Appendix

S1: MSC isolation and differentiation studies

MSC isolation. Mice were bred at the breeding center Janvier (Le Genest-Saint-Ile, France) before being sacrificed in the animal facility of the U972 laboratory according to institutional guidelines. Female or male C57Bl/6 mice, 4-8 week old, were killed by cervical dislocation. Femurs and tibias were removed, cleaned of all adherent tissues and placed in PBS (phosphate-buffered saline) with 5% selected fetal calf serum (FCS, Hyclone) and antibiotics (100 U/mL penicillin, 100 µg/mL streptomycin). Ends of bone pieces were clipped and bone marrow was collected by flushing femur and tibia with FCS. Cell aggregates were dissociated with 23/26-gauge syringes.

Cells were plated at 1.5 10^6 cells/cm² in Minimum Essential Medium with GlutaMAX™-I and without nucleosides (alpha-MEM, Gibco) supplemented with 15% FCS, 2 ng/mL basic fibroblast growth factor-2 (FGF2, AbCys) and antibiotics (100 U/mL penicillin, 100 µg/mL streptomycin, Gibco). Cultures were then incubated in fully humidified atmosphere with 5% CO_2 at 37°C. After 48 hours, non-adherent cells were removed by washing with PBS and fresh medium was added. Culture medium was changed every 2 to 3 days. When sub-confluence (80 to 90%) was reached cells were harvested using 0.05% trypsin/EDTA (Gibco). Cells were re-plated for expansion in alpha-MEM at density of 10^5 cells/cm^2. Complete medium change was carried out every 2 to 3 days. Primary layers were maintained for up to 4 passages before clone selection.

Differentiation studies. Clones were induced to differentiate into osteoblastic, adipogenic and chondrogenic pathways. Adipogenic medium consisted in Dulbecco's modified Eagle Medium (DMEM) high-glucose (Gibco) with 10% FCS (Hyclone), 1µM dexamethasone, 0.5mM isobutyl-methylxanthine (IBMX), 60 µM indomethacin, 10µM rosiglitazone and 200 ng/mL insulin (all from Sigma). Cells ($2x10^5$ cells/cm²) were incubated for 1-2 weeks with medium change every 2-3 days. For osteogenic differentiation, DMEM high-glucose was supplemented with 0.1 µM dexamethasone, 25 µg/mL ascorbic acid and 3 mM NaH_2PO_4 (Sigma). Cells (10^5 cells/cm²) were incubated for 21 days with medium change every 2-3 days. For chondrogenic differentiation, cells from a cell suspension containing 3.10^7cells/mL were aggregated in a droplet of 20µl in one well of a 24-well plate. Droplets were incubated without culture medium for 3 hours to allow micromass formation, then 250 µl of complete DMEM medium with 10% FCS were added gently to each well. After 24h, the medium was replaced by chondrogenic medium consisting in DMEM high glucose without FCS, with 1X insulin-transferrin-selenium (Lonza), 0.1 µM dexamethasone, 1 mM sodium pyruvate, 170 µM ascorbic acid-2-phosphate, 350 µM proline (Sigma), 10 ng/mL transforming growth factor (TGF-β1, AbCys) and 10ng/mL bone morphogenetic protein-6 (BMP-6, R&D systems).

Evidence for adipocytic, osteoblastic and chondrocytic differentiation was provided by histochemical stains and expression of differentiation-affiliated transcripts using Taqman Low-Density Arrays from Applied Biosystems® as described previously [29]. For immunofluorescence studies cells were grown on Lab-Tek chambers, fixed in cold methanol before staining with monoclonal anti-vimentin Vim13.2 and anti-nestin antibodies (from Dako).

