# Derivation and maintenance of mouse haploid embryonic stem cells

Ulrich Elling[1]*, Michael Woods[2,3]*, Josep V. Forment[4], Beiyuan Fu[3], Fengtang Yang[3], Bee Ling Ng[3], Jose R. Vicente[2], David J. Adams[3], Brendan Doe[3], Stephen P. Jackson[5], Josef M. Penninger[1,6#] and Gabriel Balmus[2#]

1. Institute of Molecular Biotechnology of the Austrian Academy of Science (IMBA), Vienna Biocenter (VBC), Dr. Bohr Gasse 3, Vienna, Austria.
2. UK Dementia Research Institute at University of Cambridge and Department of Clinical Neurosciences, University of Cambridge, CB2 0AH, UK.
3. Wellcome Trust Sanger Institute, Hinxton, Cambridge CB10 1SA, UK.
4. DNA Damage Response Biology, Oncology Innovative Medicines, Astra Zeneca, Cambridge CB4 0WG, UK
5. The Wellcome Trust CRUK Gurdon Institute and Department of Biochemistry, University of Cambridge, Tennis Court Road, Cambridge CB2 1QN, UK.
6. Department of Medical Genetics, Life Science Institute, University of British Columbia, Vancouver, Canada.

* These authors contributed equally to this work.

# Correspondence should be addressed to GB (gb318@cam.ac.uk) or JMP (Josef.Penninger@imba.oeaw.ac.at)

**KEYWORDS** : haploid ESC, genetic screen, Haplobank, haploid stem cell

**EDITORIAL SUMMARY This protocol describes how to derive and maintain mouse haploid embryonic stem cells from female gametes. Additional procedures that can be carried out with cell lines obtained from the mouse Haplobank are also described.**

**TWEET How to derive mouse haploid embryonic stem cells.**

**COVER TEASER Mouse haploid embryonic stem cells.**

RELATED LINKS

**Key reference(s) using this protocol**

Ullrich E. et al. *Cell Stem Cell* 9(6) (2011)

http://www.sciencedirect.com/science/article/pii/S1934590911004929

Ullrich E. et al. *Nature* 550, 114–118 (2017)

https://www.nature.com/articles/nature24027

Balmus G. et al. *Nature communications* 10(1), 87 (2019)

https://www.nature.com/articles/s41467-018-07729-2

**Key data used in this protocol**

Ullrich E. et al. *Nature* 550, 114–118 (2017)

https://www.nature.com/articles/nature24027

Balmus G. et al. *Nature communications* 10(1), 87 (2019)

https://www.nature.com/articles/s41467-018-07729-2

**Abstract**

Ploidy represents the number of chromosome sets in a cell. Whereas gametes have a haploid genome (*n*), most mammalian cells have diploid genomes (*2n*). The diploid status of most cells correlates to the number of probable alleles for each autosomal gene and makes it difficult to target these genes *via* mutagenesis techniques. Here we describe a seven-week protocol for the derivation of mouse haploid embryonic stem cells (hESC) from female gametes that also describes how to maintain the cells once derived. We also detail additional procedures that can be used with cell lines obtained from the mouse Haplobank, a biobank of over 100,000 individual mouse hESC lines with targeted mutations in 16,970 genes. hESC can spontaneously diploidize and can be maintained both in haploid and diploid states. Mouse hESC are genomically and karyotypically stable, are innately immortal, isogenic and can be derived in an array of differentiated cell types, and are thus highly amenable for genetic screens and for defining molecular connectivity pathways.

## INTRODUCTION

The packing and organization of genomes into chromosomes is central to the maintenance and transmission of genes during evolution[1]. In eukaryotes, ploidy (defined as the number of chromosome sets, *n,* of each cell) can range from a haploid genome *n=1* chromosome in the male of the Jack-jumper ant (*Myrmecia pilosula*)[2] to a staggering *~1900n=~15,600* chromosomes (29,64 x $10^6$ chromosomes) in the ciliated protozoan *Oxytricha trifallax*[3]. However, within mammals, the large majority of species maintain a diploid (*2n*) organization for most somatic cells and haploid (*n*) for gametes; an exception being made by the tetraploid vizcacha rat *Tympanoctomys barrerae*[4]. The somatic/diploid and germline/haploid configuration is beneficial and enables genetic variation to be increased through sexual reproduction[5], whilst conferring protection from deleterious mutations in the long-lived somatic tissues both by allowing the presence of back-up alleles as well as providing a template for DNA damage repair[6]. Thus, to maintain normal physiology, mammalian cells have developed mechanisms that act throughout the cell cycle to prevent the occurrence of cells with abnormal chromosome numbers (aneuploid or polyploid)[7] as these are a hallmark of genomic instability and oncogenic transformation[8].

Most somatic cells maintain their diploid chromosome number through mitotic division, although non-oncogenic mosaic polyploidy can happen in selected tissues during development (*e.g.* placenta) or in adult life (*e.g.* liver, muscle or brain)[9]. In contrast, the germ cells (oocytes or sperm) are terminally differentiated cells incapable of mitotic self-renewal *in vivo*. While methods of derivation and maintenance in cell culture for somatic cells are very well established, the derivation and culture of oocyte derived

haploid embryonic stem cells has only recently been achieved[10-13] (also see preview by Sagy and Benvenisti[14]). Here, we detail an experimental protocol for the derivation, maintenance and use of haploid mouse embryonic stem cells (hESCs).

**Development and applications of the protocol**

The haploid budding yeast (*Saccharomyces cerevisiae*) has been used for a long time as a model organism for genetic experiments as its genome can be easily manipulated so-as-to generate insertion/deletions, targeted or random-mutagenesis[15].

*In vivo*, mouse *N*-Ethyl-*N*-Nitrosourea (ENU) mutagenesis was until recently the only method used to perform random mutagenesis screens in mammalian haploid cells[16]. Although very informative, these screens require extensive resources, space and time for phenotypic mutants of interest to be identified. More recently, the discovery of KBM7, a near-haploid leukemic cancer cell line[17,18], and its further engineering to remove one extra copy of chromosome 8[19] and parts of chromosome 15[20] created the HAP1 cell line, representing the first human cell line with a haploid-genome to be propagated in culture. Limitations of the HAP1 cells relate to their transformed nature and presence of various genomic alterations, including the BCR-ABL1 translocation, confounding some studies which aim to understand physiologic genetic pathways in normal cells.

*In vitro* activation of oocytes for parthenogenetic embryogenesis and later generation of haploid ESCs was initially done in the 1920's by Gregory Pincus, who observed that some unfertilized rabbit oocytes could spontaneously undergo variable degrees of

development that were indistinguishable morphologically from the development of fertilized oocytes[21]. Subsequent work expanded on these methods in mouse models and showed that exposing a mouse oocyte to electric- or hyaluronidase-mediated activation mimics fertilization and promotes cell division towards the formation of parthenogenetic blastocysts containing a mixture of haploid and diploid cells[22,23]. Further elaboration on the methods of oocyte activation in mouse showed that this can be achieved in many ways, from spontaneous activation upon mechanical handling, to thermal, electric or chemical activation[24]. Ultimately investigators showed strontium chloride ($SrCl_2$) to be the only known parthenogenetic activating agent that induces repetitive intracellular calcium releases[25-27] in a fashion similar to those following normal fertilization by spermatozoa[28].

Moreover, it was observed that parthenogenesis may occur spontaneously in mice[29] and women[30], leading to the generation of benign ovarian teratomas composed of a mixture of diploid and haploid cells, thus highlighting how haploid cells can persist after parthenogenesis. Collectively, these observations led to the development of methods for oocyte activation, growth and expansion of parthenogenetic blastocysts and subsequent isolation and maintenance of haploid mammalian ESCs from the mouse[10,11,31-33], the monkey[34], the rat[35] and human[12,13]. Analogously, by transferring sperm into an enucleated oocyte, haploid androgenetic ESCs have also been derived[36,37]. Parthenogenetic haploid ESCs are fully competent for functional contribution to the germline[38].

The generation of mouse and human hESC has opened the door to functional random mutagenesis screens[39] that have shown that these cells can be successfully used, for example, for the identification of loss-of-function and separation-of-function mutants *via* approaches such as ethyl methanesulfonate (EMS) mutagenesis[40,41], CRISPR screens[33,42,43] or transposon induced mutagenesis[42]. The use of haploid cells in point mutagenesis screens open an important avenue for identification of separation of function mutants[40]. Using hESC in CRISPR screens allows for higher efficiency as one confounding factor in such genome wide approaches in diploid cells is the presence of heterozygous deletions. Moreover, the hESCs are amenable to differentiation in different cell types that in the future will allow for screening in a multitude of cell lineages[43]. Furthermore, the generation of such cells allowed for the creation of a cell-bank (Haplobank: www.haplobank.at) of over 100,000 publicly available individual mouse hESC lines carrying conditional and reversible disruption events in 16,970 mouse genes[42].

**Limitations of the protocol**

Mouse strain background can play a defining role in the success of haploid ESC derivation. We have derived hESCs from C57BL/6J, NOD/ShiLtJ and 129S1/SvImJ genetic backgrounds, and have observed that the success rate was higher for C57BL/6J, NOD/ShiLtJ than for 129S1/SvImJ. Another important limitation of the method is the fact that mouse and human hESC spontaneously diploidize upon serial passage. Thus, in order to maintain them a haploid format, cells need to be periodically flow sorted from time to time (see **Boxes 1 - 3**). Notably, as shown by others, we have also observed that the success of hESC derivation can be influenced

by the knockout background, for example *Tp53* knockout hESC are more stable in culture as WT cells. Another important note is that, with adaptation, hESC become more stable and it takes a longer time for them to diploidize (around three weeks depending on background and number of cells sorted). Upon diploidization, ESCs remain stable and no longer duplicate their genome (i.e. they do not readily become polyploid). From all the clones that we have karyotyped, we observed low rates of chromosomal abnormalities involving mouse chromosome 8, mouse chromosome 11 and loss of chromosome X, consistent with observations in ESC derived from fertilized embryos[44,45]. Additionally, the experimenter needs to keep in mind that the parthenogenetic hESC derived from oocytes will always contain an X chromosome and never a Y chromosome.

**Experimental design considerations**

A workflow of the procedure is outlined in **Figure 1**. Female mice should be matched by genetic background, and ideally the experimenter should start with no less than five individual mice. The use of mice should be compliant with local animal research regulations as well as the ARRIVE guidelines[46]. If the derivation is done from specific knockout lines, in order to obtain both wild type (WT) control and mutants in the same derivation, heterozygous mice should be used. Although an elaborate technique, by following the procedure described herein, the derivation of haploid ESC is highly reproducible.

## MATERIALS

REAGENTS

- Female mice, ideally between 3 to 6 weeks of age. CAUTION All procedures on mice must be carried out following National and Institutional guidelines. In the UK Home Office regulations must be followed with appropriate establishment, project and personal licensing in place. Experiments using mouse models should be done in accordance with the ARRIVE guidelines [46].
- Strontium chloride hexahydrate (Cl2Sr·6H2O; ThermoFisher UK Ltd, cat. no. 10349220). CAUTION Harmful if swallowed; irritant. Consult Safety and Hazard sheets that are provided by the seller.
- Ethylene glycol-bis(β-aminoethyl ether)-N,N,N′,N′-tetraacetic acid tetrasodium salt (C14H20N2O10Na4; Sigma-Aldrich, cat. no. E8145).
- KSOM Mouse Embryo Culture Medium w/ Phenol Red (GlobalStem, cat. no. GSM 5140).
- M2 medium with HEPES, without penicillin and streptomycin, liquid, sterile-filtered (Sigma-Aldrich, cat. no. M7167).
- Pregnant mare's serum gonadotropin (PMSG; Prospec Protein Specialists, cat. no. HOR 272).
- Hyaluronidase from bovine testes Type I-S, lyophilized powder, 400-1000 units/mg solid (Sigma-Aldrich, cat. no. H3506-100MG).
- Mettler Buffer Sachets pH4 01 Pk30 (Sigma-Aldrich, cat. no. PHM0144).
- Mettler Buffer Sachets pH7 Pk30 (Sigma-Aldrich, cat. no. PHM0142).

- Dulbecco's Phosphate-Buffered Saline (DPBS; ThermoFisher, cat. no. 14190094).
- Acidic Tyrodes Solution (Millipore, cat. no. MR-004-D).
- Hydrochloric Acid Solution, 1.0 N, Bioreagent, suitable for cell culture (Sigma-Aldrich, cat. no. H9892). CAUTION Corrosive. Harmful if swallowed; irritant. Consult Safety and Hazard sheets that are provided by the seller.
- NDiff N2B27 (StemCells Inc., NDiff N2B27; cat. no. SCS-SF-NB-02).
- GSK3 inhibitor: CHIR99021 (Abcam; cat. no. ab120890). CAUTION Acutely toxic if swallowed; irritant. Consult Safety and Hazard sheets that are provided by the seller.
- MEK 1 and 2 inhibitor: PD0325901 (Abcam; cat. no. ab120639). CAUTION Acutely toxic if swallowed. Consult Safety and Hazard sheets that are provided by the seller.
- ESGRO® Recombinant Mouse mLIF Medium Supplement (LIF) (EMD Millipore-Merk; cat. no. ESC1107).
- Dulbecco's Modified Eagle's medium (DMEM) high glucose with UltraGlutamine (Lonza; cat. no. BE12-604F/U1).
- Gelatin solution (Sigma-Aldrich; cat. no. G1393).
- Phosphate Buffer Saline (1 x PBS) (Lonza; cat. no. BE17-516F).
- 2-Mercaptoethanol (Sigma-Aldrich; cat. no. M6250). CAUTION Classified as an acute toxin; health hazard; environmental hazard; chemical and DNA-mutagen. Consult Safety and Hazard sheets that are provided by the seller; use only under the chemical cabinet. Always use hood to prepare solution.
- Dimethyl sulfoxide (DMSO) (Sigma-Aldrich; cat. no. D2650).

- Fetal Bovine Serum (FBS) (Gibco by Thermo Fischer Scientific; cat. no. A31608-02).
- Bovine Albumin Fraction V (7.5% solution) (Gibco by Thermo Fischer Scientific; cat. no. 15260037).
- MEM Non-Essential Amino Acids Solution (100X) (NEAA) (Gibco by Thermo Fischer Scientific; cat. no. 11140-050). CAUTION Irritant if swallowed. Consult Safety and Hazard sheets that are provided by the seller.
- Penicillin-Streptomycin-Glutamine (100X) (Gibco by Thermo Fischer Scientific; cat. no. 10378-016). CAUTION Health hazard if swallowed. Should be disposed by appropriate routes. Consult Safety and Hazard sheets that are provided by the seller.
- Penicillin-Streptomycin (10,000 U/ml) (Gibco by Thermo Fischer Scientific; cat. no. 15140122). CAUTION Health hazard if swallowed. Should be disposed by appropriate routes. Consult Safety and Hazard sheets that are provided by the seller.
- Sodium Pyruvate (100 mM) (Gibco by Thermo Fischer Scientific; cat. no. 11360-039).
- Trypsin-EDTA (0.05%), phenol red (Gibco by Thermo Fischer Scientific; cat. no. 25300-054).
- Trypan Blue Solution, 0.4% (w/v) in PBS, pH 7.5 ± 0.5 (Corning cat. no. 25-900-CI). CAUTION Acutely toxic if swallowed; health hazard. Consult Safety and Hazard sheets that are provided by the seller.

EQUIPMENT

- Humidified bench top incubator and stereomicroscope suitable for viewing and manipulating embryos using a mouth pipette.
- Stereo-microscope (Leica cat. no. M125).
- CO2 incubator (Eppendorf/New Brunswick Galaxy 48R incubator, cat. no. CO48R230 or Panasonic cat. no. KM-CC17RU2).
- Tissue Culture Biological Safety Cabinet (NuAire LabGard ES cat. no. NU-437).
- pH meter (Mettler Toledo Five Easy FE20, cat. no. 10385343).
- Fine forceps (Student Dumont #5 forceps, Interfocus Ltd, cat. no. 91150-20).
- Fine scissors - straight/large loop (Interfocus Ltd, cat. no. 14040-10).
- Pasteur Pipets (Scientific Laboratory Supplies Ltd, cat. no. PIP4172).
- Tissue culture grade flasks or plates (Corning cat. no. 430641U or Corning cat. no. 430167)
- 60mm center well culture dish for *in vitro* fertilization (Falcon 60mm Center well culture dish; Scientific Laboratory Supplies Ltd, cat. no. 353653 or similar).
- Aspirator tube assemblies for calibrated micro-capillary pipettes (Sigma-Aldrich, cat. no. A5177-5EA).

