Cell Stem Cell, volume *9*

Supplemental Information

Forward and Reverse Genetics through Derivation of Haploid Mouse Embryonic Stem Cells

Ulrich Elling, Jasmin Taubenschmid, Gerald Wirnsberger, Ronan O'Malley, Simon Demers, Quentin Vanhaelen, Andrey I. Shukalyuk, Gerald Schmauss, Daniel Schramek, Frank Schnuetgen, Harald von Melchner, Joseph R. Ecker, William L. Stanford, Johannes Zuber, Alexander Stark, and Josef M. Penninger

Supplementary Experimental Procedures

Parthenogenic derivation of haploid ES cells. C57BL/6x129 F1 females were super-ovulated using standard protocols and unfertilized oocytes were flushed and collected. For activation, oocytes were exposed to 5% ethanol or 25 mM $SrCl_2$ as described (Kaufman et al., 1983; Otaegui et al., 1999). Four hours post activation, viable oocytes were transferred into pseudopregnant 129 females, re-collected on embryonic day (ED) 3.5 and cells were derived according to established embryonic stem derivation protocols (Bryja et al., 2006). Parthenogenetically derived ES cells were initially maintained on feeder layers and subsequently adapted to feeder cell free culture conditions. ES cell medium consisted of DMEM with 15% FCS (Gibco), supplemented with 2 mM L-Glutamate, 1 mM sodium pyruvate, 100 U penicillin/ml, 0.1 mg streptomycin/ml, 1x non essential aminoacids, 50 mM β-mercoptoethanol (all Sigma) and ESGRO at 1,000 U/ml (Millipore).

Genome coverage analysis and SNP mapping. Genomic DNA preparations were sheared using a Covaris DNA sonicator, adaptor ligated, and subjected to Illumina sequencing (HiSeq) according to the manufacturers protocol. Reads were mapped using Bowtie (allowing for up to 3 mismatches and requiring that reads map to a single genomic position) and coverage was analyzed as reads/50 kb window relative to coverage in the parental strains 129 and C57BL/6, considering only unique genomic coordinates. We retrieved SNPs that differ between C57BL/6 and 129 from Sanger (http://www.sanger.ac.uk/resources/mouse/genomes/) and evaluated mismatches we

observed during genome mapping of the deep sequencing reads against them. Each SNP that was covered by the Solexa reads was assigned to C57BL/6 and 129 according to the majority of reads.

Alkaline phosphatase activity, immunohistochemistry, and chromosome spreads. Alkaline phosphatase activity was detected using VECTOR kit SK-5300. Chromosome spreads were performed following established protocols (Nagy et al., 2008). Immunofluorescence of cultured cells or embryoid body (EB) cultures, neural stem cell, and differentiated neural stem cell cultures was performed after fixation in 4% PFA for 1h, blocking and permeabilization in PBS supplemented with 1% Glycine, 2% BSA, 0.2% Triton, and 5% FCS for 1h. Cells were incubated with the primary antibody o/n at 4ºC (anti-Nestin, Abcam 6142, 1:300; anti-Gata4, Santa Cruz, 1:500; anti-Oct3/4, BD Transduction, 1:100; anti-Tuj1, Covance RB-435P, 1:1000; anti-cytokeratin 5, PRP160P-100, Covance; anti GFAP, DAKO, 1:200; anti-Sox2, Cell Signaling, L1D6A2 mouse mAB 1:100; anti-Nanog, Abcam ab80892, 1:100), washed, incubated with fluorescent labeled goat secondary antibodies (Molecular Probes) and visualized using a Zeiss Axioplan2 (neural stem cell differentiation), or Zeiss LSM700 confocal microscope (embryoid bodies). For analyses of *in vivo* differentiation, teratomas were collected, fixed o/n in 4% PFA, paraffin embedded, sectioned, and stained with haematoxylin and eosin (H&E), Alcian blue and nuclear fast red, or processed for immunodetection of Nestin and Cytokeratine 5. Primary Abs were detected using biotinylated secondary antibodies. Ab staining was visualized using streptavidin-HRP and DAB and sections counterstained using hematoxylin. Immunohistochemistry was performed using the Ventana automated system. Images were collected using Zeiss miraxscan. Of note, teratomas were assessed by a certified pathologist.

Flow cytometry. For FACS analyses, cells were trypsinized, washed, and then incubated with 10µg /ml Hoechst33342 while pre-plating for 30 minutes. Subsequently, cells were collected by centrifugation, and FACS sorted for DNA content (as well as FSC-A and SSC-A) using BD FACSAriaIII. Intracellular staining (for FACS subsequent to

trypsinization) was performed using the same primary and secondary antibodies as described for immunohistochemistry staining.

Gene expression analyses. RNA was purified using QIAgen RNeasy Mini Kit. Reverse transcription, DNA labeling and microarray hybridization was done according to the manufacturers protocols (Agilent) using the x44K Mouse Genexpression Array DesignID 14868. For qPCR analyses, RNA was purified using QIAgen RNeasy Mini Kit and reverse transcribed using the iScript Kit (Biorad). Amplification was monitored with iQ SYBR Green supermix using the iQ5 Real Time PCR detection System (Biorad). Expression levels were calculated using the ΔΔct method with Gapdh as housekeeping gene. The following PCR primers were used:

nanog	GCAAGAACTCTCCTCCATT	forward
	ATGCGTTCACCAGATAGC	reverse
oct-4/pou5f1	TCACTCACATCGCCAATC	forward
	CCTGTAGCCTCATACTCTTC	reverse
sox2	CTCGCAGACCTACATGAAC	forward
	CTCGGACTTGACCACAGA	reverse
klf4	TCTCTCTTCTTCGGACTCC	forward
	CTGGACGCAGTGTCTTCT	reverse
c-myc	GTACCTCGTCCGATTCCA	forward
	CATCTTCTTGCTCTTCTTCAG	reverse
sall4	AACTTCTCGTCTGCCAGT	forward
	GAGTCATGTAGTGTACCTTCA	reverse
kfl2	CTCAGCGAGCCTATCTTG	forward
	AGAGGATGAAGTCCAACAC	reverse
gata-4	GTGAGCCTGTATGTAATGC	forward
	CTGCTGGCGTCTTAGATT	reverse
hand1	CCTTCAAGGCTGAACTCA	forward
	CGCCCTTTAATCCTCTTCT	reverse
gata-6	CTCCTACTTCCTCTTCTTCTAA	forward

	CGTCTTGACCTGAATACTTG	reverse
foxa2	GAGCCGTGAAGATGGAAG	forward
	GTGTTCATGCCATTCATCC	reverse
sox17	GCCGATGAACGCCTTTA	forward
	CAACGCCTTCCAAGACTT	reverse
krt18	TTGCCGCCGATGACTT	forward
	CAGCCTTGTGATGTTGGT	reverse
zfp42/rex1	CTGCCTCCAAGTGTTGTC	forward
	GAACAATGCCTATGACTCAC	reverse
drosha	CCAAGATGATCCAACTCCTT	forward
	GGTGCTGATTCTGAACAATG	reverse
rarg	CACCATTTGAGATGCTGAG	forward
	GGCTTATAGACCCGAGGA	reverse

Differentiation of ES cells, teratoma formation, and chimeric mice. For embryoid body (EB) formation, ES cells were trypsinized and cultured in absence of LIF either in hanging drops or in bacterial dishes. For retinoic acid (RA) induced differentiation, cells were grown in presence of 0.1µM RA for 1 week, plated at density of 1 million per 10 cm dish and assayed 72 hours later. For myoblast differentiation, ES cells were cultured in hanging drops for 4.5 days in ESC medium in absence of LIF and subsequently rinsed onto gelatinized cell culture dishes. Adhering cell aggregated were fed every 3rd day by replacement of 8ml/10ml ES cell medium without LIF. Movies of beating myoblasts were recorded on days 11-13 at 36 shots/second using a Zeiss Axiovert 200M and a CoolSNAP HQ^2. Derivation of neuronal stem cells and further differentiation of neuronal stem cells into GFAP$^+$ astrocytes and Tuj1$^+$ neurons was performed as described (Pollard et al., 2006). For teratoma formation, cells were injected testicular or subcutaneously into nude mice and teratoma growth was monitored. To generate chimeric mice, the diploid fraction of HMSc2 ES cells was purified using flow cytometry, cultivated for 7 days, injected into C57BL/6 ED3.5 blastocysts, and transferred into pseudopregnant 129 females. Percentage chimerism was determined by coat colour.

