ONLINE METHODS

Generation of iPSC lines. iPSC lines were generated as described previously[2]. Briefly, iPSC-derived somatic cells were isolated from chimeras by fluorescence-activated cell sorting (FACS), plated on feeders in the presence of cytokines in ESC culture conditions. Resultant iPSC colonies were picked and expanded in the absence of doxycycline and used for subsequent analyses.

SMP isolation. Myofiber-associated cells were prepared from intact limb muscles (extensor digitorum longus, gastrocnemius, quadriceps, soleus, traverus abdominis and triceps brachii) as described previously[44,45]. Briefly, intact mouse limb muscles were digested with collagenase II to dissociate individual myofibers. These were triturated and digested with collagenase II and dispase to release myofiber-associated cells. The myofiber-associated cells were next unfractionated by FACS, using the following marker profiles for each population: (i) SMPs: $CD45^{-}Sca\text{-}1^{-}Mac\text{-}1^{-}CXCR4^{+}\beta1\text{-}integrin^{+}$; (ii) Myoblast-containing population: $CD45^{-}Sca\text{-}1^{-}Mac\text{-}1^{-}CXCR4^{-}$; (iii) $Sca1^{+}$ mesenchymal cells: $D45^{-}Sca\text{-}1^{+}Mac\text{-}1^{-}$. After the initial sort, cells were resorted by FACS using the same gating profile to increase the purity of the obtained population[46].

Blastocyst injections. For blastocyst injections, female BDF1 mice were superovulated by intraperitoneal injection of PMS and hCG and mated to BDF1 stud males. Zygotes were isolated from females with a vaginal plug 24 h after hCG injection. Zygotes for 2n injections were cultured for 3 d *in vitro* in KSOM media, blastocysts were identified, injected with ESCs or iPSCs and transferred into pseudopregnant recipient females.

Teratoma formation. iPSCs were harvested by trypsinization, preplated onto untreated culture plates to remove feeders as well as differentiating cells and injected into flanks of nonobese diabetic/severe combined immunodeficient NOD/SCID mice, using ~5 million cells per injection. The mice were euthanized 3–5 weeks after injection, teratomas dissected out and processed for histological analysis.

Cellular growth assays. To measure the clonal growth potential of iPSCs, SSEA1-positive cells from the different iPSC lines were sorted into 96-well plates by FACS (BD). After 7 d, the presence of iPSC colonies was scored based on morphology. To establish growth rates, the different bulk iPSCs lines or derivative subclones were plated in six gelatinized wells of a 12-well plates and each day the number of cells was counted in duplicate using a Countess cell counter (Invitrogen). For colorimetric measurement of growth, iPSCs lines were subcloned into 96-well plates and after 7 d, the cells were exposed to XTT (TOX-2) (Sigma) reagent overnight and the absorbance at 450 nm measured with a multiwell plate reader (Molecular Devices).

Cell culture. ESCs and iPSCs were cultured in ESC medium (DMEM with 15% FBS, L-glutamin, penicillin-streptomycin, nonessential amino acids, β-mercaptoethanol and 1,000 U/ml leukemia inhibitor factor) on irradiated feeder cells. TTF cultures were established by trypsin digestion of tail-tip biopsies taken from newborn (3–8 d of age) chimeric mice produced by blastocyst injection of iPSCs.

RNA isolation. ESCs and iPSCs grown on 35-mm dishes were harvested when they reached about 50% confluency and preplated on nongelatinized T25 flasks for 45 min to remove feeder cells. Cells were spun down and the pellet used for isolation of total RNA using the miRNeasy Mini Kit (Qiagen) without DNase digestion. RNA was eluted from the columns using 50 ml RNAse-free water or TE buffer, pH7.5 (10 mM Tris-HCl and 0.1 mM EDTA) and quantified using a Nanodrop (Nanodrop Technologies).

Quantitative PCR. cDNA was produced with the First Strand cDNA Synthesis Kit (Roche) using 1 mg of total RNA input. Real-time quantitative PCR reactions were set up in triplicate using 5 ml of cDNA (1:100 dilution) with the Brilliant II SYBR Green QPCR Master Mix (Stratagene) and run on a Mx3000P QPCR System (Stratagene). Primer sequences are listed in **Supplementary Table 4**.

mRNA profiling. Total RNA samples (RIN (RNA integrity number) > 9) were subjected to transcriptomal analyses using Affymetrix HTMG- 430A mRNA expression microarray as previously described.

Statistical analyses. Hierarchical clustering was performed using the GeneSifter software (Geospiza). Correlation distance and subsequent clustering were done using Ward's method. The differentially expressed genes (twofold) were calculated using a *t*-test ($P = 0.05$) with Benjamini and Hochberg correction. Principal component analysis was performed using the GeneSifter software. Gene ontology analysis was performed using the DAVID software[47], with the classification stringency set to 'high'.

