SUPPLEMENTAL INFORMATION

Legend for supplementary figures and movies

Figure S1 Non-stochastic reprogramming from a subset of GMPs. Related to Figure 2. **(A)** Construction of a reprogramming lineage from a single GMP. Time-lapse image series acquired at 5 min intervals were annotated with colored arrows following the cells of interest. The first image was acquired after one hour of Dox addition and the GMP was likely still in its hematopoietic state. The birth of a new cell by mitosis is followed by an additional arrow of different colors. The key images informing the structure of cell pedigree are shown. The colors of the arrows correspond to the line colors shown in Figure 2A and are the same as in Movie S1. Time stamps in hours:minutes format. Two ultrafast cell cycles that lasted 6 hours 55 minutes and 7 hours 35 minutes are indicated. **(B)** Gating strategies for sorting LKS and GMP. **(C)** Hematopoietic stem and progenitors were sorted from a transgenic mouse line that expresses an H2B-GFP fusion protein. Representative images show progeny of these HSPCs can be identified based on the presence of H2B-GFP expression. **(D)** GMP-initiated reprogramming cultures on day6 were trypsinized, stained with CD45 antibody and analyzed by FACS. About 3% of cells appeared as CD45-/Oct4:GFP- and display non-characteristic scatter behavior indicative of their feeder cell nature.

Figure S2 Assessing G1 phase duration using the FUCCI reporter. Related to Figures 2 and 3. GMPs from the Rosa26:rtTA x Oct4:GFP mice were co-transduced with the polycistronic reprogramming virus together with the G1 FUCCI reporter expressing mKO (the red fluorescence signal is diagramed in Figure 2A and Figure S1A). A representative lineage with FUCCI expression is shown. Time stamps are in hours:minutes format. The yellow arrow tracks a FUCCI cycle which started at 29:45 and ended at 34:45 (the cells happened to be aggregated for view at this time point). Thus the G1 duration of this cell cycle was estimated to be 4 hours. The white circle marks another FUCCI positive cell in the vicinity, which remained in G1 for longer than 12 hours. This same lineage is also shown in Movie S2.

Figure S3 Fractionating cells based on their cycling speed. Related to Figures 3, 5 and 6. **(A)** Schematic illustration of the dye (CFSE) dilution by cell divisions. Each time when a labeled cell divides, the dye is equally distributed to the two daughter cells resulting in halving of the CFSE intensity which could be measured by flow cytometry. After a period of cell proliferation, faster cycling cells undergo more cell divisions and retain less CFSE. **(B)** Cells with heterogeneous CFSE intensities arise from a homogeneously labeled population following different number of divisions. The retaining dye intensity serves as a sorting strategy for separating cells of different cycling speed.

Figure S4 Scoring cell divisions following single cell sorting into Terasaki plates and simulation of cell cycle lengths. Related to Figure 4. **(A)** Representative images of cells identified in Terasaki plates. A well containing a single cell and one containing two cells are

shown. The correct identity of cells was also confirmed by the presence of H2B-GFP. **(B)** The number of cells was scored at three time points: immediately (0h), 14 hours (14h) and 19 hours (19h) after sorting. The number of cells present at 14 hours and 19 hours were used to estimate the cell cycle speed as indicated in the column on the right. Only wells that contained single viable cells at 0h were used for computational simulation. **(C)** Cell cycle lengths were estimated using computational simulation based on the cell numbers scored at time 0, 14 and 19 hours post single cell sorting into Terasaki plates. Each of the cell number combinations at 14 and 19 hours were given an estimated range of cell cycle durations. For any given combination, the range was determined so that for each cell cycle length within the range, there was at least 25% probability (red dashed line) of observing such a combination. Using this approach, ~20% fresh GMPs were found to be ultrafast cycling (Figure 4A, columns in red box), which is in agreement with direct measurement by time-lapse imaging as shown in Figure 3A.

Figure S5 Reprogramming single hematopoietic stem and progenitors and definition of reprogramming efficiency. Related to Figures 2, 3 and 4. **(A)** Scheme of reprogramming fresh HSPCs. HSPCs were sorted from mouse bone marrow and transduced with the polycistronic reprogramming virus (day -1). On the following day (day 0), all cells were washed and single cell sorted into 96-well plates in reprogramming conditions (MEF feeders + mESC medium + Dox). Oct4:GFP+ colonies were scored after 5-7 days. A well scores positive if it contained at least one Oct4:GFP+ colony. A well scores as one reprogramming event even if it contained multiple Oct4:GFP+ colonies. Note the definition of reprogramming efficiency. **(B)** Scheme of reprogramming cultured HSPCs. HSPCs were sorted from mouse bone marrow and transduced with the polycistronic reprogramming virus (day -6). On the following day (day -5), all cells were washed and single cell sorted into 96-well plates in hematopoietic growth factors for activation culture. After five days (day 0), activated cells were plated in new wells in reprogramming conditions. Oct4:GFP+ colonies were scored after another 5-7 days (day 5-7). A well scores positive if it contained at least one Oct4:GFP+ colony. A well scores as one reprogramming event even if it contained multiple Oct4:GFP+ colonies. Note the definition of reprogramming efficiency. **(C)** LKS cells were sorted from the bone marrow of a Rosa26:rtTA x H2B-GFP transgenic mouse, transduced with the polycistronic reprogramming virus and single cell sorted into 96-well plates to be activated with hematopoietic growth factors. Progeny of the activated LKS cell were then transferred into reprogramming conditions in a well on a new 96-well plate. Shown are representative images at 10x (top) and 20x magnifications (bottom) demonstrating predominantly reprogrammed progeny. Note that essentially all H2B-GFP+ cells situated in or close to colonies, which were already well developed on day 5 of Dox induction. **(D)** A representative FACS plot for the cells in one of such wells after staining with antibodies against CD45 and SSEA1.

Figure S6 Experimental scheme for isolating and comparing cells of different cycling speed in MEF-initiated reprogramming cultures. Related to Figures 5 and 6. **(A)** Reprogramming medium contained mESC medium + 2 µg/ml Dox. MEFs were isolated from the Rosa26:rtTA x

Oct4:GFP embryos. **(B)** Time of the appearance of Oct4:GFP+ cells from MEF-initiated reprogramming cultures. The entire culture was trypsinized and analyzed for Oct4:GFP+ cells by flow cytometry on day6 and day8 of Dox induction. This strategy enables the detection of small numbers of Oct4:GFP+ progeny not yet visible as distinct colonies. There were no Oct4:GFP+ cells on day6 but ~1% of such cells appeared on day8. **(C)** Fast cycling cells arose in MEF-initiated reprogramming cultures in response to Yamanaka factors. The entire culture was trypsinized and analyzed for CFSE by flow cytometry after 2 days of dye dilution by division (from day 4 to 6 of Dox induction). A fast cycling population (red box) was detected only with continued Yamanaka factor induction (+Dox) but was absent in the control MEFs that underwent similar CFSE label and dilution (-Dox).

Figure S7 Molecular characterization of the fast cycling cells. Related to Figure 7. **(A)** Two gene clusters (Cluster I and Cluster VI) as described in Polo et al., 2012 were used to gauge the quality of the RNA-seq data. These two signatures were chosen given their large sizes, which include hundreds of genes showing characteristic expression between reprogramming progressive and refractory cells. Fast MEFs show enriched Cluster I expression (reprogramming progressive) while bulk MEF and slow MEF are enriched for Cluster VI (reprogramming refractory). **(B)** RNA sequencing reads mapped to the 5'UTR and 3'UTR regions of the Sox2 genes are plotted as an indication for the endogenous Sox2 expression, since the exogenously expressed Sox2 do not have the UTR sequences. **(C)** Representative GSEA plot showing enrichment of a DNA replication signature. **(D-F)** The expression level of various cyclins (D), CDKs (E) and CDKIs (F) among HSPCs. Note that only p57 expression displays major difference between the LKS and GMP cells. **(G)** Stochastic and privileged reprogramming as part of a continuum. Depending on when and which cells reach a privileged state, reprogramming occurs at different efficiency and latency across cell lineages. The earlier such cells appear, the earlier/more efficient that lineage reprograms. GMPs are privileged before exposure to reprogramming factor expression and are highly efficient in reprogramming. One striking feature of the privileged cells is an ultrafast cell cycle**. (H)** Normalized MBD3 sequence reads from RNA-seq data are plotted. **(I)** MBD3 RNA expression as measured by qRT-PCR. Normalization was done using 18S RNA as controls. **(J)** Western blot analysis for MBD3 levels in fast and slow MEFs. HSP90 and beta actin were probed as loading controls.

