SUPPLEMENTARY INFORMATION

Combined single-cell tracking and omics improves blood stem cell fate regulator identification

A. Wehling[1], D. Loeffler[1], Y. Zhang[1], T. Kull[1], C. Donato[2], B. Szczerba[2], G. Camargo Ortega[1], M. Lee[1], A. Moor[1], B. Göttgens[3], N. Aceto[2,4], T. Schroeder[1]*

[1]Department of Biosystems Science and Engineering (D-BSSE), ETH Zurich, Mattenstrasse 26, 4058 Basel, Switzerland.

[2]Cancer Metastasis Lab, Department of Biomedicine, University of Basel, 4058 Basel, Switzerland

[3]Wellcome-MRC Cambridge Stem Cell Institute, Department of Haematology, Jeffrey Cheah Biomedical Centre, University of Cambridge, Cambridge, UK.

[4]Department of Biology, Institute of Molecular Health Sciences, ETH Zürich, Otto-Stern-Weg 7, 8093 Zürich, Switzerland

*Correspondence to: timm.schroeder@bsse.ethz.ch

Supplementary figure legends

Supplementary Fig. 1: Gene count in picked cells and their mapping to freshly sorted HSCs from previous studies. A, Gating strategy for HSC isolation from murine bone marrow. **B,** Single-cell cDNA libraries passing/failing QC can clearly be distinguished. Histogram of number of genes detected per cell of all cells sequenced before and after quality control. **C,** Culture has little impact on HSC transcriptomes: Isolated HSC daughters (blue) in G1 map closely to HSCs (green) in reference landscape of early adult hematopoiesis. **D,** CC progression of in vitro cultured HSC daughter cells. Transcriptionally inferred CC stages shown in black (G1), blue (S) and orange (G2M). Grey: freshly isolated stem and progenitor cells from reference data. **E,** HSCs in G1 co-localize with freshly isolated HSCs expressing highest levels of stem cell signature MolO [1]. F – I, HSC daughters and granddaughters transcriptionally map between HSCs and early MPPs. **F,** Cell types (defined by protein surface marker expression via FACS) within the transcriptional landscape of early hematopoiesis by Nestorowa et al., 2016. **G,** The landscape is dominated by the CC and cell type differences, which are almost orthogonal to each other. CC stages inferred with Seurat (Stuart et al., 2019). **H,** HSC daughter cells analyzed by trackSeq, map between HSCs and early MPPs. I, HSC granddaughters also still remain mostly in the pre-MPP transcriptome space.

Supplementary Fig. 2: trackSeq reveals expression differences from intra-clonal differences that correlate with CFD. A, trackSeq identifies expression differences between daughter cells with the same directionality as the observed CFD (here relative lysosome inheritance). Expression of every gene is tested with a paired-t-test if expression is higher or lower in LysoLow than in LysoHigh daughter cell. The resulting t-statistic is used to create a bi-polar ranking. The top and bottom 250 genes (LysoHigh/LysoLow candidates, respectively) are the top 500 trackSeq candidates. **B**, trackSeq uncovers subtle expression differences between ACD HSC daughter cells. Sister expression differences of top 250 LysoHigh and LysoLow candidates. Rows sorted per candidate, columns sorted by mean expression difference. Note the enrichment of sister pairs without difference (white) in LysoLow candidates. This is caused by lower expression and thus higher dropout rate of LysoLow candidates. However, trackSeq is still able to identify these candidates despite being close to the detection threshold. **C**, Most trackSeq candidate differences remain stable over CC. trackSeq candidate differences between all daughter pairs, sorted by average CC progression. Top 500 trackSeq candidate differences between all daughter pairs, sorted by each pairs' average CC progression. 250 unregulated genes added as visual aid. **trackSeq candidate expression or sister cell difference is not correlating with absolute lysosome numbers but with ACD**. **D,** Binning of HSC daughter pairs by absolute lysosome levels after division (z-norm). Lysosome levels for all 203 HSC daughter pairs used by trackSeq. **E,** Differential gene expression analysis between lysosome high and low group find genes correlating with absolute lysosome levels. Top 500 differentially expressed (DE) genes in blue. **F,** 85% of trackSeq candidates are independent of absolute lysosome levels. Venn diagram between top 500 trackSeq candidates and top 500 differentially expressed genes between lysosome high and low group. **Benchmarking trackSeq candidates. G**, Test of trackSeq candidates for ACD- versus SCD- sister expression differences. trackSeq provides expression differences for all daughter pairs, plotted as mean Log_2 fold change (Log2FC) in a-(ACD, lysosome ratio > 1.6x, 59 pairs) and symmetrically (SCD, lysosome ratio < 1.3x, 94 pairs) divided HSC daughter pairs. For differences to be ACD specific, we require LysoHigh candidates to have an ACD Log2FC > 0 (expressed

higher in LysoHigh than LysoLow sister) and ACD Log2FC > SCD Log2FC (expression difference is greater in ACD than in SCD pairs). We also require LysoLow candidates to have an ACD Log2FC < 0 (expressed higher in LysoLow than LysoHigh sister) and ACD Log2FC<SCD Log2FC (expression difference is greater in ACD than in SCD pairs). trackSeq candidates that pass filter in black. Orange candidates selected for experimental validation. **H**, Top scoring trackSeq candidates are more ACD-specific. Incremental expansion of trackSeq LysoLow (left) and LysoHigh (right) candidate list from top 50 to top 2000 genes. Dashed lines indicate trackSeq candidate set used for Figure 2 (250 candidates, Data S4) and trackSeq network analysis (500 candidates). **I**, **J** Paired expression between SCD and ACD pairs for LysoLow **(I)** and LysoHigh **(J)** validation candidates. Mean Log2FC indicated.

Supplementary Fig. 3: Validation of trackSeq candidates at the protein level. A, Selected candidates range from weak to strong expression in HSCs. Top 500 trackSeq candidates in blue, selected LysoLow (Map2k4, Bace1) and LysoHigh candidates (Tgm2, Ecd) in orange, rest of HSC genes in grey. **B**, Experimental workflow. Time lapse culture identical to trackSeq but endpoint scSeq is replaced by immunostainings of candidates after 44 hours of culture. **C**, Protein expression validates trackSeq candidate identification. Scale bar = 5 μm. Total protein content of all candidates increases with CC progression. For comparison we therefore binned cells in TSD intervals and SCD (lysosome ratio < 1.3x, center-right) or ACD (lysosome ratio > 1.6x, right) daughter pairs. ACD LysoHigh daughters express higher ECD levels than their LysoLow sisters in Interval II and III. Interval I - II - III: 13/22 - 39/30 – 28/20 ACD/SCD pairs. 5 replicates. P-values from Wilcoxon signed-rank test throughout figure. MAP2K4 exhibits lower protein levels in ACD LysoLow daughter cells in Interval I, to a lesser extent II and then equilibrates in III. Interval I – II - III: 16/16 – 24/31 – 15/11 ACD/SCD pairs. 3 replicates. TGM2 does not exhibit protein expression asymmetries. Interval I – II- III: 18/18 – 61/64 – 12/12 ACD/SCD pairs. 4 replicates. BACE1 does not exhibit protein expression asymmetries. Interval I- II - III: 16/14 – 45/37 – 31/28 ACD/SCD pairs. 4 replicates.

