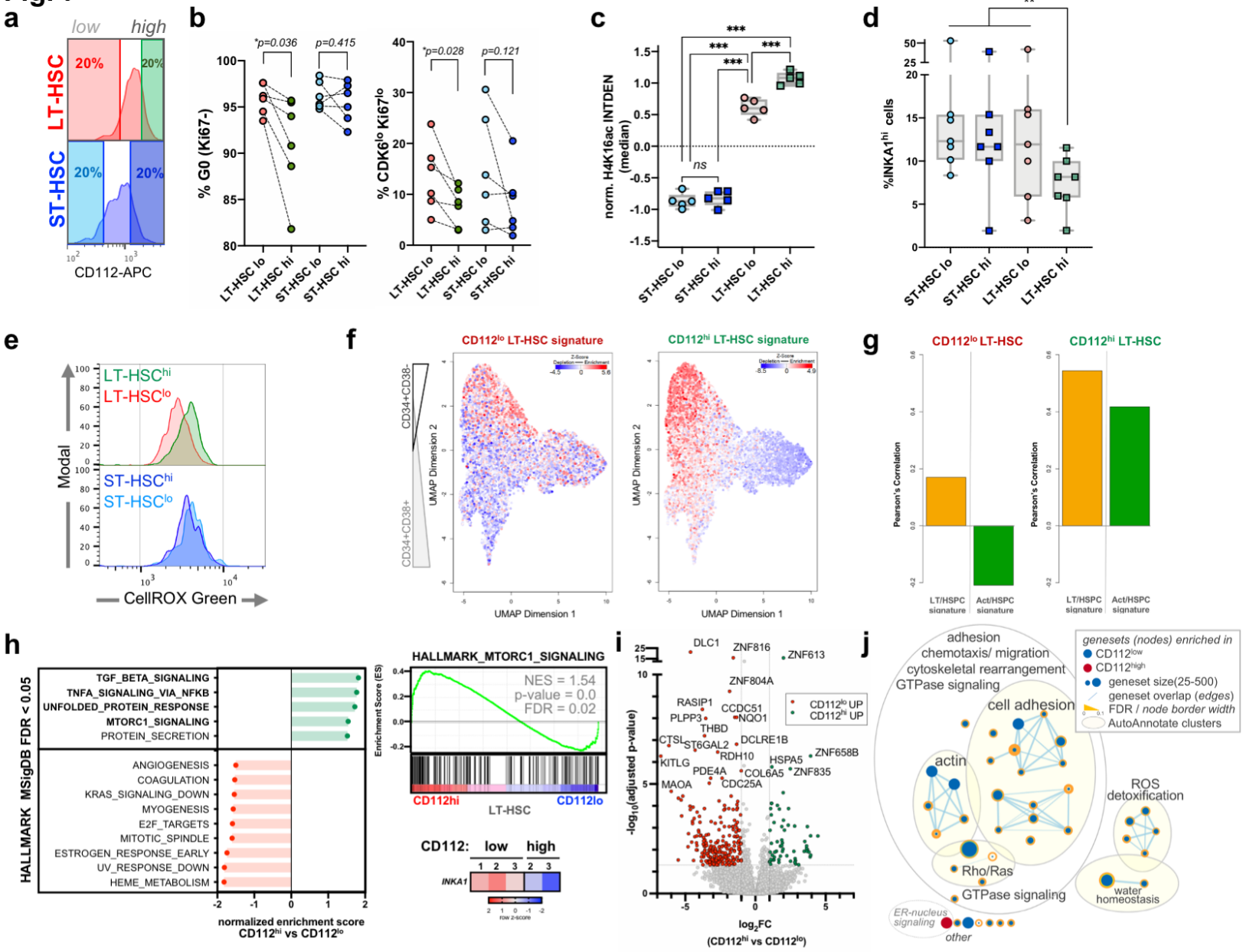
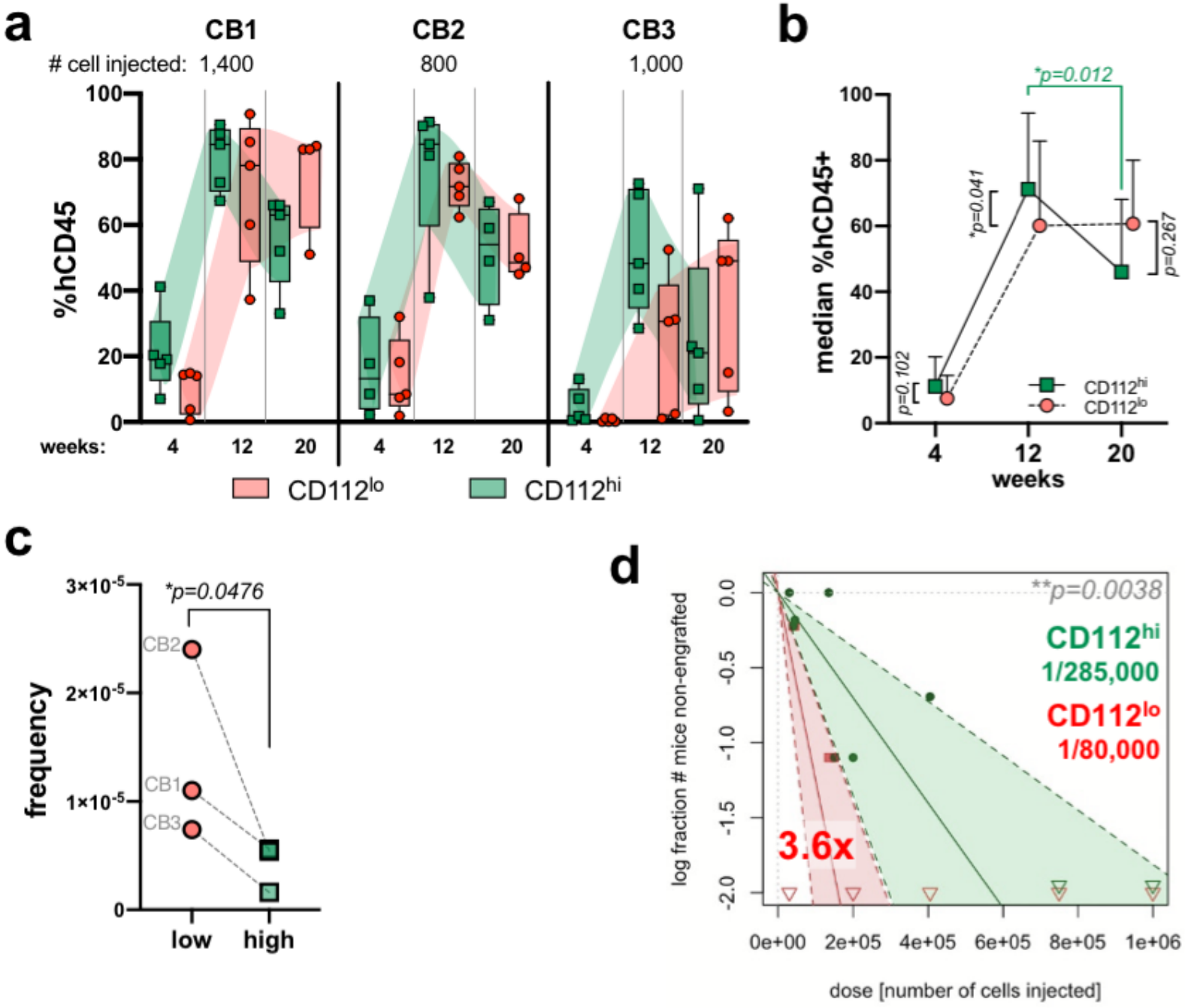