Retroviral infection of ES cells. Oct4 enhanced gene trap retroviruses carrying a splice acceptor followed by a neomycin resistance gene in 3 reading frames and Oct4 binding sites to enhance transcription (Schnutgen et al., 2008) were packaged in Platinum E cells (Cell Biolabs), concentrated by centrifugation (25,000 rpm, 4^0C, 4h) and applied to ES cells with 2µg polybrene per ml for 8 hours. Selection for gene trap insertions was done using G418 (Gibco) at 0.2mg/ml. To estimate numbers of integrations 500.000 cells were plated on 15 cm dishes, selected for integrations using G418 selection and colonies counted after 10 days. For comparison, 5.000 cells were plated without selection.

Inverse PCR. The protocol for inverse PCR was adapted based on Carette et al (Carette et al., 2011). In brief, genomic DNA preparations were digested using DpnII or MseI, purified using the QIAquick Gel Extraction Kit, and fragments ligated at a concentration of 3µg/ml over night. The ligase was then heat inactivated and rings were re-digested in ligase buffer using the enzymes NheI and PvuII. Linearized fragments were purified using the QIAquick Gel Extraction Kit and subjected to PCR using Accuprime Taq polymerase (Invitrogen), primers FS Solexa upstream and FS Solexa downstream, and a BioRad Thermal Cycler. The program of 95ºC for 30 sec, 60ºC for 30 sec, and 68ºC for 105 sec was repeated 36 times. Amplicons were loaded on agarose gels, eluted and subjected to deep sequencing using an Illumina Genome Analyzer and primer FS flowcell. The following iPCR primers were used:

upstream primer

AATGATACGGCGACCACCGAGATCGCCAGTCCTCCGATTGA

downstream primer

CAAGCAGAAGACGGCATACGAGTTCCTATTCCGAAGTTCCTATTCTCTA

flowcell sequencing primer

TGATTGACTACCCGTCAGCGGGGGTCTTTCA

Mapping of viral integration sites. Solexa reads were mapped to the mouse genome using Bowtie and requiring a unique best match to the genome. ENSEMBL gene annotations were used to determine the fraction of integration sites in introns, exons, UTRs, promoters (defined as 2kb upstream of the transcriptional start site), and the

remaining intergenic regions. ENSEMBL transcripts were split into 10 bins according to their expression levels in our haploid ES cells as measured by their absolute signal on the Agilent array used for transcriptome analysis and the fraction of transcripts with viral integrations was assessed in each bin. For the equivalent gene-based analysis, the most highly expressed transcript was considered for each gene. The analysis was repeated to estimate the coverage of viral integrations with sub-samples of the total insertions sites.

For confirmation of mapped integration sites the following primers were used:

madcam1-F	AGTCTCTCCTTTGCCCTGCTACTGG
madcam1-R	CACAGGCATTGAACAGTTTTGTTGG
drosha-F	TTCGAGTTATAGACTGTAATGAGCC
drosha-R	CCTACACTCTCTAGCAACGGAAGCC
RARG-F	GCTGTTGTCACCCTTGTGCAT AAGCC
RARG-R	AGATGCTGGGAATGGAACCCTGGTCC
Ap4s1-F	GTAGCTTAGAAACTCTGGCCACTGG
Ap4s1-R	CAGTGAAGTCTGAATACAGAGAATGG
Arap1-F	GTCCATGCAGGTTTGAGTGACTCC
Arap1-R	GACCTCCAGCTACAGAGGACAGAGCC
Evx1-F	TGTCAAGGGCAAGAGCTGCGAAGG
Evx1-R	CCAATGTCAAACCGGAAGGGAGAAGG
Bcl2l1-F	GAGTTACAGATGACTGCGAGCTGCC
Bcl2l1-R	GAAGCATTGAGTAGCTTTACCTGCC
2210012G12Rik-F	GTAGACCTGACTTGACTGGCTTGG
2210012G12Rik-R	GATGCTCATCTTACCAAACGCATCTC
Tit-F	CTTCGACCGTCTGGTCCTCAAGAGG
Tit-R	GAAACCAGCCTGATCTACATAGTGG
chr2:50928851-F	ACTTCCGACAAGAT TCTCAGTCC
chr2:50928851-R	CGTGACCTTTGGGTGTGTAATGCC

Protein quantification and differentiation in single ES cells. Differentiation analysis was carried out using a modification of high content screening (HCS) protocols we have previously published analyzing the loss of Oct4 expression as a measure of differentiation (Walker et al., 2010; Walker et al., 2007). All HCS experiments were carried out with the feeder free subclone HMSc2-27 at >50 passages. Cells were cultured in parallel in separate rows of 96-well tissue-culture plates (Greiner). Cells were trypsinized to a single-cell suspension and plated at 6000 cells/well into wells that had been pre-coated with a fibronectin/gelatin mixture (12.5 μg/mL fibronectin, Sigma and 0.02% gelatin in

water, Millipore). Cells were fixed at 24h, 48h, and 72h time-points with 4% paraformaldehyde for 15 min at room temperature. Immunostaining for Oct4 was carried out as follows: fixed cells were blocked for 30 min at room temperature, permeabilized with 0.1% Triton X-100 / PBS for 1 h at 4C, washed once with permeabilization buffer (PB, 5 % FBS and 0.3% Triton X-100 in PBS), then incubated with mouse monoclonal anti-Oct4 antibody (BD, 1:100) for 1 h at room temperature. Then, cells were washed four times with PBS and incubated with anti-mouse IgG1 AlexaFluor488 (Invitrogen, 1:100) for 1h in a 1:5000 dilution of DAPI in PB. Plates were then imaged on a ThermoFisher Cellomics ArrayScan VTi automated fluorescence microscope. Data were acquired using a minimum of 9 wells for each time point and condition. An algorithm was designed in the R language based upon nuclear size and DAPI intensity (DNA content) to distinguish haploid cells from diploid cells within mixed cultures of the haploid cell line HMSc2-27. To ensure the fidelity of our haploid vs diploid analyses, the R algorithm called cells "haploid" only if their DAPI intensity was below the mean DAPI intensity of the haploid controls and called cells "diploid" only if their DAPI intensity was above the mean DAPI intensity of the diploid controls (see Supplementary Fig. S7).

Reversion of the splice acceptor element in the retrovirus. Clones carrying the gene trap vector were transiently transfected with a plasmid encoding for Cre recombinase as well as GFP, FACS sorted for GFP positive cells, plated at clonal density, picked, expanded and analyzed for inversion using PCR analysis.

MicroRNA Sensor experiments. The pSIN-TRE-dsRed-miR30/shRNA-PGK-Venus-shRNA target site vector was used. This vector is a derivative of pSENSOR that enables fluorescence (dsRed)-based monitoring of shRNA expressing cells (Fellmann et al., 2011). Two variants of this vector harboring a potent shRNA targeting Firefly Luciferase (shLuc.1309) with or without its specific target site in the 3'UTR of Venus were each co-transfected with MSCV-rtTA3-PGK-Puro into control wild type ES cells as well as ES cells with antisense (AS) and sense (S) integrations of the mutagenesis vector using Lipofectamine 2000 according to manufacturers protocols. 8 hours after transfection, transfected cells were treated with doxycyline (1 µg/ml) to

induce shRNA expression, and after 48 hours, shRNA expressing (dsRed+) cells were analyzed for Venus reporter expression level on a FACS-Aria-III flow cytomter (BD).

Ricin screen. Ricin crude extracts in cell culture medium was generated as in (Simmons and Russell, 1985) and concentration was titrated to kill all cells efficiently within 3-4 days. In order to identify genes involved in ricin toxicity, we plated 25 million cells of the mutagenized library described above (Suppl. Figure S23) in five 15cm dishes. The library used had a complexity of about 7.5 million different, genetically independent mutations. On a sixth 15cm plate, 5 million non-mutagenized cells of subclone HMSc2-27 were plated. Cells were maintained in ES cell culture medium in presence of ricin for 2 weeks. At this point, several hundred distinct colonies had appeared on otherwise empty plates while the control plate was completely free of colonies with typical ES cell morphology. To purify ricin insensitive cells further, all cells were trypsinized and replated on an equal surface. Ricin selection in control and library was extended by 1 week. While control plates were entirely free of colonies now, library plates begun to overgrow. We lysed all cells in one pool, purified DNA and subjected the DNA to inverse PCR and deep sequencing (described above) to retrieve all viral integration sites in cells that remained after 3 weeks of constant ricin treatment.