Movie S1. Time-lapse image series acquired at 5 minutes intervals. Related to Figures 2 and 3. Arrows follow the movement of each cell that developed into Oct4:GFP+ cells at the end of image acquisition. The same shaped arrows denote cell progeny from the same founder cell. The blunt arrows denote cells migrated into view from outside of the field. The narrow/sharp arrows on the bottom right corner denote a second independent founder cell and its progeny. At each division, an arrow of a new color is added while the old color is inherited by one of the daughter cells. With cell clusters, arrows denote such clusters rather than single cells. One founder cell gave rise to multiple colonies due to splitting. Their active migrative behavior brought the arrows

of different shapes and colors into forming mixed colonies. Time stamps are shown in hours:minutes format. Overlayed images of DIC and GFP channels are shown.

Movie S2. Time-lapse image series acquired at 15 minutes intervals. Related to Figure 3. GMPs were transduced with the polycistronic reprogramming vector together with the FUCCI reporter which displays red fluorescence in G1 phase of the cell cycle. The founder cell (divided on the second frame) is followed by the yellow arrows, which split and yielded multiple Oct4:GFP+ cell clusters at the end of image acquisition. At later half of the movie, the arrow denotes the cell cluster rather than any single cells. For easier viewing, only one branch is followed throughout (the other cells of this lineage temporarily migrated outside of this field). The FUCCI signal appeared several times before the cells begin to form compact colonies, with very low and short signals for the first cell cycle. Three still images of the second cell cycle from this movie are shown in Figure S2. Time stamps are shown in hours:minutes format. Overlayed images of DIC, GFP and monomeric Kusabira Orange (FUCCI) are shown.

Table S1 Early Oct4:GFP+ cells from GMP-initiated reprogramming culture progress through reprogramming highly efficiently. Related to Figure 2.

(A) Day3 cultures were trypsinized into single cell suspensions and defined numbers of Oct4:GFP+ cells were sorted into 96-well plates to allow further reprogramming. Oct4:GFP+ colonies were scored after another 7 days (for a total of 10 days on Dox). **(B**) Established iPS cells (Dox independent) and **(C)** ES cells derived from inner cell mass were sorted and plated similarly as controls. **(D)** Poisson statistics was used to calculate the estimated frequency of cells that yielded final Oct4:GFP+ colonies. Day3 Oct4:GFP+ cells were at least as efficient as the established iPS and ES cells in giving rise to Oct4:GFP+ colonies.

A. Oct4:GFP+ cells sorted from day3 cultures initiated from GMPs

Number of cells sorted/well	Number of wells plated	Number of wells with at least one colony
10	60	60
5	60	55
2	60	45
1	60	17

B. Oct4:GFP+ cells sorted from established iPS cells

Number of cells sorted/well	Number of wells plated	Number of wells with at least one colony
10	60	59
5	60	44
2	60	24
1	60	17

C. Oct4:GFP+ cells sorted from inner cell mass derived ES cells

Number of cells sorted/well	Number of wells plated	Number of wells with at least one colony
10	60	60
5	60	50
2	60	22
1	60	10

D. Estimated frequency of Oct4:GFP+ cells leading to productive reprogramming

Cell type	Frequency	*p*	
Day3 Oct4:GFP+ cells	1 in 1.9	<0.01	
iPSCs	1 in 3.4		0.78
ESCs	1 in 3.2		

Table S2 Faster cycling GMPs reprogram at higher efficiency. Related to Figure 3.

A. Single cell reprogramming from GMPs of different cycling speed using the polycistronic virus. Same data is shown in Figure 3D.

GMP cycling phenotype	Wells plated	Positive wells	Efficiency (%)
CFSE Low (Fast)	360	23	6.4
CFSE High (Slow)	360	7	1.9

B. Bulk reprogramming from GMPs of different cycling speed using the polycistronic virus.

GMP cycling phenotype	Cells plated	# of colonies	Efficiency (%)
CFSE Low (Fast)	3925	120	3.05
CFSE Int. (Medium)	18182	324	1.8
CFSE High (Slow)	5080	18	0.35

C. Bulk reprogramming from GMPs of different cycling speed from the iPS mouse.

GMP cycling phenotype	Cells plated	# of colonies	Efficiency (%)
CFSE Low (Fast)	5333	5	0.94
CFSE High (Slow)	3234	0	0

Table S3 The emergence and disappearance of ultrafast cycling cells following culture in hematopoietic growth factors from LKS and GMP cells, respectively. Related to Figure 4.

Ultrafast cycling cells emerged from LKS cells after culture but disappeared from GMPs following the same treatment. Fraction of cells in each bin of cycling speed is shown. This data is plotted in Figure 4A. See also Figure S4.

Cell types & culture condition	Cycling speed (hours/cycle)					Total events scored
	4-8	8-11	9-15	>12	Dead	
Fresh LKS	0.00	0.04	0.55	0.36	0.06	139
Cultured LKS	0.16	0.14	0.43	0.22	0.04	181
Fresh GMP	0.07	0.24	0.46	0.15	0.08	112
Cultured GMP	0.01	0.03	0.27	0.56	0.14	181

Table S4 Dramatically altered reprogramming behavior of LKS and GMP cells following culture in hematopoietic growth factors. Related to Figure 4.

Single cell reprogramming from LKS and GMPs with or without culture. Same data is plotted in Figure 4B.

Initial phenotype	Culture condition	Wells plated	Positive wells	Efficiency (%)
GMP	Fresh	420	77	18.3
	Cultured	420	2	0.5
LKS	Fresh	420	10	2.4
	Cultured	420	*102	24.3

*15% of these wells (or 3.6% of total wells) showed privileged reprogramming as shown in Figure 4C and Figure S5C.

EXTENDED EXPERIMENTAL PROCEDURES

Mice, ESC and MEF derivation

All mouse work was approved by the Institutional Animal Care and Use Committee of Yale University. The Oct4:GFP x Rosa26:rtTA mice were derived by crossing Oct4:GFP mice with Rosa26:rtTA mice as described previously (Eminli et al., 2009; Megyola et al., 2013). Embryonic day 3.5 blastocysts were used to derive embryonic stem cells at Yale Animal Genomic Services. The reprogrammable mouse ($R26^{rtTA}$;$Col1a1^{4F2A}$) (Carey et al., 2010) with the Yamanaka factors targeted into the Col1a1 locus (stock number 011004) and the H2B-GFP transgenic line (stock number 006069) were purchased from the Jackson Laboratory. All mouse strains were crossed with the Rosa26:rtTA strain to achieve Dox-inducible factor expression.

For reprogramming source/founder cells, MEFs were prepared from day 13.5 embryos of the Rosa26:rtTA x Oct4:GFP mice. For feeder cells, MEFs were either derived from CF1 mice (Jackson Lab) and inactivated by gamma irradiation or purchased from Millipore.