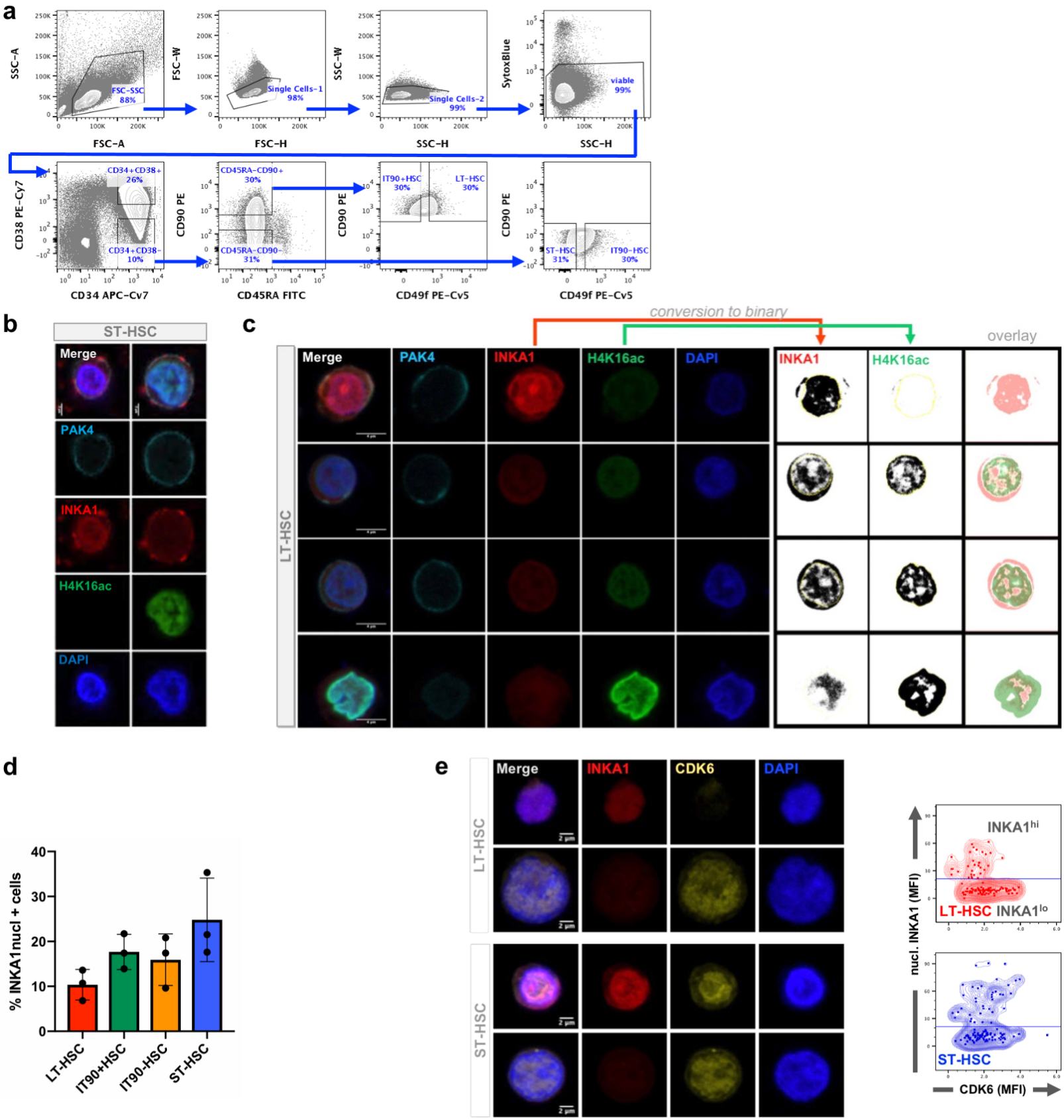