Gpr107 knockdown. A shRNA targeting Gpr107 was designed (TGCTGTTGACAGTGAGCGCAACTAGCTTATTCATAGCCAATAGTGAAGCCAC AGATGTATTGGCTATGAATAAGCTAGTTTTGCCTACTGCCTCGGA) and cloned into LMN (MSCV-miR30-PGK-NeoR-IRES-GFP) as described (Zuber et al., 2010; Zuber et al., 2011). LMN vectors harboring the Gpr107 shRNA or a control shRNA targeting Renilla Luciferase were retrovirally transduced into HMSc2-27 and NIH3T3 cells in triplicates**.** After 72 hours, cells were split onto 2 10cm plates, one plate was left untreated and one plate was treated with ricin (1:250 of crude extract for NIH 3T3, 1:2000 of crude extract for HMSc2-27). Analysis of survival was performed after 48hours following addition of ricin. Cells were analyzed by flow cytometry for absolute number, eGFP expression and viability (propidium iodine staining) using an LSR Fortessa (BD).

Supplementary Figure and Table Legends

Supplementary Figure S1. DNA content analyses during derivation of haploid cells, chromosome spread, and SNP analyses; refers to Figure 1.

A) The first FACS analysis upon derivation of our two cell lines showed a small subpopulation of haploid cells in G1 of cell cycle (1n) derived from parthenogenotes. The initial sorts are shown. Upon repeated FACS purification (sort 7 is shown), a population of HMSc1 was enriched for haploidy (bottom panel). Of note, since HMSc1 cells always exhibited a larger number of diploid cells we primarily focused on HMSc2 cells. **B)** The vast majority of chromosome spreads displayed a precisely haploid genome (20 chromosomes). Rare spreads with changed chromosome number can be due to overlapping or washed off chromosomes. **C)** SNP comparison between HMSc1 or HMSc2 to an independent sequencing run of HMSc2 based on discriminatory SNPs between the 129 and C57BL/6 mouse strains. Haploid cell lines were derived from C57BL/6x129 F1 intercrosses. Data are shown as identical SNPs to HMSc2 confirming genetic independence of the two haploid clones (Student's t-test: $p<5.788E-08$).

Supplementary Figure S2. Transcriptome analysis and lineage markers in embryoid bodys; refers to Figure 2.

A) Haploid HMSc1 and HMSc2 display a gene expression profile that clusters with that of anestablished control ES cell line (diploid IB10/C ES cells). The transcriptome profile of MEFs is shown as control. Values are relative to the reference pool of all 4 RNA samples. **B)** Clustering shows that the transcriptional profile of both haploid HMSc1 and HMSc2 cells closely resembles that of diploid IB10/C ES cells. For clustering, the 100 most up- or down-regulated genes between MEFs and diploid IB10/C ES cells were selected. Three prototypical ES cell genes, namely Nanog, Oct4, and Klf2 are indicated. Supplementary Table 2 lists the 100 genes included in the analysis. For both A) and B), upregulated genes are shown in blue, downregulated genes are shown in red (see color keys), and hierarchical clustering of genes is shown as tree on the side of the heat maps.

C) qPCR analysis reveals down-regulation of the ES cell markers Nanog, Rex1, Oct4, Sox2, Klf2, Klf4, and Sall4 in EBs (analysed on day 7) derived from the haploid ES cell line HMSc1 accompanied by expression of the indicated lineage commitment markers (see text). mRNA expression was normalized to undifferentiated haploid ES cells (set at 1).

Supplementary Figure S3. Haploid ES cells can contribute to various tissues in adult mice and teratomas; refers to Figure 3.

A) Contribution of the HMSc2 derived cells to multiple tissues was determined by PCR followed by BamHI digest using primers GAATGTGAGCGCACAGGGTGATGTGCC and CCCACAGAACACAGTCACAGGGTCC. The indicated tissues were harvested from mice displaying coat color chimerism, processed, and analyzed for the presence C57BL6 (BL6) and 129 specific bands. The presence of a 129 band indicates tissue contribution by HMSc2 cells. The specificity of the primers in the BamHI digests are shown in the bottom panel. **B-I)** Histological examination of teratomas stained with hematoxylin & eosin **(B-G)** or Alcian blue, counterstained with alizarin red **(H,I)**. Arrowheads point towards **B)** keratinized stratified epithelium, **C)** pigmented epithelium, **D)** ciliated respiratory epithelium, **E)** glandular tubules with goblet cells, **F)** adipocytes, **G)** neurotubules, **H)** cartilage tissue, and **I)** sweat gland structures. Scale bars are 100µm.

Supplementary Figure S4. Oct4 and Sox2 expression on haploid ES cell subclones; refers to Figure 4.

Immunostaining for Oct4 protein expression (red, top panels) and Sox2 protein expression (red, bottom panels) on 9 different subclones that were established by plating single haploid cells directly after FACS purification. Subclones were derived from both HMSc1 and HMSc2 haploid ES cells. The subclones HMSc1-N1 (A, J), HMSc1-N3 (B, K), HMSc2-N3 (C, L), HMSc2-N4 (D, M), HMSc2-N6 (E, N), HMSc2-1 (F, O), HMSc2-15 (G, P), HMSc2-17 (H, Q) and HMSc2-27 (I, R) were seeded on gelatin coated coverslips and immunostained for Oct4 and Sox2 expression. Data are from cells that were subcloned after > 30 passages of the parental line confirming stability of expression

of bona fide ES cell markers. Cells were counter-stained with DAPI (blue). Scale bars are 50μm.

Supplementary Figure S5. Developmental potential of haploid ES cell subclones; refers to Figure 4.

Immunostaining for Oct4 (red) and Tuj1 (green) expression and expression of the endodermal marker Gata4 (red) in attached embryoid bodies (EBs, day 10) derived from the indicated subclones. Cells were counterstained with DAPI (blue). Data are from cells that were subcloned after > 30 passages of the parental line. Scale bars are 50μm.

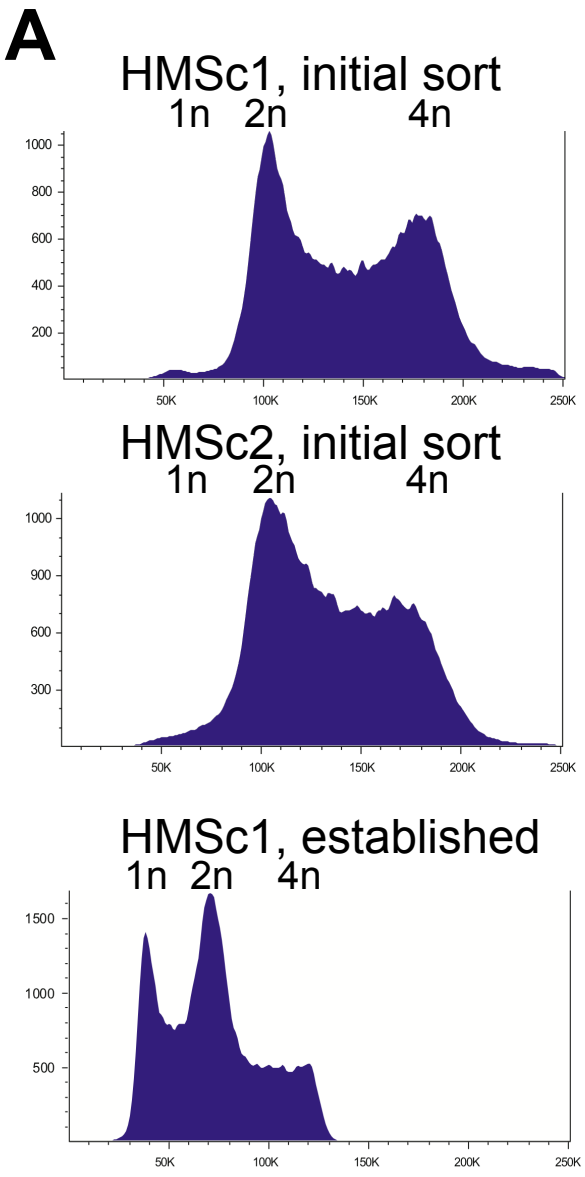
Supplementary Figure S6. Proliferation and stability of the haploid cell fraction in different ES cell subclones; refers to Figure 4.