Hematopoietic stem and progenitor cell isolation, culture and reprogramming

Hematopoietic populations were stained and FACS sorted on a BD Aria instrument as described previously (Akashi et al., 2000; Guo et al., 2010; Megyola et al., 2013). Briefly, mouse bone marrow cells were stained with an antibody cocktail containing hematopoietic lineages antibodies, c-Kit APC, Sca PE or pacific blue, CD34 FITC and CD16/32 PEcy7 (all from BD Biosciences or eBiosciences). GMPs are defined as lineage negative, c-Kit positive, Sca negative, CD34 and CD16/32 double positive. LKS are defined as lineage negative, c-Kit and Sca double positive. MEPs are lineage negative, c-Kit positive, Sca negative, CD34 and CD16/32 double negative. CMPs are lineage negative, c-Kit positive, Sca negative, CD34

positive and CD16/32 negative. For single cell sorting into 96 well plates, only the center 6x10 wells were used to avoid excessive drying of wells on the periphery. 80-90% efficiency was routinely achieved with single cell sorting into 96 well plates. The hematopoietic growth factors included 100 ng/ml mSCF, 50 ng/ml mIL3, 50 ng/ml Flt3L and 50 ng/ml mTPO (all from PeproTech) and were added to a serum free base medium X VIVO15 (BioWhittaker). Reprogramming conditions included inactivated MEF feeders, complete mESC medium (DMEM with 15% Hyclone FBS, Non-essential amino-acids, Pen/Strep, L-glutamine, β-mercaptoethanol and 1000U/ml LIF (Chemicon) and Dox (Sigma) at 2 μg/ml. Reprogramming efficiency from single cell reprogramming experiments was calculated by the number of wells containing Oct4:GFP+ colonies divided by the number of sorted wells.

CFSE labeling and sorting

CFSE was purchased from Life Technologies. A working concentration of 7.5-8 μM was used following manufacturer's instruction. CFSE labeled GMPs were cultured in hematopoietic growth medium X VIVO15 (supplemented with growth factors as described above) for 24 hours before sorting on CFSE levels. The brightest and dullest 15-20% of cells were sorted as slow and fast GMPs, respectively. To label MEFs, cultures were trypsinized and passed through a 70μM filter to generate single cell suspensions. CFSE labeled MEFs were cultured in mESC medium with 2 μg/ml of Dox for another 48 hours before sorting on CFSE levels. MEFs cultured in the absence of Dox were used as controls.

Time-lapse imaging and data analysis

All image acquisition and analysis were performed as described previously (Megyola et al., 2013). To ensure accuracy of tracing, all iPSC lineages were inspected manually and annotated using a custom program coded in Matlab.

Lentivirus production and transduction

Dox inducible polycistronic vector expressing the Yamanaka factors was purchased from Addgene (construct 20321). Lentiviral vector encoding the FUCCI reporter (monomeric Kusabira orange, mKO) was obtained from Amalgaam. Lentiviral constructs for control shRNAs or those that target p53 (sc-29436-SH) and p57 (sc-37621-SH) were purchased from Santa Cruz Biotechonolgy. Lentivirus production was performed as described by the RNAi Consortium at the Broad Institute (http://www.broadinstitute.org/rnai/public/static/protocols).

Computational simulation and estimation of cell cycle length

Hematopoietic progenitor cells were single cell sorted into 15µl of reprogramming medium (mESC medium supplemented with 2 µg/ml of Dox) in Terasaki plates (Keller et al., 1990) to facilitate scoring of the small number of cells confined to each wells. The wells were scored blindly immediately after sorting (time 0), 14 hours or 19 hours later. Cell cycle lengths were estimated using computational simulation with the following assumptions: (1) a single cell was sorted into each well at time 0, (2) the cell starts at a random point in a cell cycle at time 0, (3) the cell cycle length of the starting cell and its progeny follow a normal distribution with mean of the given cell cycle length, and standard deviation set to 10% or 20% of the mean and (4) the level of cell death is negligible. Because of assumptions 1 and 4, we excluded wells that had no cells or more than one cell at time 0, and grouped apoptotic wells to a separate category

(Dead). After simulation, each of the cell number combinations at 14 and 19 hours were given an estimated range of cell cycle duration. Specifically, for a given mean cell cycle length between 4 and 30 hours, we calculated the probability of observing the specific combinations of cell numbers at 14 hours and 19 hours. These combinations include: (1) >4 cells at 19 hour, (2) >2 cells at 14 hours, but <=4 cells at 19 hour, (3) =2 cells at 14 hour and >2 cells at 19 hour, and (4) <=2 cells at 19 hour. For any given combination, the range was determined so that for each cell cycle length within the range, there was at least 25% probability of observing such a combination.

Transcriptome analysis by RNA sequencing

Cells were directly sorted into an RNase-free 96-well PCR plate (USA Scientific) followed by cell lysis, cytoplasmic content separation and total RNA preparation. Prior to collecting sorted cells, the 96-well PCR plate was kept on ice and each of wells was loaded with 20U RNaseOut (Invitrogen) and 0.5ul dithiothreitol (DDT) (Invitrogen). 1,000 cells of each population of interest were sorted in each well using a SONY Reflection or BD Aria instrument. The cells were quickly spun down (2,000rpm, 2min) in the wells and then RNase-free phosphate-buffered saline (PBS) was added to bring the total volume to 2.5ul. Next, 2.5ul of 2x cell membrane selected lysis buffer was added and the cells were lysed by pipetting up and down for 5 times. The entire lysate solution was transferred to a 1.5ml Eppendorf tubes and spun at 8,000rpm for 5 min in a chilled centrifuge (4℃). The supernatant, which contains the total mature RNA from cytoplasm, was transferred to PCR tubes on ice. Finally 5μl of total RNA was obtained from each sorted cell sample.

The mRNA selection and reverse transcription was performed as described previously with some modifications (Pan et al., 2013). Briefly, 5'-phosphorylated oligo-GdT24 (pGdT24) primer was used to selectively reverse transcribe mRNAs. Before reverse transcription, the total RNA sample was mixed with dNTPs (0.5 mM) and pGdT24 (4 μM) and denatured on a pre-heated 68℃ PCR machine for 5 minutes, then rapidly cooled down by immediately placing it on ice. Subsequently, the first strand cDNA was synthesized with addition of the cocktail containing the First Strand Buffer, DDT, RNaseOut, and 10u/μl Superscript Reverse Transcriptase III (all from Invitrogen). Then, the double-stranded cDNA (dscDNA) was generated in a 50μl of reaction mix. The product was purified with the Genomic DNA Clean and Concentrator kit (Zymo). Afterwards, several steps including DNA end-blunting, 5'-end phosphorylation and ligation were performed with the End-It DNA End-Repair Kit (Epicentre) with T4 DNA ligase (Epicentre). With no further purification or deactivation of enzyme, the product was directly amplified by adding a mixture of the Reaction Buffer and DNA polymerase (15:1) from REPLI-g UltraFast Mini Kit (QIAGEN). When the reaction was completed, the amplification product was purified using the Genomic DNA Clean and Concentrator kit (Zymo) column as mentioned above. Finally, 3 to 5 μg (up to 8 μg) of amplified cDNA derived from mature mRNA was obtained, which was then evaluated by PCR using a set of primers as described in (Pan et al., 2013). This product was fragmented for sequencing library construction as described below.

Standard Illumina HiSeq2000 protocols were used to construct the sequencing library from the aforementioned cDNA amplicons. Briefly, the amplicons were sonicated 6 times with high power for 5 minutes using a Bioruptor Sonicator (Diagenode) and fragmented to an approximately 200-500 bp size range. After end-repairing, 3'-A tailing and ligation, the product was analyzed by gel electrophoresis (E-gel EX 2%, Invitrogen) and a 50bp range (300-350bp)

was selected for library construction. The gel patch in this range was cut out and the DNA was eluted. 5µl of eluent was further amplified by PCR using Phusion High-Fidelity DNA polymerase (NEB) for 8 cycles. The product was size selected again and the DNA concentration was quantitated by a Bioanalyzer (Agilent). Multiple samples were indexed and loaded to the Hi-Seq2000 for sequencing performed with 50bp single end reads. Base-calling was performed using the Illumina Casava 1.8.2 pipeline. All data has been deposited at the GEO database with accession number GSE53074.