a
low
high
LT-HSC
ST-HSC
20%
20%
20%
20%
CD112-APC
b
*p=0.036
p=0.415
*p=0.028
p=0.121
% G0 (Ki67-)
% CDK6lo Ki67lo
LT-HSC lo
LT-HSC hi
ST-HSC lo
ST-HSC hi
c
norm. H4K16ac INTDEN (median)
***
***
***
***
ns
ST-HSC lo
ST-HSC hi
LT-HSC lo
LT-HSC hi
d
%INKA1hi cells
**
ST-HSC lo
ST-HSC hi
LT-HSC lo
LT-HSC hi
e
LT-HSChi
LT-HSClo
ST-HSChi
ST-HSClo
Modal
CellROX Green
f
CD112lo LT-HSC signature
CD112hi LT-HSC signature
CD34+CD38-
CD34+CD38+
UMAP Dimension 2
UMAP Dimension 1
UMAP Dimension 1
g
CD112lo LT-HSC
CD112hi LT-HSC
Pearson's Correlation
LT/HSPC signature
Act/HSPC signature
LT/HSPC signature
Act/HSPC signature
h
HALLMARK MSigDB FDR < 0.05
TGF_BETA_SIGNALING
TNFA_SIGNALING_VIA_NFKB
UNFOLDED_PROTEIN_RESPONSE
MTORC1_SIGNALING
PROTEIN_SECRETION
ANGIOGENESIS
COAGULATION
KRAS_SIGNALING_DOWN
MYOGENESIS
E2F_TARGETS
MITOTIC_SPINDLE
ESTROGEN_RESPONSE_EARLY
UV_RESPONSE_DOWN
HEME_METABOLISM
normalized enrichment score
CD112hi vs CD112lo
HALLMARK_MTORC1_SIGNALING
NES = 1.54
p-value = 0.0
FDR = 0.02
Enrichment Score (ES)
CD112hi
CD112lo
LT-HSC
CD112: low high
INKA1
i
DLC1
ZNF816
ZNF613
ZNF804A
RASIP1
CCDC51
PLPP3
NQO1
THBD
CTSL
ST6GALT2
DCLRE1B
RDH10
ZNF658B
KITLG
HSPA5
PDE4A
COL6A5
ZNF835
MAOA
CDC25A
CD112lo UP
CD112hi UP
-log10(adjusted p-value)
log2FC
(CD112hi vs CD112lo)
j
adhesion
chemotaxis/ migration
cytoskeletal rearrangement
GTPase signaling
cell adhesion
actin
ROS detoxification
Rho/Ras
GTPase signaling
water homeostasis
ER-nucleus signaling
other
genesets (nodes) enriched in
CD112low
CD112high
geneset size(25-500)
geneset overlap (edges)
FDR / node border width
AutoAnnotate clusters