REAGENT SETUP

- **Strontium chloride.** Prepare 100mM $SrCl_2$ using an analytical balance with precision of 0.1mg to weigh 26mg and add to 1 ml of KSOM. Make up fresh and use within 24hrs. Store at 4C. CAUTION Strontium chloride may cause serious eye damage or respiratory irritation.

- **Ethylene glycol-bis(β-aminoethyl ether)-N,N,N',N'-tetraacetic acid (EGTA).** Prepare 0.5 M EGTA using an analytical balance with precision of 0.1mg to weigh 234mg and add to 1 ml of KSOM. pH adjustment of the EGTA solution is important, pH will initially be ~10. Add 1M HCL to the 1 ml EGTA solution in increments of 200$\mu$l, adding a total of 600$\mu$l using the pH meter to ensure pH reading = 8. Make up fresh and use within 24hrs. Store at 4C.
- **Oocyte Activation Media**. Add 150$\mu$l of 100mM $SrCl_2$ and 12$\mu$l of 0.5M EGTA to 3 ml of KSOM and filter sterilise. Store at 4˚C for use the following day or put 2 ml in 60 mm centre well culture dish, surrounded by PBS in outer well, and pre-equilibrate to 37 ˚C for use on the same day.
- **Hyaluronidase.** Make a stock solution of hyaluronidase by adding 3.53 ml of M2 to 100 mg lyophilised powder. Gently agitate the solution ensuring powder is fully dissolved. Make working concentrations by adding 100 $\mu$l of 30 mg/ml stock hyaluronidase solution to 900 $\mu$l M2 to create a 3mg/ml solution. Sterile filter and aliquot into 200 $\mu$l aliquots and store at -20 ˚C for up to 1 year. For use in removal of oocyte cumulus cells, add 1 aliquot of 200 $\mu$l solution to 1.8 ml M2.
- **Pregnant mare's serum gonadotropin (PMSG).** Prepare PMSG in a clean class 2 safety cabinet. Reconstitute the 5000 international units (IU) lyophilised powder by mixing with 80ml PBS ensuring powder is completely dissolved. Aliquot into 1.5 ml vials and store at -20 ˚C to use for up to 3 months.
- **Human chorionic gonadotropin (HCG).** Prepare HCG in a clean class 2 safety cabinet. Reconstitute the 1500 IU lyophilised powder by mixing with 30ml

PBS ensuring powder is completely dissolved. Aliquot into 1.5 ml vials and store at -20 ˚C to use for up to 3 months.

- **NDiff-N2B27 culture media.** To 94 ml NDiff N2B27 media add: 5 ml 7.5% BSA, 1 ml Penicillin-Streptomycin (100X), 10 µl LIF, 1µl 2-Mercaptoethanol and 2i (10 µl PD (10 mM stock; stored at 4˚C) and 30 µl CHIR (10 mM stock; stored at -20˚C)). NDiff-N2B27 culture media can be stored at 4˚C for up to one month.

  CAUTION 2-Mercaptoethanol is classified as a severely toxic chemical and DNA-mutagen. Consult Safety and Hazard sheets that are provided by the seller, use only under the chemical cabinet.

- **DMEM culture media.** To 500 ml DMEM high glucose media (1 bottle) add: 75 ml FBS, 6 ml NEAA (100X), 6 ml Sodium Pyruvate (100 mM), 6 ml Penicillin-Streptomycin-Glutamine (100X), 60 µl LIF, 6 µl 2-Mercaptoethanol. Optional: if 2i addition is required (see Method) add 60 µl PD (10 mM stock; stored at 4˚C) and 180 µl CHIR (10 mM stock; stored at -20˚C). DMEM culture media can be stored at 4˚C for up to one month.

  CAUTION 2-Mercaptoethanol is classified as a severely toxic chemical and DNA-mutagen. Consult Safety and Hazard sheets that are provided by the seller; use only under the chemical cabinet.

- **0.02% Gelatin solution.** Allow 2% gelatin solution to completely liquefy at 37°C and make 0.02% gelatin working solution with 1 x PBS (*e.g.* 5 ml 2% gelatin in 500 ml 1xPBS). Can be stored for up to one year at room temperature.

EQUIPMENT SETUP

- **Stereo microscope with heated stage.** Manipulating embryos is done using a stereomicroscope with heated stage and zoom capacity up to 100x magnification.
- **Humidified incubator.** Oocytes in activation medium and all subsequent culturing is done in a humidified incubator set to 37˚C, 5% CO2.

**PROCEDURE**

**Superovulation**

*Timing: 30-40 minutes*

1 l Identify female mice to use as donors for oocyte production to be superovulated between the ages of 3-4 weeks and set aside in advance. CAUTION All procedures on mice must be carried out following appropriate National regulations. In the UK this means in accordance with Home Office regulations and with appropriate establishment, project and personal licensing in place. CRITICAL STEP Note that the best age for superovulation varies from strain to strain but usually lies between 3 and 6 weeks of age, during the prepubescent stage of development[47].

2 l Inject the mice at 3pm, removing the PMSG hormone (6.25 IU, 0.1ml for each female mouse) from the freezer 15-20 minutes before the injections are scheduled to allow the hormone to defrost and warm to room temperature.

3 l Once completely defrosted draw up the hormone into a 1 ml syringe making sure there are no air bubbles in the syringe. Attach a 27 G needle to the syringe, hold

with the needle facing up and depress the syringe plunger to expel air from the dead space created.

4 l Scruff the mouse and turn it so its stomach is facing upwards and its head is slightly tilted back.

5 l With your free hand, hold the syringe so the needle is bevel side up. Carefully insert the needle into the peritoneal cavity of the mouse, to the left or right of the midline of the mouse making sure not to hit the bladder.

6 l Following your approved technique for mouse intra peritoneal (IP) injection, administer 6.25 IU PMSG, 0.1 ml. Briefly, after preparing the PMSG solution draw it up in the syringe, with keeping in mind dead volumes in the needle. Gently remove the animal from the cage and restrain appropriately in the head-down position. Disinfect lower right quadrant of the abdomen with 70% ethanol and inject by inserting the needle with bevel facing up at approximatively 30$^{o}$ angle to horizontal towards the head in order to avoid damage to the urinary bladder, cecum and other abdominal organs. Wait briefly before withdrawing the needle to make sure the liquid does not seep out.

7 l Place the mouse into an empty cage and then repeat steps 5 – 6 for each mouse until finished. Once all mice have been injected they can then be returned to their original cage.

8 l Dispose of the needle and syringe in a sharps bin. CAUTION Extreme caution is required at all times when handling used sharps. Dispose sharps directly without capping.

9 l Repeat steps 2-8 46-48 hours after PMSG administration. but inject with 0.1 ml, 5 IU HCG rather than PMSG.

**Dissection and oocyte harvest**

*Timing: 50 – 60 minutes*

10 l Set up 4 single well dishes per line; 1 containing activation medium and 3 with KSOM for washing. There should be 1 ml of culture media in the centre well with 2-3 ml of DPBS in the outer well.

11 l Filter 10 ml of M2 into a Falcon tube using a Millex - GS 0.22 $\mu$m filter and place in an incubator to warm to 37˚C.

12 l Remove 1 vial of hyaluronidase (working stock solution) from the -80 freezer and thaw in an incubator, warming to 37˚C.

13 l Place a 50mm dish containing 2-3 ml of M2 or 0.9% saline on a heat blanket or warming stage set to 37˚C.

14 l Cull a female by cervical dislocation, lay the mouse on its back exposing the abdomen.

15 l Make a midline incision up and though the abdominal wall followed by the peritoneal sac to expose the internal organs, pushing the gut up and out toward the chest revealing the reproductive organs.

16 l Grasp the upper end of one of the uterine horns with fine forceps and gently pull the uterus, oviduct, ovary, and fat pad taut and away from the body cavity. This will reveal a fine membrane, the mesometrium, which connects the reproductive tract to the body wall and carries a prominent blood vessel. Make a

hole in the membrane close to the oviduct with the closed tips of a pair of fine forceps or scissors.

17 l Reposition the forceps and cut the uterus near the oviduct.

18 l Transfer the oviduct and attached segment of uterus to the 50 mm dish containing DPBS at 37 ˚C.

**CRITICAL** It is important to keep oocytes at 37˚C at all stages to maintain oocyte viability and chances of successful parthenogenesis.

19 l Repeat steps 14-18 for all females; oviducts from several mice can be collected in the same dish.

20 l Prepare a 35 mm dish containing 2 ml of filtered M2 for washing oocytes after exposure to hyaluronidase. Keep the dish at 37˚C.

21 l Pipette 1.8 ml of the pre-warmed M2 into a 35 mm culture dish.

22 l Move 1 x oviduct into the dish with the 1.8 ml filtered M2 and 'pop' the swollen ampulla to release the cumulus mass, then discard the oviduct. Repeat until all oviducts have been processed. (Supplementary **Video 1**).

23 l Add all the hyaluronidase working stock solution from the vial and gently swirl/agitate the dish to help dispersal of cumulus mass or pipette up and down gently several times with a 1 ml tip, then wait for the cumulus cells to start falling away leaving individual oocytes.

**CRITICAL** Hyaluronidase can have a detrimental effect on the oocyte, so the oocytes should be collected and washed quickly. Reducing the amount of hyaluronidase added to the M2 will increase the time it takes embryos to become free from cumulus cells but will allow more time for users less familiar with the protocol to collect the oocytes.

24 l Pick up the individual oocytes with a transfer pipette and place into the 2 ml of warmed M2, repeating until all oocytes are collected.

Troubleshooting

25 l Wash oocytes through 6 x drops of pre-warmed, filtered M2. Once washed the oocytes can be kept together in a drop of M2 and assessed for viability.

Troubleshooting

**Activation**

*Timing: 2-3 hours*

26 l Remove oocytes which have lysed, fragmented or have a large perivitelline space (shrunken cell). A lysed cell of an oocyte will look darker in comparison to a healthy oocyte. Fragmentation of a single cell can occur and can vary from a section to the entire cell. A larger perivitelline space can be obvious when the oocyte is placed next to healthy examples. Sometimes the increased perivitelline space is all around the circumference of the cell, other times it is more obvious on one side of the oocyte. If in doubt, the oocytes can be retained and put in activation medium to be removed later if necessary. Aspirate the oocytes around a drop of M2 with a mouth pipette, rolling the oocytes to allow viewing from a variety of angles under a bench top stereomicroscope to make the assessment thorough.

27 l Move any oocytes identified as sub-viable to a separate drop of M2 for final assessment. Discard sub-viable oocytes (Supplementary **Video 2**).

28 l Transfer all viable oocytes to the pre-incubated activation medium and place in tissue culture for 90 minutes at 5% CO2, 37˚C.

**CRITICAL** Viable oocytes must be placed in to the activation medium for the stated time and without delay to optimise chances of successful parthenogenic activation.

29 l Remove oocytes from activation medium by gently mouth pipetting and wash through 3 dishes of pre-equilibrated KSOM.

30 l Observe the oocytes for polar body extrusion, indicating successful activation. Once the oocytes are in the activation media it takes up to 90 minutes for all the oocytes to be activated (polar body extrusion).

31 l Culture oocytes for 72 hours.

**Selection of blastocysts**

*Timing: 50 – 60 minutes*

**CRITICAL This part is outlined in Figure 2.**

32 l On day 3 post activation, embryos will be at a variety of developmental stages: some embryos will look sub viable and some may have lysed (**Figure 2a**). Remove lysed embryos but retain those where cells have not divided cleanly as they may still progress to blastocyst.

Troubleshooting

33 l Move 8-cell embryos to KSOM supplemented with CHIR99021 (3 $\mu$M) and PD0325901 (1 $\mu$M), pre-equilibrated to 5% CO2, 37˚C and continue to culture.

34 l On day 4, there will be some embryos that have developed to blastocyst that are suitable for zona pellucida removal (**Figure 2b**). Only select embryos that have a well expanded blastocoel for zona removal.

35 l Repeat steps 32 to 34(the check and selection of blastocysts for zona removal) on day 5 and 6 (**Figure 2c**).

Troubleshooting

**Zona pellucida removal**

*Timing: 30 minutes*

**CRITICAL** When handling embryos, process one line at a time to make sure that strains do not get mixed and label all dishes and vials containing embryos with strain information or a unique identifier.

36 l Pre-warm acidic Tyrodes solution to 37˚C.

37 l Put two ~200 µl drops of acidic Tyrodes and two ~200 µl drops of M2 on to a dish suitable for working under a stereo microscope.

38 l Select the blastocysts for zona removal and move to one of the M2 drops.

39 l Transfer the blastocysts to the first drop of acidic Tyrodes and observe the dissolving of the zona. The zona should take 30-70 seconds to dissolve.

Troubleshooting

40 l Briefly aspirate the blastocysts up and down the mouth-pipette to ensure the zona is completely removed and transfer to the second drop of M2 (Supplementary **Video 3** and Supplementary **Video 4**).

41 l Observe the blastocysts in M2 and manipulate using a new mouth-pipette to check that the zona has been completely removed (**Figure 2d;** Supplementary **Video 4**). If the zona remains intact after inspection, repeat steps 38 to 41 using a second drop of acid.

Troubleshooting

42 l When the zona's have been removed, immediately transfer the blastocysts to a fresh dish of M2.

Troubleshooting

43 l Wash blastocysts through an additional 2 dishes of M2.

44 l If transporting blastocysts, pipette 50 μl of M2 into a 1.5 ml microcentrifuge tube and label with strain and embryo number. Using a mouth pipette and M2, load an air bubble to be positioned behind the blastocyst and transfer embryos to the vial. The bubble behind the blastocysts should come from the pipette indicating blastocysts are deposited into the vial.

Troubleshooting

**ESC cell line establishment**

*Timing: 10 - 20 days*

45 l Coat plates with 0.02% gelatin for 5 minutes (100 μl/well). It is not necessary to wash the plates after removal of gelatin solution.

46 l Transfer denuded embryos to individual wells of 96-well plates containing 250μl NDiff-N2B27 culture media (see Reagents set-up).

47 l Culture embryos until they grow into visible cellular colonies (around 1 mm in size; this takes up to 10 days in culture) (**Figure 2e**).

**CRITICAL** Pay particular attention to retaining the colony if media change is required (if media starts to become yellow in colour), as the colony will not attach strongly to the bottom of the dish.

48 l Transfer a unique colony to 15 μl trypsin and, after 2-3 minutes, pipette up and down using 15 μl DMEM culture media to inactivate the trypsin and transfer to another well of a 96-well gelatin coated plate containing 250 μl NDiff-N2B27 culture media.

CAUTION Haploid ESC diploidize faster in DMEM culture media than in NDiff-N2B27 culture media. Do not try to spin down the cells after trypsinization, as it is very easy to lose the lines when doing this.

49 l In 3 days or upon confluency, trypsinise as described in step 48 and re-plate in a gelatin coated 24 well plate; at this point the hESC will form colonies (**Figure 2e**).

50 l In 3 days or upon confluency, trypsinise as described in step 48 (increasing the trypsin volume accordingly to cover the bottom of the well; *e.g.* 50 μl trypsin) and block trypsinization with an equal volume of DMEM culture media. Transfer cells to a gelatin coated 6 well plate; at this point flow cytometry cell cycle analysis can be performed in order to assess the haploid status of each clone (see **Box1**)

Troubleshooting

51 l In 3 days or upon confluency, trypsinise cells as described in step 48 and transfer them to a gelatin coated 10 cm dish or T25 flask. The first sorting for haploid enrichment can also be performed (see **Box 3**). Plate sorted cells in a gelatin coated T25. Troubleshooting

52 l Upon regrowth, trypsinise (Step 48) and split into several gelatin coated flasks or 10cm dishes, 1 million/flask and in 2 days freeze down some cells in 1 ml 1:9 DMSO : FBS freezing media and transfer to a liquid nitrogen storing unit. From

this point onwards, one of the clones can be adapted to DMEM with 2i and LIF. If desired, cells can be adapted in DMEM with LIF alone.