Fig. S1: Differentiation potential of the MSC clones.
Upper panel: differentiation into adipocytes (Nile red, staining, Day 14, bar: 20µm), osteoblasts (alizarin red staining, Day 21) and chondrocytes (safranin O staining, Day 21, bar: 100µm). Ctl: control; Tt: treated. Lower panel: QRT-PCR of typical marker expression of the respective lineages. Left: adipocytic lineage, middle: osteoblastic lineage, right: chondrocytic lineage. Example shown for clone C1 at P4.

S2: Self-organizing-maps (SOMs)

After preprocessing, the expression data are analyzed using self-organizing maps (SOMs) as described previously [43]. The SOM method transforms the expression pattern of more than 34,000 genes of each sample into an intuitive mosaic image using a pixelation of 20x20 tiles by iterative machine learning. Each of the pixels represents a microcluster of a few up to hundreds of genes with similar expression profiles using the Euclidean distance as similarity metrics. Their mean expression is assigned to a so-called metagene serving as representative prototype of each pixel. The expression state of each sample is defined by the expression values of all 400 metagenes and visualized as expression landscape by appropriate color coding of the meta-gene pixels (red-green-blue indicates over- to under-expression relative to the mean expression of the respective genes in all samples studied). These expression portraits typically show a smooth blurry texture with a certain number of red overexpression and blue under-expression spots due to the fact that neighbored meta-genes possess similar expression profiles and thus tend to be colored similarly. Each spot consequently represents an expression module of genes co-expressed in terms of strongly correlated expression profiles.

Fig. S2: SOM-gallery of the regeneration of global gene expression profiles following Sca-1 sorting.
Shown are the SOM-portraits of the mother population (WCP) and of Sca-1 sorted subpopulations (Sca-1L, Sca-1M, and SCa-1H) immediately after sorting (D0) and at day 4 (D4) and day 7 (D7) after sorting for 3 different clones (AC5, AC3 and C1). Differences between the subpopulations are hidden behind the large clone specific differences and the expression changes during regeneration.

S3: Chromatin states of additional MSC markers

Beside the chromatin state of the Ly6a promoter, also the chromatin states of the promoters of CD105, CD146, CD44 and CD140a were characterized by chromatin immunoprecipitation of H3K4me3 and H3K27me3 as described in the Material and Methods section. These MSC markers were selected according to their gene expression in Sca-1 sorted populations (see Fig. S3A). The primer sequences used are given in Table S1.

Fig. S3: Chromatin states of MSC markers in Sca-1H vs- Sca-1L populations.
A) Fold changes (in Sca-1H vs. Sca-1HL cells) in expression of 13 MSC markers commonly used in literature. Shown are averages of the 3 clones analyzed (see Fig. S2). Boxes: 4 markers have been selected for ChIP.

B) H3K4me3 modification level at the promoter of the selected MSC markers in Sca-1H and Sca-1L populations (n=3, error bars: SD). For comparison the modification of the Ly6a promoter in the same sorting experiments is given as well. A significant H3K4me3 enrichment is observed for the Sca-1 promoter only. C) H3K27me3 modification at the promoters of the genes shown in (B) in the same populations. No significant enrichment is observed. (*=0.05, T-test, measurements are independent from those shown in Fig. 3)

S4: Model version.

In our simulations we used a revised version of the previous model [35] assuming a positive feedback between gene expression and H3K4me3 modification of the gene promoter. In contrast to the original model, transcription T_i of a gene i is calculated according to:

$$\frac{dT_i}{dt} = P_{max}(m + m_0)\theta_{pro,i} - \delta T_i,$$

where m is the H3K4me3 modification level and m_0=0.1 is a small constant ensuring a baseline gene expression at m=0. This linear dependence on the modification level instead on the binding probability of the H3K4me3 modification complex ensures a more robust systems behavior, which is even improved by increasing the degradation constant δ from 0.1 to 1.0. P_{max} is the maximal promoter activity of the gene and $\theta_{pro,i}$ is the occupancy of the promoter by polymerase II.