Analysis of RNA-seq data

The first 8 bases of each fastq sequence read were removed using the fastx-toolkit (http://hannonlab.cshl.edu/fastx_toolkit/) to remove bases introduced during random priming. Sequences were mapped to the mouse mm9 genome using the splice aware tophat version 2.0.4 alignment software. RNA-sequencing quality control analysis and generation of RPKM data was performed using the RNA-SeQC software (http://www.ncbi.nlm.nih.gov/pubmed/22539670). The number of reads in each mouse gene was counted using the htseq-count software (http://www-huber.embl.de/users/anders/HTSeq/doc/overview.html). Differentially expressed genes were identified using the R bioconductor edgeR exactTest function (http://www.ncbi.nlm.nih.gov/pubmed/19910308) after TMM normalization (http://www.ncbi.nlm.nih.gov/pubmed/20196867). Only genes with 3 or more samples with more than 1 count per million reads were analyzed to exclude poorly expressed genes. Multidimensional scaling analysis was performed using the edgeR plotMDS function to assess overall sample relatedness. Normalized coverage bigwig files were created using the

bedtools genomeCoverageBed command, using a scaling factor of number of reads/1000000 (http://www.ncbi.nlm.nih.gov/pubmed/20110278). Coverage files were visualized using the IGV genome browser (http://www.ncbi.nlm.nih.gov/pubmed/22517427). Read counts in coding and UTR regions of selected genes were determined using the bedtools coverageBed command.

To compare RNA-seq data with published gene signatures, Cluster I and VI genes from Polo et al. (Polo et al., 2012) were mapped with gene symbol to the RNA-seq data. Samples from iPSC, MEF, fastMEF and slowMEF were row centered and normalized (by subtracting with mean and dividing with standard deviation for each gene), and plotted with heatmap.

To perform PCA analysis, a custom code was written in Matlab, and the "pca" function within matlab statistics toolbox was used.

For GSEA analysis, first, fastGMP was compared to slowGMP, and second, fastMEF and fastGMP were compared to slowMEF and MEF samples. GSEA version 2.0 program was used with v4.0 gene sets within MSigDB. To analyze GO category, c5 collection of gene sets were run. To perform focused analysis on metabolism, EMT or apoptosis, c2 and c5 gene set collections were filtered so that at least one of the following words was found in the gene set name: apoptosis, metabolism, EMT and mesenchymal. This resulted in a total of 136 gene sets. GSEA was then performed with 5000 permutations on the genes to derive statistics.

Quantitative RT-PCR

Superscript III (Invitrogen) was used for reverse transcription. Primer pairs for analyzing Sox2 3'UTR are the following: pair1 forward GGACCATGTATAGATCTGGAGGA and pair 1 reverse GCCCGGAGTCTAGCTCTAAATA; pair 2 forward CAAGGAAGGAGTTTATTCGGATTTG and pair 2 reverse

CGTTTGCCTTAAACAAGACCAC. 18S RNA (ABI) was used as a control to normalize expression data. Relative expression values were calculated as 2-ΔCt. Ct cycle values were determined in triplicates from quantitative PCR (qPCR). For MBD3, the Taqman assay Mm00488961_m1 (Life Technologies) was used.

Western blot analysis

Western blot analysis was performed as described before (Guo et al., 2012). Cells were lysed with RIPA buffer containing protease inhibitors and analyzed using 10% polyacrylamide gels. Antibodies against p53 (#2524), p57 (#2557) and MBD3 (#3896) were from Cell Signaling. Hsp90 antibody was from Stressgen (SPS-771) and β-actin antibody was from Abcam (AB8227).

Statistics

All *p* values were derived using student's t test, unless otherwise noted. Bar graphs are represented as mean +/- standard deviations.

REFERENCES

Akashi, K., Traver, D., Miyamoto, T., and Weissman, I.L. (2000). A clonogenic common myeloid progenitor that gives rise to all myeloid lineages. Nature 404, 193-197.

Carey, B.W., Markoulaki, S., Beard, C., Hanna, J., and Jaenisch, R. (2010). Single-gene transgenic mouse strains for reprogramming adult somatic cells. Nat Methods 7, 56-59.

Eminli, S., Foudi, A., Stadtfeld, M., Maherali, N., Ahfeldt, T., Mostoslavsky, G., Hock, H., and Hochedlinger, K. (2009). Differentiation stage determines potential of hematopoietic cells for reprogramming into induced pluripotent stem cells. Nat Genet 41, 968-976.

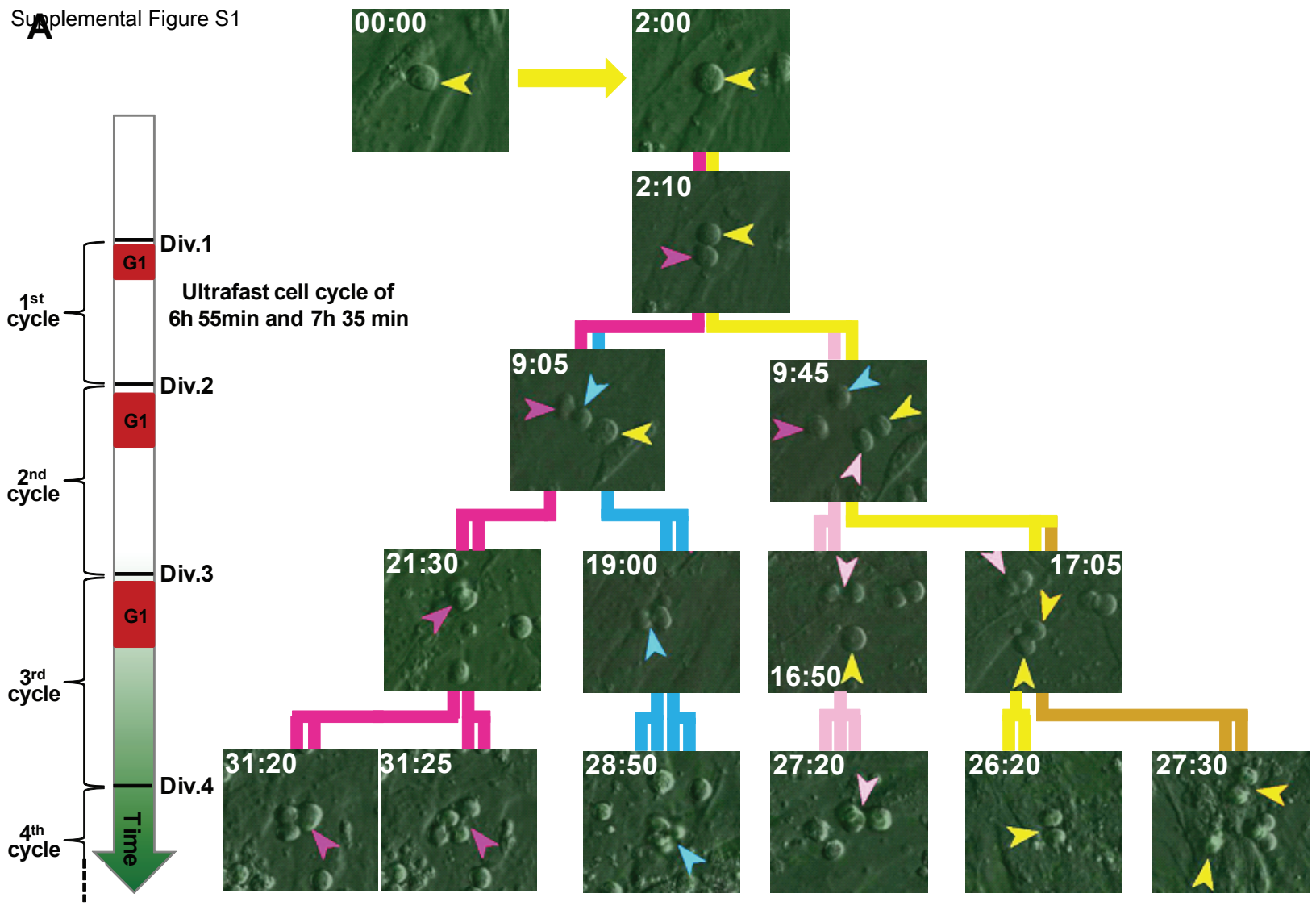
Guo, S., Lu, J., Schlanger, R., Zhang, H., Wang, J.Y., Fox, M.C., Purton, L.E., Fleming, H.H., Cobb, B., Merkenschlager, M., et al. (2010). MicroRNA miR-125a controls hematopoietic stem cell number. Proc Natl Acad Sci U S A 107, 14229-14234.