**TIMING**

It is recommended that the superovulation dates are on Tuesday and Thursday, as this allows mouse dissection, oocyte harvest and activation on a Friday followed by embryo culture over the weekend and assessing for blastocysts during the weekdays of the following week.

Day 1 (Steps 1 to 8): Inject female donors: 0.1 ml, 6.25 IU PMSG, 3 pm

Day 3 (Step 9): Inject female donors: 0.1 ml, 5 IU HCG, 3 pm

Day 4 (Steps 10 to 31): Dissection of female donors, harvesting and activation of oocytes.

Day 7 (Steps 32 and 33): Remove lysed embryos and transfer 8 cell embryos to KSOM supplemented with 2i.

Day 8 (Step 34 and 35): Some embryos will have developed to blastocyst; select those with the largest blastocoel for zona removal.

Day 9 (Step 36 to 44): Most blastocysts should be available for zona removal on this day. Move the remaining embryos into fresh, pre-equilibrated KSOM for further culture.

Day 10 (Repeat steps 37 to 46): Zona removal of the remaining blastocysts.

Day 11 (Steps 45 to 47): Transfer to tissue culture room in NDiff-N2B27 culture media (about 3 minutes/blastocyst on average).

Day 18 (Step 48): Passage to 96 wells. Most blastocysts will expand to cellular colonies by this point but some might need additional days (about 15 minutes/cell line on average).

Day 21 (Step 49): Passage to 24 wells (about 15 minutes/cell line on average).

Day 24 (Step 50): Passage to 6 well plate (about 15 minutes/cell line on average).

Day 27 (Step 51, 52 and **Box 1** and **Box 2**): Passage to T25 flask or 10 cm dish and cell cycle profile, genotyping and mycoplasma testing (about 15 minutes/cell line on average).

Day 30 (Optional. Repeat Steps 51 and 52): Expansion to T75 flasks (about 15 minutes/cell line on average) to increase the number of sorted cells.

Day 33 (**Box 3**): Sorting for haploid enrichment (about 2-3 hours/cell line on average - depending on the haploid content).

Day 36 to 46 (Repeat Steps 50 to 52): Depending on how many cells have been sorted, expansion of haploid populations and freeze down for banking of cell lines (about 15 minutes/cell line on average).

**TROUBLESHOOTING**

See Table 1 for troubleshooting guidance.

**ANTICIPATED RESULTS**

The protocols presented here will allow derivation and maintenance of mouse haploid ESCs. Although the protocol for derivation of hESC is lengthy (about seven weeks), it allows reliable and reproducible derivation of hESCs. In our hands mouse hESCs can be successfully established in gelatin coated plates, with only minor improvements

when deriving them on plates containing irradiated feeders. The bottleneck for successful cell line production is the quality of the embryos selected. This can vary depending on mouse strain, mouse age, handling of oocytes, decision on what oocytes to select, successful activation, subsequent decision on what blastocysts to select and successful zona pellucida removal. In our hands, from 1534 oocytes, we had overall success rates of 1 blastocyst for each 10.5 oocytes (1 blastocyst for every 3.95 oocytes if we do not include oocytes that do not progress within 24 hours and that were unlikely to be viable in the first place). Our activation success rate is 1 per 2.65 oocytes and our average success rates for cell line production is 1 cell line for every 1.4 blastocysts used. Although there might be a minor advantage to using feeders, for users that do not have much experience with handling ESCs on feeders, culture without feeders is the most straightforward to undertake as it allows easier identification of the haploid ESCs colonies. We recommend that before freezing down the cellular clones a mycoplasma test be performed. In our hands we have never had any mycoplasma contamination.

The associated procedures allow for the screening of clones that contain haploid populations (see **Box 1**), karyotyping of the isolated hESC clones (see **Box 2**) and sorting for enrichment of hESCs (see **Box 3**). We also describe considerations for use of hESCs imported from the Haplobank (see **Box 4**). Haploid cells can be subsequently successfully used for genetic screens (see **Box 5** for one such example). Similar principles can be used for the maintenance and propagation of human haploid ESCs.

**Competing financial interests**

JMP is a founder of JLP Health.

## Data availability

All data presented in the manuscript are available from the corresponding authors upon reasonable request.

## Author contributions

EU, MW and GB performed experimental analysis and procedures through-out and wrote the manuscript. BD and DJA helped MW with set-up of blastocyst work. BLN assisted with flow-cytometry with help from JVF and SPJ. BF and FY performed the FISH and karyotyped the cell lines. JRV wrote the transposon induced mutagenesis protocol with help from EU. GB and JMP conceived the idea of this manuscript. All authors commented on the manuscript.

## Acknowledgments

MW, BD, BF, BLN and DJA are supported by the Wellcome Trust through core funding to the Wellcome Trust Sanger Institute (WT098051). Research in the SPJ laboratory is funded by Cancer Research UK (Programme grant C6/A18796), Wellcome Trust Investigator Award (206388/Z/17/Z). Core funding is provided by CRUK (C6946/A14492) and the Wellcome Trust (WT092096). Research in JMP laboratory is funded by Advanced ERC and Era of Hope/DoD grants. Research in the GB laboratory is funded by a UK Dementia Research Institute fellowship (MC_PC_17111).

**Box 1 Quantification of cell cycle profile of newly derived ESC**

**Reagents**

- Propidium iodide (PI) solution (Sigma-Aldrich; cat. no. P4864). CAUTION Health hazard if swallowed; mutagen. Consult Safety and Hazard sheets that are provided by the seller.
- Sphero™ Rainbow Fluorescent particles, Spherotech (cat. no. RCP-20-5).

**Equipment**

- BD Falcon™ Round-Bottom Tubes (BD, Polypropylene, 5 ml, snap, cat. no. 352063).
- 30 $\mu$m filters (Sysmex, CellTrics 30 $\mu$m green; cat. no. 25004-0042-2316)
- Flow cytometer (Beckman Coulter; CytoFLEX S Flow Cytometer or similar, Model B75442)

**Reagent setup**

- **FACS buffer.** 0.5% BSA in 1 x PBS. Can be stored for up to 1 month at 4 ºC.
- **PI solution.** 10 $\mu$g/ml PI, 250 $\mu$g/ml RNase in FACS buffer. Make fresh.

**Sample preparation**

*Timing: 1 hour*

1 l (Optional) Collect media from the plate/well into a 15 ml conical tube and place it on ice.

2 l (Optional) Wash the plate/well with 2 ml of 1 x PBS and collect them in the same tube.

**CRITICAL** Steps 1 and 2 are important if floating cells (i.e. dead or mitotic) are to be taken in account, otherwise the experimenter can just remove medium, wash with 1 x PBS and proceed to step 3.

3 l Add 1 ml trypsin (for a 6 cm dish) and incubate for a few minutes at 37 °C.

4 l Harvest trypsinized cells into the tube from step 2, if available.

5 l Filter cells through a 30 $\mu$m filter to remove any cell clumps and count cells. **CRITICAL** Proceed with 1 million cells up to 5 million cells depending on how many cells are available. For consistency of comparison with diploid control across the gates, it is important to keep all the samples at the same cell concentration (so that voltage settings on the flow cytometer do not need to be changed).

6 l Centrifuge 5 min at 450 g and discard supernatant.

7 l Resuspend cells in 5 ml of ice-cold 1xPBS containing 1 mg/ml BSA.

8 l Centrifuge 5 min at 450 g, discard the supernatant and resuspend cells in the residual liquid.

9 l Add 4 ml of ice-cold 70% ethanol drop wise using the vortex.

10 l Incubate samples at -20 °C for at least 30 min.

Pause point: Cells can be kept at -20 °C for weeks.

11 l Centrifuge 5 min at 450 g and discard supernatant.

12 l Resuspend cells in 0.5 ml PI reagent and transfer cells to a FACS tube.

Pause point: Cells can be kept at 4 °C overnight.

**Flow cytometric analysis**

*Timing: 30 minutes for 10 samples*

13 l Using the diploid control, acquire events with SSC-A-Lin *versus* FSC-A-Lin axes. The SSC-A-Lin *versus* FSC-A-Lin plot should reveal one population. Centre the main population using the voltage and gain settings and create the first gate (R1: ESCs) (**Figure 3a**).

14 l Isolate doublets by plotting FSC-W-Lin *versus* FSC-A-Lin and set gate R2 around single cells (**Figure 3b**).

**CRITICAL** In the case of the haploid/diploid cells the doublet cell population will reveal two overlapping populations, haploid (*n*; below) and diploid (*2n*; above); when the first experimental sample is run doublecheck that the haploid population is not excluded).

15 l Isolate the live cells by plotting FSC-W-Lin *versus* PI stain to identify dead cells (left side) and events caused by auto-fluorescence (right side) and remove them by setting gate R3 on the PI population (**Figure 3c**).

CRITICAL In the case of the haploid/diploid cells the doublet cell population will reveal two populations overlapping at the 100 K mark: haploid (*n*; at 50 K to 100 K) and diploid (*2n*; 100 K to 225 K); when the first experimental sample is run doublecheck that the haploid population is not excluded.

16 l On Count-Lin *versus* PI-Lin axes resolve the cell cycle profile. Acquire at least 10,000 events. On a cell line that contains a mix of haploid/diploid cells there will be evident (**Figure 3d**):

- Haploid G1 phase of the cell cycle (*hG1*) at the 50K mark;
- Haploid S phase of the cell cycle (*hS1*) between 50K to 100K marks;

- Overlapping haploid G2 phase of the cell cycle (*hG2*) and diploid G1 phase of the cell cycle (*dG1*) populations at the 100K mark;
- Diploid S phase of the cell cycle (*dS*) between 100K and 200K mark;
- Diploid G2 phase of the cell cycle (*dG2*) at the 200K mark.

**Figure 4a** shows a depiction of the cell cycle profiles of 3 random WT clones.

CRITICAL It is critical to run first the diploid only control to recognize the diploid cell cycle peaks. Some experimental samples might be pure haploid populations and without a diploid control it will be impossible to tell them apart.

17 l Save the experiment settings and experimental analysis on the flow cytometer.

**Box 2 Chromosomal analysis of hESC**

**Reagents**

- KaryoMAX® Colcemid™ Solution in HBSS (ThermoFisher cat. no. 15210040; 10 µg/ml (stock)). CAUTION Acutely toxic if swallowed; health hazard. Consult Safety and Hazard sheets that are provided by the seller.
- KCl (Sigma-Aldrich; cat. no. P9333).
- Acetic acid (>99.7%) (Sigma-Aldrich; cat. no. 695092-500ML-D). CAUTION Acutely toxic if swallowed; flammable; corrosive. Consult Safety and Hazard sheets that are provided by the seller.
- Methanol (>99.8%) (Sigma-Aldrich; cat. no. 179337-2.5L-D) (methanol: acetic acid; 3: 1). CAUTION Acutely toxic if swallowed; flammable. Consult Safety and Hazard sheets that are provided by the seller.
- Antifade Mounting Medium with DAPI (Vectashield laboratories; cat. no. H-1200). CAUTION Health hazard if swallowed; mutagen. Consult Safety and Hazard sheets that are provided by the seller.

**Equipment**

- Glass histology slides (75x25) (ThermoFisher; cat. no. 4951PLUS4).
- Glass coverslips (60x24) (ThermoFisher; cat. no. Q10143263NR).
- Water bath (Grant; JB Nova Unstirred Water Bath, cat. no. SAP12)

**Reagent setup**

- **Hypotonic solution.** 5.6 g KCl in 100 ml $H_2O$ (0.75M, stock); 10X dilution of stock (working solution). Working solution should be made fresh; stock can be stored at room temperature for one year.
- **Methanol:acetic acid fixative.** Mix 3:1 methanol:acetic acid (i.e. mix 30 ml methanol with 10 ml acetic acid). Always make on the day. CAUTION. Work under the chemical hood to prepare the fixative. Acutely toxic if swallowed; flammable.

**Sample preparation**

*Timing: 1 day and 3 hours*

1 l One day prior to karyotyping prepare a gelatinised plate by adding 5 ml 0.02% gelatin and leaving for 5 minutes prior to plating. Then passage 1 million hESC onto a 10 cm gelatinised plate.

2 l Add microscope slides to a jar containing 70% ethanol (this will degrease the slides and allow optimal adherence).

3 l 24 hours later add 200 *µ*l Colcemid to the culture to achieve a final concentration of 0.2 µg/ml. Return the flask to the incubator and continue incubating for 45 minutes.

4 l Prepare working hypothonic solution (10X stock) 4 ml for each sample and place it in a 37 °C water bath.

**CRITICAL** The hypotonic solution needs to be 37 °C for optimal hypotonization.

5 l Prepare the Methanol : Glacial Acetic Acid (3 : 1) fixative solution and place on ice.

6 l Wash the plate vigorously 2-3 times with the media covering the cells and save it in a 15 ml conical tube;

**CRITICAL** Most metaphases will be loosely attached to the plate and will be harvested at this step.

7 l Wash the plate gently with 1 x PBS and trypsinise the hESCs using 1.5 ml trypsin for 1 - 2 minutes in the incubator.

8 l Wash the cells with the media from the 15 ml conical tube (step 6) and harvest them in the same conical tube; this media will also stop trypsin action.

9 l Centrifuge the cells at 450 g for 5 minutes and discard the supernatant by inverting the conical tube.

10 l Add 6 ml working hypotonic solution and mix well by pipetting up and down.

11 l Incubate cells for 6 minutes at 37 °C.

**CRITICAL** Optimal hypotonisation is highly dependent on this incubation step. Under-hypotonisation will not break the cell wall. Over-hypotonisation will lead to chromosomes dissociating from the metaphase plate. If doing more than 4 samples, stagger them at this step so only 4 samples are processed at a time.

12 l Stop hypotonization by adding 1 ml cold fixative to each sample and mix by gentle inversion 2 - 3 times.

13 l Centrifuge the cells at 450 g for 5 minutes and discard the supernatant by inverting the conical tube.

14 l Resuspend the cell pellet in the residual liquid by flicking the tube.

**CRITICAL** It is important to resuspend the cells well at this step otherwise clumps will form and it will be hard to resolve single metaphase spreads.

15 | Using a Pasteur pipette, carefully add 2 ml of fixative solution dropwise, with gentle mixing to avoid clumping. Add an additional 6 ml of fixative and mix by gentle inversion of the tube.

**CRITICAL** It is important to gently add the fixative at the first fixation as this will have a big impact on the metaphases (chromosomes can dissociate from the metaphase plate). The fixative needs to be freshly made.

16 | Incubate the metaphase preparations on ice for 30 minutes.

17 | Centrifuge cells (450 g, 5 minutes) and discard the supernatant.

Pause step: At this step the metaphase preparations can be stored at -20 for years.

18 | Repeat steps 13 - 16 three times to clear the cytoplasm.

19 | Before spotting on slides resuspend the pellet in a small amount (0.5 ml) of fixative (this volume may need to be adjusted slightly according to cell number).

20 | Spot 2 - 3 drops in a row on each slide. Allow slide to dry. Humidity will affect spreading of metaphases (higher humidity will increase spreading).

If FISH analysis is desired follow the protocol by Rens W. et al. (2006)[48].

21 | Mount slides using antifade mounting medium with DAPI, cover with a coverslip and seal. A representative karyotype of a hESC clone is presented in **Figure 4b**.

**Box 3 Sorting of haploid ESC**

Before preparing the sort bear in mind you should always have a fully diploid control ESC cell line available to stain side by side to be able to calibrate the cell cycle profiles on the sorting machine. Before starting the cell harvest procedure fill to the edge two BD Falcon round-bottom tubes with warm ESC DMEM culture media and leave at room temperature in the hood for at least 30 minutes (the serum from the media will prevent cells sticking to the walls and not being collected in the collection media) and gelatinize T75 flasks using 5 ml 0.02% gelation solution for 5 minutes. Prepare a bucket of ice to store and transport the cells to and from the sorter.