Supplementary Fig. 4: trackSeq LysoLow and LysoHigh candidate set analysis. A, LysoHigh candidates display a varied expression pattern, while LysoLow candidates are enriched for lowly expressed genes. Top 500 LysoLow and LysoHigh candidates plotted on density plot of gene expression landscape. **B,** Most interactions in trackSeq candidate set originate from one large network. Top 500 LysoHigh (black) and LysoLow (white) genes mapped together to STRING database. **C,** Enrichments of Reactome pathways, Ras and Rap1 signaling and Wikipathways, Wnt and Vegf signaling shown within big network. **D,** Cell cycle GO process is enriched in the large network. FDR of enrichment indicated.

Supplementary Fig. 5: The systematic CC difference after ACD is neutralized by targeting trackSeq candidate proteins. A, Experimental setup. HSCs are isolated and cultured. After 24h of culture, shortly before the earliest HSC divisions, inhibitors are added and time lapse imaging started to capture divisions and entire daughter CC. **B,** Inhibiting lysosomes elongates the CC but inhibition of most trackSeq candidate proteins does not. ACD (lysosome sister ratio > 1.6x throughout figure) daughter cells shown per condition. P-values from Wilcoxon rank sum test against, < 0.05 highlighted throughout figure. Blue line indicates median of None condition (n=82 cells, 3 replicates). DMSO (n=56, 5 replicates); NH4Cl (1 μM: n = 38, 10 μM: n = 46; 2 replicates),

Bafilomycin (0.5 nM: n = 42; 1 nM: n = 32; 2 replicates). Compound 5 (0.2 μM: n = 54, 2 replicates, 2 μM: n = 122, 4 replicates); NSC15520 (5 μM: n = 120; 50 μM: n = 106, 4 replicates); Pyrcoumin (0.5 μM: n = 74; 5 μM: n = 122, 4 replicates); NSC663246 (34 nM: n = 20, 1 replicate; 340 nM: n= 102, 4 replicates). **C,** Inhibiting trackSeq candidate proteins neutralizes the systematic CC difference after ACD while unspecific inhibition does not. Orange line indicates lack of ACD CC effect. Note that daughter pairs treated with lysosome inhibitors causing CC elongation still display CC effect. (Left) Systematic CC difference after ACD removed by inhibition of trackSeq candidates Rad9a, Dctpp1 and Cdc25a. Daughter pairs of cells from panel a shown. (Right) Relative CC difference not existing between between SCD (lysosome sister ratio < 1.3x) daughter cell pairs. None (n = 41, 3 replicates), DMSO (n=287, 5 replicates); NH4Cl (1 μM: n = 183, 10 μM: n = 203; 2 replicates), Bafilomycin (0.5 nM: n = 225; 1 nM: n = 196; 2 replicates). Compound 5 (0.2 μM: n = 34, 2 replicates, 2 μM: n = 70, 4 replicates); NSC15520 (5 μM: n = 120; 50 μM: n = 106, 4 replicates); Pyrcoumin (0.5 μM: n = 76; 5 μM: n = 87, 4 replicates); NSC663246 (24 nM: n = 20, 1 replicate; 340 nM: n= 73, 4 replicates). P-values from Wilcoxon signed rank test of LysoHigh versus LysoLow daughter CC.

Supplementary Fig. 6: Sister difference convergence later in CC occurs mostly in subset of CC dependent LysoHigh transcripts, while LysoLow candidates mostly diverge. A, trackSeq identified daughter differences do not generally converge later in CC. Mean of absolute sister expression differences in Log2(counts+1) throughout figure, across all trackSeq top 1000 candidates and sister pairs plotted. Line = loess fit. **B**, Identification of trackSeq genes converging and diverging later in CC. Pearson correlation computed between absolute sister expression difference and CC progression for each trackSeq candidate and binned into four bins (bin color code used throughout figure)**. C**, Absolute sister expression differences across the CC in the different bins. Means plotted as colored lines. Converging genes of bin 1 are mostly CC-related transcripts enriched in LysoHigh candidates. In contrast, most LysoLow candidates diverge with CC progression (see pie charts)**.** P-values from hypergeometric test. Line = loess fit (left). Bin 1 genes are centered around the interaction region of DNA metabolic processes (CC-related)**.**