Embryoid body formation. Before plating embryoid bodies, the iPSCs were depleted of mouse embryonic fibroblasts by splitting the cells 1:3 onto gelatin-coated plates on each day, for 2 consecutive days. On the 3rd day (designated day 0), iPSCs were trypsinized and plated at a density of 5,000 cells/ml in Isocove's Modified Dulbecco's Medium (IMDM) with 15% FCS (Atlanta Biologicals), 10% protein-free hybridoma medium (PFHM-II; Gibco), 2 mM L-glutamine (Gibco), 200 μg/ml transferrin (Roche), 0.5 mM ascorbic acid (Sigma) and 4.5 × 10–4 M monothioglycerol (MTG; Sigma). Differentiation was carried out in 60-mm ethylene oxide–treated Petri grade dishes (Parter Medical). The embryoid bodies were left to differentiate until day 6, when the cells were harvested to assay for hematopoietic colonies.

Hematopoietic colony formation assays. Day 6 embryoid bodies were collected by gravity, dissociated with trypsin and then passed several times through a 20 gauge needle to ensure dissociation. For the growth of hematopoietic progenitors, the cells were then seeded at a density of 100,000 cells/ml in IMDM containing 1% methylcellulose (Fluka Biochemika), 15% plasma-derived serum (PDS; Animal Technologies), 5% PFHM-II and specific cytokines as follows: primitive erythrocytes (erythropoietin (EPO, 2 U/ml)); macrophages (IL-3 (10ng/ml), M-CSF (5 ng/ml)); megakaryocytes (IL-3 (10 ng/ml), IL-11 (5 ng/ml), thrombopoietin (TPO, 5 ng/ml)); mixed colonies (SCF (5ng/ml), IL-3 (10 ng/ml), G-CSF (30 ng/ml), GM-CSF (10 ng/ml), IL-11 (5 ng/ml), IL-6 (5 ng/ml), TPO (5 ng/ml), and M-CSF (5 ng/ml)). All cytokines were purchased from R&D Systems. Primitive erythroctye colonies (eryPs) were counted on day 10 (4 d after embryoid body harvest). Macrophage colonies were counted on day 13 (7 d after embryoid body harvest). Mixed colonies were counted on day 14 (8 d after embryoid body harvest) and consist of a layer of macrophages, a layer of granulocytes, and a central core of red erythroid cells. Statistical analysis was performed using the Krward software. *P* values were calculated using the nonparametric Wilkinson test.

HELP DNA methylation analysis. High molecular weight DNA was isolated from iPSCs using the PureGene kit from Qiagen and the HELP (HpaII tiny fragment enrichment by ligation-mediated PCR) assay was carried out as previously described[1,2]. Briefly, 1 μg of genomic DNA was digested overnight with either HpaII or MspI (New England Biolabs). On the following day, the reactions were

doi:10.1038/nbt.1667

extracted once with phenol-chloroform and resuspended in 11 µl of 10 mM Tris-HCl pH 8.0 and the digested DNA was used to set up an overnight ligation of the JHpaII adaptor using T4 DNA ligase. The adaptor-ligated DNA was used to carry out the PCR amplification of the HpaII- and MspI-digested DNA as previously described[48]. All samples for microarray hybridization were processed at the Roche-NimbleGen Service Laboratory. Samples were labeled using Cy-labeled random primers (9 mers) and then hybridized onto a mouse custom-designed oligonucleotide array (50-mers) covering 25,720 HpaII amplifiable fragments (HAF) (>50,000 CpGs), annotated to 15,465 unique gene symbols (Roche NimbleGen, Design name: 2006-10-26_MM5_HELP_Promoter Design ID = 4803). HpaII-amplifiable fragments are defined as genomic sequences contained between two flanking HpaII sites found within 200–2,000 bp from each other and is represented on the array by 15 individual probes, randomly distributed across the microarray slide. HAF were first realigned to the MM9 July 2007 build of the mouse genome and then annotated to the nearest transcription start site (TSS), allowing for a maximum distance of 5 kb from the TSS. Scanning was performed using a GenePix 4000B scanner (Axon Instruments) as previously described[49]. Quality control and data analysis of HELP microarrays was performed as described[50].