**Reagents**

- 70% Ethanol (Fisher Bioreagents, Cat. No. BP8201-1). CAUTION Flammable. Irritant if swallowed. Consult Safety and Hazard sheets that are provided by the seller.
- Hoechst 33342, Trihydrochloride, Trihydrate - 10 mg/ml Solution in Water (Invitrogen; ThermoFischer Scientific cat. no. H3570). CAUTION Health hazard if swallowed; mutagen. Consult Safety and Hazard sheets that are provided by the seller.
- Verapamil hydrochloride (Sigma-Aldrich; cat. no. V4629); working solution 100 µmol/L. CAUTION Acutely toxic if swallowed; environmental hazard. Consult Safety and Hazard sheets that are provided by the seller.
- Sphero™ Rainbow Fluorescent particles, Spherotech (cat. no. RCP-20-5).

**Equipment**

- BD Falcon™ Round-Bottom Tubes (BD, Polypropylene, 5 ml, snap, cat. no. 352063).
- 30 $\mu$m filters (Sysmex, CellTrics 30 $\mu$m green; cat. no. 25004-0042-2316).
- Flow cytometer (Beckman Coulter; CytoFLEX S Flow Cytometer or similar, Model B75442).
- Cell sorter (BD Biosciences; BD Influx or similar cat. No. 646500).

**Reagent setup**

- **FACS buffer.** 0.5% BSA in 1 x PBS. Can be stored for up to 1 month at 4 ºC.
- **Hoechst working solution.** 1 $\mu$g/ml Hoechst in DMEM culture media. (e.g. add 1 $\mu$l of 10mg/ml Hoechst 33342 stock for each 100$\mu$l DMEM culture media). Make fresh.

**Sample preparation**

*Timing: 1 hour*

1 l Wash cells with 1 x PBS, trypsinize using 1 ml Trypsin and take up cells in DMEM culture media (10 ml for a T75 flask) by pipetting up and down a couple of times.

2 l Filter using 30 $\mu$m filters to remove cell clumps in a new 15 ml conical tube.

3 l Transfer cells to a 15 ml conical tube and spin down at 450 g for 5 minutes; discard the supernatant and re-suspend in appropriate volume of NDiff-N2B27 culture media (0.5 ml minimum for a T25 or 1 ml for a T75 flask).

4 l Prepare Hoechst working solution in DMEM culture media (*e.g.* 3*μ*l in 300 *μ*l DMEM culture media).

**CRITICAL** For some applications such as point mutagenesis screens, it is advised to avoid the use of Hoechst dye for setting up the gates and instead sort by size only.

5 l For each ml of cell suspension, add 100 *μ*l of 1 *μ*g/ml Hoechst working solution to a final Hoechst concentration of 10 *μ*g/ml. Incubate in the tissue culture incubator for up to 30 minutes with gentle resuspension every 10 minutes.

6 l (Optional) If you so wish, add verapamil (an inhibitor of calcium pumps and hence any G1 to S transition) to staining buffer to promote the identification of hG1 phase cells.

7 l Filter into a new flow cytometry tube using 30 *μ*m filters to remove cell clumps. Aspirate media from the collection tubes allowing 0.5 ml residual and transport the collection tubes, the diploid control tube and the sample tubes to the sorting machine for sorting. Sorting can take 2 – 3 hours (depending on how many cells are sorted).

**CRITICAL** As with any cell sorting protocol make sure the sorting machine is clean before sorting so that contamination with other cells or pathogens is avoided.

8 l Align the optical light path of the flow cytometer before cell analysis using 3 *μ*m beads (Sphero™ Rainbow Fluorescent particles, Spherotech) with minimum peak coefficient of variance (CV) for all fluorescence channels.

**CRITICAL** In parallel, stain as above the same (isogenic) cell line that has been allowed to diploidize previously. Using small aliquots (*e.g.* 100 *μ*l) prepare

a 1: 1 mixture from the haploid line for sorting and the isogenic diploid line. This step will allow the clear differentiation between haploid/diploid populations.

9 l Sort the G1 haploid population following the gating strategy depicted in **Figure 5** and collect as many cells as possible. Haploid cells can be sorted as single cells in 96 wells. The clones selected this way have different haploid latency times (see **Figure 5d**). If for downstream experimental purposes (*e.g.* point mutagenesis) the use of Hoechst dye is not desired, establish the gates with a small subpopulation of Hoechst stained cells, and then sort the bulk of the cells by size (mouse haploid ESCs are approximatively half the size of diploid isogenic ESCs). Precaution must be taken not to mistakenly gate out the diploids as doublets.

**CRITICAL** Sort the bulk of the hG1 population and avoid the sub-hG1 population that might contain cells that have lost chromosome X (mESCs are reported to randomly loose the X chromosome[45]).

10 l Centrifuge the sorted cells down (450 g) and resuspend in hESC media.

11 l Plate cells in appropriate flasks or dish containing hESC media and allow for recovery.

The haploid ESC can be sorted as a bulk population or as single cell clones. While the sorting of a bulk population reliably obtains almost pure haploid populations, sorting individual cells has different rates of success due to the long time these cells have to be in culture before they can be expanded for banking. An example of the haploid status of 3 individual clones from sorting 3 individual WT cells is presented in **Figure 5e**.

**Box 4 Haplobank**

The *Haplobank* (www.haplobank.at) contains over 100,000 cell lines covering about 17,000 genes in the mouse genome and thus represents the largest homozygous ESC library available to the scientific community to date[42]. These cell lines are on/off systems that allow for conditional induction of gene disruption and subsequent reversal allowing for fast complementation to be performed for validation. Moreover, these mutants were engineered in single cell derived clones and are thus isogenic, therefore allowing for improved reproducibility across different comparisons not only in one laboratory but across the scientific community. Such a resource will permit the broad and robust interrogation of the functional genome.

Below we highlight information to help with the decision making when selecting clones from the Haplobank website and the protocol for gene trap flipping. The constructs used for disrupting are presented in **Supplementary Figure 1a** and their sequences are available as **Supplementary Data**.

**Decision points: Selection of clone**

- Sense: disruptive (knockout; WT conditional ready).
- Antisense: non-disruptive (WT; knockout conditional ready).
- Clones highlighted in green: homozygous integration sites are confirmed.
- Clones highlighted in blue: only one gene trap insertion (should be first choice).
- Clones highlighted in grey: more than one gene trap insertion (should be second choice).

- 1Intron, Intron, 5'UTR, CDS, and ncExon: disruptive mutations. If possible, we would recommend going for an insertion in the first Intron (1Intron). Keep in mind that a mutation in the coding sequence is phenotypically not reversible.
- Upstream, InterDown and InterUp carry the gene trap in an intergenic region and cannot be regarded as knockouts.
- The choice of the mutagen is then the last thing: Tol2GT/Retro/Lenti are enhanced gene traps while Tol2 is a polyA- enhanced gene trap that carries EGFP.
- The insertion site is predicted with high certainty to be disruptive for all RefSeq transcripts of this gene. The splice acceptor is integrated in a genomic locus common to all RefSeq transcripts (intronic and exonic) and before 50% of the longest open reading frame.
- Integration sites were identified using three different approaches: two inverse PCRs using the two different restriction enzymes E1 (**Supplementary Figure 1b, Table 2 and Table 3**) and a barcode PCR, all followed by next generation sequencing. If the mapping strategy was successful, it has a green light.

**Gene trap flipping: Infection of murine ESC**

*Timing: 3 - 4 days*

**CRITICAL** Gene trap cassettes harbouring disruptive splice acceptor sites are integrated either in sense or antisense relative to corresponding gene transcripts. This will lead to disruptive or non-disruptive integrations, respectively. Inverted LoxP and FRT sites flank the gene trap cassette, enabling the orientation to be reversed twice, by transiently infecting cells with Cre or Flp recombinase expressing plasmids (double-

flex system). The thereby generated isogenic sister clone can be used as a perfect internal control. For the production of retrovirus containing the Cre or Flp expressing vector the Cre/Flp-expressing plasmid should be linked with a selection marker (e. g. mCherry; Cre-puro), to evaluate infection efficiency and/or select for infected cells. CRITICAL Before infection the ESCs should be propagated by standard procedures and should be passaged several times prior to infection and or sorted for haploid content if desired. The maps of these constructs are presented in Supplementary Figure 1 and respective sequences in Supplementay Data.

1 l 3 - 4h prior to infection seed about 250,000 cells per 6- well or 50,000 cells per 24 well for infection in triplicates;

2 l Distribute viral supernatant on ESCs. Infection efficiency can be increased by adding 2$\mu$g/ml Polybrene to the culture medium. Alternatively, if you use a frozen virus stock thaw aliquot (on ice) and add media according to previously determined dilution factor.

3 l 8 - 12h post infection exchange media. Check infection efficiency by flow cytometry 48h post infection or change to selection media and select until selection control wells are empty (*e.g.* with puromycin it takes 2-3 days).

4 l Confirm the successful inversion of the gene trap cassette by PCR or western-blot for the target gene.

**Box 5 Transposon mutagenesis in haploid ES cells**

*Timing: 10 to 15 days*

Various vectors can be used to stably integrate gene trap cassettes into genomic DNA of haploid cells. Amongst them, virus is most commonly used. However, integration sites for viral integrations cluster strongly and thus result in poor genome saturation[42]. Moreover, integration sites relative to gene bodies vary for different vector systems. Transposon mutagenesis is an attractive alternative method to integrate gene trap vectors into genomes as it not only achieves better genome saturation but also circumvents virus work and associated risks.

Single transposon integrations are only generated by strongly compromising on integration efficiency. When electroporating for complex library preparation (*e.g.* to generate 50 million independently mutated cells), the experimenter needs to electroporate at least 50 independent reactions. Using Tol2 transposons achieves good single integration events at 0.5 $\mu$g transposon plasmid and 10 $\mu$g transposase plasmid, however this may vary depending on transposon system and transposase variant, electroporator, cell line, and media conditions.

The challenge for transposon mutagenesis is to achieve single integration events for each cell. Electroporation of DNA is best suited for single integrations, nevertheless careful titration of transposon containing plasmids must be undertaken. The following protocol describes one possible approach, namely using an AMAXA 2D electroporator (see Equipment) and a Mouse ES cell Nucleofection Kit (see Materials). The workflow is outlined diagrammatically in **Figure 6**.

**Reagents**

- Embryonic Stem Cell Nucleofector™ Kit (or similar), Lonza, mouse (cat. no. VPH-1001) or human (cat. no. VAPH-5022).
- DNeasy Blood & Tissue Kits (50 DNeasy Mini Spin Columns – cat. no. 69504; 4 DNeasy 96 Plates – cat. no. 69581). CAUTION Buffer AL and AL/E: irritant if swallowed. Buffer AW1: acutely toxic if swallowed; irritant. Proteinase K: acutely toxic if swallowed; health hazard.
- Isopropanol (AppliChem, Cat. No. A3928.0500PE). CAUTION Flammable. Irritant if swallowed. Consult Safety and Hazard sheets that are provided by the seller.
- Quant-iT PicoGreen dsDNA Assay Kit (Invitrogen, Cat. No. P7589)
- T4 DNA Ligase (Roche, Cat. No. 10716359001)
- Phusion High-Fidelity DNA Polymerase (New England Biolabs, Cat. No. 174M0530S)
- QIAquick PCR Purification Kit (Qiagen, Cat. No. 28104). Buffer PB CAUTION Flammable. Irritant if swallowed. Consult Safety and Hazard sheets that are provided by the seller.
- QIAquick Gel Extraction Kit (Qiagen Cat. No. 28706). Buffer PB CAUTION Flammable. Irritant if swallowed. Consult Safety and Hazard sheets that are provided by the seller.
- NlaIII - restriction Enzyme (New England Biolabs, Cat. No. R0125S).
- TaqI - methyltransferase Enzyme (New England Biolabs, Cat. No. M0219S).
- MseI - restriction Enzyme (New England Biolabs, Cat. No. R0525S).
- PacI - restriction Enzyme (New England Biolabs, Cat. No. R0547S).

- SbfI - restriction Enzyme (New England Biolabs, Cat. No. R0642S).
- NextSeq 550 v2.5 Reagent Kits (for more details go on www.illumina.com).
- Tris-HCl 10 mM, pH 8.0 (Teknova, Cat. No. T1173)
- EDTA Dissodium Salt, 0.005 M (Astral Scientific, Cat. No. E-155-500ML)
- NaCl 100mM (Jena Bioscience, Cat. No. BU-114-100)
- Sodium Dodecyl Sulfate 10% (Fisher Bioreagents, Cat. No. BP2436-1) CAUTION Irritant if swallowed. Causes serious eye damage. Consult Safety and Hazard sheets that are provided by the seller.
- Proteinase K from Tritirachium album (Sigma, Cat. No. P2308-25MG). CAUTION Irritant if swallowed. Consult Safety and Hazard sheets that are provided by the seller.
- Agencourt AMPure XP beads (Beckman Coulter; cat. no. #A63882).

**Equipment**

- Electroporator (Nucleofector™ 2b Device, Lonza, cat. no. AAB-1001).
- Covaris Sonicator; 1 × 96 microtube plate, Adaptive Focused Acoustics (AFA) fibre (Covaris, cat. no. 520069).
- Next Generation Sequencer (Illumina NextSeq 550 System or higher; for further details go on www.illumina.com). NGS requires extensive expertise. We recommend the use of an NGS facility.

**Reagent setup**

- **Genomic DNA lysis buffer.** 10 mM Tris-HCl, pH 8.0; 5 mM EDTA ;100 mM NaCl; 1.0% SDS 1 mg/ml; Proteinase K. Can be stored at room temperature for up to 6 months.

**Electroporation of gene traps and screening**

1 l Expand haploid ESCs using standard conditions and sort for haploid cells (**Box 3**) a few days prior to electroporation.

2 l Harvest cells by trypsinization, stop trypsinization using culture medium and centrifuge cells using a 15 or 50 ml centrifuge tube depending on cell volume for 5 minutes at 20 g.

3 l In the meantime, prepare electroporation cuvettes and set the electroporator to program A-013. (Alternative programs suggested by the supplier can also be tested.)

4 l Prepare and label cell culture dishes, fill them with ES cell medium.

5 l Aspirate supernatant, take up cell pellet in 1 x PBS, count cells.

6 l Add 7 million cells to independent 15 ml centrifuge tubes and fill up with 1 x PBS to 15 ml. Use one tube per planned electroporation plus one tube for control.

7 l Centrifuge cells for 5 minutes at 20 g. In the meantime, prepare DNA solutions.

8 l In separate 1.5 ml tubes, mix 90 $\mu$l of electroporation mix (also see suppliers notes), 10 $\mu$g of transposase containing plasmid, and various amounts of transposon containing plasmid. We recommend to test (0, 0.25 $\mu$g, 0.5 $\mu$g, 1 $\mu$g, 2.5 $\mu$g, and 5 $\mu$g). Ensure that the total added volume does not exceed 10

µl. If the volume is between 10 µl and 20 µl, add 1/9 of the volume of DNA 10 x PBS. 20 µl should not be exceeded.

9 l Aspirate 1 x PBS in ES cell containing tubes. Briefly centrifuge at 100 g for 10-20 seconds and fully aspirate the 1 x PBS.

10 l Take up DNA solution and use it to carefully resuspend cell pellet, immediately place cells in electroporation cuvette.

11 l Electroporate cells.

12 l Carefully aspirate cells from the cuvette using the provided plastic Pasteur pipette directly into the cell culture medium of a prepared dish.

13 l 24 hours post electroporation begin antibiotic selection for integration events. Upon completion of selection cells can be used for cryopreservation and usage at a later time point.

**Library preparation and sequencing**

For library preparation enzyme 1 (E1) is used to fragment the genome (see **Table 2**). Because the recognition sequence for E1 is also present in the terminal repeat of the gene trap vector, it is possible to retrieve the exact integration site of the gene trap cassette within the genome by circularizing E1-digested genomic DNA (ring ligation) and subsequently amplifying the genomic region by inverse PCR (iPCR) using primers "US" and "DS" **(see Tables below** and **Supplementary Figure 1b)**. To improve iPCR efficiency, a linearization step using E2 should be used, which re-opens the rings generated previously. Moreover, each integration site can be mapped by using two different E1 enzymes. We recommend to split the samples and use both E1 enzymes on each sample in parallel. The directionality of the mapping strategies for different

gene traps is important, as this affects the assignment of a particular insertion to the sense and antisense strand in the genome, respectively. The DS primer is common for all PCR reactions, and binds in the gene trap. The US primer contains a sequence that will bind to the oligo present on the next generation sequencing (NGS) flow cell surface, an index of 4 to 8 bases, and then a sequence that binds in the gene trap.