Supplementary figures

Figure S1

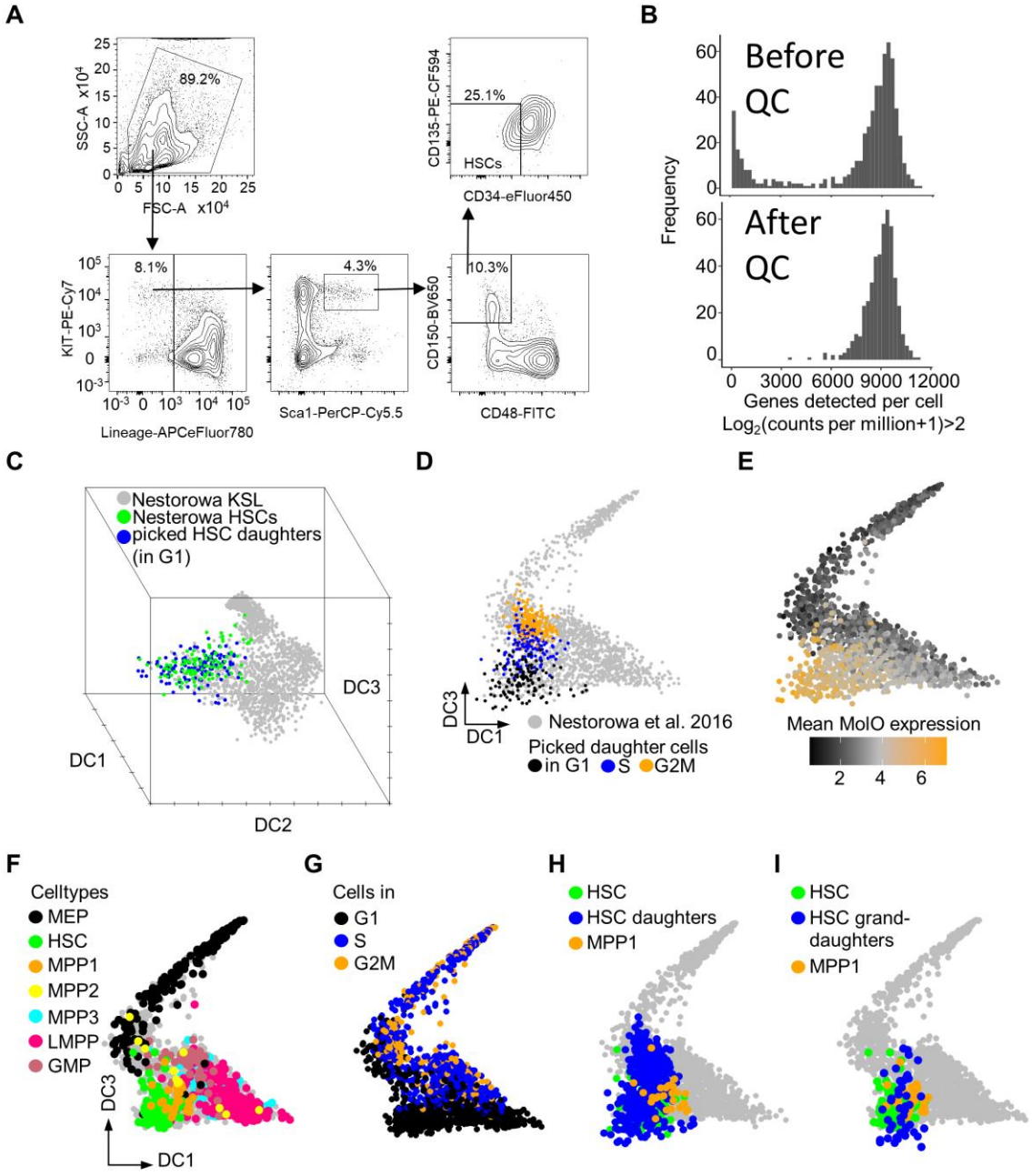


Figure S2

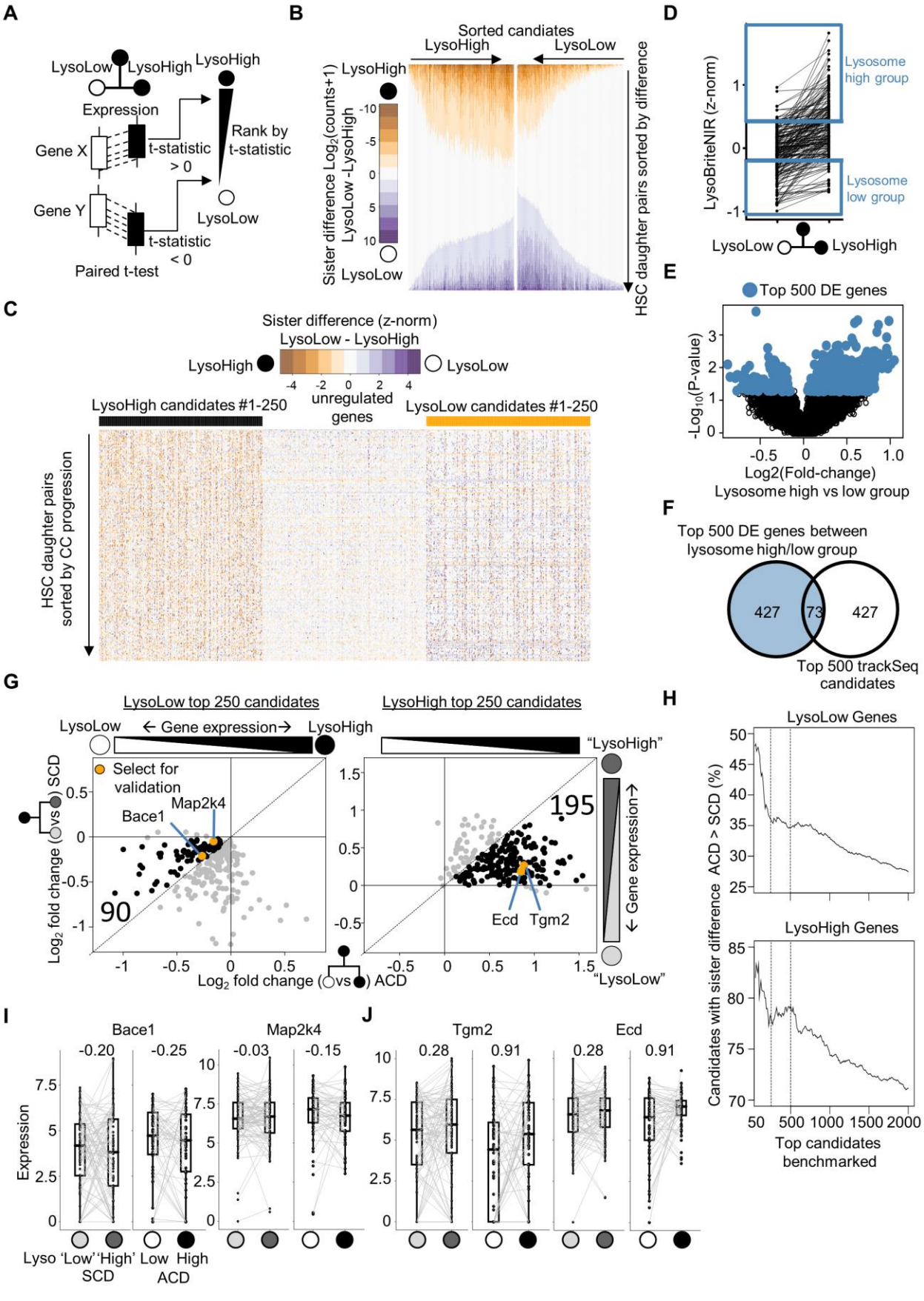

A
LysoLow
LysoHigh
LysoHigh
Expression
Gene X
t-statistic
> 0
Rank by
t-statistic
Gene Y
LysoLow
t-statistic
< 0
Paired t-test
B
Sorted candiates
LysoHigh
LysoLow
LysoHigh
Sister difference Log2(counts+1)
LysoLow -LysoHigh
LysoLow
HSC daughter pairs sorted by difference
C
Sister difference (z-norm)
LysoLow - LysoHigh
LysoHigh
LysoLow
unregulated
genes
LysoHigh candidates #1-250
LysoLow candidates #1-250
HSC daughter pairs
sorted by CC progression
D
LysoBriteNIR (z-norm)
Lysosome
high group
Lysosome
low group
LysoLow
LysoHigh
E
Top 500 DE genes
-Log10(P-value)
Log2(Fold-change)
Lysosome high vs low group
F
Top 500 DE genes between
lysosome high/low group
427
73
427
Top 500 trackSeq
candidates
G
LysoLow top 250 candidates
LysoHigh top 250 candidates
LysoLow
← Gene expression→
LysoHigh
"LysoHigh"
Select for
validation
Bace1
Map2k4
90
195
Ecd
Tgm2
Log2 fold change (vs) SCD
Log2 fold change (vs) ACD
← Gene expression→
"LysoLow"
H
LysoLow Genes
LysoHigh Genes
Candidates with sister difference ACD > SCD (%)
Top candidates
benchmarked
I
Bace1
Map2k4
-0.20
-0.25
-0.03
-0.15
Expression
Lyso 'Low' 'High'
SCD
Low
High
ACD
J
Tgm2
Ecd
0.28
0.91
0.28
0.91