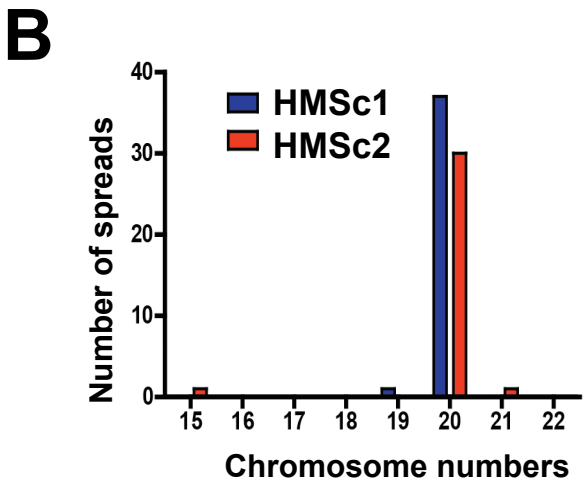
A) Proliferative kinetics was monitored in 7 subclones derived from haploid HMSc2. $1x10^6$ cells were plated on a 10cm dish and cell numbers were determined 72 hours later. All clones displayed robust growth. Clones HMSc2-15 and HMSc2-27 reached >30 million cells within a 72 hour timeframe while the HMSc2-N4 and HMSc2-N6 completed more than 1 cell cycle less. **B)** Percentages of haploid cells present in the populations of seven different subclones derived from the haploid HMSc2 ES cell line. Upon FACS purification of a purely haploid population for all subclones (1n peak as defined by Hoechst33342 histogram), relative populations of haploid versus diploid cells were determined after ten days of culture by FACS analysis on DNA content and quantified using the ModFit software. **C)** Morphology of the HMSc2 subclone HMSc2-27 cultured under feeder cell free conditions. HMSc2-27 cells remain a large haploid fraction for several weeks even without FACS sorting and exhibit the typical morphology of ES cells forming rounded, compact colonies. Representative DIC (differential interference contrast, Nomarski) images are shown. **D)** Expression of the prototypical murine embryonic stem cell markers Oct4, Nanog, and Sox2. Note absence of Phalloidin-positive feeder cells. Nanog and Oct4 were co-stained and are shown separately in the red channel. Scale bar: 50μm. **E)** Flow cytometry of DNA content in HMSc2-27 cells. DNA content was determined using Hoechst33342. The 1n and 2n chromosomes for haploid

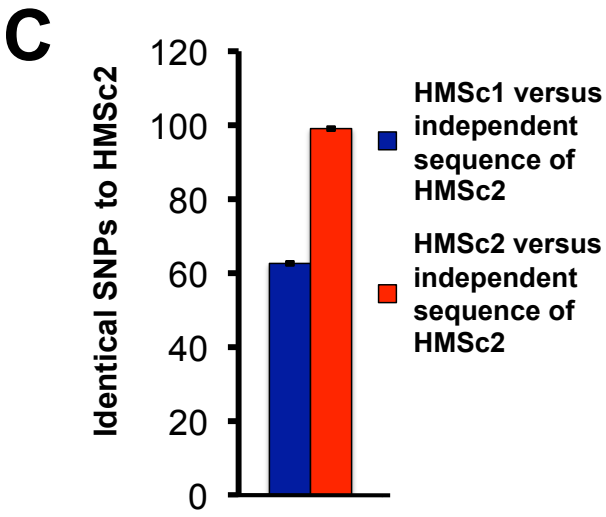
and 2n and 4n chromosomes for diploid ES cells are indicated. **F)** Chromosome spreads of HMSc2-27 cell confirming a haploid genome.

Supplementary Figure S7. Separation of haploid and diploid cells using high content imaging analysis and quantitative assessment of Oct4 levels; refers to Figure 5.

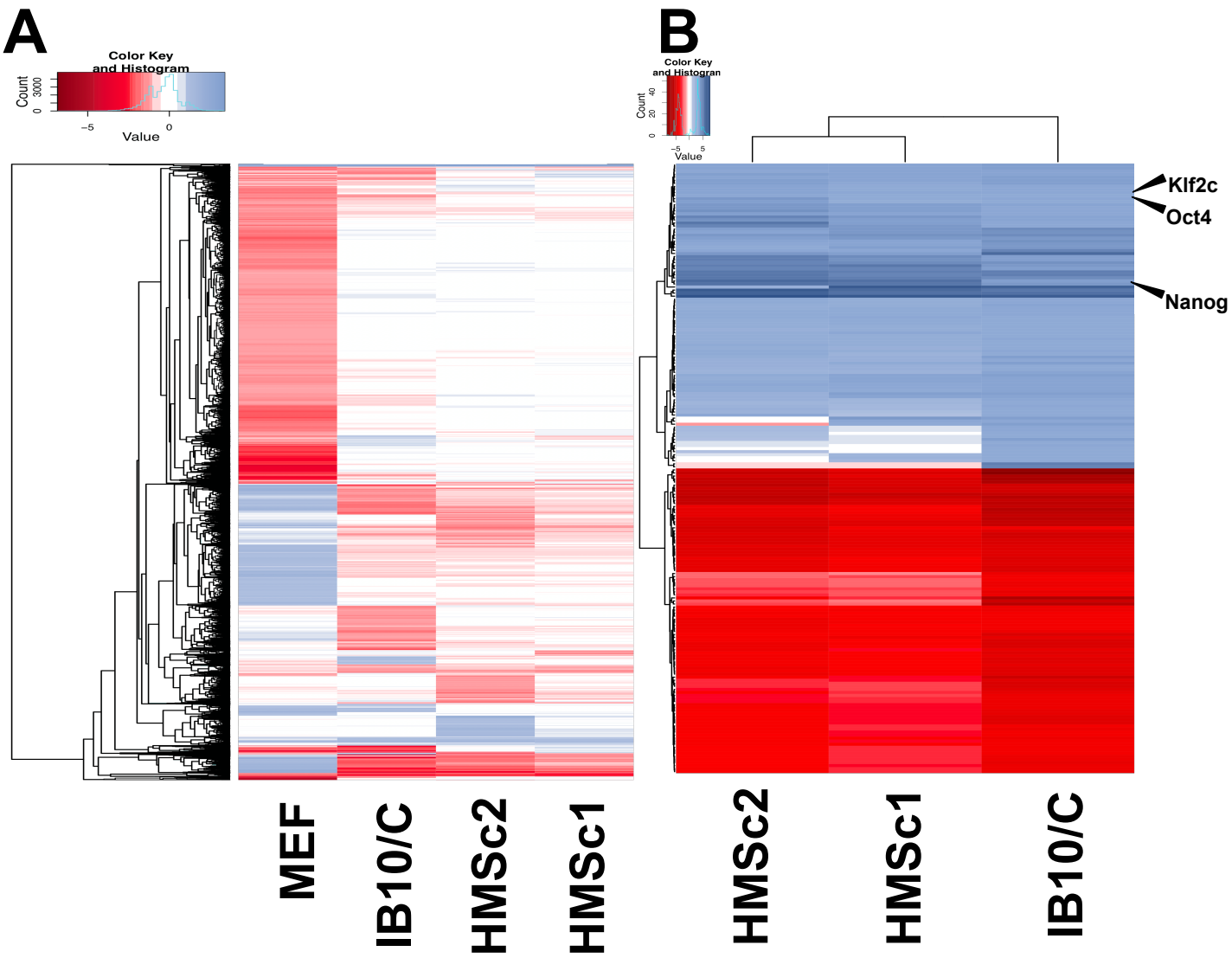
A) Representative high content scanning images of haploid HMSc2-27 and diploid CCE cells from 24 hour LIF withdrawal cultures. Parallel cultures of haploid HMSc2-27 and diploid CCE cells were used to develop an algorithm to distinguish haploid versus diploid cells on nuclear area (nuclear mask) and DNA content (DAPI staining intensity). The nuclear masks defined by DAPI staining are depicted in blue. Orange masks were rejected as nuclei from the algorithm as they are too small. The green lines in the Oct4 stained section cells depict nuclear masks as defined by DAPI staining. Of note, as expected, staining is less intense in the haploid cells for both DAPI and Oct4. **B)** Relative frequency plots showing separation of haploid (yellow) and diploid (blue) HMSc2-27 cells as a function of total DAPI intensity (x-axis) in the presence of LIF (top panels), the absence of LIF (bottom panels), and following treatment with 0.5μM retinoic acid (bottom) at 24h, 48h and 72h. To avoid inclusion of diploid cells in the haploid group and vice versa, only cells lying below the means of haploid controls were considered haploid, and only cells lying above the means of diploid controls are considered diploid by our algorithm. **C)** Baseline Oct4 expression in haploid versus diploid HMSc2-27 ES cells cultured in the presence of LIF. **D)** Oct4 expression of haploid versus diploid HMSc2-27 ES cells cultured in absence of LIF or in the presence of 0.5 μM retinoic acid to induce differentiation. Representative frequency histograms of Oct4 intensity corresponding to the data in Fig. 5C are shown. Note that at the time-points analyzed haploid cells (yellow) display nearly identical differentiation dynamics to control diploid cells (blue). Oct4 intensity decreases over time under both differentiation conditions but particularly rapidly under retinoic acid treatment. Note that Oct4 expression decreases in haploid as well as diploid HMSc2-27 cells at comparable rates upon initiation of differentiation.