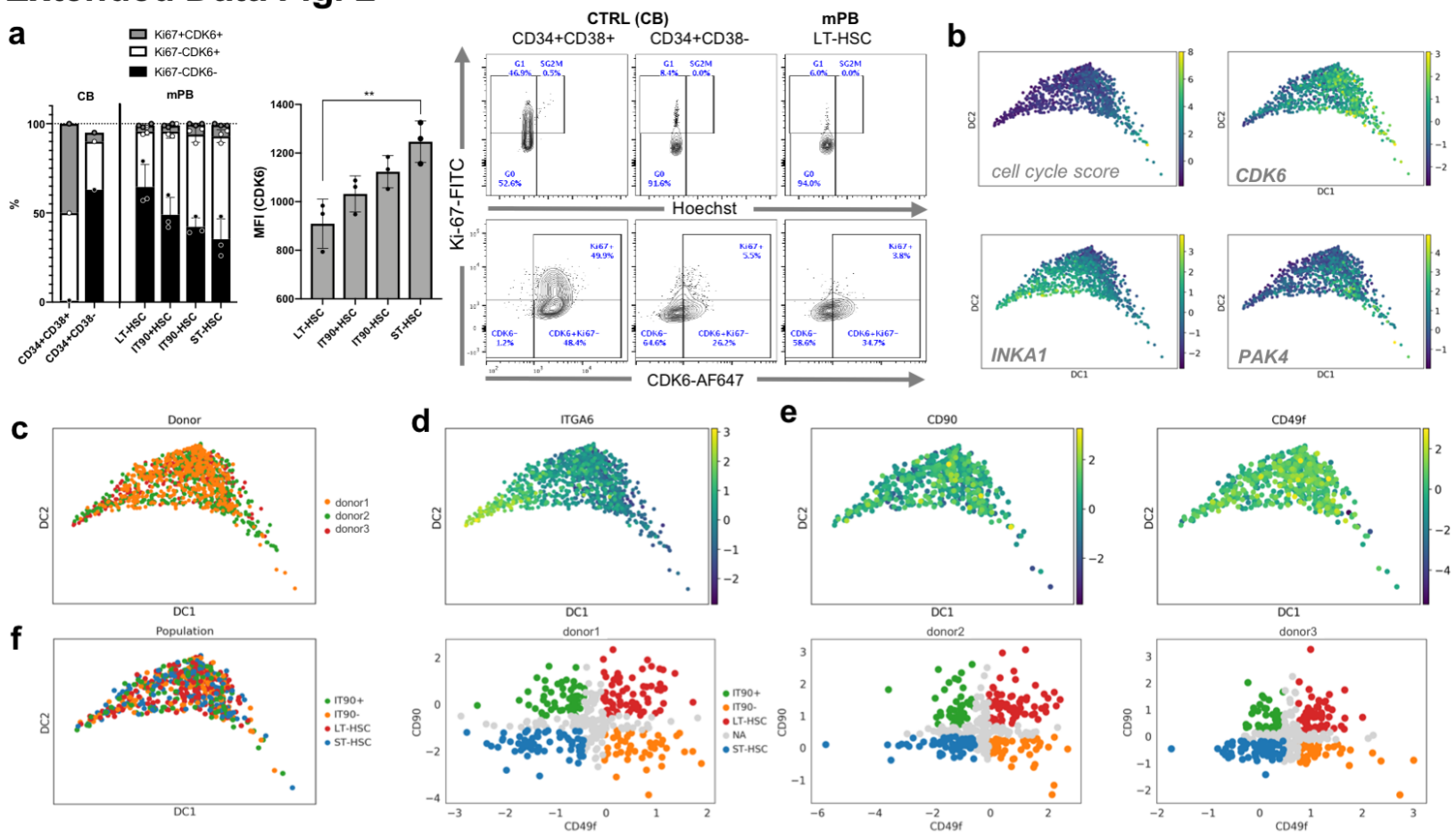
Fig. 8


a
CB1
CB2
CB3
cell injected: 1,400
800
1,000
%hCD45
weeks:
4
12
20
CD112lo
CD112hi
b
median %hCD45+
*p=0.012
*p=0.041
p=0.267
p=0.102
CD112hi
CD112lo
weeks
c
frequency
*p=0.0476
CB2
CB1
CB3
low
high
d
**p=0.0038
CD112hi
1/285,000
CD112lo
1/80,000
3.6x
log fraction # mice non-engrafted
dose [number of cells injected]