Mutagenesis systems and enzymes used. (EGT; enhanced gene trap):

| **Name** | **Enzymes 1 (E1-1; E1-2)** | **Enzyme 2 (E2)** | **Mapping strategy** |
|---|---|---|---|
| **L**enti**-EGT** | NlaIII; TaqI | PacI | 5’ |
| **R**etro**-EGT** | NlaIII; MseI | SbfI | 5’ |
| **T**ol**2-EGT** | NlaIII; TaqI | PacI | 3’ |
| **T**ol**2-**poly**A**-EGT | NlaIII; TaqI | PacI | 3’ |

iPCR primer sequences used for library amplification and sequencing:

| Name | Sequence |
|---|---|
| Primer DS | AATGATACGGCGACCACCGAGATCTACACGAGCCAGAACCAGAAGGAACTTGA |
| Primer US | CAAGCAGAAGACGGCATACGAGAT***INDEX***GTGACTGGAGTTCAGACGTGTGCTC TTC |
| Lenti-EGT | CAGACCCTTTTAGTCAGTGTGGAAAATCTCTAGCA |
| Retro-EGT (1:1 mix) | GAGTGATTGACTACCCGTCAGCGGGGGTCTTTCA and TGAGTGATTGACTACCCACGACGGGGGTCTTTCA |
| Tol2-EGT Tol2-polyA-EGT | CACTTGAGTAAAATTTTTGAGTACTTTTTACACCTCTG |

The INDEX is a custom barcode (4-8 base pairs) that allows the unique identification of all the PCR reactions from one complex sample. For each sample two different US

primers with two different indices need to be used: one for the PCR reaction of the genomic DNA digested by E1-1 and one for the PCR reaction of the genomic DNA digested by E1-2. For NGS, all samples can be mixed and loaded into one NGS flow cell. The index sequence will allow the identification of the original samples. The workflow for library preparation is as below.

1 l Lyse cell pellet according to pellet size in 2-10 ml genomic DNA lysis buffer (see REAGENT SET UP).

2 l Incubate at 55˚C over night.

3 l Add 1:1000 RNAse A for 1h at 37˚C.

4 l Precipitate the genomic DNA using isopropanol and spool the DNA using heat sealed Pasteur pipettes.

5 l Wash with 70% ethanol twice and resuspend in 0,1-2 ml Tris/EDTA - TE Buffer.

6 l Digest samples with enzymes E1 (**Table 2**) in parallel (two separate reactions) using the following master mix:

| Component | Volume | Final concentration |
|---|---|---|
| gDNA 100ng/µl | 40 µl | 50ng/µl |
| 10x CutSmart | 8 µl | 1x |
| Enzyme1 | 3 µl | 150U/µl |
| dH2O | 29 µl | |

7 l Incubate at 37˚C (65˚C TaqI) over night.

8 l Purify restriction digests (Qiagen PCR kit or Agencourt AMPure XP beads).

9 l Elute in 100 µl in total.

10 l Perform ring ligation using the following master mix:

| Component | Volume | Final concentration |
|---|---|---|
| E1-gDNA | 100 µl | 8.3% |
| 10x Ligation Buffer | 120 µl | 1x |
| T4 DNA Ligase | 4 µl | 1.6 U/µl |
| dH2O | 976 µl | |

11 l Incubate over night at 16˚C.

12 l Heat-inactivate T4 DNA Ligase at 65˚C for 15min.

13 l Linearize samples with enzyme E2 at 37˚C for 2h using the following master mix:

| Component | Volume | Final concentration |
|---|---|---|
| RL-E1-gDNA | 1200 µl | |
| Enzyme 2 | 2 µl | ~0.016 U/µl |

14 l Purify restriction digest (Qiagen PCR kit or Magnetic Speedbeads) and elute in 100 µl H2O in total.

15 l Perform iPCR reaction. Use 5-10 reactions per sample and process half or all of the digest according to the expected DNA amount.

| Component | Volume | Final concentration |
|---|---|---|
| E2-RL-E1-gDNA from step 7 | 10 µl | 20% |
| primer US 100 µM | 0.1 µl | 0.2 µM |
| primer DS 100 µM | 0.1 µl | 0.2 µM |
| 10 mM dNTP mix | 1 µl | 0.2 mM |
| 10x pol. buffer | 5 µl | 1x |
| 20x polymerase | 3 µl | 1.2x |
| dH2O | 30.8 µl | |

Use the PCR cycler parameters given below:

| Cycles | Denature | Anneal | Extend |
|---|---|---|---|
| 1 | 95 ºC, 3 min | | |
| 2-38 | 95 ºC, 15 sec | 61 ºC, 25 sec | 72 ºC, 75 sec |
| 39 | | | 72 ºC, 5 min |

**CRITICAL** Because after amplification there will be big molar amounts of amplicons, there is a very high danger of contaminating other experiments with post PCR samples. We recommend using different pipette sets, rooms and equipment for post PCR samples.

16 l Pool all the PCR products from one sample (5-10 reactions) and analyse 20 µl on an agarose gel. A smear band beginning at around 400 base pairs expected.

17 l Quantify the DNA amount and mix all PCR samples (both separated digests for all 5 samples) in a way that you use the same DNA amount from each.

18 l Load that mixture on another agarose gel and cut the part of the smear between 400 and 800 base pairs.

19 l Extract the DNA from the gel using Qiagen's Gel Extraction Kit.

20 l Perform next generation sequencing.

**Statistical analysis**

To increase reliability of the analysis for each biological replicate two sequencing libraries should be constructed: one for sequencing into the adjacent genome 5' of the

insertion site, and one into the 3'adjacent genome. Pearson correlation should be used to compare between the read counts obtained by sequencing the two libraries and should allow reliable identification for each insertion. The statistical analysis can be done using the R framework[49] to map the reads for each insertion-site and Gaussian kernel convolution[50] can be used to identify common insertion sites. By using co-occurrence analysis in R and using a Fisher's exact test the overlap of common insertion sites across the biological repeats can be performed to identify strong gene candidates. For each gene the insertion sites can be mapped to identify position and direction of insertions across the gene body (*e.g.* gene X). These candidates can be subsequently used for DAVID analysis[51] to generate a network of genes nodes, with the node sizes being representative of the total number of common insertion sites **(Figure 6d)**.

**Anticipated results**

Typically, cells will undergo cell division prior to expression of transposase. Therefore, individual colonies appearing in selection will very often be mixed colonies with individual cells harboring independent integration events. Thus, to test for multiplicity of infection (MOI), pick colonies, trypsinize cells (of each condition) subsequent to completion of selection (days 5-7) and reseed them at clonal density (100 cells/10 cm dish, 500 cells/15 cm dish). Pick 20-50 colonies per condition at day 10-12 after reseeding and expand them in 96 well dishes. Genomic DNA can be prepared in bulk or in 96 wells depending on screen set-up using either genomic DNA lysis buffer or Quiagen DNeasy Blood & Tissue Kits. Optionally, acoustic shearing of DNA can be carried out using the Covaris system. It is most convenient to perform a PCR across

a random barcode integrated within the transposon and subject it for Sanger sequencing. Clean sequencing tracks are indicative of single integrations, while overlaying sequence tracks indicate multiple integration events. Please note that in the presence of transposase, transposons may be remobilized and integrated in a new genomic site. Even though these DNA transposons work *via* a cut-paste mechanism, mobilization in G2 phase of cell cycle can in some cases result in 2 independent transposon integration sites within a cell that carry the same random genetic barcode. Alternatively perform inverse PCR (iPCR) to map integration sites. Please note, that efficiency of iPCR is only around 80%, so some double integrations may be missed.

For generation of complex pools of cells proceed with screening for phenotypes of interest. Note that loss of protein may take up to 7 to 10 days for stable proteins in cases of late integration. The number and location of integrated transposons per cell can be identified by library sequencing and analysis. Library preparation can be performed in various ways (see review by Elling and Penninger 2014 [39]). We recommend the iPCR method described here and that we have successfully used[42]. Alternatively the transposon insertion site sequencing can be done also by *QiSeq* method as described by MJ Friedrich and collaborators[52].

Before executing NGS, please consider the expected complexity. You will receive 100 - 200 million reads from one single lane. As this is usually multiple times more than the complexity of your input, you might be oversampling the experiment. This can – by misalignment – cause artefacts. For that reason, it can be advantageous to combine several experiments in 1 lane by indexing. Subdividing single experiments

is not advisable, as the data are not simply additive. In order to optimally use NGS, consider the scale of input and experiment. If you wish to quantitatively use the NGS data, the 200 million reads must stem from 200 million independent genomes. If the complexity of input is significantly lower, the experiment can be scaled accordingly.

## Figure legends

**Figure 1. Workflow diagram of the haploid ESC (hESC) derivation. (a)** Female mice of selected genotypes are selected, superovulated and used for the harvest of oocytes. The oocytes are then activated by the use of strontium chloride ($SrCl_2$) and passed through rounds of selection for identification of blastocysts that can be used for zona pellucida removal and later transferred individually in NDiff-N2B27/2i/LIF culture media in 96 well plates (steps 1 to 52). **(b)** The individual clones are allowed to expand to form defined colonies that are disrupted by trypsinisation and grown and expanded into individual cell lines (steps 53 to 62). The established cell lines are then quality checked including: haploid status (for cell cycle analysis see **Box 1**), karyotyped (see **Box 2**), mycoplasma tested and banked. **(c and d)** The haploid cell lines can be adapted to DMEM/2i/LIF culture or only DMEM/LIF culture while enriching for the haploid content by sorting (see **Box 3**).

**Figure 2. Morphological selection during the haploid ESC (hESC) derivation. (a)** On day 4 post harvesting and activation, embryos will be at a variety of developmental stages with some embryos looking sub-viable and some lysed (left panel: pre-sort). Upon removal of sub-viable or lysed embryos only the embryos that show viable characteristics are retained (right panel: post-sort). Scale bar 250 $\mu$m. **(b)** On day 5, the embryos that have developed to blastocyst and are suitable for zona pellucida removal are identified. Only select embryos that have a well expanded blastocoel for zona removal. Scale bar 250 $\mu$m. **(c)** The check and selection of blastocysts for zona removal is repeated on day 6 (representative images). Scale bar 250 $\mu$m. **(d)** Representative image depicting the morphologic aspect of blastocyst post-zona

removal (days 6 to 9). These individual embryos are transferred to 96 well plates at this point. Scale bar 250 $\mu$m. **(e)** Representative image of a colony that has formed from a post-denuding blastocyst and that is ready for dissociation using trypsin. Scale bar 250 $\mu$m. **(f)** Representative image of an established hESC clone/cell line that is ready for quality control. Scale bar 250 $\mu$m.

**Figure 3. Representative flow cytometric analysis of the cell cycle profile. (a)** The samples harvested and fixed as described in the protocol from **Box 2** are acquired on SSC-A-Lin *versus* FSC-A-Lin plot to reveal the ESCs population. The gate R1 is centred around the main population representing the bulk of the ESCs. **(b)** Doublets are isolated by setting up gate R2 around the single cells by on an FSC-W-Lin *versus* FSC-A-Lin plot. In the case of the haploid/diploid cells the doublet cell population will reveal two overlapping populations, haploid (*n*; below; note that as explained in the text using the right controls this population can be sorted by size only without the use of a dye) and diploid (*2n*; above). **(c)** The live cells are identified by plotting gate R3 on FSC-W-Lin *versus* PI to identify and isolate dead cells (left side) and events caused by auto-fluorescence (right side). **(d)** Representative cell cycle profile of a clone that contains a mix of haploid and diploid cells. On Count-Lin *versus* PI-Lin axes the following cell cycle phases can be identified: the haploid G1 phase of the cell cycle (*hG1*), the haploid S phase of the cell cycle (*hS1*), the overlapping haploid G2 phase of the cell cycle (*hG2*) and diploid G1 phase of the cell cycle (*dG1*), the diploid S phase of the cell cycle (*dS*) and the diploid G2 phase of the cell cycle (*dG2*).

**Figure 4. Qualitative control of the haploid ESCs. (a)** Representative examples of the cell cycle profiles of three individual ESC cell lines at the first cell cycle analysis post derivation. While some clones might be more stable and thus highly enriched for haploid ESCs (left side profile), most will be a mix of haploid/diploid cells (middle profile) while some will be more unstable and be fully diploid (right side profile). **(b)** Fluorescent in situ hybridisation analysis of a representative haploid clone shows a stable *n*=20 karyotype (100X magnification).

**Figure 5. Sorting of haploid cells.** The sorting of the haploid ESCs follows similar set-up as for the cell cycle analysis on fixed cells. **(a)** The haploid/diploid cell line is stained with Hoechst 33342 and gate R1 is centred around the main population representing the bulk of the ESCs (SSC-H-Lin vs. FSC-H-Lin). **(b)** Doublets or clumps are removed by setting gate R2 around the main cell population (FL4-W: Hoechst-Lin vs. FL4-H: Hoechst-Lin); care must be taken so that diploids are not mistaken as doublets. A 1:1 haploid: diploid isogenic mix is recommended to be used always to set up the gates. **(c)** The cells gated on R2 are subsequently resolved on an FSC-H-Lin vs. FL4-H: Hoechst-Lin plot and haploid G1 (hG1) population is identified and sorted using gate R3. Gate R3 is built tightly around the hG1 peak to avoid sorting of cells with lower DNA content (that might be aneuploidy). **(d)** The profile of the hG1 (gate R3) on the Count-Lin vs. FL4: Hoechst-Lin representing the cell cycle profile. **(e)**. Graph presenting the overtime %hG1 status of 3 individual WT clones derived by individually sorting from the same WT line. Data was obtained by flow-cytometry profiling using PI. 10,000 cells (events) were acquired for each cell line at each time point.

**Figure 6. Representative diagram of a transposon screen in haploid ESCs. (a)** Diagram showing a map of one gene trap construct. The red arrow depicts a random integration event into a random gene (Gene X) and points to the direction of integration. As the integration is achieved in haploid ESC, upon diploidization the insertion will also be duplicated thus resulting always in homozygous representation. In screening strategies using transposons but also CRISPR-Cas9 this represents an important advantage. **(b)** Screening using transposon approaches rely on the electroporation of the transposon into the haploid ESC and selection of cells that have successfully integrated the constructs. After around 10 days in culture to allow the respective proteins to be depleted, these cells can be subsequently used for different screening strategies where a comparison between control untreated samples and treated sample is made. We recommend an n=4 technical replicates over 2 biological replicates for each group. **(c)** Library preparation for inverse PCR (iPCR) requires digestion of the genomic DNA using enzymes E1. Upon digestion the DNA fragments containing part of the enhanced gene trap containing DNA fragments (pEGT; red) are ring ligated followed by digestion of the ring DNA constructs containing EGT's followed by next generation sequencing (NGS). See also **Supplementary Figure 1b**. **(d)** Schematic representation of a possible screening outcome. The overlap between the common insertion sites across the 4 biological replicates can be represented as a Venn diagram. The enriched transposons (presumably one transposon on each cell) can be mapped to the respective genome locations for each gene. Based on this enrichment and based upon statistical analysis across the technical repeats by comparing the treatment groups to untreated groups, a gene ranking statistical score can be generated. The top candidate genes can subsequently be used for DAVID

pathway analysis for example, where each gene circle size can be representative of the number of insertions. Moreover, mapping the insertions over the gene bodies can inform on specific clustering around the promoter or maybe around a relevant domain region. For the statistical analysis see also the protocol by Friedrick et al. 2017[52].

**SI:**

Supplementary **Video 1.** Removing cumulus oocyte complex from ampulla.

Supplementary **Video 2.** Identification and isolation of sub-viable (pathogenic) embryos.

Supplementary **Video 3.** Zona pellucida removal (denuding) – first 20 seconds.

Supplementary **Video 4.** Zona pellucida removal (denuding) – making sure zona is gone up to 70

Supplementary Figure 1 Gene trap vectors and library preparation.