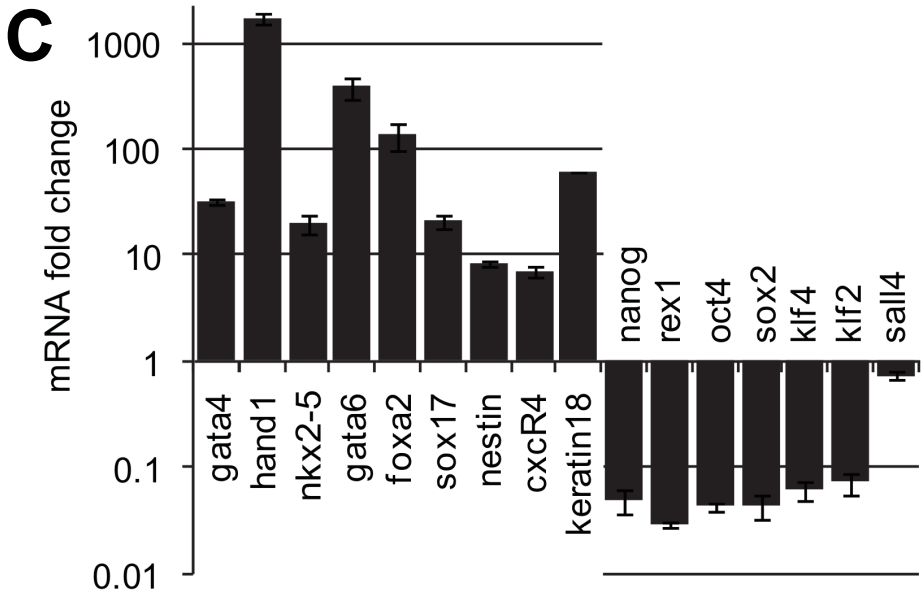




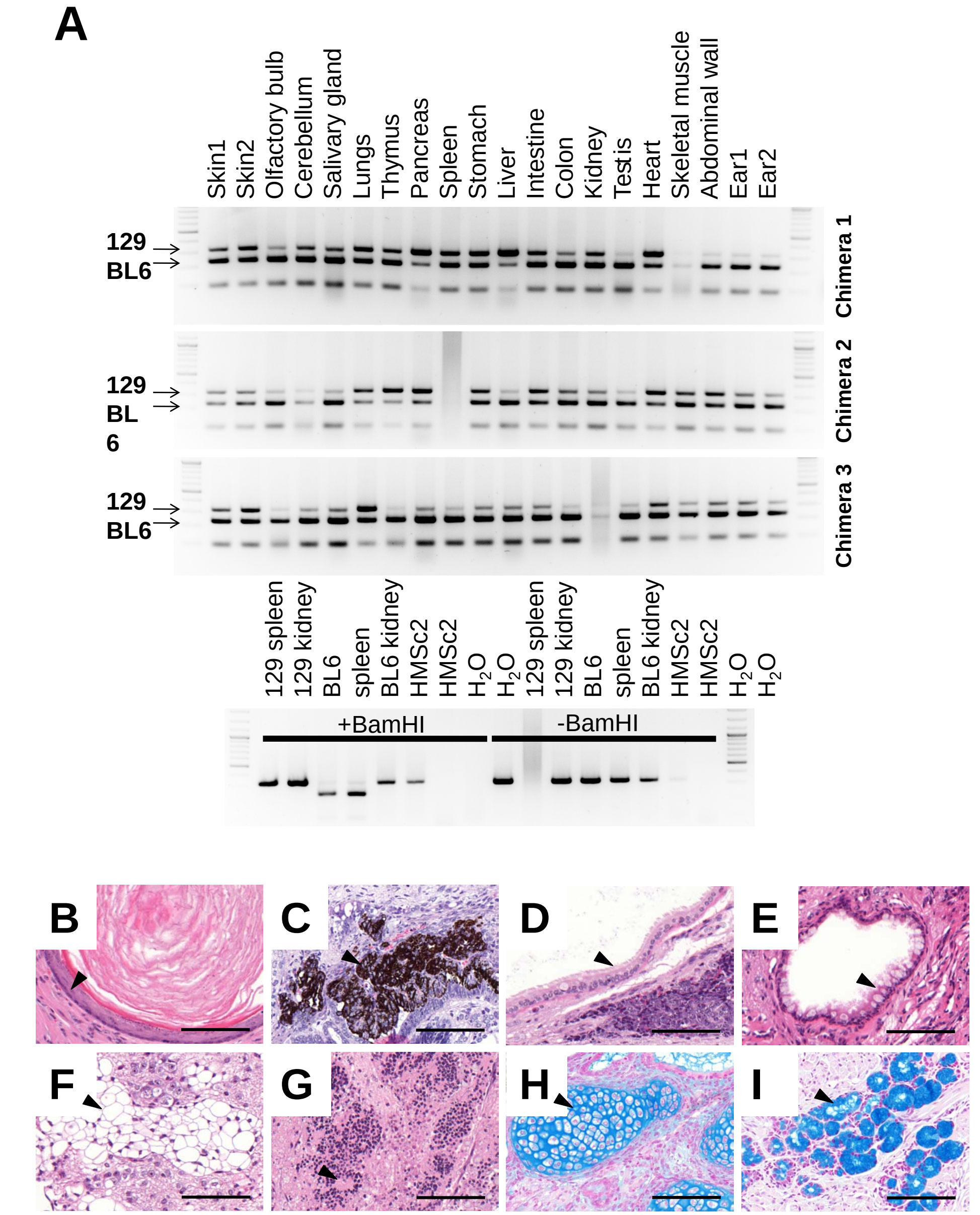


Suppl. Figure S1

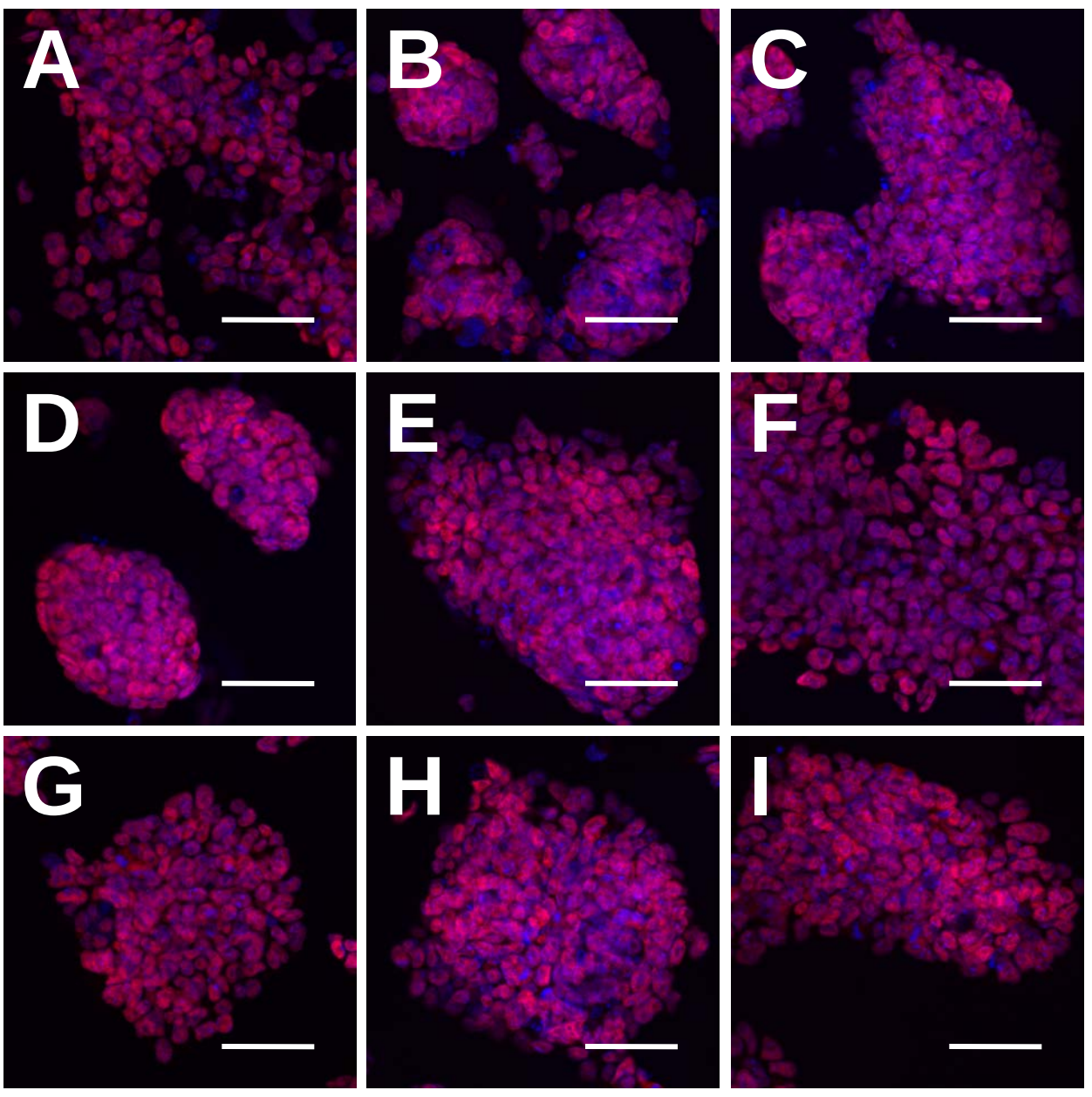




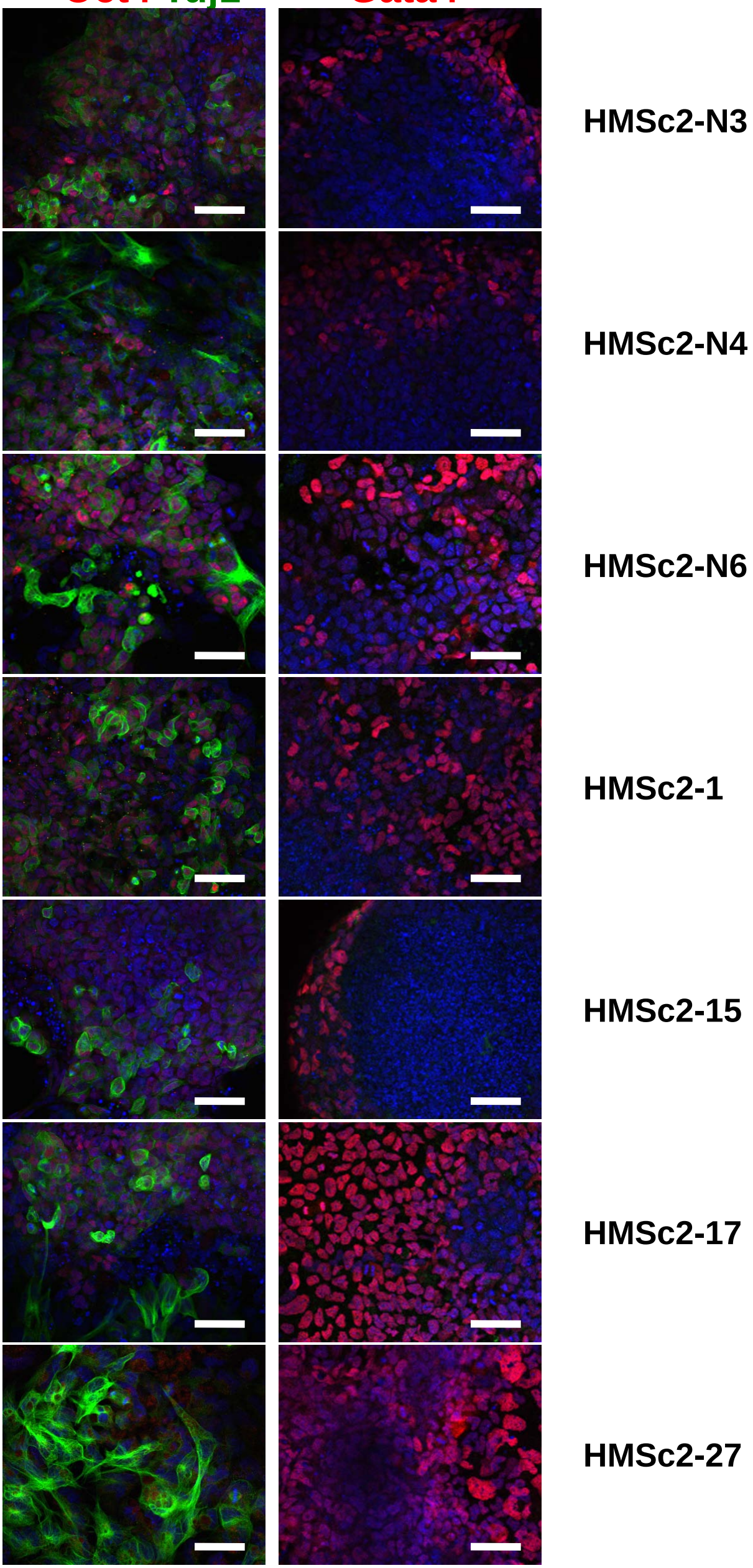
Suppl. Figure S2



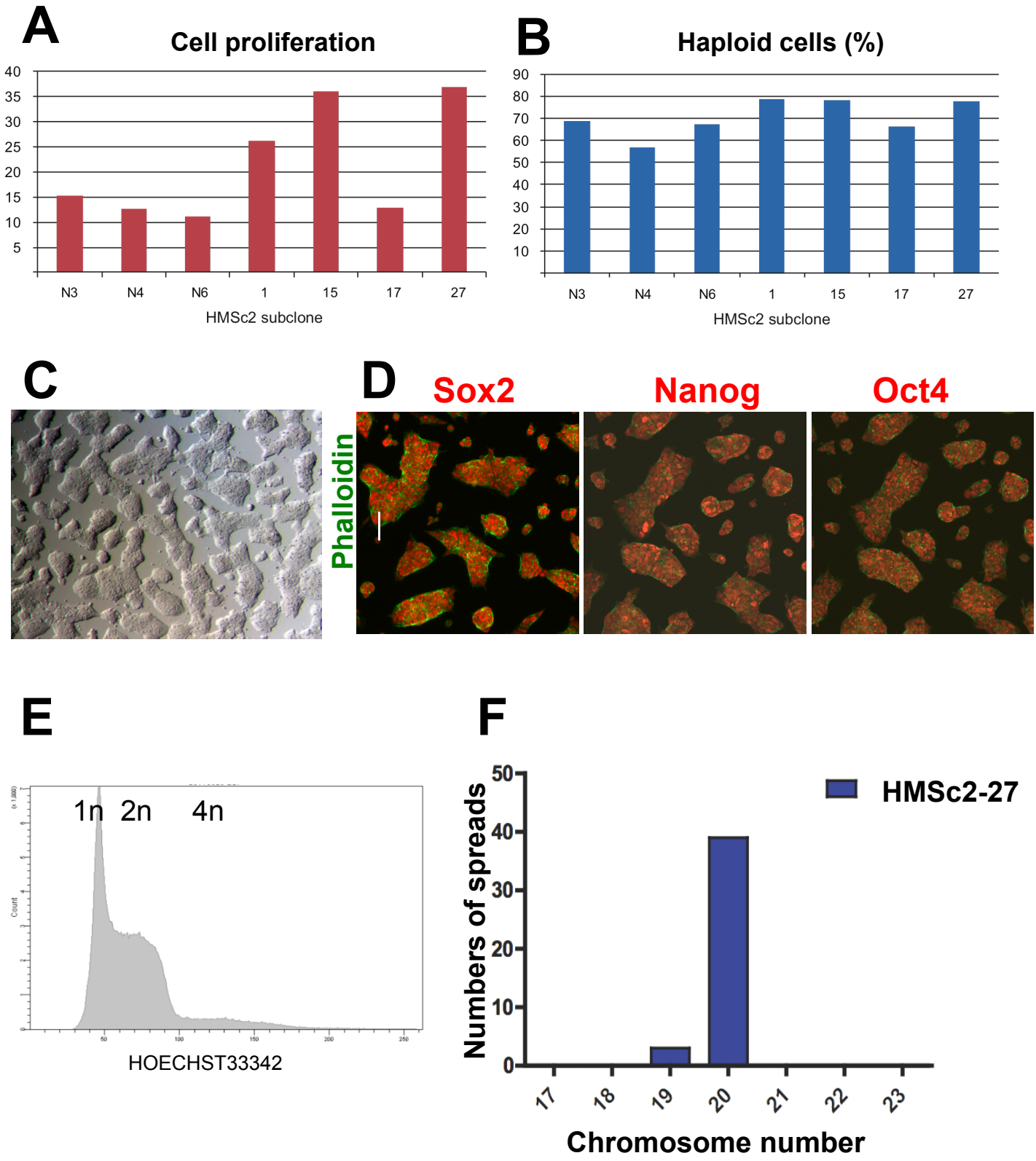
Suppl. Figure S3



Suppl. Figure S4

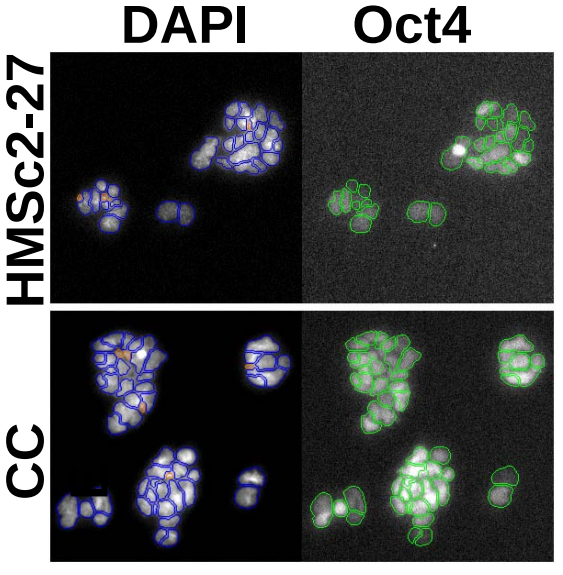


Suppl. Figure S5

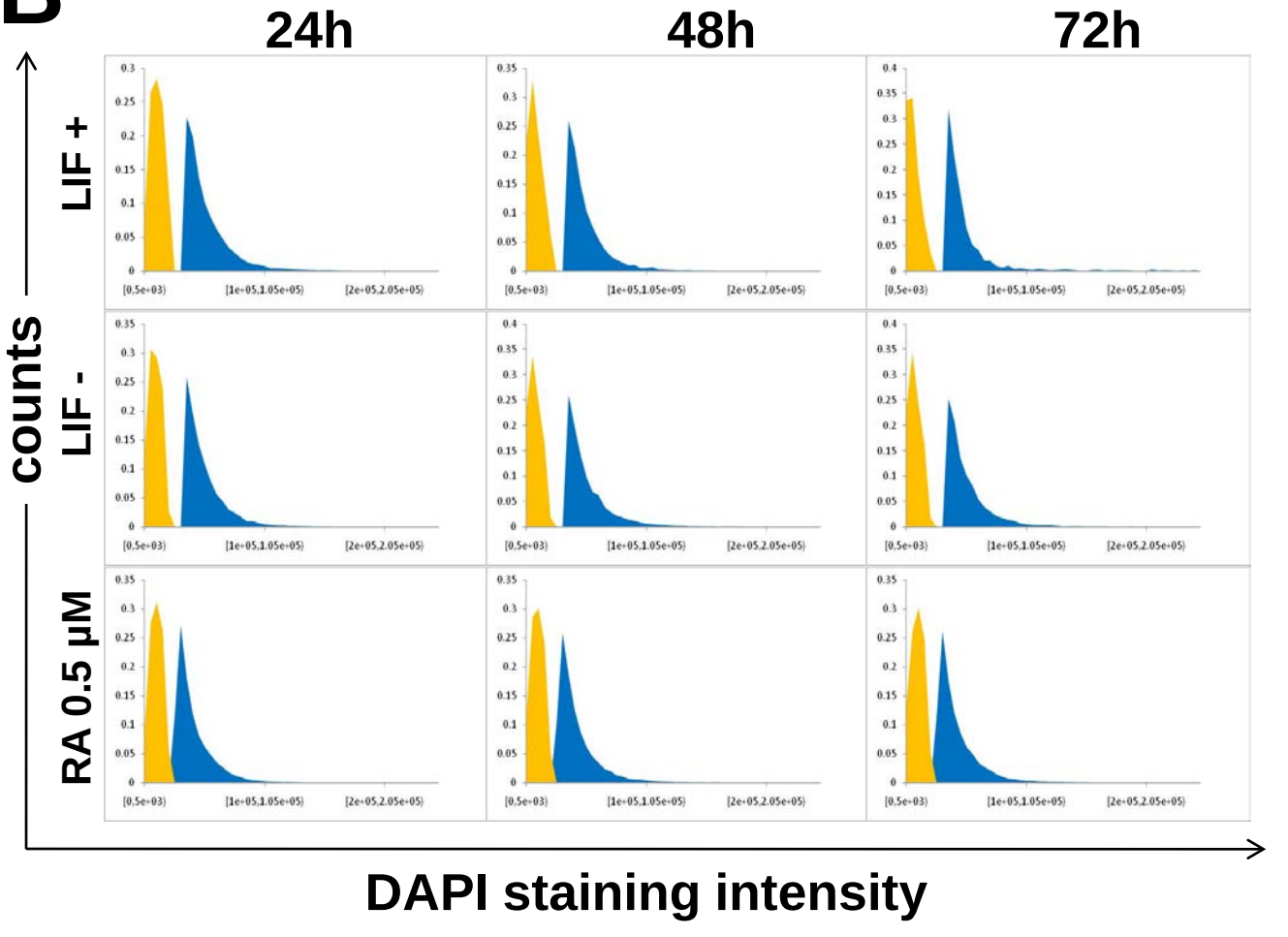


Suppl. Figure S6

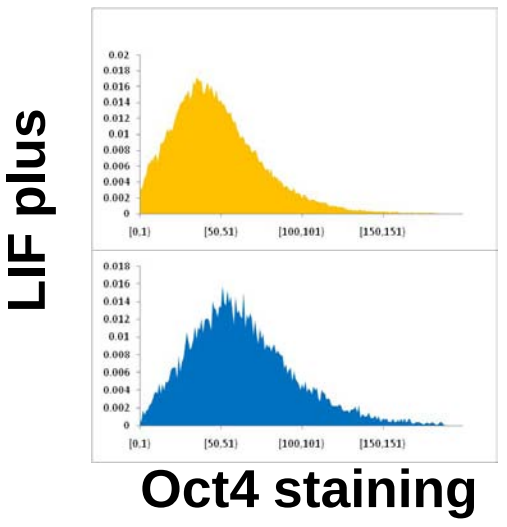
A



B



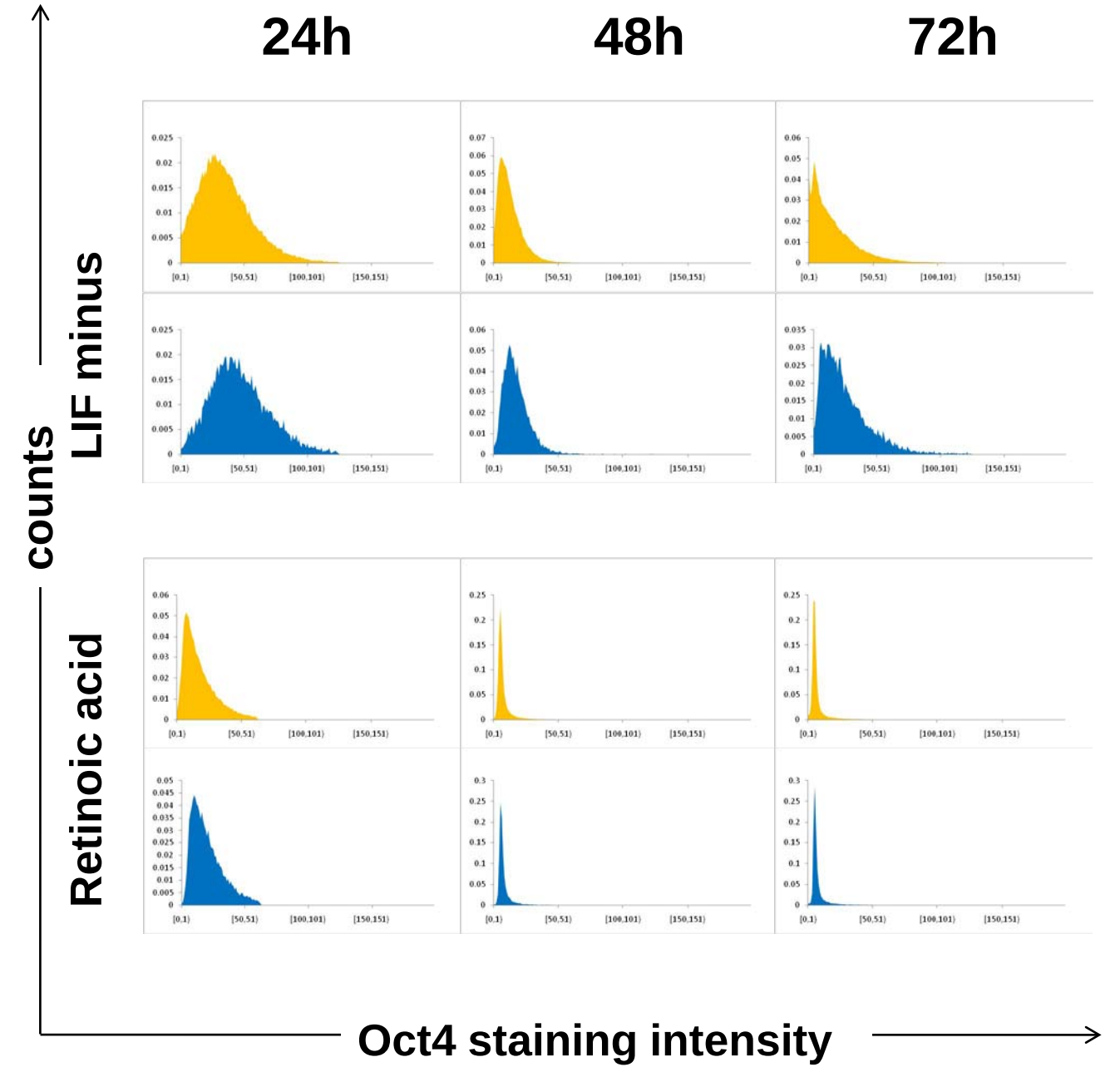
C



Haploid cells

Diploid cells

D



Suppl. Figure S7

Supplementary Table S1. Genome coverage; refers to Figure 1.

List of genomic regions with CNVs in either haploid HMSc1 or HMSc2 stem cells compared to the parental strains. For each cell-line individually, we selected overlapping genomic windows with CNVs (50kb, 10kb offset; normalized read-count difference ≥2-fold; multiple-testing corrected p-value≤10^{-3}) and merged adjacent windows within 50kb into regions (columns A-C; regions are either defined in HMSc1 or HMSc2 stem cells and they can overlap). Additional columns are: the number of individual windows that contributed to the region (columns H and M); the averages (columns D and I) and maxima (columns E