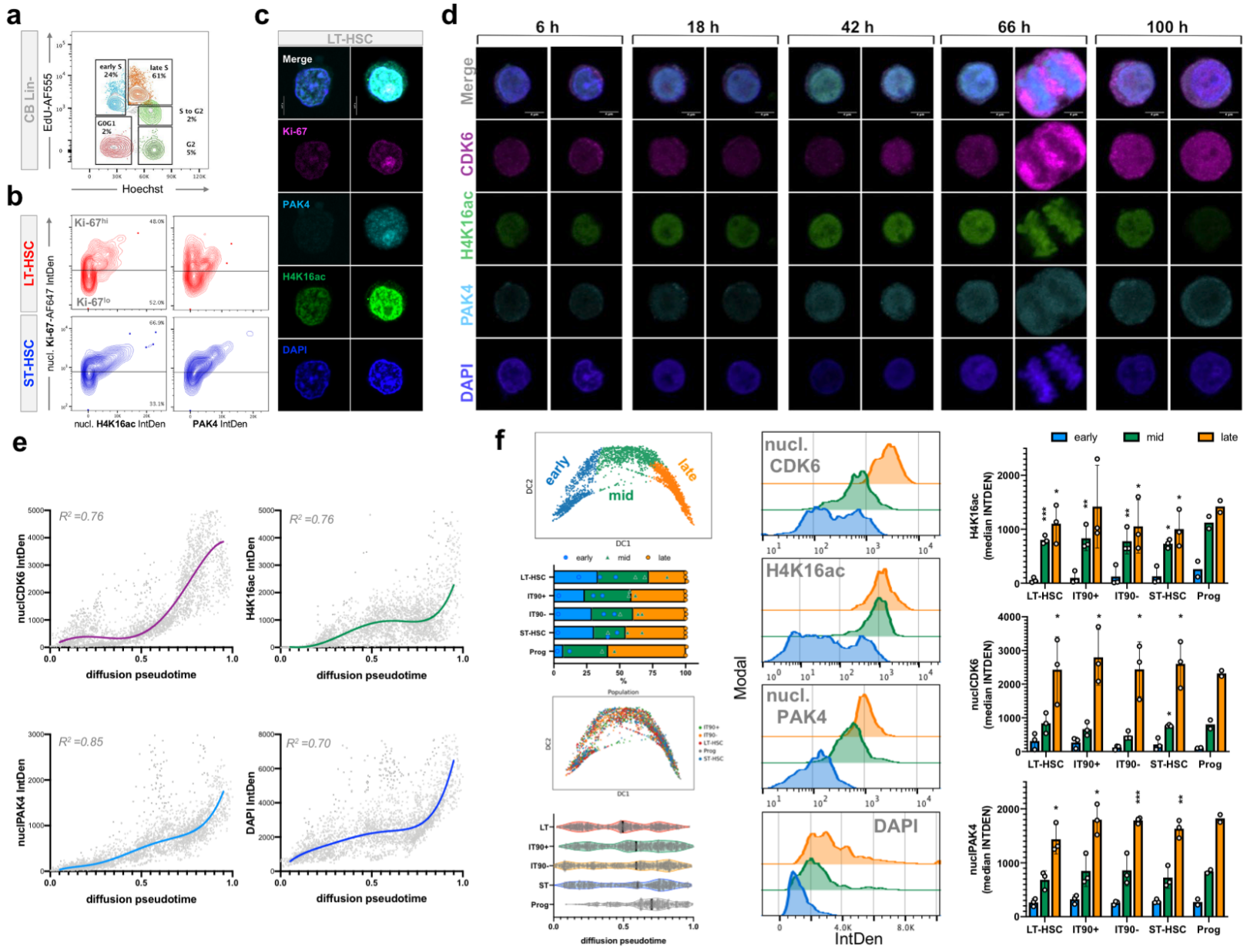
Extended Data Fig.1



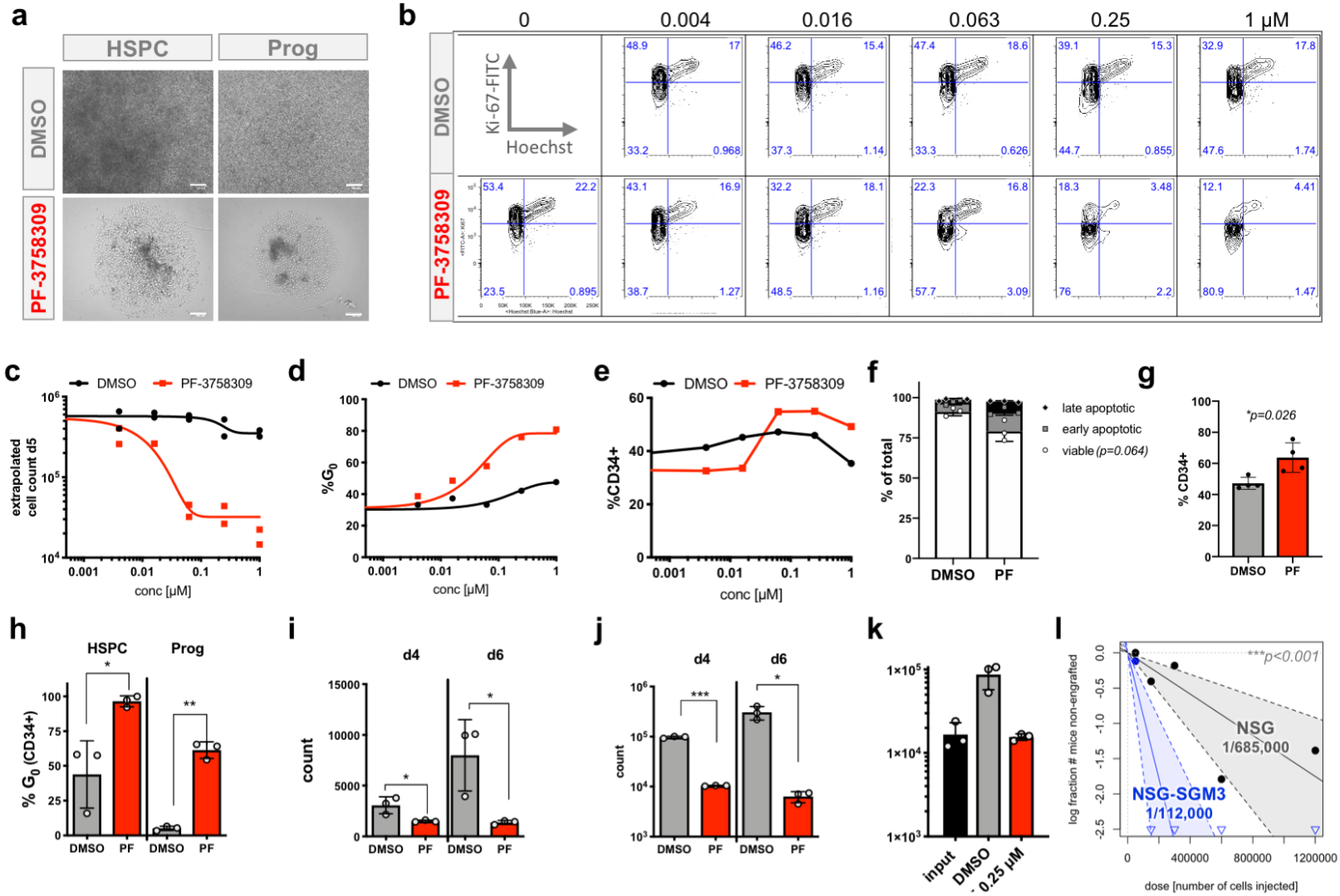
Extended Data Fig. 2



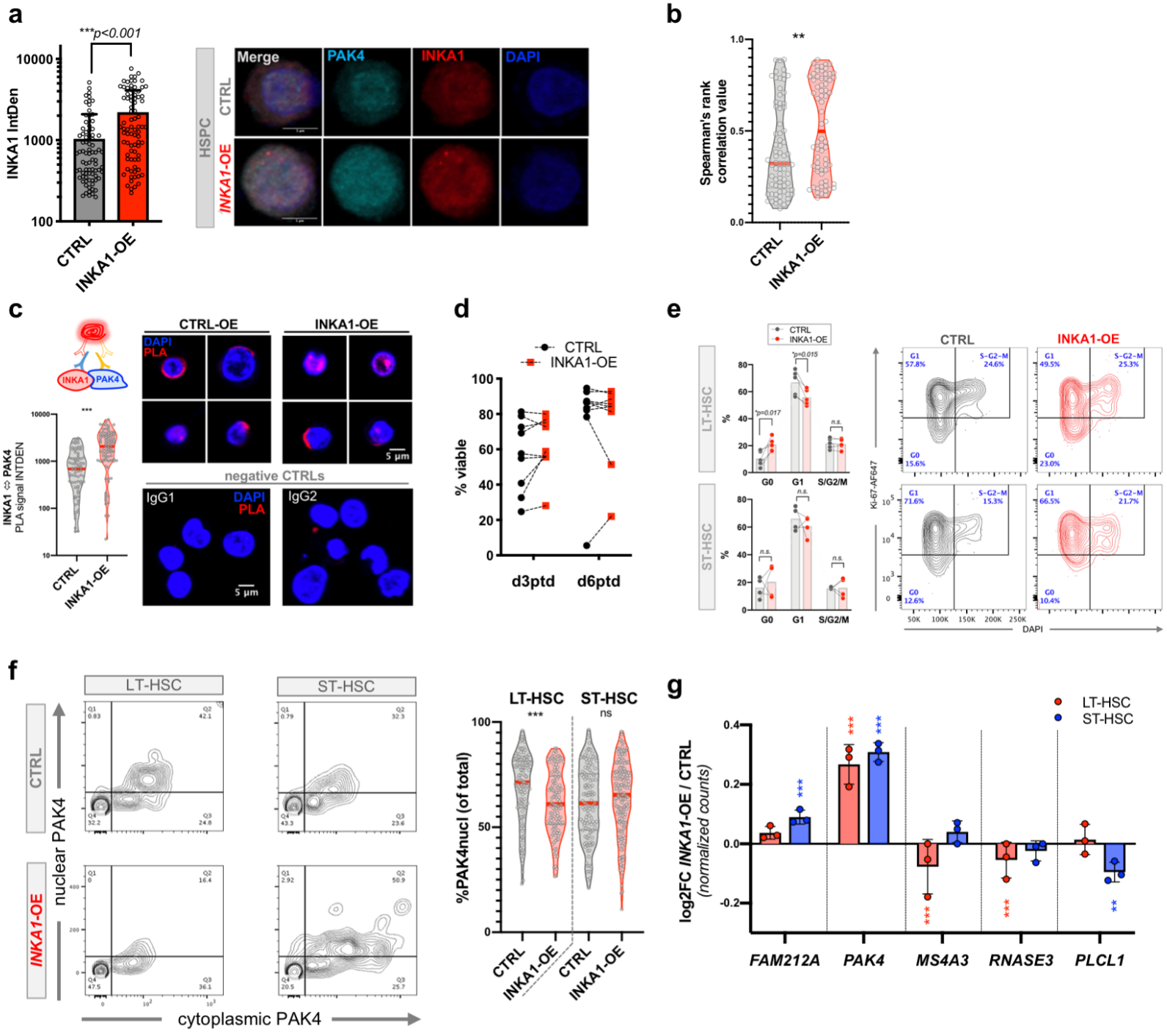
Extended Data Fig. 3