(a) Schematic representation of the gene trap vectors as presented in Elling et al. 20171 and on the Haplobank website (www.haplobank.at). Retroviral enhanced gene trap (Retro-EGT; related sequence provided as Supplementary Data file 1). Lentiviral enhanced gene trap (Lenti-EGT; related sequence provided as Supplementary Data file 2). Tol2 autonomous transposon enhanced gene trap (Tol2-EGT; related sequence provided as Supplementary Data file 3). Tol2 autonomous transposon polyadenylation enhanced gene trap (Tol2-polyA-EGT; related sequence provided as

Supplementary Data file 4). Abbreviations used: LTR, long terminal repeat; 6xOPE, six osteopontin enhancer elements; FRT/F3, heterotypic improved flippase target sequences; LoxP/Lox5171, heterotypic target sequences for the Cre-recombinase; SA, splice acceptor; βgal, β-galactosidase; NeoR, neomycin phosphotransferase fusion gene; polyA, bovine growth hormone polyadenylation sequence; L200/R175, left and right Tol2 transposon elements; IRES, internal ribosome entry site; EGFP, enhanced green fluorescent protein; RPB1, DNA-directed RNA polymerase II subunit rpb1; SD, splice donor. (b) Schematic representation of library preparation of gene trap vectors integration site. Tol2 – EGT is shown as example. Following fragmentation of the genome with enzyme 1 (E1, NlaIII in the example), the gene trap end containing the barcode and a genomic DNA portion is circularized (ring ligation). Prior to PCR amplification, linearization with enzyme 2 (E2, PaqI in the example) is needed. Each integration site can be mapped by using two different E1 enzymes. The genomic region is then amplified by PCR using US and DS primers.

Supplementary Data: Four sequences for gene trap cassettes harbouring disruptive splice acceptor sites.

**Table 1: Troubleshooting guidance.**

| Step | Problem | Possible Reason | Solution |
|---|---|---|---|
| 26 | There are too many embryos to easily process in 50-60 minutes | More mice superovulated than is necessary | Only superovulate 10-15 mice per session |
| 27 | Embryos are being lost when manipulated using a mouth pipette | Pipette tip is flicked against cell culture dish | Use a cell culture dish with low edges or an unpturned dish with drops of media |
| | | Embryos are sucked up beyond the tip of the pipette | Aspirate a small amount of M2 into the pipette tip and include a small air bubble to sit behind the embryos when manipulating |
| 39 | Embryos are not lysed but cell division looks abnormal | Embryo patterning differs in the absence of sperm entry | Allow the embryos to culture for longer |
| 42 | Blastocyst is close to breaching it's zona and hatching | The blastocysts were at an optimal stage for acid treatment over night | Perform the zona removal, being prepared to take them out of the acid sooner |
| 46 | The zona does not dissolve rapidly | The acid Tyrodes is not sufficiently warm | Warm the acid in an incubator at 37 ºC and use a heat plate on the microscope |
| 49 | It is hard to visualise the dissolving of the zona | Magnification of the microscope is insufficient | Use x100 magnification capable objective |
| 50 | Zona disappears but then reappears when in M2 | Variation in osmotic gradient between medias may give the impression the zona is shrinking away | Ensure the acid is suitably warmed |
| 52 | Embryos are lost when retrieving them from the vial | Too much media in vial | Use no more than 50 $\mu$l in the vial. |
| | | | Flush the inner sides and lid of the vial with M2 to dislodge and collect embryos that may have got attached to these surfaces |

**References**


1. Maderspacher, F. Theodor Boveri and the natural experiment. *Current Biology* **18,** R279–R286 (2008).
2. Crosland, M. W. J. and Crozier, R. H. Myrmecia pilosula, an Ant with Only One Pair of Chromosomes. *Science* **231,** 1278–1278 (1986).
3. Swart, E. C. *et al.* The Oxytricha trifallax Macronuclear Genome: A Complex Eukaryotic Genome with 16,000 Tiny Chromosomes. *PLOS Biology* **11,** e1001473 (2013).
4. Gallardo, M. H., Bickham, J. W., Honeycutt, R. L., Ojeda, R. A. & Köhler, N. Discovery of tetraploidy in a mammal. *Nature* **401,** 341–341 (1999).
5. Marston, A. L. & Amon, A. Meiosis: cell-cycle controls shuffle and deal. *Nat. Rev. Mol. Cell Biol.* **5,** 983–997 (2004).
6. Jackson, S. P. & Bartek, J. The DNA-damage response in human biology and disease. *Nature* **461,** 1071–1078 (2009).
7. Hustedt, N. & Durocher, D. The control of DNA repair by the cell cycle. *Nature Cell Biology* **19,** 1–9 (2017).
8. Hanahan, D. & Weinberg, R. A. Hallmarks of Cancer: The Next Generation. *Cell* **144,** 646–674 (2011).
9. Davoli, T. & de Lange, T. The Causes and Consequences of Polyploidy in Normal Development and Cancer. *Annual Review of Cell and Developmental Biology* **27,** 585–610 (2011).
10. Elling, U. *et al.* Forward and reverse genetics through derivation of haploid mouse embryonic stem cells. *Cell Stem Cell* **9,** 563–574 (2011).
11. Leeb, M. & Wutz, A. Derivation of haploid embryonic stem cells from mouse embryos. *Nature* **479,** 131–134 (2011).

12. Sagi, I. *et al.* Derivation and differentiation of haploid human embryonic stem cells. *Nature* **532,** 1–19 (2016).

13. Zhong, C. *et al.* Generation of human haploid embryonic stem cells from parthenogenetic embryos obtained by microsurgical removal of male pronucleus. *Cell Research* **26,** 743–746 (2016).

14. Sagi, I. & Benvenisty, N. Haploidy in Humans: An Evolutionary and Developmental Perspective. *Dev. Cell* **41,** 581–589 (2017).

15. Botstein, D. & Fink, G. R. Yeast: An Experimental Organism for 21st Century Biology. *Genetics* **189,** 695–704 (2011).

16. Cordes, S. P. N-ethyl-N-nitrosourea mutagenesis: boarding the mouse mutant express. *Microbiol. Mol. Biol. Rev.* **69,** 426–439 (2005).

17. Kotecki, M. Isolation and Characterization of a Near-Haploid Human Cell Line. *Experimental Cell Research* **252,** 273–280 (1999).

18. Carette, J. E. *et al.* Haploid Genetic Screens in Human Cells Identify Host Factors Used by Pathogens. *Science* **326,** 1231–1235 (2009).

19. Carette, J. E. *et al.* Generation of iPSCs from cultured human malignant cells. *Blood* **115,** 4039–4042 (2010).

20. Essletzbichler, P. *et al.* Megabase-scale deletion using CRISPR/Cas9 to generate a fully haploid human cell line. *Genome Research* **24,** 2059–2065 (2014).

21. Pincus, G. Observations on the living eggs of the rabbit. *Proc. R. Soc. Lond. B* **107,** 132–167 (1930).

22. Tarkowski, A. K., Witkowska, A. and Nowicka, J. Experimental Parthenogenesis in the Mouse. *Nature* **226,** 162–165 (1970).

23. Graham, C. F. Parthenogenetic Mouse Blastocysts. *Nature* **226,** 165–167 (1970).

24. Kaufman, M. H. *Early Mammalian Development: Parthenogenetic studies*. **14,** (1983).

25. Fraser, L. R. Strontium supports capacitation and the acrosome reaction in mouse sperm and rapidly activates mouse eggs. *Gamete Research* **18,** 363–374 (1987).

26. Kline, D. & Kline, J. T. Repetitive calcium transients and the role of calcium in exocytosis and cell cycle activation in the mouse egg. *Developmental Biology* **149,** 80–89 (1992).

27. Mikich, A. B. *et al.* Calcium oscillations and protein synthesis inhibition synergistically activate mouse oocytes. *Mol. Reprod. Dev.* **41,** 84–90 (1995).

28. Cuthbertson, K. S. R., Whittingham, D. G. & Cobbold, P. H. Free Ca2+ increases in exponential phases during mouse oocyte activation. *Nature* **294,** 754–757 (1981).

29. Stevens, L. C. & Varnum, D. S. The development of teratomas from parthenogenetically activated ovarian mouse eggs. *Developmental Biology* **37,** 369–380 (1974).

30. Linder, D., McCaw, B. K. & Hecht, F. Parthenogenic Origin of Benign Ovarian Teratomas. *N. Engl. J. Med.* **292,** 63–66 (1975).

31. Leeb, M., Perry, A. C. F. & Wutz, A. Establishment and Use of Mouse Haploid ES Cells. *Curr Protoc Mouse Biol* **5,** 155–185 (2017).

32. Shuai, L. *et al.* Generation of Mammalian offspring by haploid embryonic stem cells microinjection. *Curr Protoc Stem Cell Biol* **31,** 1A.6.1–15 (2014).

33. Balmus, G. *et al.* ATM orchestrates the DNA-damage response to counter toxic non-homologous end-joining at broken replication forks. *Nat Commun* **10,** 646 (2019).

34. Yang, H. *et al.* Generation of haploid embryonic stem cells from Macaca fascicularis monkey parthenotes. *Cell Research* **23,** 1187–1200 (2013).

35. Li, X. *et al.* Generation and Application of Mouse-Rat Allodiploid Embryonic Stem Cells. *Cell* **164,** 279–292 (2016).

36. Yang, H. *et al.* Generation of Genetically Modified Mice by Oocyte Injection of Androgenetic Haploid Embryonic Stem Cells. *Cell* **149,** 605–617 (2012).

37. Li, W. *et al.* Androgenetic haploid embryonic stem cells produce live transgenic mice. *Nature* **490,** 407–411 (2012).

38. Leeb, M. & Wutz, A. Germline potential of parthenogenetic haploid mouse embryonic stem cells. *Development* **139,** 3301–3305 (2012).

39. Elling, U. & Penninger, J. M. Genome wide functional genetics in haploid cells. *FEBS Lett.* **588,** 2415–2421 (2014).

40. Forment, J. V. *et al.* Genome-wide genetic screening with chemically mutagenized haploid embryonic stem cells. *Nat. Chem. Biol.* **13,** 12–14 (2017).

41. Herzog, M. *et al.* Detection of functional protein domains by unbiased genome-wide forward genetic screening. *Sci Rep* **8,** 6161 (2018).

42. Elling, U. *et al.* A reversible haploid mouse embryonic stem cell biobank resource for functional genomics. *Nature* **550,** 114 (2017).

43. Gao, Q. *et al.* Derivation of Haploid Neural Stem Cell Lines by Selection for a Pax6-GFPReporter. *Stem Cells and Development* **27,** 479–487 (2018).

44. Kim, Y. M., Lee, J.-Y., Xia, L., Mulvihill, J. J. & Li, S. Trisomy 8: a common finding in mouse embryonic stem (ES) cell lines. *Mol Cytogenet* **6,** 3 (2013).

45. Morey, R. & Laurent, L. C. Getting off the ground state: X chromosome inactivation knocks down barriers to differentiation. *Cell Stem Cell* **14,** 131–132 (2014).

46. Kilkenny, C., Browne, W. J., Cuthill, I. C., Emerson, M. & Altman, D. G. Improving bioscience research reporting: the ARRIVE guidelines for reporting animal research. *PLoS biology* **8,** e1000412 (2010).

47. Brownstein, D. G. Manipulating the Mouse Embryo: A Laboratory Manual. Third Edition. By Andras Nagy, Marina Gertsenstein, Kristina Vintersten, and Richard Behringer. Cold Spring Harbor Laboratory Press. *The Quarterly Review of Biology* **78,** 365–365 (2003).

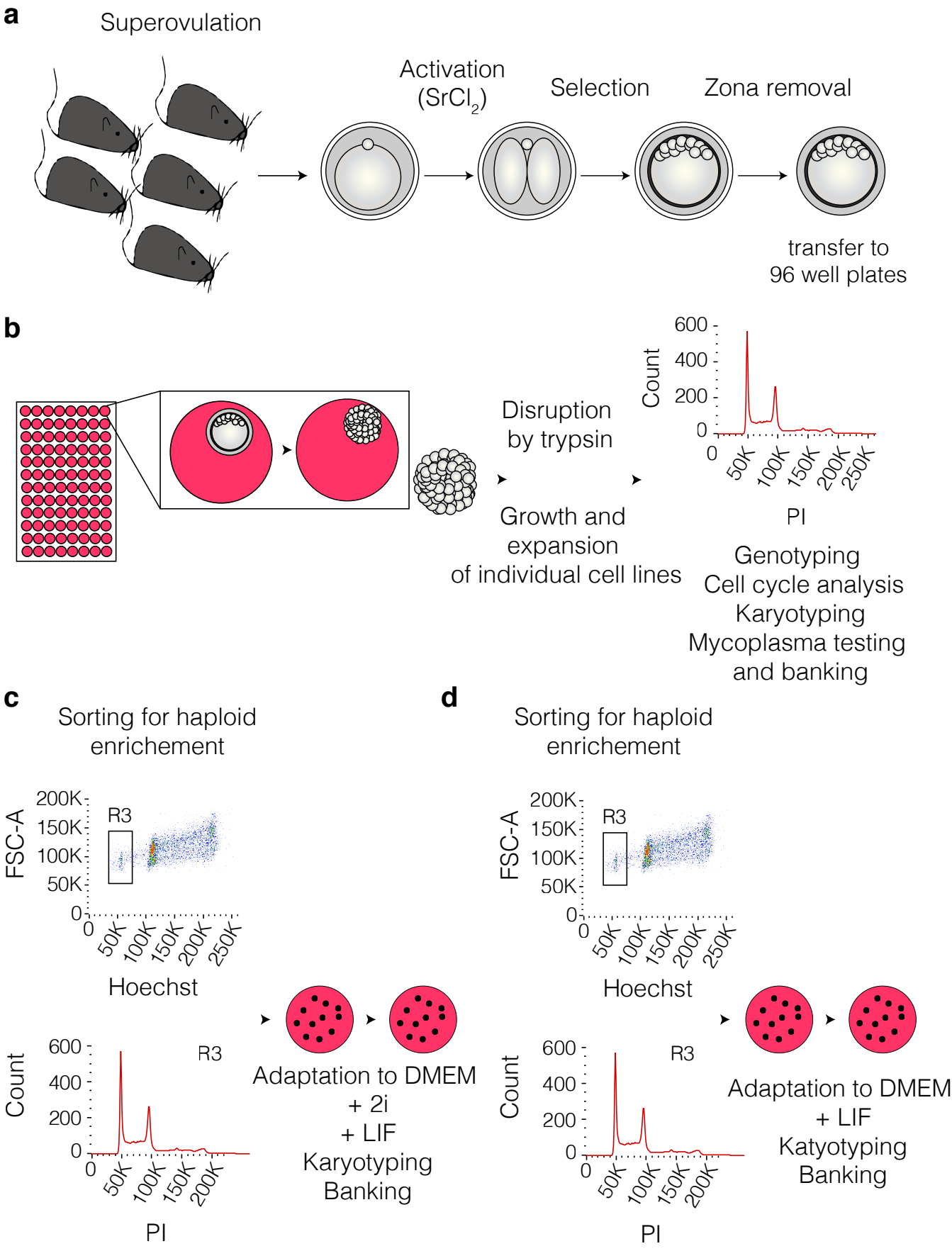
48. Rens, W., Fu, B., O'Brien, P. C. M. & Ferguson-Smith, M. Cross-species chromosome painting. *Nat Protoc* **1,** 783–790 (2006).

49. Team, R. C. R: A language and environment for statistical computing. (2013).

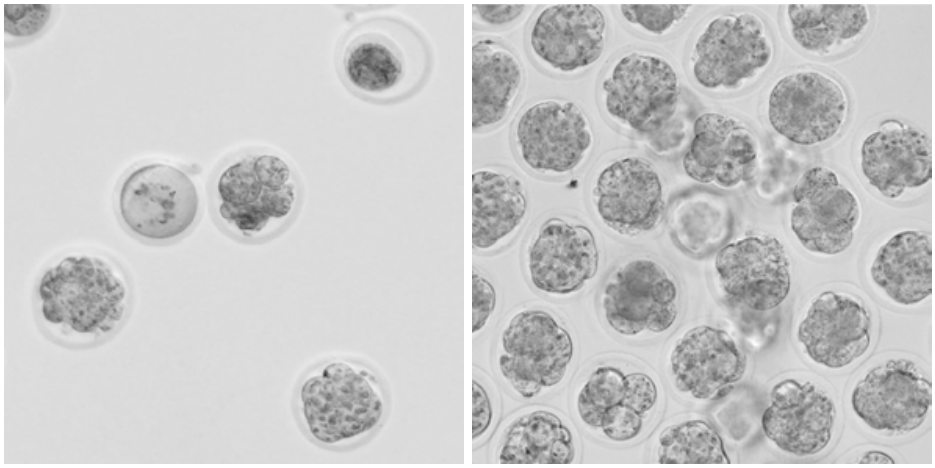
50. de Ridder, J., Uren, A., Kool, J., Reinders, M. & Wessels, L. Detecting Statistically Significant Common Insertion Sites in Retroviral Insertional Mutagenesis Screens. *PLoS Comput Biol* **2,** e166 (2006).