Figure S3

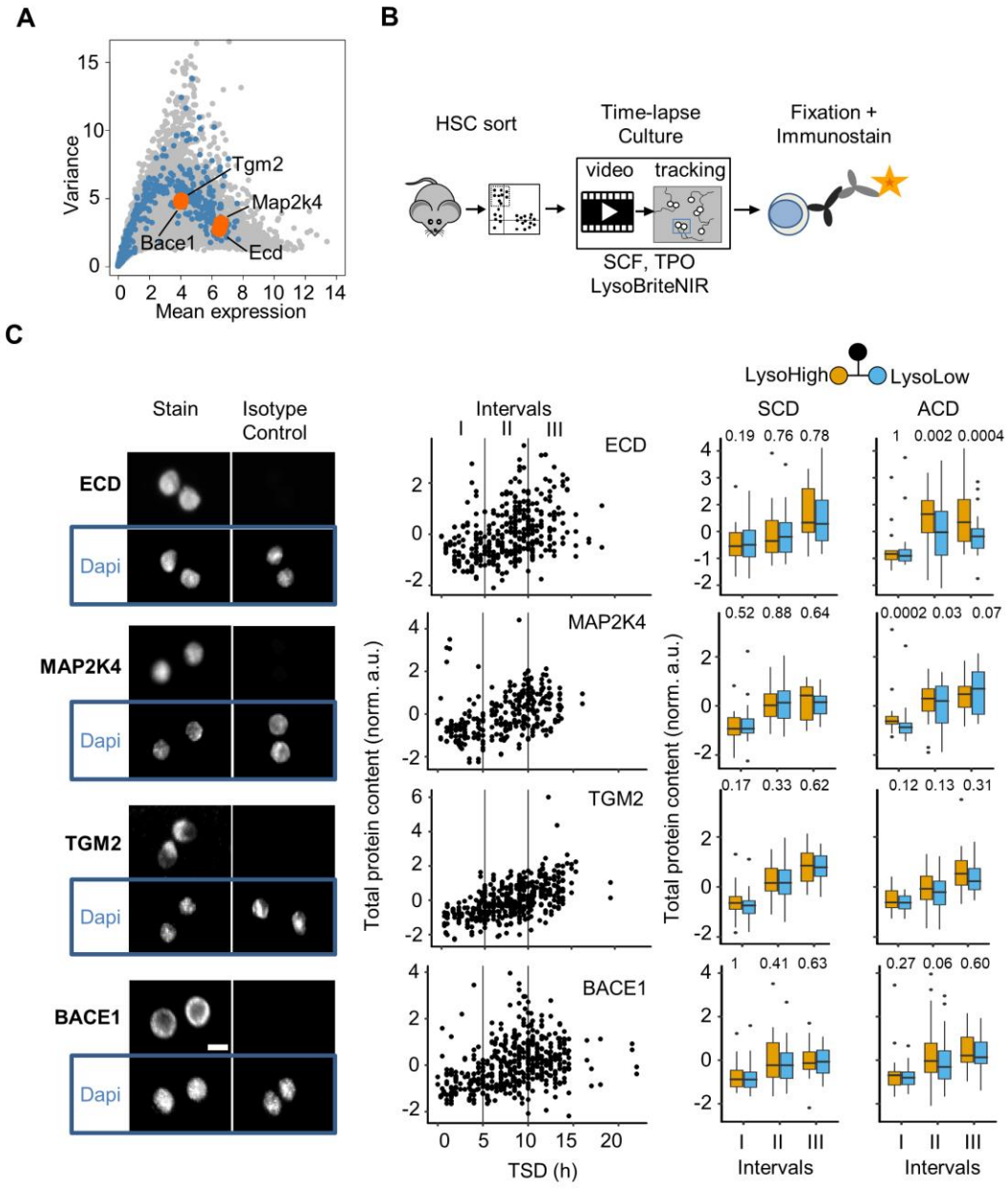

A
Variance
Mean expression
Tgm2
Map2k4
Bace1
Ecd
B
HSC sort
Time-lapse Culture
video
tracking
Fixation + Immunostain
SCF, TPO
LysoBriteNIR
C
Stain
Isotype Control
ECD
MAP2K4
TGM2
BACE1
Dapi
LysoHigh
LysoLow
Intervals
SCD
ACD
Total protein content (norm. a.u.)
TSD (h)
0.19 0.76 0.78
1 0.002 0.0004
0.52 0.88 0.64
0.0002 0.03 0.07
0.17 0.33 0.62
0.12 0.13 0.31
1 0.41 0.63
0.27 0.06 0.60

Figure S4

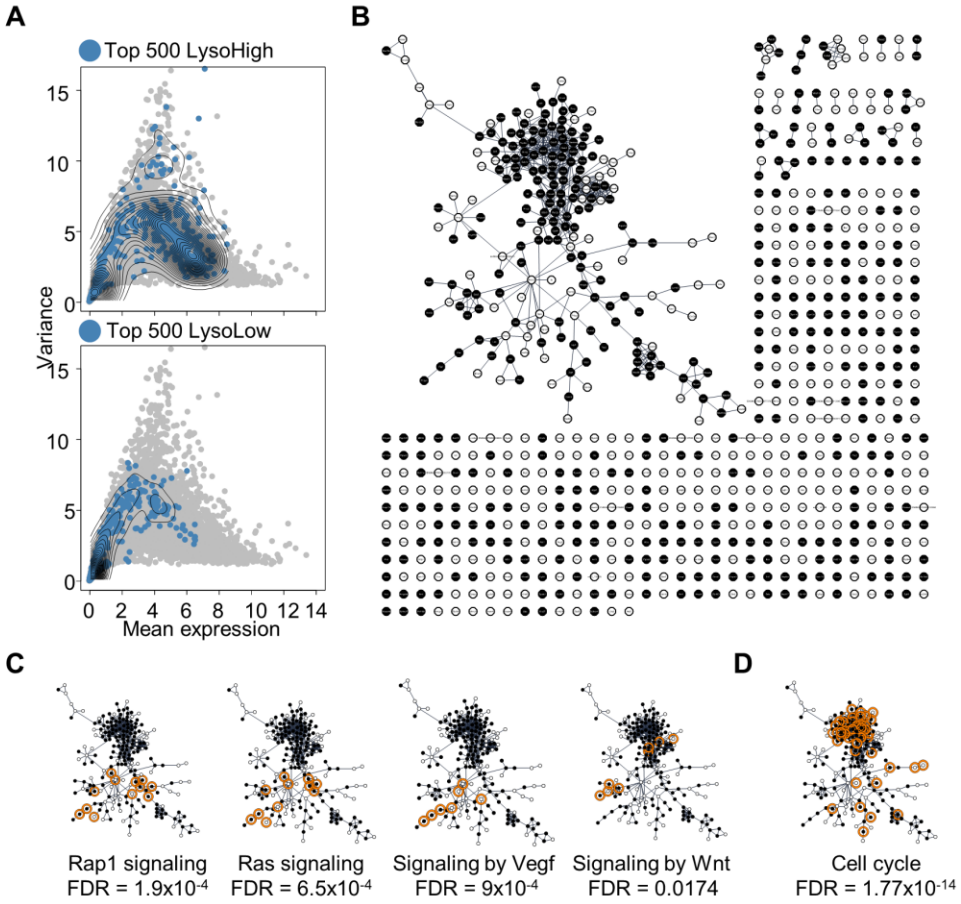

A
Top 500 LysoHigh
Top 500 LysoLow
Variance
Mean expression
B
C
Rap1 signaling
FDR = 1.9x10-4
Ras signaling
FDR = 6.5x10-4
Signaling by Vegf
FDR = 9x10-4
Signaling by Wnt
FDR = 0.0174
D
Cell cycle
FDR = 1.77x10-14