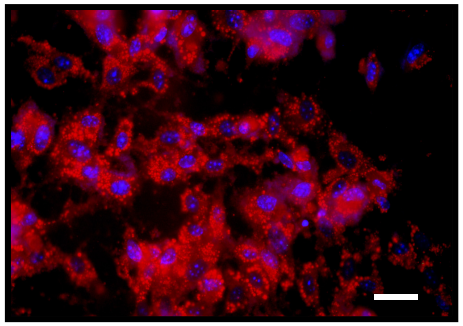
Table S1: Additional primer sequences used in ChIP.

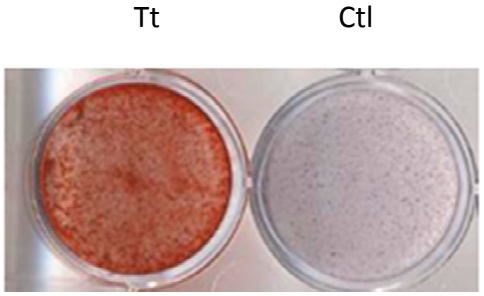
Gene	Primer sequences
CD105	mCd105-ChIP-for 5'- ACC CTG GGT GTT ACA GGT CT-3' mCd105-ChIP-rev 5'- GAA GGT GCC AAC GCT AAC T-3'
CD146	mCd146-ChIP-for 5'- TGG GAG TTT GGG GAT GTC TT-3' mCd146-ChIP-rev 5'- GGG ACT CTG GCA AGA AGG AG-3'
CD44	mCd44-ChIP-for 5'- CTG GAT GCC TCT GGG TTG A-3, mCd44-ChIP-rev 5'- CAG CAC TTC CCA GGC GTA -3'
CD140a	mCd140a-ChIP-for 5'- ACT TTC CTC CCA ACG CTC AG-3' mCd140a-ChIP-rev 5'- CCA AGA GAT CGC TTC CTG CT-3'

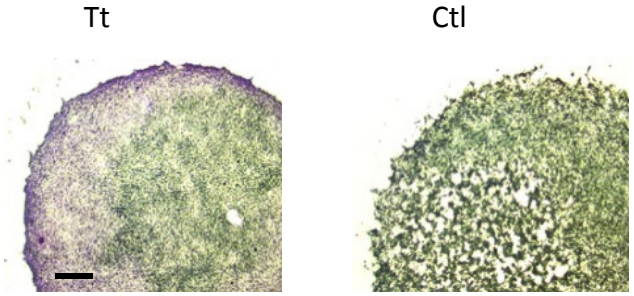
Table S2: Model parameters.

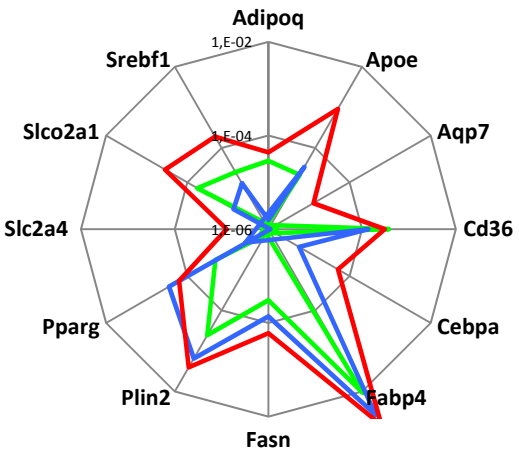
The unit k_bT refers to the Boltzmann unit. Simulation time steps are denoted by Δt. Thereby, 100 Δt refer to about 1 effective cell doubling time τ=22.8h. Parameters of the artificial genome and the TF networks were chosen as in Binder et al. [34]. The energy constants were chosen from realistic ranges for protein binding of some k_bT. Simulations were started assuming that all gene promoters are fully H3K4me3 modified and that all associated CpGs are un-methylated.