51. Da Wei Huang, Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4,** 44–57 (2009).

52. Friedrich, M. J. *et al.* Genome-wide transposon screening and quantitative insertion site sequencing for cancer gene discovery in mice. *Nat Protoc* **12,** 289–309 (2017).

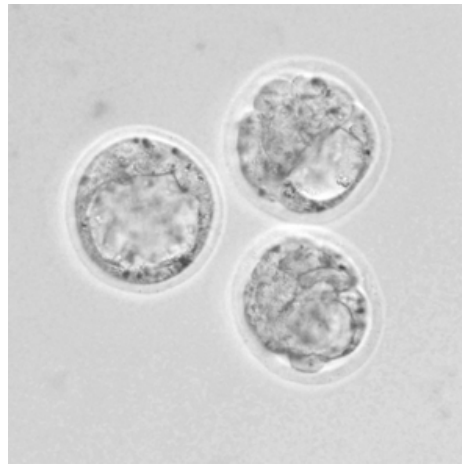

a
Superovulation
Activation
($SrCl_2$)
Selection
Zona removal
transfer to
96 well plates
b
Disruption
by trypsin
Growth and
expansion
of individual cell lines
Count
600
400
200
0
0
50K
100K
150K
200K
250K
PI
Genotyping
Cell cycle analysis
Karyotyping
Mycoplasma testing
and banking
c
Sorting for haploid
enrichement
FSC-A
200K
150K
100K
50K
0
R3
Hoechst
Count
R3
PI
Adaptation to DMEM
+ 2i
+ LIF
Karyotyping
Banking
d
Sorting for haploid
enrichement
FSC-A
R3
Hoechst
Count
R3
PI
Adaptation to DMEM
+ LIF
Katyotyping
Banking

**a** Pre-sort Post-sort



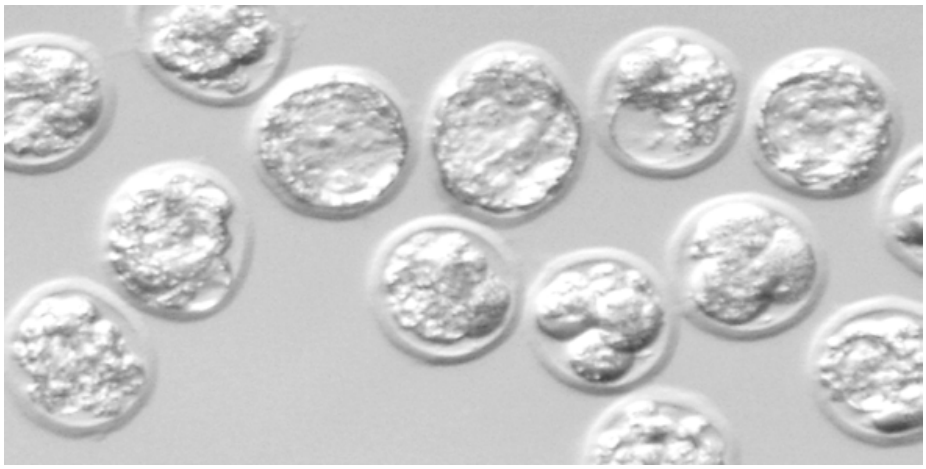
Day 4

**b** Post-sort



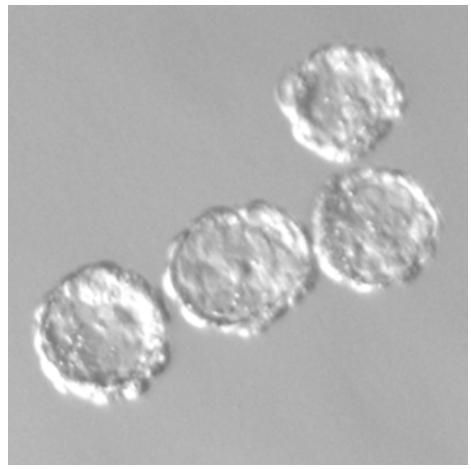
Day 5

**c** Post-sort (Pre-denuding)



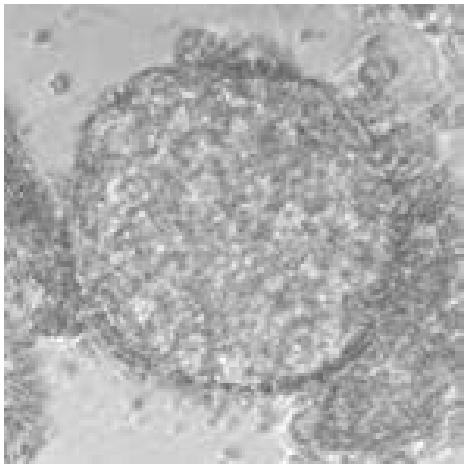
Day 6

**d** Post-denuding



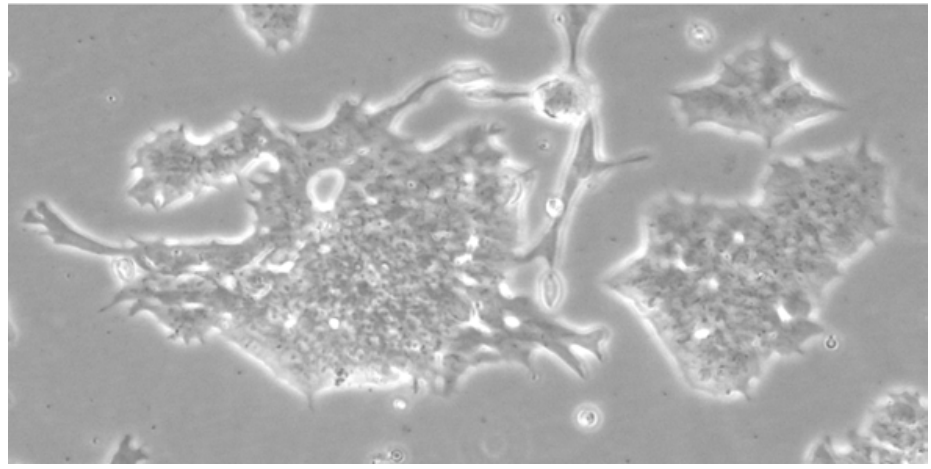
Day 6 to 9

**e** Post-denuding colony

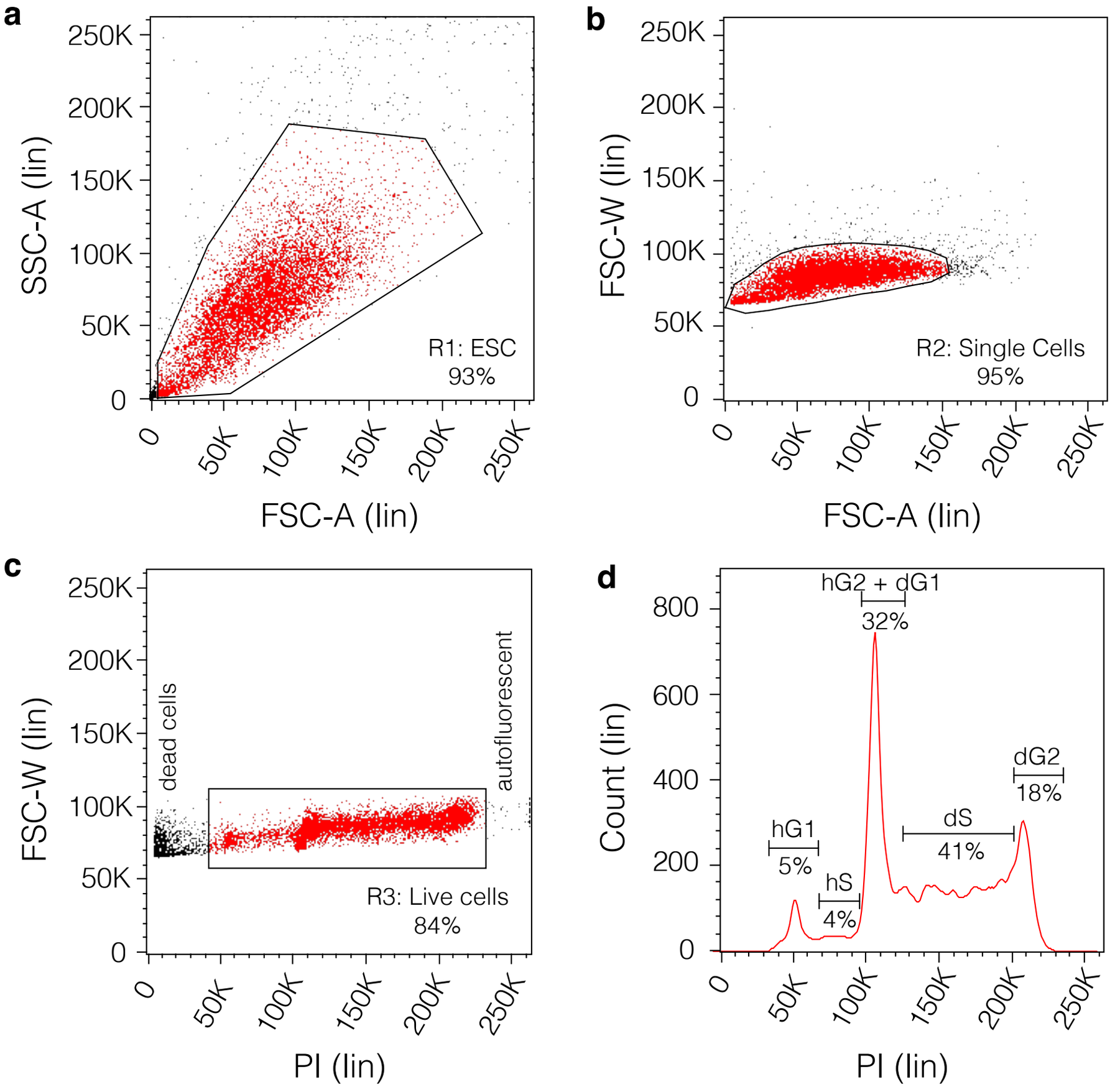


Day 15 to Day 20

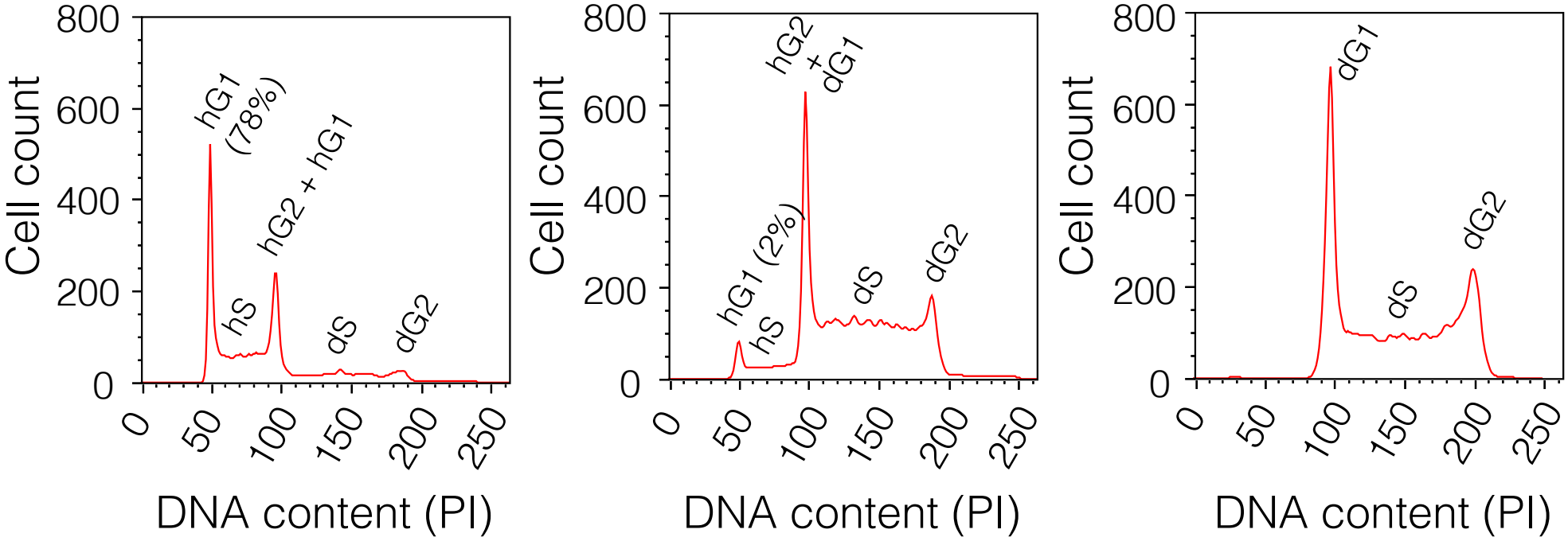
**f** Established cell line



Day 18 to Day 23

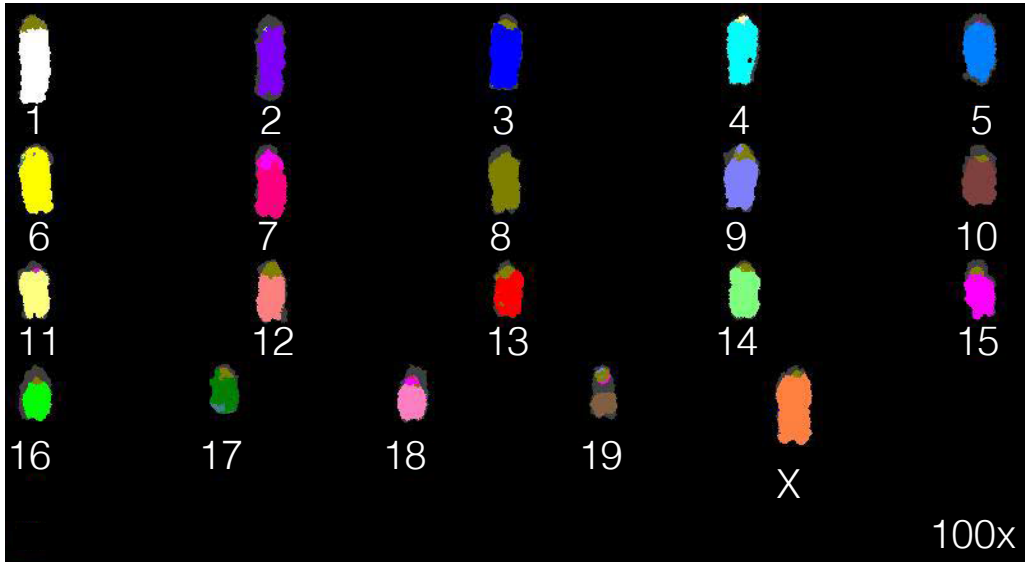

a
SSC-A (lin)
FSC-A (lin)
R1: ESC
93%
b
FSC-W (lin)
FSC-A (lin)
R2: Single Cells
95%
c
FSC-W (lin)
PI (lin)
dead cells
autofluorescent
R3: Live cells
84%
d
Count (lin)
PI (lin)
hG2 + dG1
32%
hG1
5%
hS
4%
dS
41%
dG2
18%
0
50K
100K
150K
200K
250K
200
400
600
800