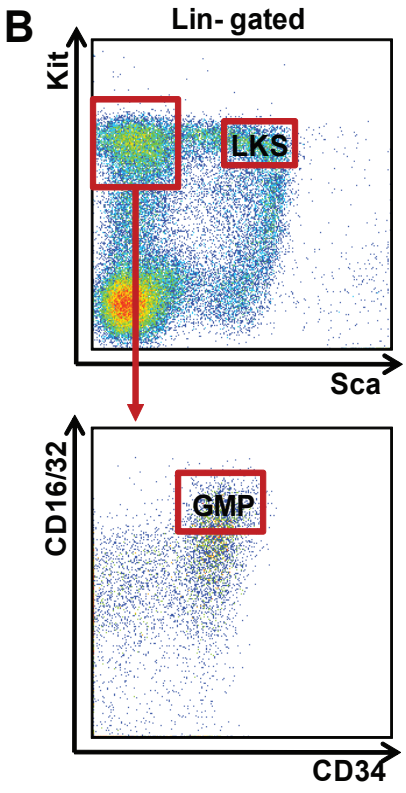
Keller, J.R., McNiece, I.K., Sill, K.T., Ellingsworth, L.R., Quesenberry, P.J., Sing, G.K., and Ruscetti, F.W. (1990). Transforming growth factor beta directly regulates primitive murine hematopoietic cell proliferation. Blood 75, 596-602.

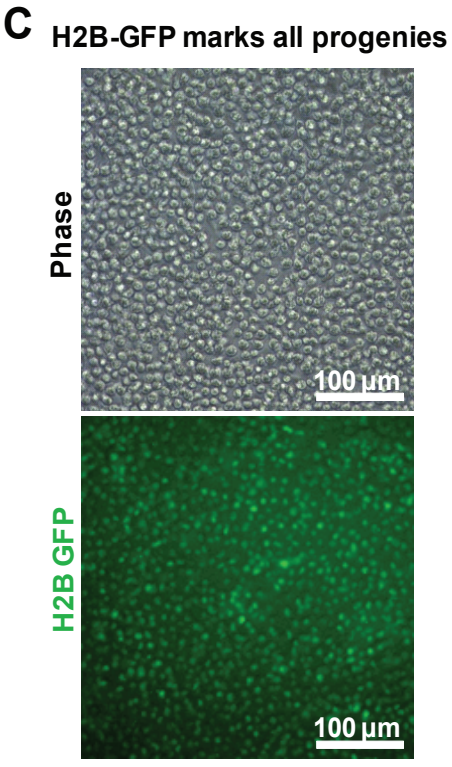
Megyola, C.M., Gao, Y., Teixeira, A.M., Cheng, J., Heydari, K., Cheng, E.C., Nottoli, T., Krause, D.S., Lu, J., and Guo, S. (2013). Dynamic Migration and Cell-Cell Interactions of Early Reprogramming Revealed by High Resolution Time-lapse Imaging. Stem Cells.

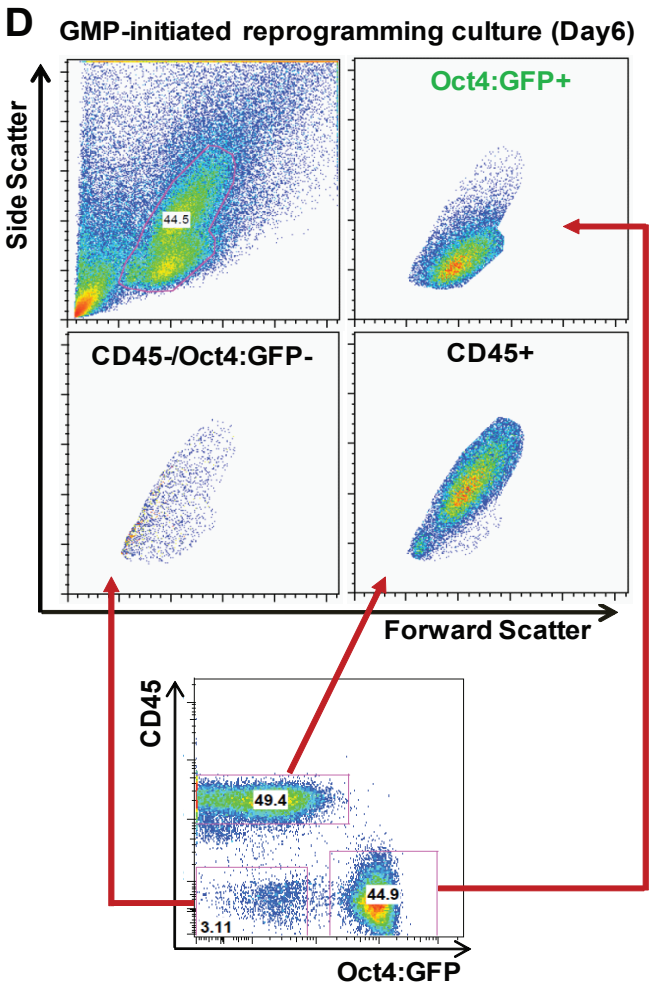
Pan, X., Durrett, R.E., Zhu, H., Tanaka, Y., Li, Y., Zi, X., Marjani, S.L., Euskirchen, G., Ma, C., Lamotte, R.H., et al. (2013). Two methods for full-length RNA sequencing for low quantities of cells and single cells. Proc Natl Acad Sci U S A 110, 594-599.

Polo, J.M., Anderssen, E., Walsh, R.M., Schwarz, B.A., Nefzger, C.M., Lim, S.M., Borkent, M., Apostolou, E., Alaei, S., Cloutier, J., et al. (2012). A molecular roadmap of reprogramming somatic cells into iPS cells. Cell 151, 1617-1632.