and J) across all windows of the significance of enrichment or depletion of sequencing reads (log10 of hypergeometric P-values for depletion minus the value for enrichment such that significant enrichment has large positive and significant depletion has large negative values that correspond to the respective log10 values); average (columns F and K) and maxima (columns G and L) of the read-count ratios corrected to the conservative end of a 99% confidence interval to correct for artifacts stemming from low read-counts in some windows.

Supplemental Table S2. List of genes with the highest discriminatory values between MEFs and diploid IB10/C ES cells; refers to Figure 2 and Supplemental Figure S2.

This list of two hundred genes (100 most upregulated and 100 most downregulated genes) was used as the basis for Supplementary Fig. 3b. Of note, the gene list includes the bona fide ES cell markers Nanog, Oct4, and Klf2.

Supplementary References

Bryja, V., Bonilla, S., and Arenas, E. (2006). Derivation of mouse embryonic stem cells. Nat Protoc *1*, 2082-2087.

Carette, J.E., Guimaraes, C.P., Wuethrich, I., Blomen, V.A., Varadarajan, M., Sun, C., Bell, G., Yuan, B., Muellner, M.K., Nijman, S.M.*, et al.* (2011). Global gene disruption in human cells to assign genes to phenotypes by deep sequencing. Nat Biotechnol *29*, 542-546.