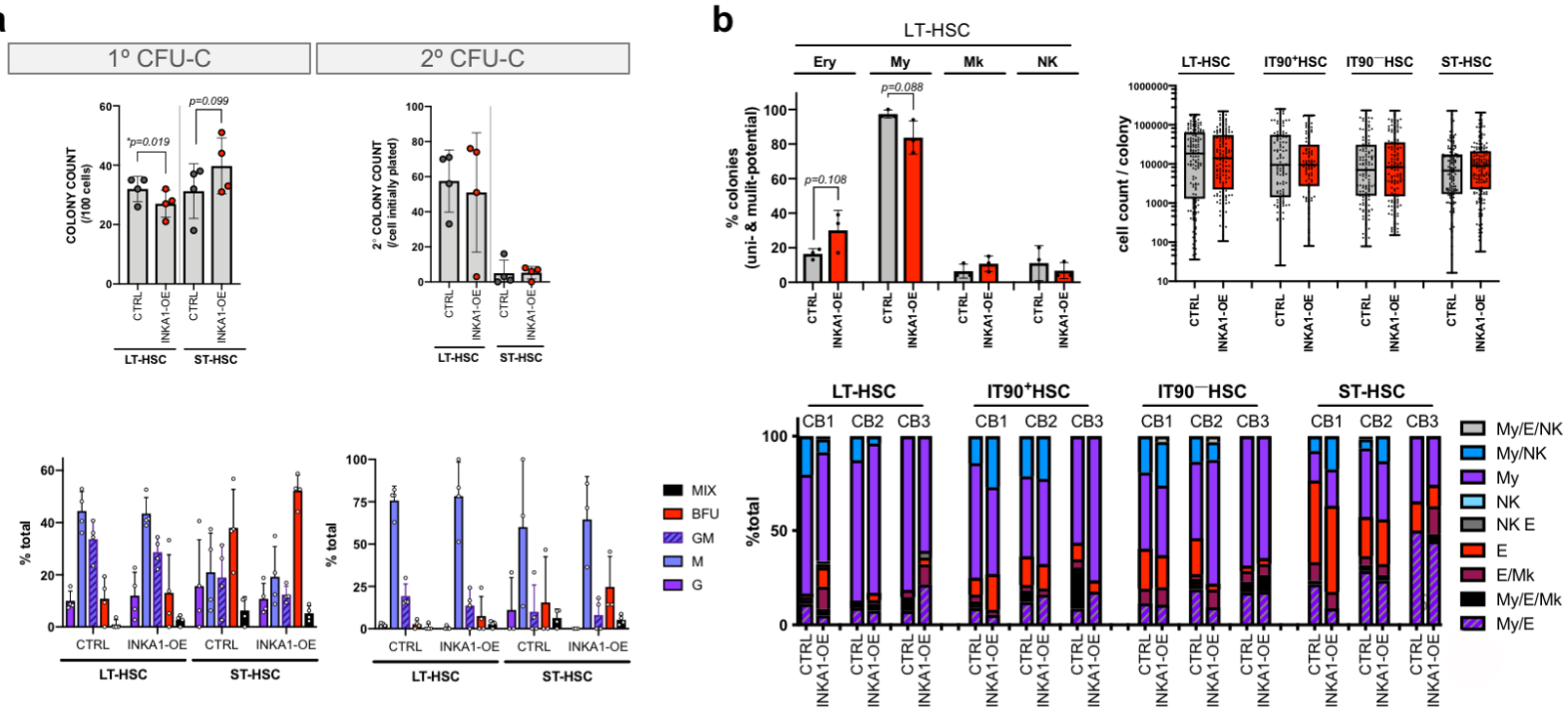
Extended Data Fig. 4



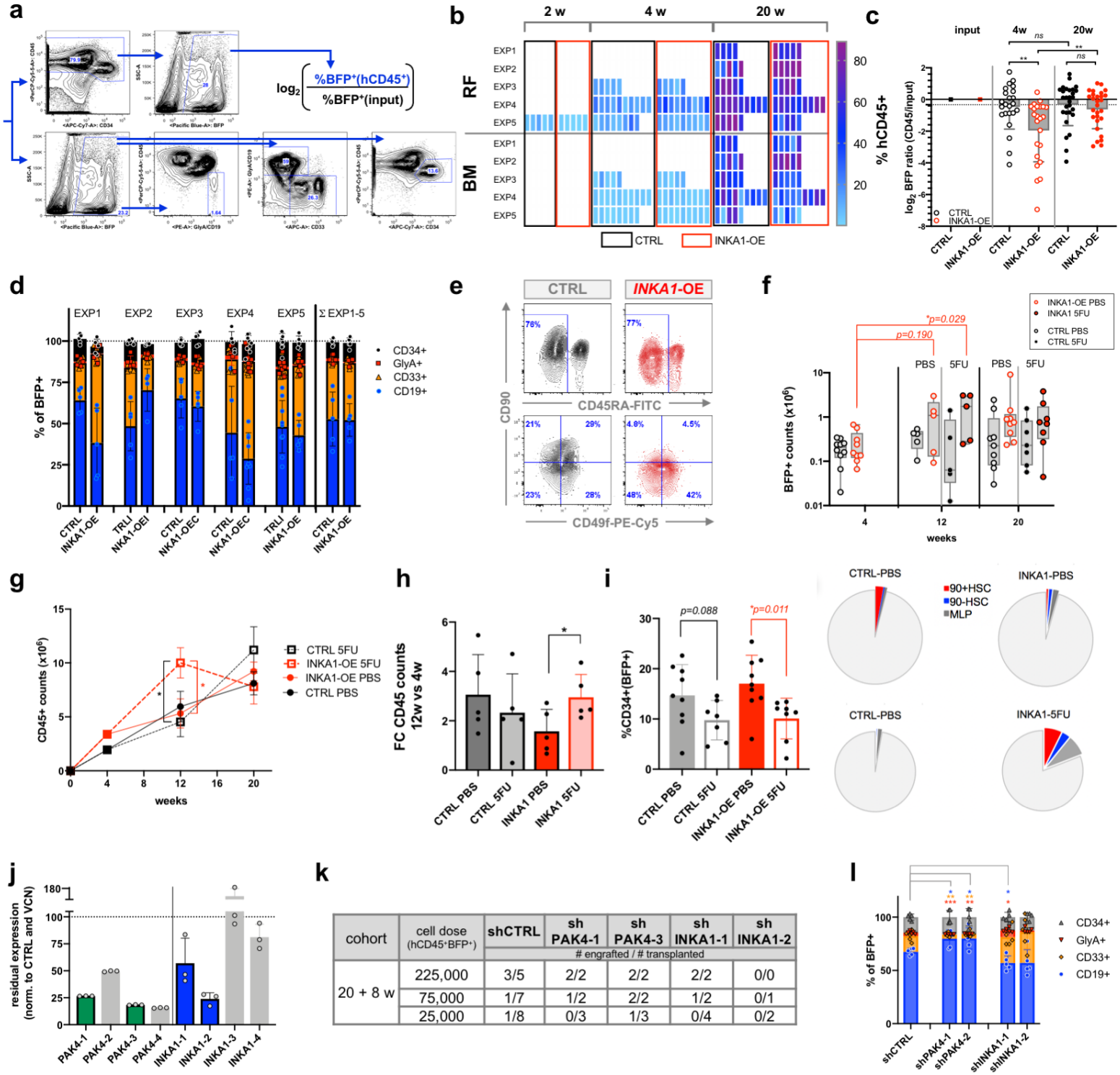
Extended Data Fig. 5



Extended Data Fig. 6