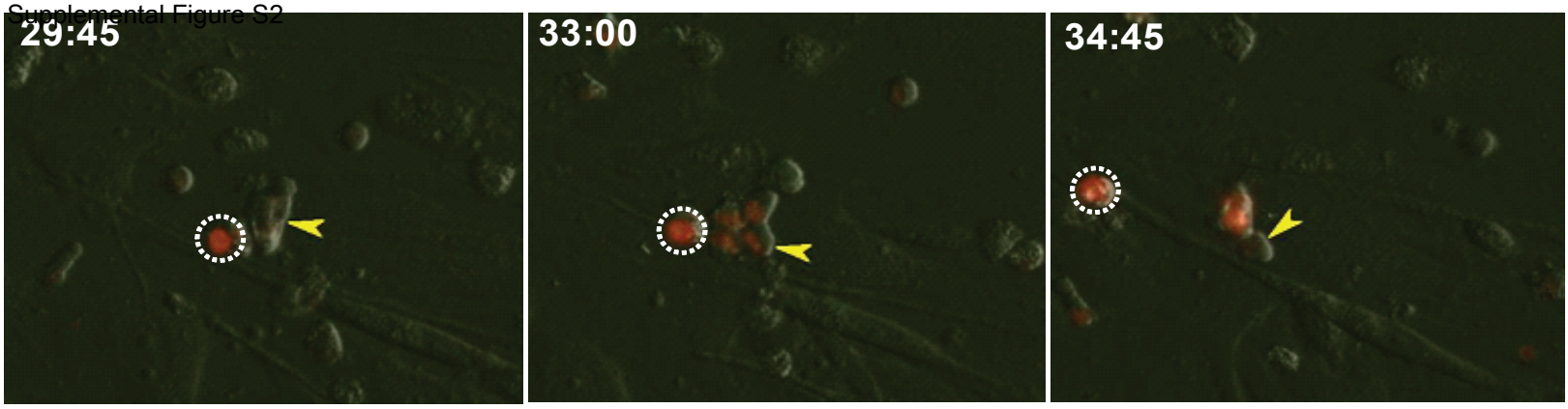






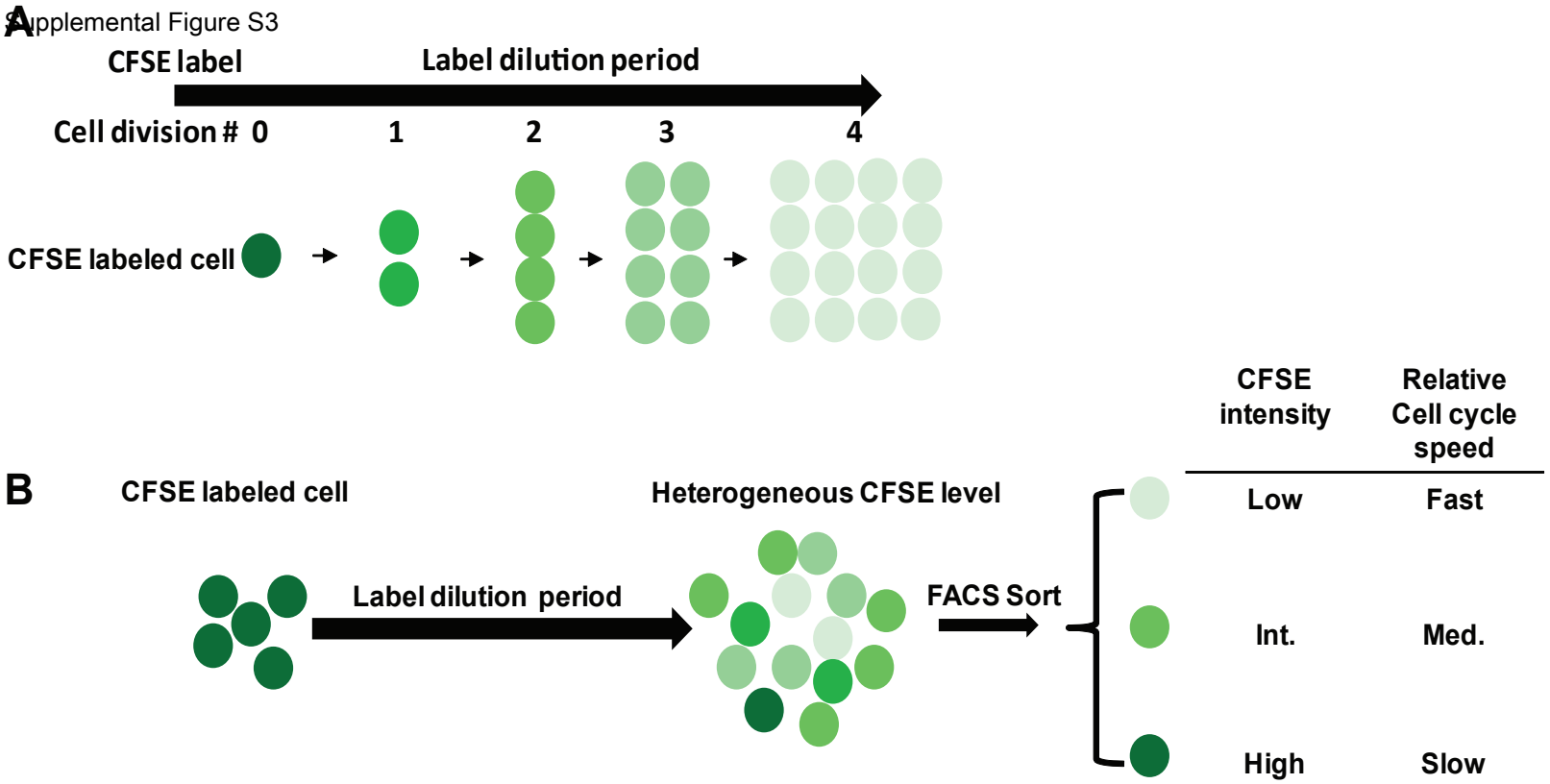
Guo et al. Fig S1

Supplemental Figure S2



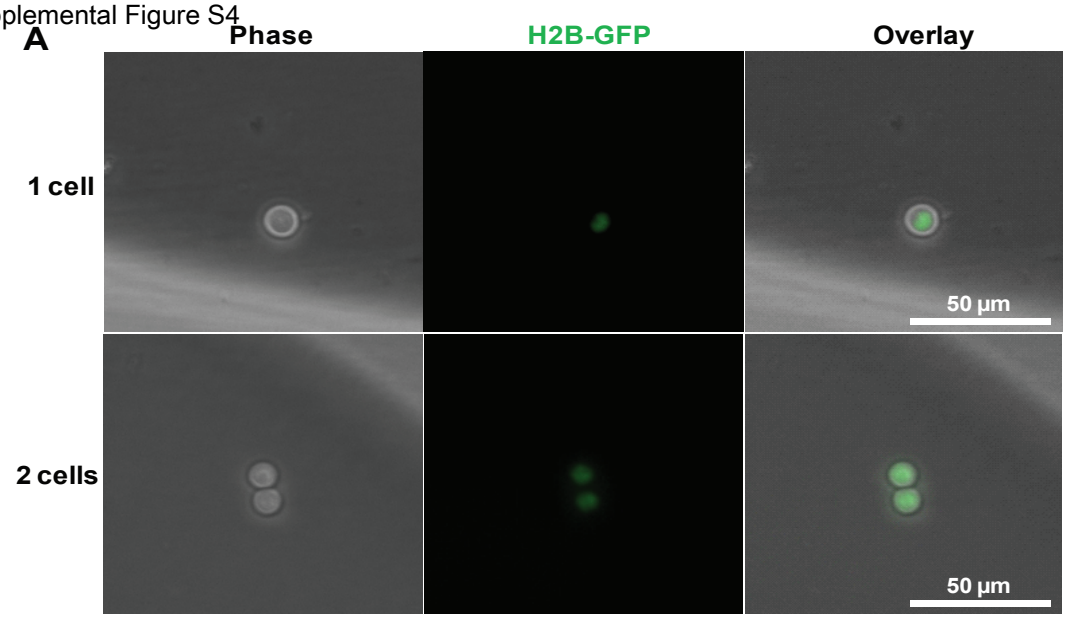
Guo et al. Fig S2

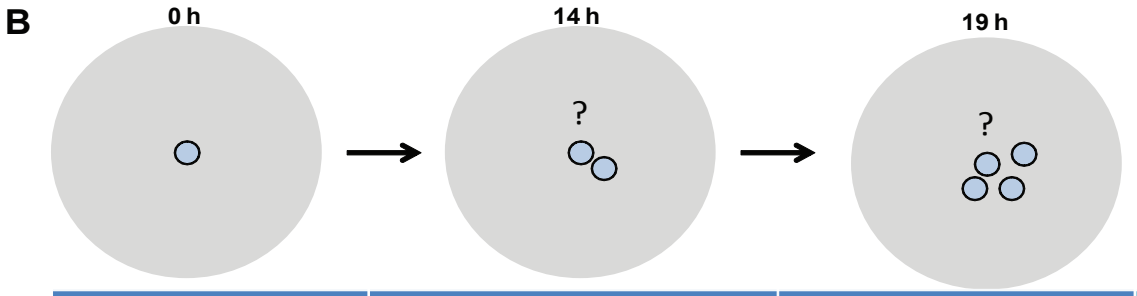
Supplemental Figure S3



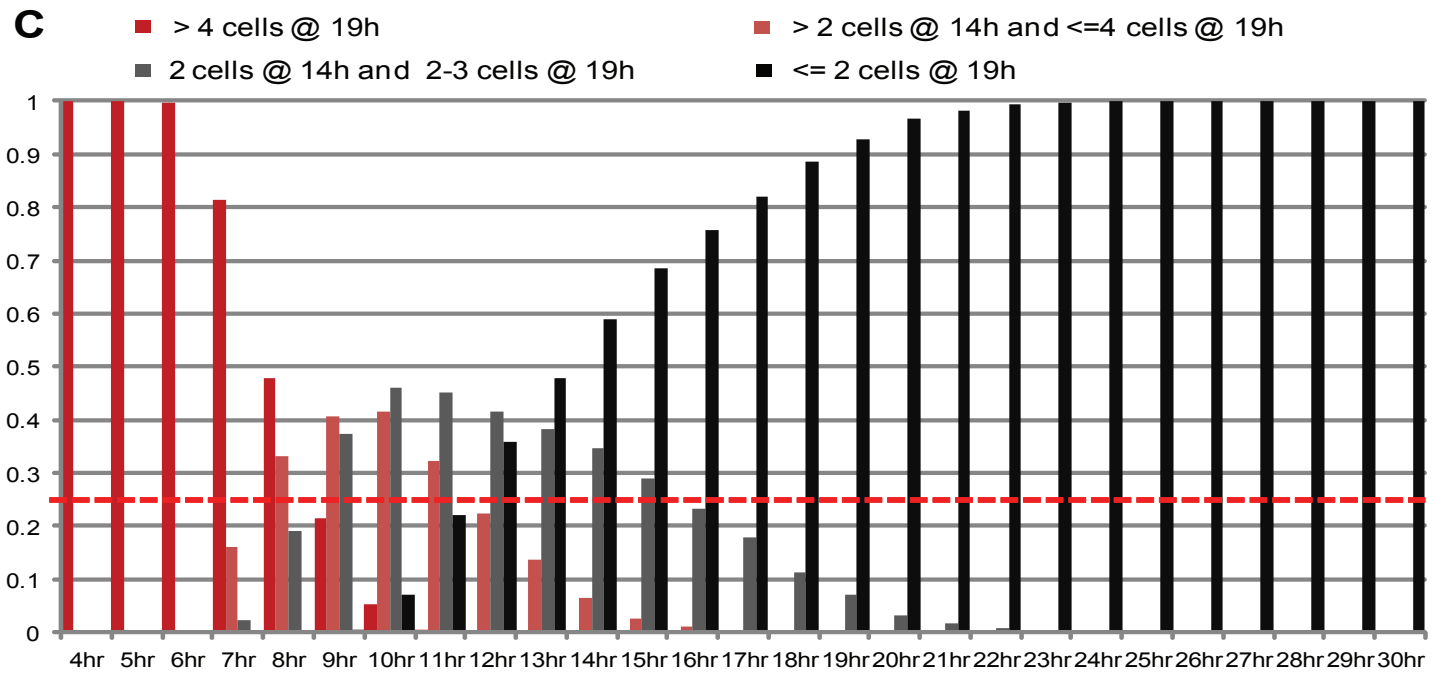
Guo et al. Fig S3

Supplemental Figure S4





Cell number @ 0h	Cell number @ 14h	Cell number @ 19h	Estimated cycle speed
1	0		Dead
1	1	1 or 2	>=12
	2	2	
1	2	3 or more	9-15
1	3 or more	Less than 5	8-11
1		5 or more	4-8

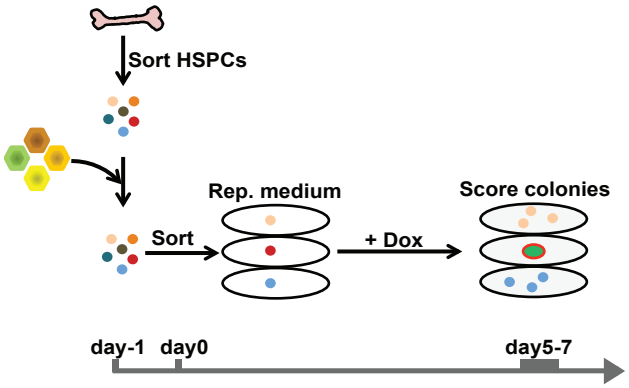