Fellmann, C., Zuber, J., McJunkin, K., Chang, K., Malone, C.D., Dickins, R.A., Xu, Q., Hengartner, M.O., Elledge, S.J., Hannon, G.J.*, et al.* (2011). Functional identification of optimized RNAi triggers using a massively parallel sensor assay. Mol Cell *41*, 733-746.

Kaufman, M.H., Robertson, E.J., Handyside, A.H., and Evans, M.J. (1983). Establishment of pluripotential cell lines from haploid mouse embryos. J Embryol Exp Morphol *73*, 249-261.

Nagy, A., Gertsenstein, M., Vintersten, K., and Behringer, R. (2008). Karyotyping mouse cells. CSH Protoc *2008*, pdb prot4706.

Otaegui, P.J., O'Neill G, T., and Wilmut, I. (1999). Parthenogenetic activation of mouse oocytes by exposure to strontium as a source of cytoplasts for nuclear transfer. Cloning *1*, 111-117.

Pollard, S.M., Benchoua, A., and Lowell, S. (2006). Neural stem cells, neurons, and glia. Methods Enzymol *418*, 151-169.

Schnutgen, F., Hansen, J., De-Zolt, S., Horn, C., Lutz, M., Floss, T., Wurst, W., Noppinger, P.R., and von Melchner, H. (2008). Enhanced gene trapping in mouse embryonic stem cells. Nucleic Acids Res *36*, e133.