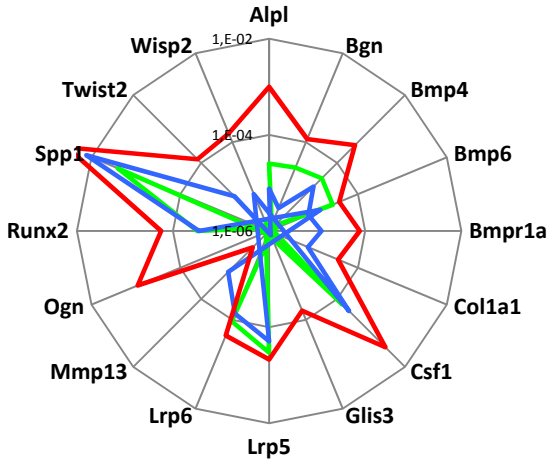
Parameter	**Symbols**	**Values**
Ground enthalpy per bound interaction complex	ε_1	6 k_bT
Free enthalpy change of CpG binding	ε_{BS}	-5 k_bT
Free enthalpy change of H3K4me3 binding	ε_{HM}	-2 k_bT
Histone modification rate	$k_{M,0}$	0.05 / Δt
Histone de-modification rate	k_D	0.005 / Δt
DNA maintenance methylation probability	D_{main}	0.933
DNA de novo methylation probability	D_{novo}^{0}	0.1
Interaction energy between DNA- and histone- methyltransferase	ε_{methyl}	6 k_bT
Cell growth rate (after 10 growth steps the cell gets divided)	R_0	0.1 / Δt
Maximal promoter activity	P_{max}	1000/ Δt
Transcript degradation rate	δ	1.0 / Δt




Tt
Ctl


Tt
Ctl


Adipoq
Apoe
Aqp7
Cd36
Cebpa
Fabp4
Fasn
Plin2
Pparg
Slc2a4
Slco2a1
Srebf1
1,E-02
1,E-04
1,E-06


Alpl
Bgn
Bmp4
Bmp6
Bmpr1a
Col1a1
Csf1
Glis3
Lrp5
Lrp6
Mmp13
Ogn
Runx2
Spp1
Twist2
Wisp2
1,E-02
1,E-04
1,E-06

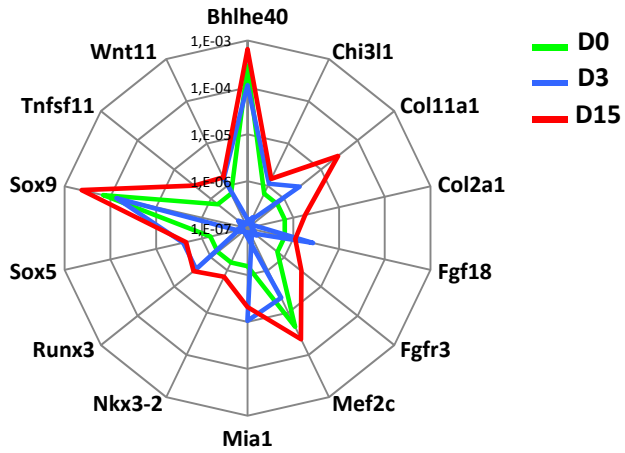

Bhlhe40
Chi3l1
Col11a1
Col2a1
Fgf18
Fgfr3
Mef2c
Mia1
Nkx3-2
Runx3
Sox5
Sox9
Tnfsf11
Wnt11
1,E-03
1,E-04
1,E-05
1,E-06
1,E-07
D0
D3
D15