Figure S5

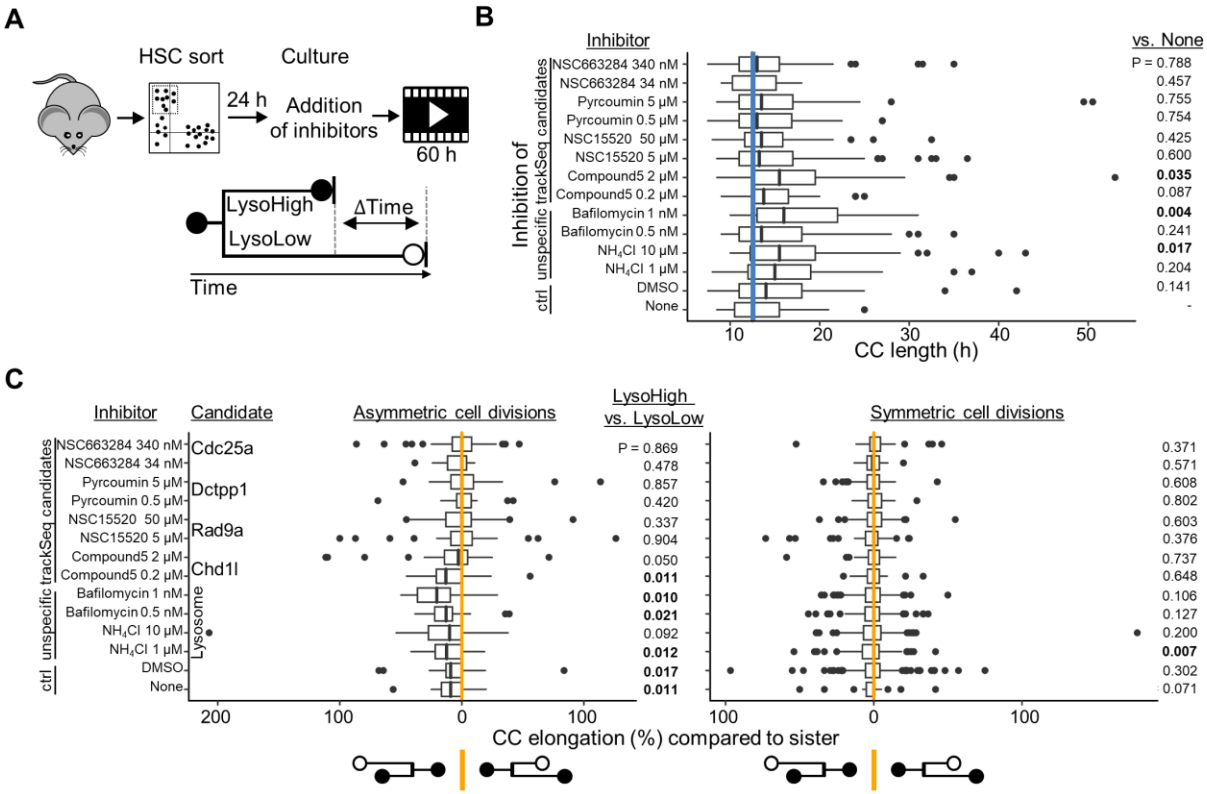
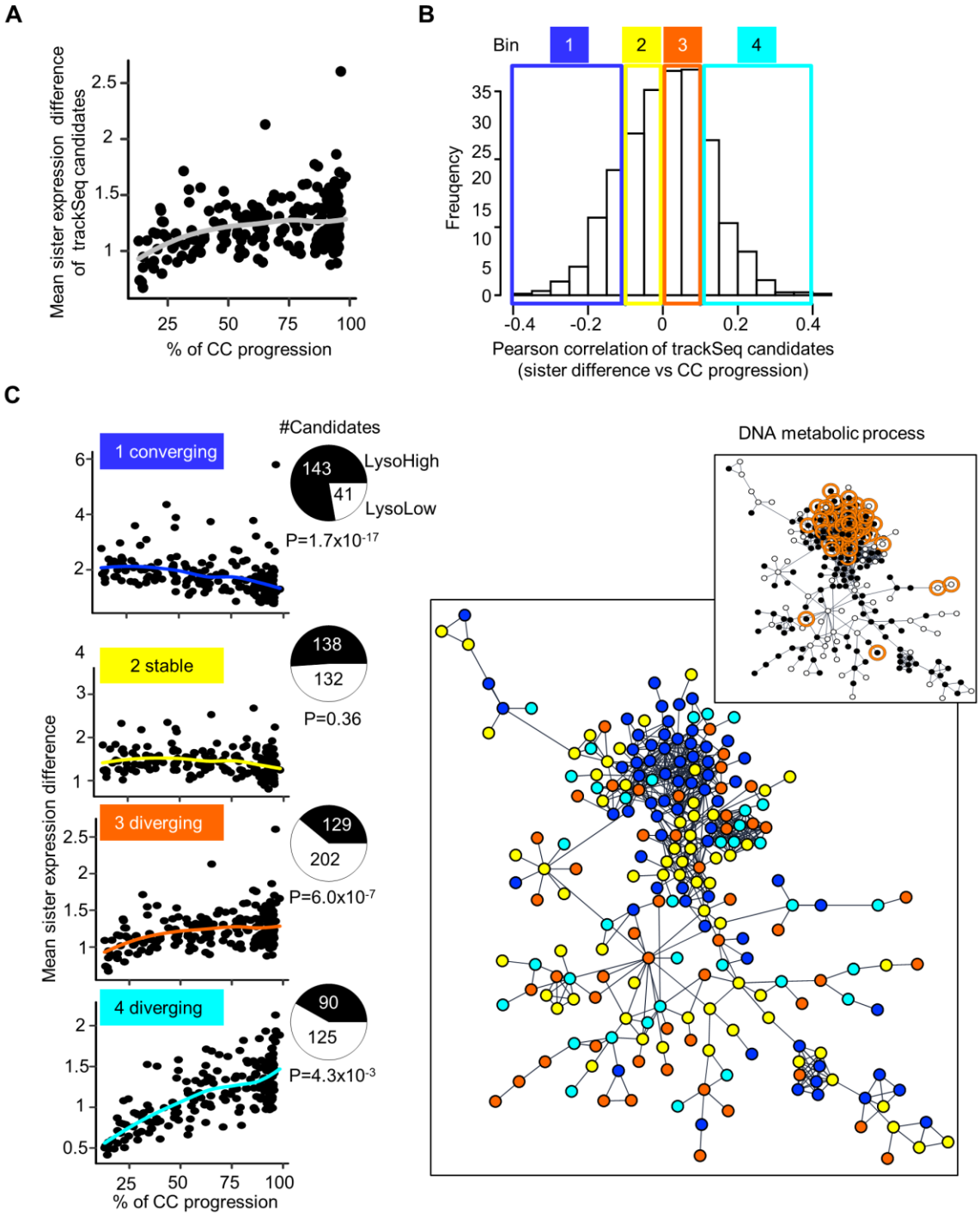