Signal intensities at each HpaII-amplifiable fragment were calculated as a robust (25% trimmed) mean of their component probe-level signal intensities. Any fragments found within the level of background MspI signal intensity, measured as 2.5 mean-absolute-differences (MAD) above the median of random probe signals, were categorized as 'failed'. These failed loci therefore represent the population of fragments that did not amplify by PCR, whatever the biological (e.g., genomic deletions and other sequence errors) or experimental cause. On the other hand, 'methylated' loci were so designated when the level of HpaII signal intensity was similarly indistinguishable from background. PCR-amplifying fragments (those not flagged as either methylated or failed) were normalized using an intra-array quantile approach wherein HpaII/MspI ratios are aligned across density-dependent sliding windows of fragment size–sorted data. DNA methylation was therefore measured as the $\log_2$(HpaII/MspI) ratio, where HpaII reflects the hypomethylated fraction of the genome and MspI represents the whole genome reference. Analysis of normalized data revealed the presence of a bimodal distribution. For each sample, a cutoff was selected at the point that more clearly separated these two populations and the data were centered around this point. Each fragment was then categorized as either methylated, if the centered log HpaII/MspI ratio < 0, or hypomethylated if on the other hand the log ratio > 0.

HELP data analysis. Statistical analysis was performed using R 2.9 and BioConductor[51]. Unsupervised hierarchical clustering of HELP data was performed using the subset of probe sets (n = 3745) with s.d. > 1 across all cases. We used 1– Pearson correlation distance, followed by a Lingoes transformation of the distance matrix to a Euclidean one and subsequent clustering using Ward's method. Correspondence analysis was performed using the BioConductor package MADE4. The top 100 genes whose methylation status varied the most across the different groups were identified as those with the greatest s.d. across all samples.

Quantitative DNA methylation analysis by MassARRAY EpiTyping. Validation of HELP findings was performed by matrix-assisted laser desorption ionization/time-of-flight (MALDI-TOF) mass spectrometry using EpiTyper by MassARRAY (Sequenom) on bisulfite-converted DNA following manufacturer's instructions[52] but using the Fast Start High Fidelity Taq polymerase from Roche for the PCR amplification of the bisulfite-converted DNA. MassArray primers were designed to cover the promoter regions of the indicated genes. (Primer sequences available as **Supplementary Table 5**).

Chromatin immunoprecipitation (ChIP). Cells were fixed in 1% formaldehyde for 10 min, quenched with glycine and washed three times with PBS. Cells were then resuspended in lysis buffer and sonicated 10 × 30 s in a Bioruptor (Diagenode) to shear the chromatin to an average length of 600 bp. Supernatants were precleared using protein-A agarose beads (Roche) and 10% input was collected. Immunoprecipitations were performed using polyclonal antibodies to H3K4trimethylated, H3K27trimethylated, H3 pan-acetylation and normal rabbit serum (Upstate). DNA-protein complexes were pulled down using protein-A agarose beads and washed. DNA was recovered by overnight incubation at 65 °C to reverse cross-links and purified using QIAquick PCR purification columns (Qiagen). Enrichment of the modified histones in different genes was detected by quantitative real-time PCR using the primers in the **Supplementary Table 4**.

44. Conboy, I.M., Conboy, M.J., Smythe, G.M. & Rando, T.A. Notch-mediated restoration of regenerative potential to aged muscle. *Science* **302**, 1575–1577 (2003).
45. Sherwood, R.I. *et al.* Isolation of adult mouse myogenic progenitors: functional heterogeneity of cells within and engrafting skeletal muscle. *Cell* **119**, 543–554 (2004).
46. Cheshier, S.H., Morrison, S.J., Liao, X. & Weissman, I.L. In vivo proliferation and cell cycle kinetics of long-term self-renewing hematopoietic stem cells. *Proc. Natl. Acad. Sci. USA* **96**, 3120–3125 (1999).
47. Huang, D.W. *et al.* Systematic and integrative analysis of large gene lists using DAVID Bioinformatics Resources. *Nat. Protoc.* **4**, 44–57 (2009).
48. Figueroa, M.E., Melnick, A. & Greally, J.M. Genome-wide determination of DNA methylation by Hpa II tiny fragment enrichment by ligation-mediated PCR (HELP) for the study of acute leukemias. *Methods Mol. Biol.* **538**, 395–407 (2009).
49. Selzer, R.R. *et al.* Analysis of chromosome breakpoints in neuroblastoma at sub-kilobase resolution using fine-tiling oligonucleotide array CGH. *Genes Chromosom. Cancer* **44**, 305–319 (2005).
50. Thompson, R.F. *et al.* An analytical pipeline for genomic representations used for cytosine methylation studies. *Bioinformatics* **24**, 1161–1167 (2008).
51. Culhane, A.C., Thioulouse, J., Perriere, G. & Higgins, D.G. MADE4: an R package for multivariate analysis of gene expression data. *Bioinformatics* **21**, 2789–2790 (2005).
52. Ehrich, M. *et al.* Quantitative high-throughput analysis of DNA methylation patterns by base-specific cleavage and mass spectrometry. *Proc. Natl. Acad. Sci. USA* **102**, 15785–15790 (2005).