Figure S1

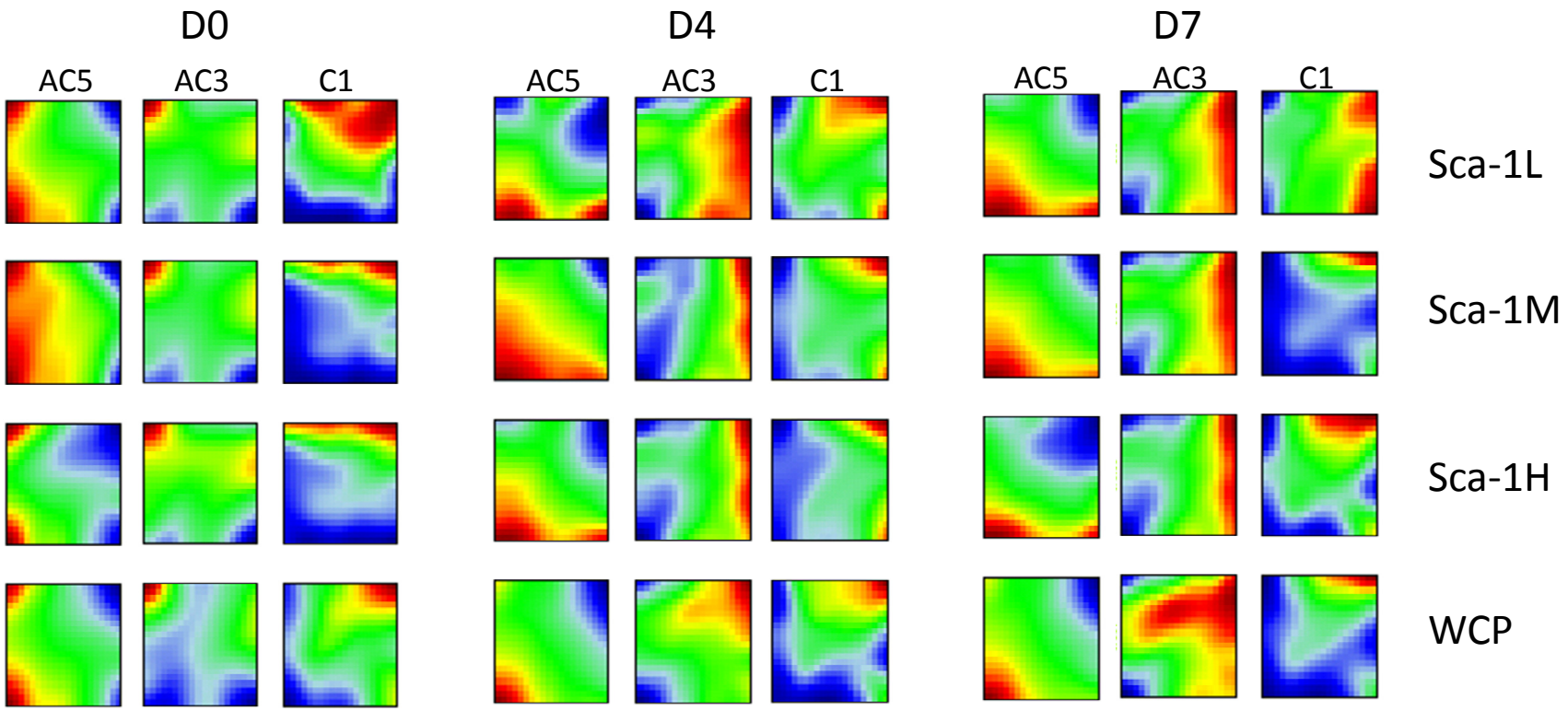
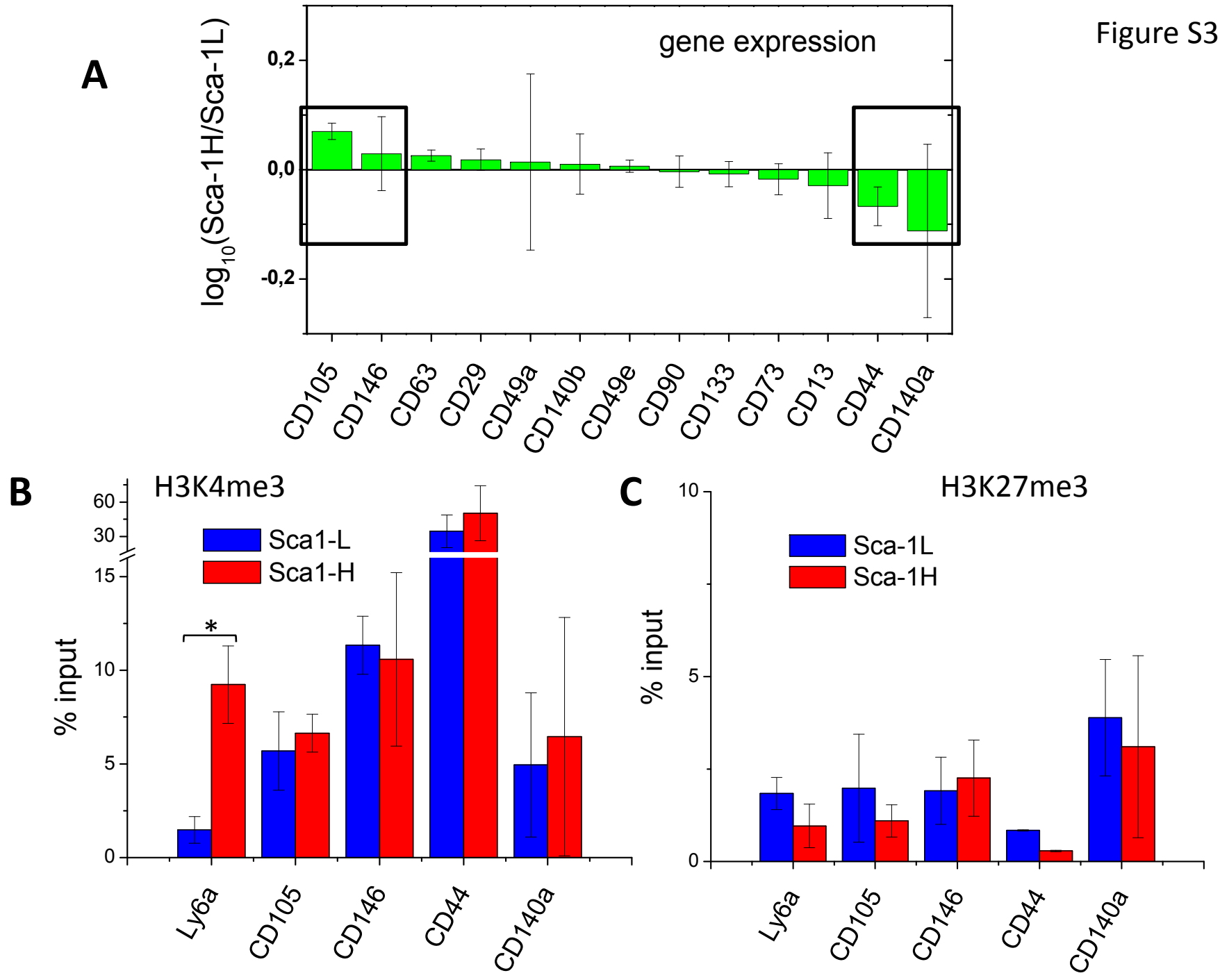


Figure S2

Figure S3


A
gene expression
log10(Sca-1H/Sca-1L)
0,2
0,0
-0,2
CD105
CD146
CD63
CD29
CD49a
CD140b
CD49e
CD90
CD133
CD73
CD13
CD44
CD140a
B
H3K4me3
Sca1-L
Sca1-H
% input
60
30
15
10
5
0
*
Ly6a
CD105
CD146
CD44
CD140a
C
H3K27me3
Sca-1L
Sca-1H
% input
10
5
0
Ly6a
CD105
CD146
CD44
CD140a