A
HSC sort
Culture
24 h
Addition of inhibitors
60 h
LysoHigh
LysoLow
ΔTime
Time
B
Inhibition of unspecific trackSeq candidates
ctrl
Inhibitor
vs. None
NSC663284 340 nM P = 0.788
NSC663284 34 nM 0.457
Pyrcoumin 5 µM 0.755
Pyrcoumin 0.5 µM 0.754
NSC15520 50 µM 0.425
NSC15520 5 µM 0.600
Compound5 2 µM 0.035
Compound5 0.2 µM 0.087
Bafilomycin 1 nM 0.004
Bafilomycin 0.5 nM 0.241
NH4Cl 10 µM 0.017
NH4Cl 1 µM 0.204
DMSO 0.141
None -
10 20 30 40 50
CC length (h)
C
Inhibitor
Candidate
Asymmetric cell divisions
LysoHigh vs. LysoLow
Symmetric cell divisions
Cdc25a
Dctpp1
Rad9a
Chd1l
Lysosome
NSC663284 340 nM P = 0.869 0.371
NSC663284 34 nM 0.478 0.571
Pyrcoumin 5 µM 0.857 0.608
Pyrcoumin 0.5 µM 0.420 0.802
NSC15520 50 µM 0.337 0.603
NSC15520 5 µM 0.904 0.376
Compound5 2 µM 0.050 0.737
Compound5 0.2 µM 0.011 0.648
Bafilomycin 1 nM 0.010 0.106
Bafilomycin 0.5 nM 0.021 0.127
NH4Cl 10 µM 0.092 0.200
NH4Cl 1 µM 0.012 0.007
DMSO 0.017 0.302
None 0.011 0.071
200 100 0 100
100 0 100
CC elongation (%) compared to sister

Figure S6


A
Mean sister expression difference of trackSeq candidates
% of CC progression
B
Bin
1
2
3
4
Freuqency
Pearson correlation of trackSeq candidates
(sister difference vs CC progression)
C
#Candidates
1 converging
143
41
LysoHigh
LysoLow
P=1.7x10-17
2 stable
138
132
P=0.36
3 diverging
129
202
P=6.0x10-7
4 diverging
90
125
P=4.3x10-3
Mean sister expression difference
% of CC progression
DNA metabolic process

Supplementary Methods

Ethical statement

Experiments using murine bone marrow cells were approved by the animal ethics committee Basel-Stadt (approval number 2655) and were done in agreement with Swiss federal law and institutional guidelines of ETH Zurich. The general well-being of the mice was regularly monitored by animal facility caretakers by visual inspections. Mice were euthanized if symptoms of pain and/or distress were observed.

Mice

Male C57BL/6JRj (Janvier Labs) that acclimatized for at least 1 week before an experiment. Mice were housed in improved hygienic conditions in individually ventilated cages with 2–5 mice per cage, supplied with environmental enrichment. The housing facility had an inverse 12 h day–night cycle with controlled temperature (21±2 °C) and humidity (55±10%). Animals had ad libitum access to standard diet and drinking water at all times. The general well-being of the mice was routinely monitored by animal facility caretakers by daily visual inspections. Mice were euthanized if symptoms of pain and/or distress were observed.

Hematopoietic stem cell isolation

Femurs, tibiae, coxae, humeri, ulnae and vertebrae were isolated from male 12- to 19-week-old mice, crushed in PBS (2% FCS, 2 mM EDTA) and then filtered through a 100-µm nylon mesh, while on ice. Erythrocytes were lysed for 3 min at room temperature in ACK lysing buffer (Lonza), stained with biotinylated lineage antibodies against CD3ε (145-2C11), CD19 (eBio1D3), TER-119 (TER-119), B220 (RA3-6B2), Ly-6G (RB6-8C5) and CD11b (M1/70) to be depleted with streptavidin-conjugated magnetic beads (Roti-MagBeads, Roche). Subsequently, cells were stained for fluorescence-activated cell sorting with anti-CD34-eFluor450 (RAM34), SCA1–

PerCP–Cy5.5 (D7), KIT–PE–Cy7 (2B8), streptavidin–APCeFluor780, CD48–APC (all eBioscience) or CD48–FITC (HM48-1; Biolegend), CD135–PE–CF594 (A2F10.1; BD) and CD150–BV650 (TC15-12F12.2; Biolegend) for 90 min on ice. Sorting took place with a BD FACS Aria III using a 70-µm nozzle, single-cell purity mode and sorting purities ≥ 98%. Sorting scheme in Fig. S1.

Cultivation media

Cells were cultured in phenol-red-free IMDM supplemented with 20% BIT (Stemcell Technologies), 100 ng/mL murine SCF, 100 ng/mL murine TPO (PeproTech), 2 mM L-glutamine, 50 µM 2-mercaptoethanol, 50 U/mL penicillin and 50 µg/mL streptomycin (all Gibco) and 16.6 µM LysoBrite-NIR (22641, AAT Bioquest).

HSC granddaughter isolation

To isolate HSC granddaughters, we used Cell Trace Violet (CTV, C34557 ThermoFisher Scientific). After lineage depletion, bone marrow cells were washed (2x 10 mL PBS, 4 ^{o}C) to remove proteins in supernatant that inhibit CTV labelling. After washing, cells were suspended in 1 mL of 2 µM CTV and incubated for 5 minutes at 37 ^{o}C. To stop the labelling, we added 1 mL Fetal Calf Serum and washed twice (2x 10 mL PBS 4 ^{o}C). Hereafter, cells were stained for fluorescence-activated cell sorting as before and the initial CTV peak (no divisions) was recorded. With each division, CTV intensities will half. Thus, generations can be traced and after 60 hours in culture, we isolated HSC granddaughters for scSeq.

Time-lapse imaging

We used Nikon-Ti Eclipse microscopes with linear-encoded motorized stages, Orca Flash 4.0 V2 (Hamamatsu) cameras, and Spectra X fluorescent light sources (Lumencor). We used collimated white light emitted by the Spectra X via a custom-made motorized mirror controlled by an Arduino

UNO Rev3 (Arduino) as transmitted light for bright-field illumination.

Single cell RNA sequencing (scSeq)

After single cell isolation by picking, we performed scSeq in an adaption from previously described protocols [2]. In brief, for cDNA library production, we first thawed and centrifuged (350 RCF, 1 min at 4 °C) the PCR plates with the previously isolated HSCs. Then, we added 2 µL of annealing mix (dH_2O (Invitrogen) with 1:3,000,000 dilution of ERCC spike-ins (Thermo Scientific), 1 µM Oligo-dT (Sigma-Aldrich/Merck Kgaa) and 5 mM dNTP (Takara)) to each well, centrifuged the plates again for liquid collection and incubated for 72°C for 3 minutes. Hereafter, plates were put on ice immediately.