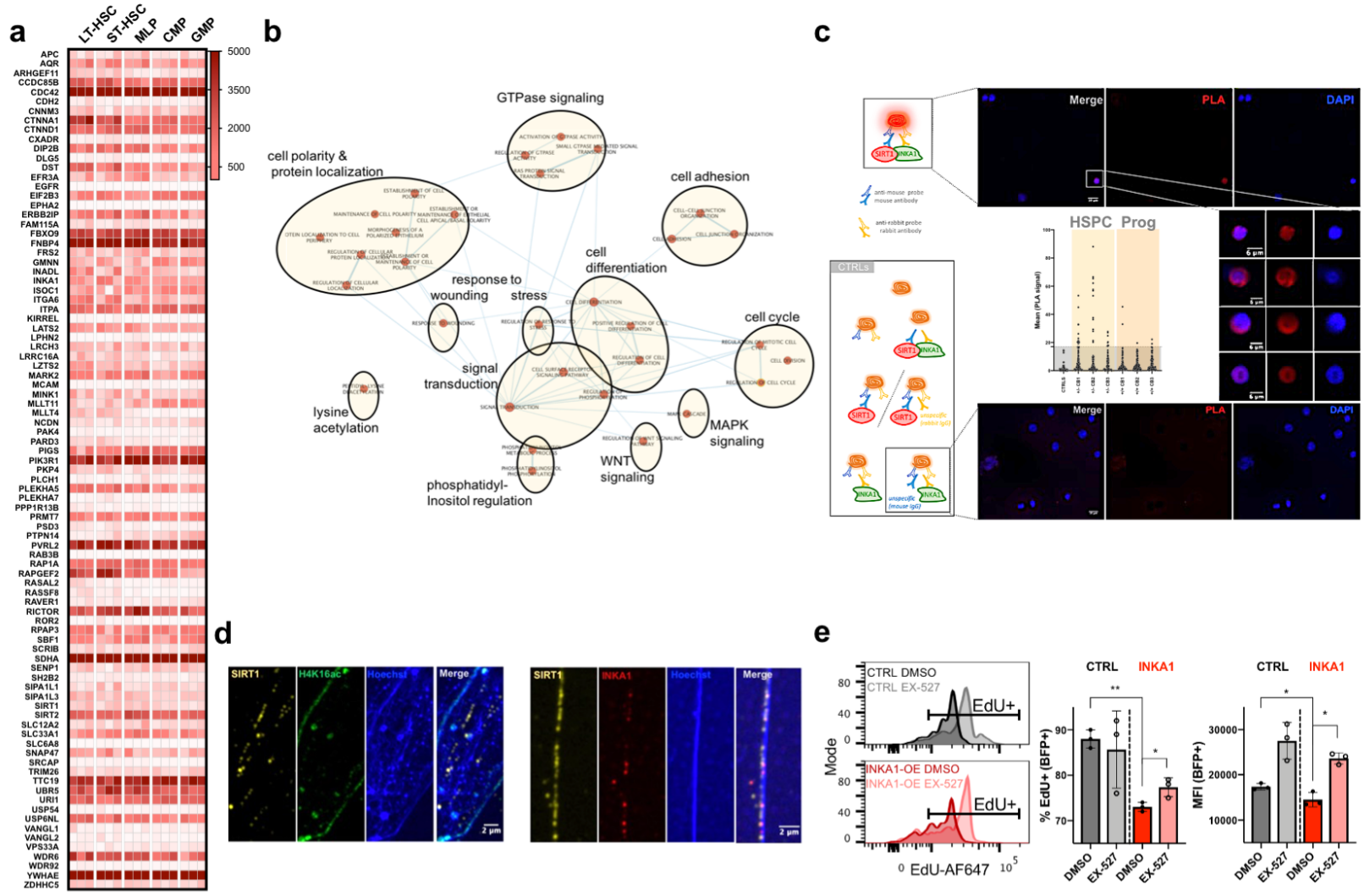
Extended Data Fig. 7



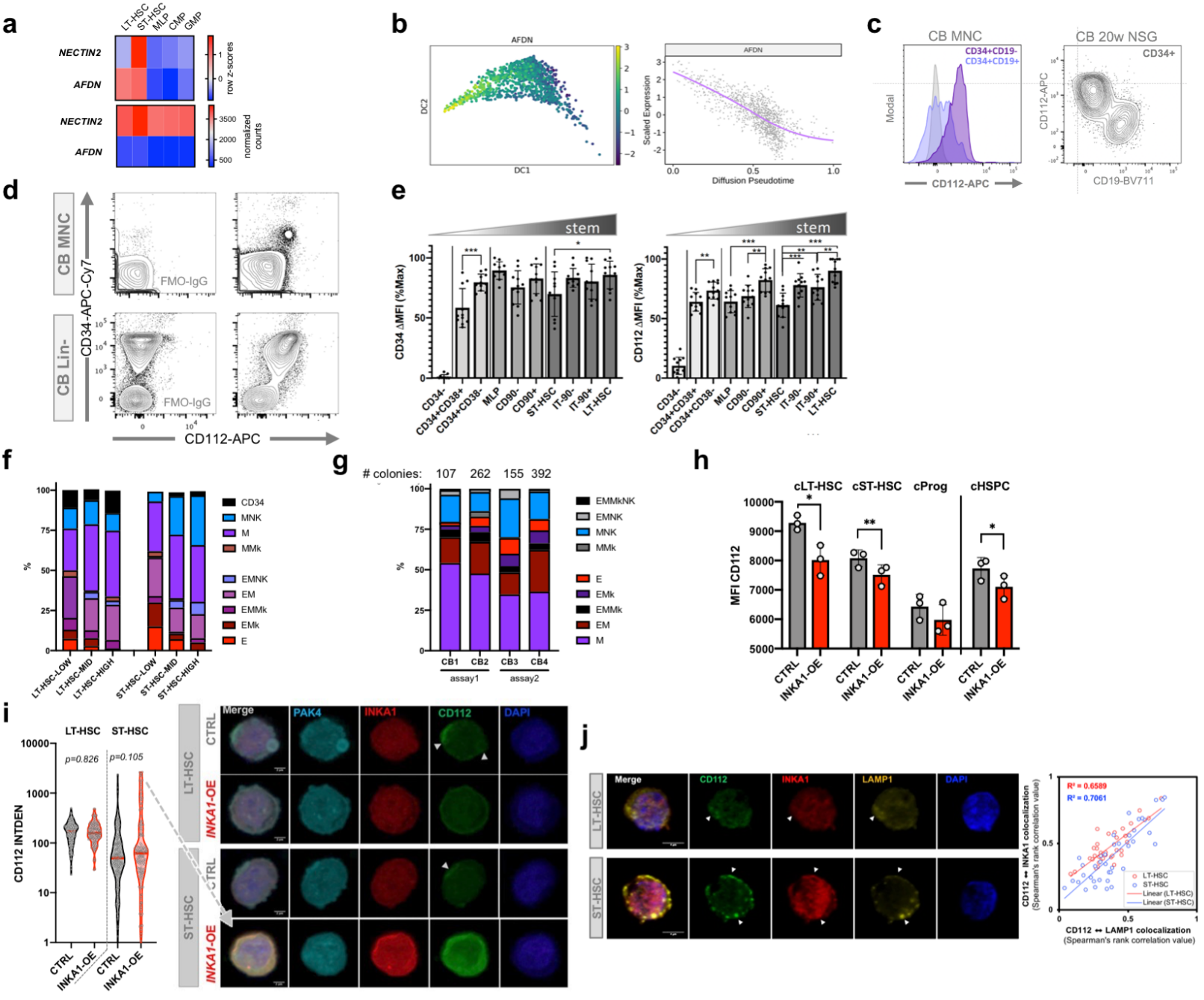
k

cohort	cell dose (hCD45⁺BFP⁺)	shCTRL	sh PAK4-1	sh PAK4-3	sh INKA1-1	sh INKA1-2
		# engrafted / # transplanted				
20 + 8 w	225,000	3/5	2/2	2/2	2/2	0/0
	75,000	1/7	1/2	2/2	1/2	0/1
	25,000	1/8	0/3	1/3	0/4	0/2

Extended Data Fig. 8



Extended Data Fig. 9



Extended Data Fig. 10