Guo et al. Fig S4

Supplemental Figure S5

A. Reprogramming from single cells (fresh cells)

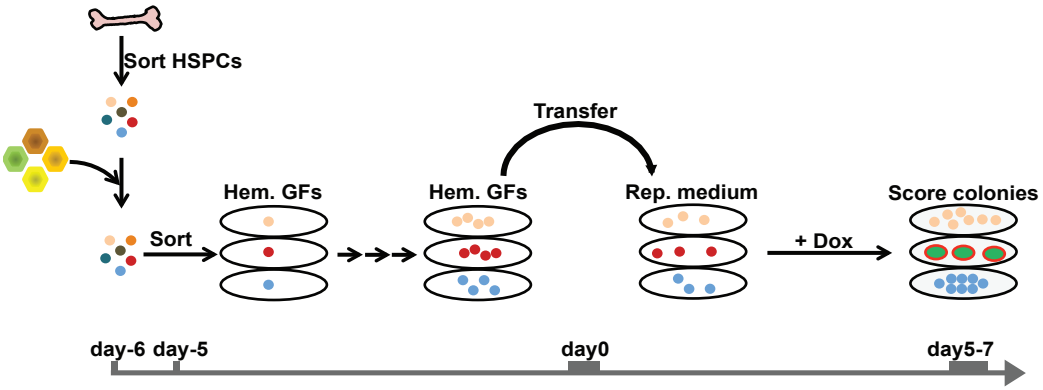
Reprogramming efficiency = $\frac{\text{Number of wells containing Oct4:GFP+ colonies}}{\text{Number of wells sorted}}$



d-1: Sort from BM and transduce with 4F
d0: Single cell sort into 96 wells containing reprogramming medium
d5-7: Score wells containing Oct4 GFP+ colonies

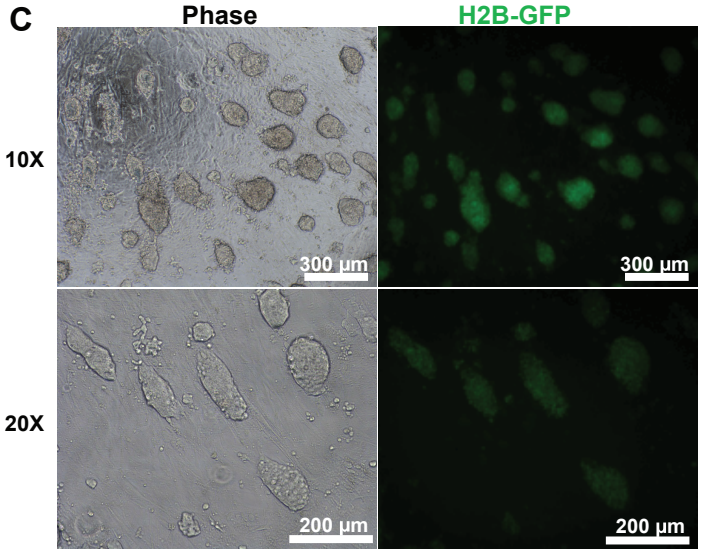
B. Reprogramming from single cells after hematopoietic culture (cultured)