For reverse transcription, we added 5.7 µL of reverse transcription mix (dH_2O (Invitrogen) with 3.51 U/µL Maxima H Minus (Thermo Scientific), 0.88 U/µL RNase inhibitor (Takara), 1:2.85 dilution of 5x Maxima RT buffer (Thermo Scientific), 3.51 µM TSO (Sigma-Aldrich/Merck Kgaa) and 13.16% v/v PEG 8000 (Sigma-Aldrich/Merck Kgaa)) to each well and after a quick centrifugation step for volume collection (350 RCF, 1 min at 4 °C), we incubated the plates for 90 minutes at 42°C and another 15 minutes at 70°C to complete reverse transcription.

After another centrifugation step (350 RCF, 1 min at 4 °C), we added 40 µL pre-amplification mix (dH_2O (Invitrogen) with 0.03 U/µL Terra PCR Direct Polymerase (Takara) and 0.25 µM ISPCR primer (Sigma-Aldrich/Merck Kgaa)) to each well, to complete the production of cDNA libraries. Again, volumes were collected by a quick spin down. For the pre-amplification PCR, plates were first incubated at 98°C for 3 minutes, followed by 19 cycles of 98°C incubation for 15 seconds, 60°C for 30 seconds and 68°C for 4 minutes. Pre-amplification was concluded by a final incubation at 72°C for 10 minutes. Afterwards samples were spun down for volume collection.

For PCR clean-up we used a 0.6:1 of Ampure XP magnetic beads to sample ratio. Ampure beads

were at least 15 minutes equilibrated at room temperature. After mixing with the Ampure beads and incubation at room temperature for eight minutes, samples were washed three times with 80% freshly prepared ethanol solution on a magnetic stand, and then dried on air for 5 minutes. We eluted cDNA libraries with 26 µL of elution buffer (Quiagen). After quality control of size distribution with a fragment analyser (Agilent) and cDNA quantification, we used 150 pg cDNA of each well for library preparation.

The final cDNA library preparation for sequencing was done using the Nextera XT DNA sample preparation kit 96 samples (Illumina). Single cell libraries were tagged using the Nextera XT 96-Index kit, 384 samples (Illumina) and sequenced with the NextSeq500 from Illumina.

Primary analysis and quality control of scSeq data

Reads were aligned with *STAR* [3] to the M 21 release mouse genome (ftp://ftp.ebi.ac.uk/pub/databases/gencode/Gencode_mouse/release_M21). *Samtools* [4] was used to create sorted .bam files to extract counts with *featureCounts* [5]. Parsed count data were analyzed in the R 3.6.2 programming environment.

The quality control and normalization of primary scSeq count data was executed with *R* (https://www.r-project.org). Gene names were mapped from ensemble identifiers to symbols using the *biomart* library [6]. Quality control was performed with the *R* package *SingleCellExperiments* and the function *perCellQCMetrics* from the *scater* [7] package. In addition to pass quality control, we required cells to have at least 100,000 reads mapped to nuclear genes, no more than 20% mitochondrial reads detected, below 50% of all reads mapped to External RNA Controls Consortium (ERCC) spike-ins and at least 3000 nuclear genes detected. We normalized for cell-specific count biases using ERCC spike-ins with the *scran* [8] library function *computeSpikeFactors*, followed by log count normalization with the *scater* function *logNormCounts*.

Lower dimensional representations and pseudotime

To capture the cell cycle with pseudotime, we used the S and G2M phase specific genes provided by Seurat [9] as features and the destiny package [10] to compute a diffusion map and pseudotime. We also used the S and G2M phase genes by *Seurat* to create our Umap representation with the *Seurat* function RunUMAP. Mappings onto the Nestorowa dataset were done as described on their paper's website (http://blood.stemcells.cam.ac.uk/projection_demo.html) [11].

Cell cycle inference and regression

We used the *R* library *Seurat* [9] to infer cell cycle (CC) progression and CC stages at the time of isolation from single cell scSeq profiles after quality control and normalization. To regress out S- and G2M-scores from our data, we used Seurat as presented here https://satijalab.org/seurat/archive/v3.1/cell_cycle_vignette.html**.** See the script *DealWithCellCycle_Seurat.R.* For comparison with the NOstrat and TSDstrat set, we performed the NOstrat ranking on the CC regressed count data.

Cell mapping

We used the *R* library *destiny* [10] to map our picked HSC daughter cells onto freshly isolated bone marrow cells from http://blood.stemcells.cam.ac.uk/single_cell_atlas.html [11].

Protein-protein network analysis

We used the *R* package *STRINGdb* to query the *STRING* database [12] version 10 for protein-protein interactions among candidate sets of *Mus musculus* genes with a conservative *score_threshold* of 800. To test for enrichment of interactions among the trackSeq dataset, we created a baseline null distribution of protein-protein interactions by querying 1000 candidate sets of 500 randomly sampled (drawn without replacement) detected genes. This is the same gene pool used by all our stratification schemes, described above, to rank genes. The retrieved sets of randomly chosen

genes were queried for their numbers of genes mapped to STRING and found numbers of interactions. The distribution of interaction counts among a gene set of 500 randomly chosen genes is then used to judge the empirical likelihood to achieve the observed number of interactions found in the trackSeq candidate set. For network visualization and process enrichment analysis we used the *Cytoscape* version 3.7.2 *StringApp* plugin [13].

Lentivirus production and transduction

To transduce cells with PCNAVENUS, we cloned PCNA and VENUS cDNAs into third generation lentivirus [14] for production in human 293T cells. Concentration was done by ultracentrifugation at 68000 RCF and titrated with mouse 3T3 fibroblasts. The resulting virus was stored at −80 °C for later transductions. HSCs were infected for 24h in 96 round-bottom well plates (Corning) and IMDM media containing 20% BIT, 2 mM L-glutamine, 50 µM 2-mercaptoethanol, 50 U/mL penicillin and streptomycin each, and 100 ng/mL murine TPO and murine SCF each. The multiplicity of infection was 300.

Endpoint Immunostainings

Cells were fixed with 10% formalin solution (Sigma-Aldrich, Cat# HT5011-15ML) for 10 minutes at room temperature (20-25°C) in µ-slide $VI^{0,4}$ channel slides (IBIDI) coated with 10 µg/mL anti-CD43 antibody. They were blocked and permeabilized with TBS (0.1M Tris, 0.15 M Nacl, pH 7.5) containing 10% donkey serum (Jackson ImmunoResearch Cat# 017-000-121), 0.05% Tween-20 (Merck, Cat# P9416) and 0.1% Triton X-100 (PanReac AppliChem, Cat# A1388) for 2 hours. Primary antibodies, secondary antibodies (see Data S1), DAPI (Merck, Cat# D9542), and streptavidin conjugates were diluted in the same buffer described above. Primary antibodies were incubated with the cells for 15 hours at 4 °C, secondary antibodies were incubated for 1 hours at room temperature and streptavidin conjugates were incubated for 30 minutes at room temperature.