Simmons, B.M., and Russell, J.H. (1985). A single affinity column step method for the purification of ricin toxin from castor beans (Ricinus communis). Anal Biochem *146*, 206-210.

Walker, E., Chang, W.Y., Hunkapiller, J., Cagney, G., Garcha, K., Torchia, J., Krogan, N.J., Reiter, J.F., and Stanford, W.L. (2010). Polycomb-like 2 associates with PRC2 and regulates transcriptional networks during mouse embryonic stem cell self-renewal and differentiation. Cell Stem Cell *6*, 153-166.

Walker, E., Ohishi, M., Davey, R.E., Zhang, W., Cassar, P.A., Tanaka, T.S., Der, S.D., Morris, Q., Hughes, T.R., Zandstra, P.W.*, et al.* (2007). Prediction and testing of novel transcriptional networks regulating embryonic stem cell self-renewal and commitment. Cell Stem Cell *1*, 71-86.

Zuber, J., McJunkin, K., Fellmann, C., Dow, L.E., Taylor, M.J., Hannon, G.J., and Lowe, S.W. (2010). Toolkit for evaluating genes required for proliferation and survival using tetracycline-regulated RNAi. Nat Biotechnol *29*, 79-83.

Zuber, J., Shi, J., Wang, E., Rappaport, A.R., Herrmann, H., Sison, E.A., Magoon, D., Qi, J., Blatt, K., Wunderlich, M.*, et al.* (2011). RNAi screen identifies Brd4 as a therapeutic target in acute myeloid leukaemia. Nature.