Reprogramming efficiency = $\frac{\text{Number of wells containing Oct4:GFP+ colonies}}{\text{Number of wells sorted}}$



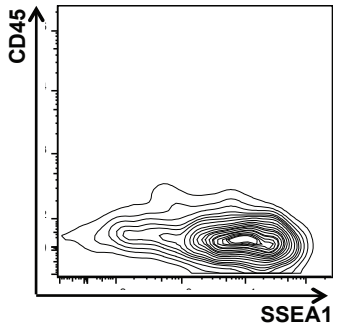
day-6: Sort from BM and transduce with 4F
day-5: Single cell sort into 96 wells containing hematopoietic GFs
day0: Transfer from hematopoietic medium into reprogramming medium
day5-7: Score wells containing Oct4:GFP+ colonies

C

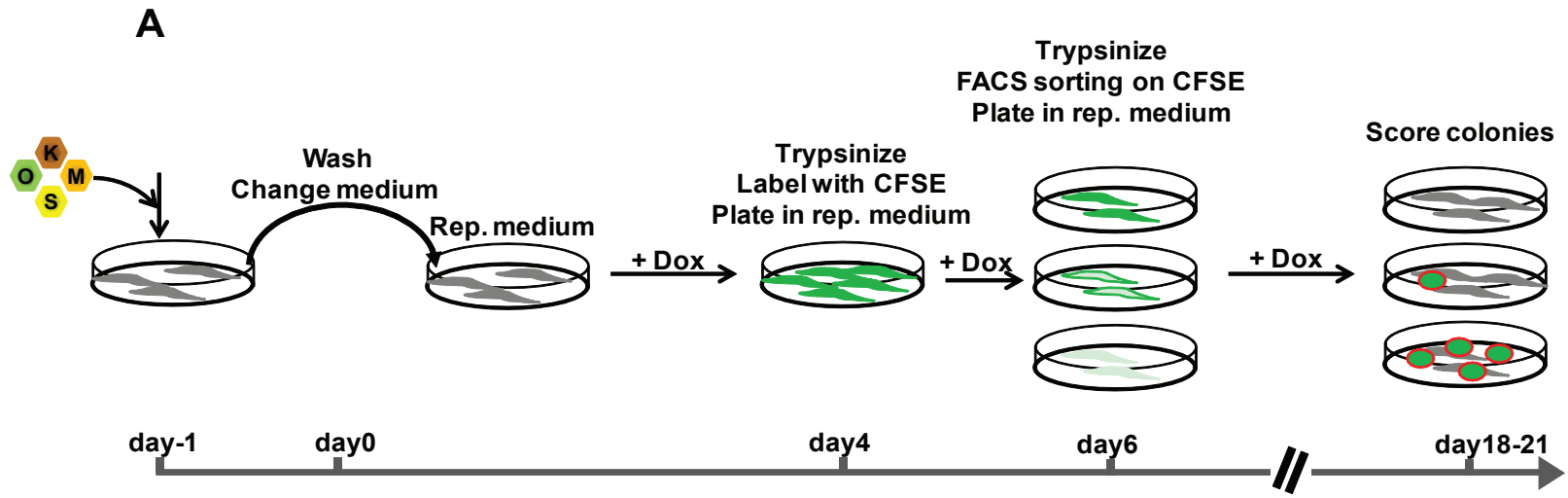


Progeny of a single LKS cell
Day 5 on Dox

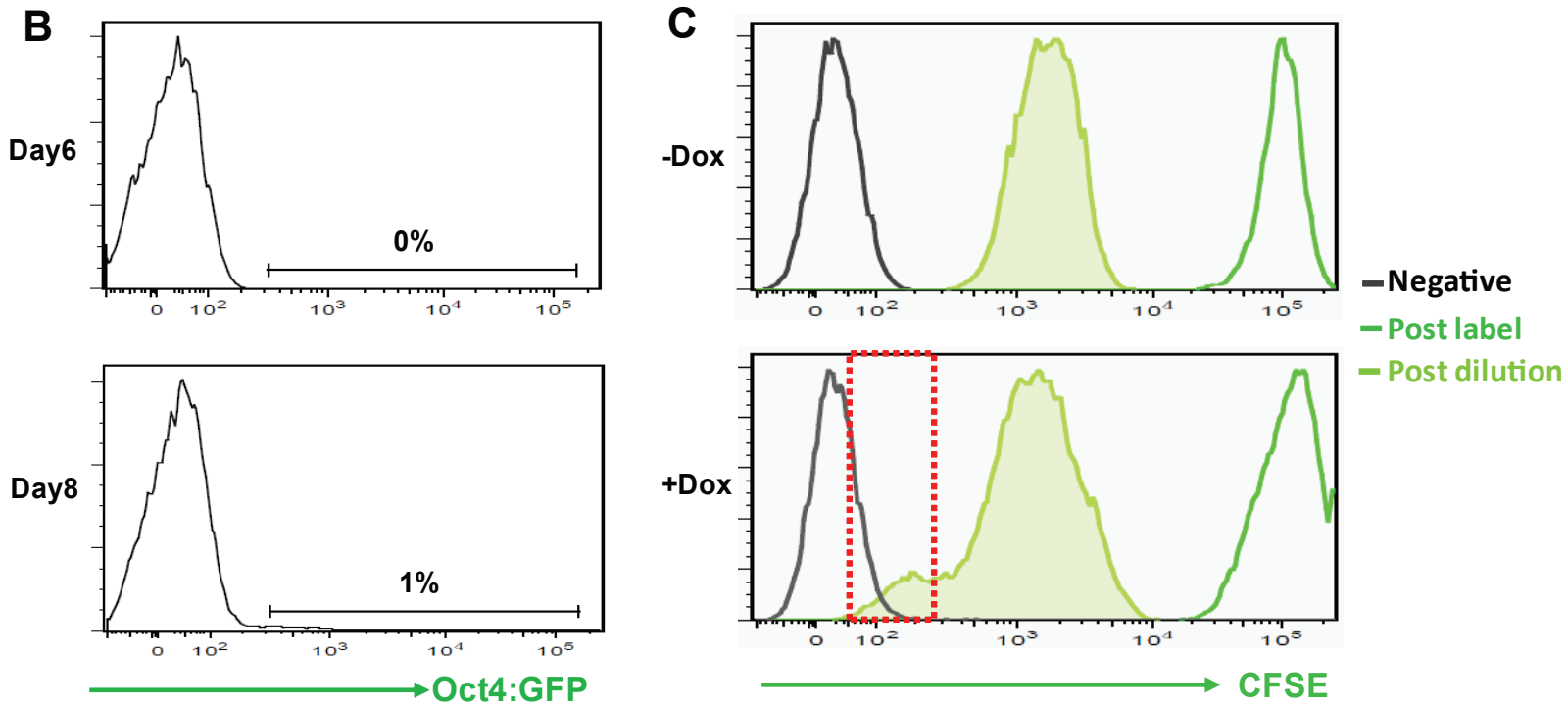
D



Supplemental Figure S6



d-1: Transduce with 4F
d0: Wash away virus and add reprogramming medium
d4: Trypsinize culture and label with CFSE. Plate labeled cells back in reprogramming medium
d6: Trypsinize culture and FACS sort for fast/intermediate/slow cells based on CFSE intensity. Plate sorted cells back in reprogramming medium
d8-18: Feed cells with fresh reprogramming medium every 2-3 days
d18-21: Score for Oct4:GFP+ and/or alkaline phosphatase+ colonies



Guo et al. Fig S6

Supplemental Figure S7