TBS containing 0.05% Tween-20 and 0.1% Triton X-100 were used to wash the cells in between the staining steps for at least four times.

Coating for cell adhesion assay

HSCs were seeded on coated µ-slide VI^{0,4} channel slides (IBIDI). Slides were either coated with 5 µg/cm^2 LAMININ (#L2020 Merck) in culture media for one hour at room temperature or 10 µg/mL LAMININ 511 or 421 (#LN511 and #LN421, BioLamina) at 4 °C overnight. The time lapse imaging interval was 20 minutes. Adhesion was measured as mean cell displacement between time lapse imaging frames.

Proximity ligation assay (PLA)

HSCs were seeded on anti-CD43 coated (10 µg/mL, 1 hour at room temperature) µ-slide VI^{0,4} channel slides (IBIDI) and cultured identical to trackSeq time lapse cultures, with alerT reporting ACD events. After 44 hours, cells were fixed and treated as for endpoint immunostainings. We performed PLA with one primary antibody against ITGB4 (Biolegend Cat#123602), and subsequently two PLA oligonucleotides conjugated donkey anti-rat secondary antibodies (Merck DUO96020; Jackson ImmunoResearch 712-005-150) against the primary antibody [15]. We used "multicolor" reagent pack (Merck DUO96000-100RXN) for signal detection. The PLA reaction was performed as stated in the published PLA Muticolor Detection protocol (Merck).

Confocal microscopy

We performed confocal microscopy on a Leica TSC SP8 Falcon equipped with three photomultiplier tubes, two HyD detectors, four lasers (405 nm, 442 nm, Argon Laser (458, 476, 488, 496 and 514 nm), White Light Laser (470-670 nm)) using a 63x water immersion lens (NA 1.2, FWD 0.22 mm). Scannings were acquired at 20-25°C, 400 Hz, in the bidirectional mode, with z-slices of 1 µm in height at 1024x1024 resolution in 8-bit with only HyD detectors as described

previously [15].

trackSeq candidate inhibition

HSCs were isolated and cultured as described above. Time lapse imaging was started 24 h after isolation and analyzed using alerT. Inhibitors were solved in DMSO (Merck KgAa) and diluted with media to final concentrations indicated in Fig. 4H. DMSO control concentration corresponded to maximal DMSO concentration in inhibitors (0.01% v/v). Pyrcoumin was kindly provided by COMAS from the Max-Planck-Institute in Dortmund. Compound 5" was acquired from MOLPORT (MolPort-003-177-841), NSC663284 from Bio-Techne AG (1867/10), NSC15520 from Chemie Brunschwig AG (26020), Bafilomycin A1 from abcam (ab120497) and NH_4Cl (A9434) from Merck.

Additional information

Data S1. (separate file)

DataS1_AntiBodiesUsed.xlsx: List of antibodies used in study including producer, catalogue numbers, clones, fluorophores and Research Resource Identifiers.

Data S2. (separate file)

DataS2_MetadataPickedHSCdaughters.csv: Metadata of scSeq data analyzed in this study. Contains all single cell processed data (except counts) used in this study, including identifiers, time lapse quantifications, UMap projections, pseudo time mapping, inferred CC phases and scoring, time since divisions, sister differences, and HSC scorings.

Data S3. (separate file)

DataS3_CandidatesOfStratifications.xlsx: Contains top 500 gene symbols of all candidate sets identified in this study and the top 10 GO biological processes enriched from STRING.

Data S4. (separate file)

DataS4_Candidates_trackSeq.csv: The trackSeq top 500 gene candidate set with top 250 (LysoHigh) and bottom 250 (LysoLow) candidates, ranked by comparing paired HSC daughter cells with known fate directionality from their relative lysosome inheritance. Ranks indicated.

Data S5. (separate file)

DataS5_Top500LysoHighLysoLow.csv: The trackSeq top 500 LysoLow and top 500 LysoHigh genes including information on their mapping to the STRING database to create the identified network in Fig. 3.

Data S6. (separate file)

DataS6_NetworkEnrichments_GOProcesses.csv: The GO processes found to be enriched by STRING within the identified network within the top 500 LysoLow and top 500 LysoHigh genes.

Data S7. (separate file)

DataS7_PCNAVENUSdata.csv: The tracked PCNAVENUS time lapse data. Contains 311 HSC daughter pairs from 5 technical replicates, with their lysosomal inheritance ratio (column "LysoRatio"), their CC length and CC phase durations. "Daughter1" indicates LysoHigh and "Daughter2" the LysoLow daughter.

Supplementary References

1. Wilson NK, Kent DG, Buettner F, et al. Combined Single-Cell Functional and Gene Expression Analysis Resolves Heterogeneity within Stem Cell Populations. *Cell Stem Cell*. 2015;16(6):712–724.
2. Bagnoli JW, Ziegenhain C, Janjic A, et al. Sensitive and powerful single-cell RNA sequencing using mcSCRB-seq. *Nat. Commun.* 2018;9(1):1–8.
3. Dobin A, Davis CA, Schlesinger F, et al. STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29(1):15–21.
4. Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009;25(16):2078–2079.
5. Liao Y, Smyth GK, Shi W. FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*. 2014;30(7):923–930.
6. Durinck S, Spellman PT, Birney E, Huber W. Mapping identifiers for the integration of genomic datasets with the R/ Bioconductor package biomaRt. *Nat. Protoc.* 2009;4(8):1184–1191.
7. McCarthy D, Campbell K, Lun A, Wills Q. scater: pre-processing, quality control, normalisation and visualisation of single-cell RNA-seq data in R. 2016;545–552.
8. Lun ATL, McCarthy DJ, Marioni JC. A step-by-step workflow for low-level analysis of single-cell RNA-seq data with Bioconductor. *F1000Research*. 2016;5:2122.
9. Stuart T, Butler A, Hoffman P, et al. Comprehensive Integration of Single-Cell Data. *Cell*. 2019;177(7):1888-1902.e21.
10. Angerer P, Haghverdi L, Büttner M, et al. destiny: diffusion maps for large-scale single-cell data in R. *Bioinformatics*. 2016;32(8):1241–1243.
11. Nestorowa S, Hamey FK, Pijuan Sala B, et al. A single-cell resolution map of mouse hematopoietic stem and progenitor cell differentiation. *Blood*. 2016;128(8):e20–e31.
12. Szklarczyk D, Gable AL, Lyon D, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* 2018;47:607–613.
13. Doncheva NT, Morris JH, Gorodkin J, Jensen LJ. Cytoscape StringApp: Network Analysis and Visualization of Proteomics Data. *J. Proteome Res.* 2019;18(2):623–632.
14. Endele M, Loeffler D, Kokkaliaris KD, et al. CSF-1-induced Src signaling can instruct monocytic lineage choice. *Blood*. 2017;129(12):1691–1701.
15. Kunz L, Schroeder T. A 3D Tissue-wide Digital Imaging Pipeline for Quantitation of Secreted Molecules Shows Absence of CXCL12 Gradients in Bone Marrow. *Cell Stem Cell*. 2019;25(6):846-854.e4.