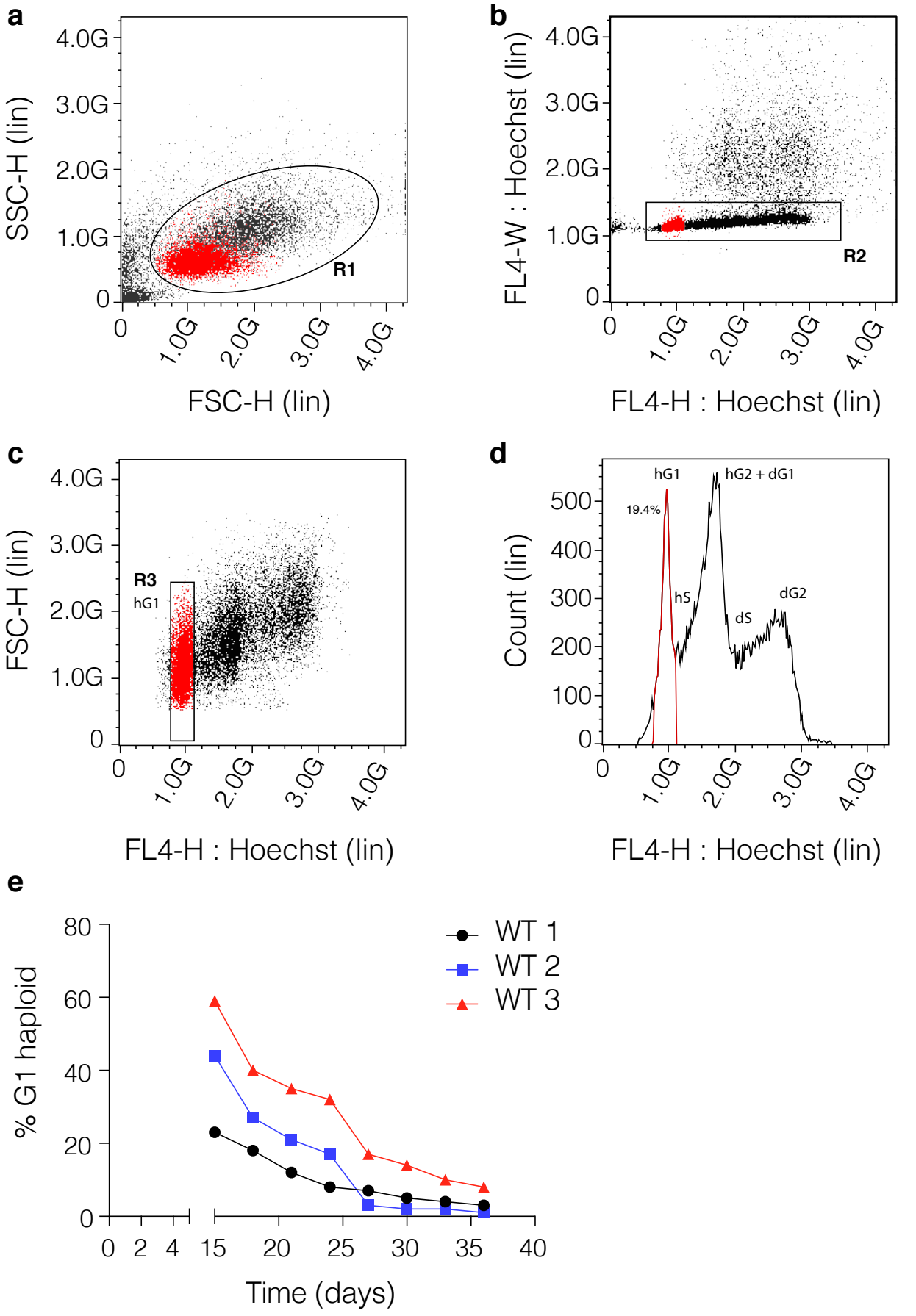
**a**

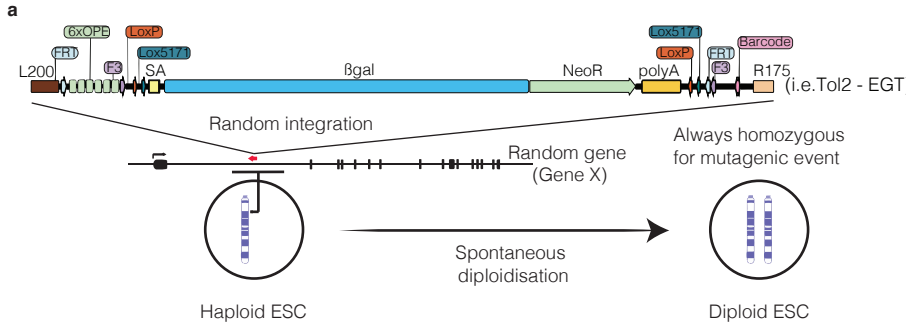

Cell count
800
600
400
200
0
hG1 (78%)
hS
hG2 + hG1
dS
dG2
0
50
100
150
200
250
DNA content (PI)
hG1 (2%)
hS
hG2 + dG1
dS
dG2
DNA content (PI)
dG1
dS
dG2
DNA content (PI)

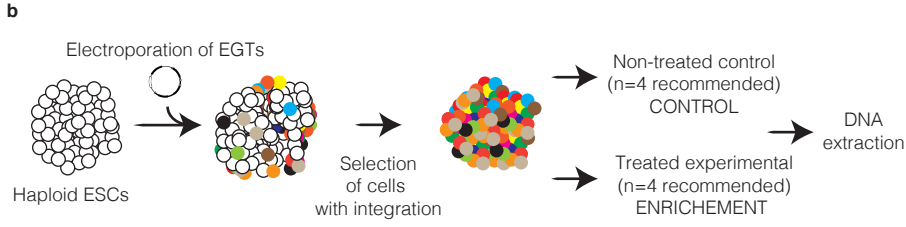
**b**

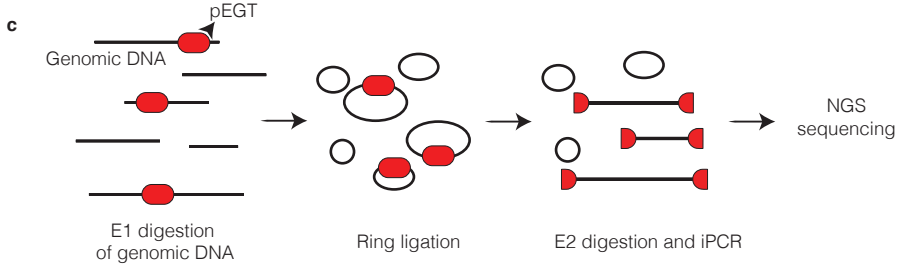

100x

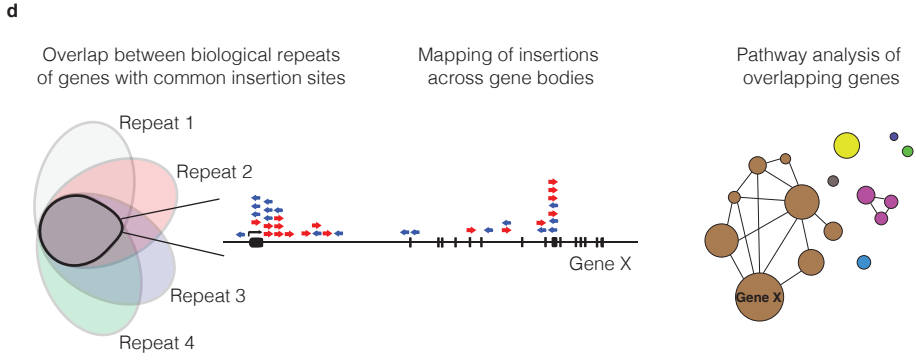

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
X
100x


a
SSC-H (lin)
FSC-H (lin)
R1
b
FL4-W : Hoechst (lin)
FL4-H : Hoechst (lin)
R2
c
FSC-H (lin)
FL4-H : Hoechst (lin)
R3
hG1
d
Count (lin)
FL4-H : Hoechst (lin)
hG1
19.4%
hS
hG2 + dG1
dS
dG2
e
% G1 haploid
Time (days)
WT 1
WT 2
WT 3


a
6xOPE
LoxP
FRT
Lox5171
L200
F3
SA
ßgal
NeoR
polyA
Lox5171
LoxP
FRT
F3
Barcode
R175
(i.e.Tol2 - EGT)
Random integration
Random gene
(Gene X)
Always homozygous
for mutagenic event
Spontaneous
diploidisation
Haploid ESC
Diploid ESC


b
Electroporation of EGTs
Haploid ESCs
Selection
of cells
with integration
Non-treated control
(n=4 recommended)
CONTROL
Treated experimental
(n=4 recommended)
ENRICHEMENT
DNA
extraction


c
pEGT
Genomic DNA
E1 digestion
of genomic DNA
Ring ligation
E2 digestion and iPCR
NGS
sequencing


d
Overlap between biological repeats
of genes with common insertion sites
Mapping of insertions
across gene bodies
Pathway analysis of
overlapping genes
Repeat 1
Repeat 2
Repeat 3
Repeat 4
Gene X
Gene X

**Supplementary Materials for**

# Derivation and maintenance of mouse haploid embryonic stem cells

**Elling, Woods et al.**

**Supplementary Figure 1**

**Supplementary Data files 1 to 4**

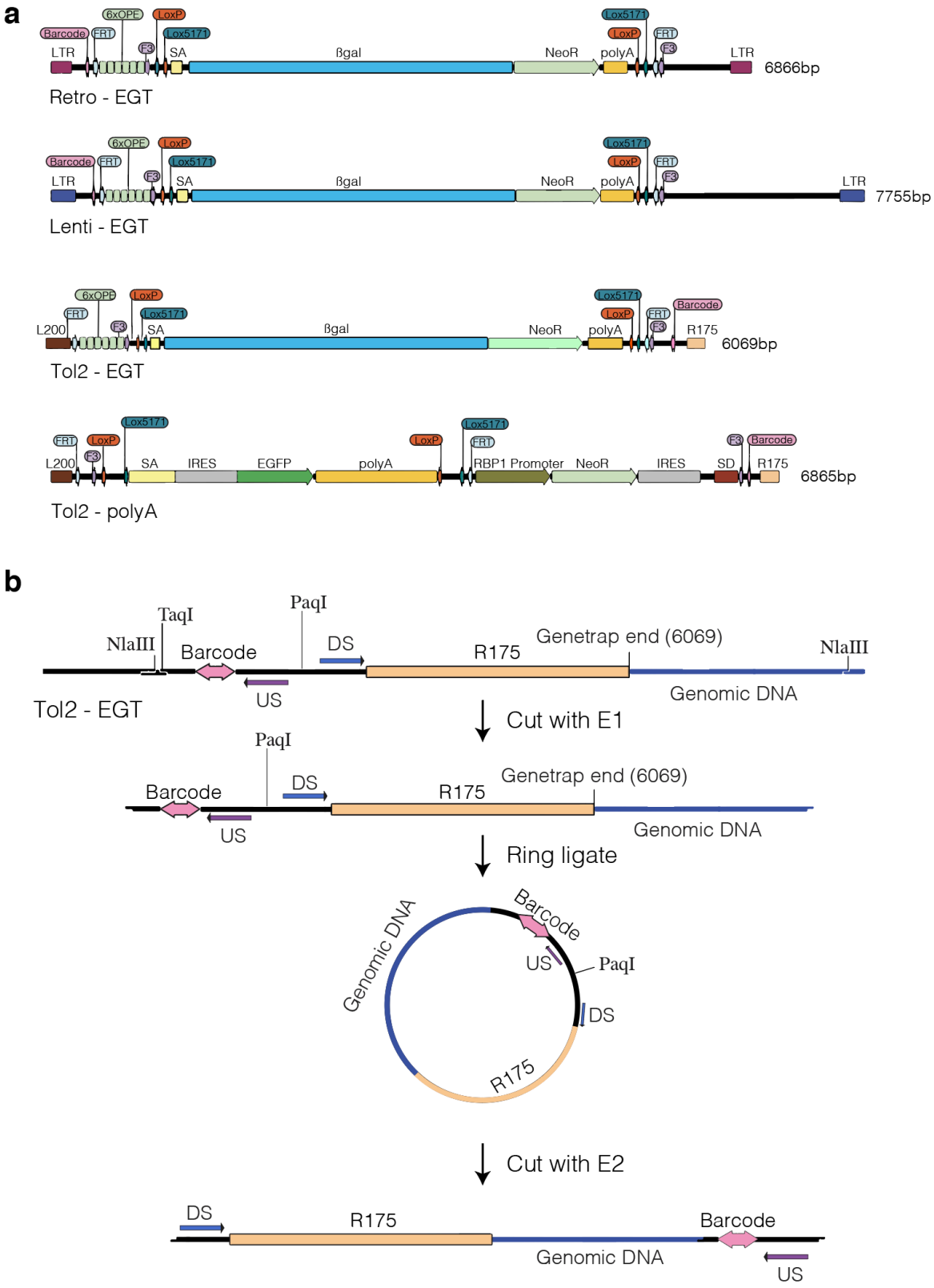
# Derivation and maintenance of mouse haploid embryonic stem cells

Ulrich Elling[1]*, Michael Woods[2,3]*, Josep V. Forment[4], Beiyuan Fu[3], Fengtang Yang[3], Bee Ling Ng[3], Jose R. Vicente[2], David J. Adams[3], Brendan Doe[3], Stephen P. Jackson[5], Josef M. Penninger[1,6#] and Gabriel Balmus[2#]

1. Institute of Molecular Biotechnology of the Austrian Academy of Science (IMBA), Vienna Biocenter (VBC), Dr. Bohr Gasse 3, Vienna, Austria.
2. UK Dementia Research Institute at University of Cambridge and Department of Clinical Neurosciences, University of Cambridge, CB2 0AH, UK.
3. Wellcome Trust Sanger Institute, Hinxton, Cambridge CB10 1SA, UK.
4. DNA Damage Response Biology, Oncology Innovative Medicines, Astra Zeneca, Cambridge CB4 0WG, UK
5. The Wellcome Trust CRUK Gurdon Institute and Department of Biochemistry, University of Cambridge, Tennis Court Road, Cambridge CB2 1QN, UK.
6. Department of Medical Genetics, Life Science Institute, University of British Columbia, Vancouver, Canada.

* These authors contributed equally to this work.

# Correspondence should be addressed to GB (gb318@cam.ac.uk) or JMP (Josef.Penninger@imba.oeaw.ac.at)


a
6xQPE
LoxP
Barcode
FRT
Lox5171
LTR
F3
SA
βgal
NeoR
polyA
LTR
6866bp
Retro - EGT
7755bp
Lenti - EGT
L200
R175
6069bp
Tol2 - EGT
IRES
EGFP
RBP1 Promoter
SD
6865bp
Tol2 - polyA
b
TaqI
PaqI
NlaIII
DS
US
R175
Genetrap end (6069)
Genomic DNA
Tol2 - EGT
Cut with E1
Ring ligate
Cut with E2

**Supplementary Figure 1**. **(a)** Schematic representation of the gene trap vectors as presented in Elling et al. 2017[1] and on the Haplobank website (www.haplobank.at). Retroviral enhanced gene trap (Retro-EGT; related sequence provided as Supplementary Data file 1). Lentiviral enhanced gene trap (Lenti-EGT; related sequence provided as Supplementary Data file 2). Tol2 autonomous transposon enhanced gene trap (Tol2-EGT; related sequence provided as Supplementary Data file 3). Tol2 autonomous transposon polyadenylation enhanced gene trap (Tol2-polyA-EGT; related sequence provided as Supplementary Data file 4). **Abbreviations used**: LTR, long terminal repeat; 6xOPE, six osteopontin enhancer elements; FRT/F3, heterotypic improved flippase target sequences; LoxP/Lox5171, heterotypic target sequences for the Cre-recombinase; SA, splice acceptor; βgal, β-galactosidase; NeoR, neomycin phosphotransferase fusion gene; polyA, bovine growth hormone polyadenylation sequence; L200/R175, left and right *Tol2* transposon elements; IRES, internal ribosome entry site; EGFP, enhanced green fluorescent protein; RPB1, DNA-directed RNA polymerase II subunit rpb1; SD, splice donor. **(b)** Schematic representation of library preparation of gene trap vectors integration site. Tol2 – EGT is shown as example. Following fragmentation of the genome with enzyme 1 (E1, NlaIII in the example), the gene trap end containing the barcode and a genomic DNA portion is circularized (ring ligation). Prior to PCR amplification, linearization with enzyme 2 (E2, PaqI in the example) is needed. Each integration site can be mapped by using two different E1 enzymes. The genomic region is then amplified by PCR using US and DS primers.

**Supplementary Data file 1.** Sequence file for the retroviral construct (Retro-EGT) presented in Supplementary Figure 1a.

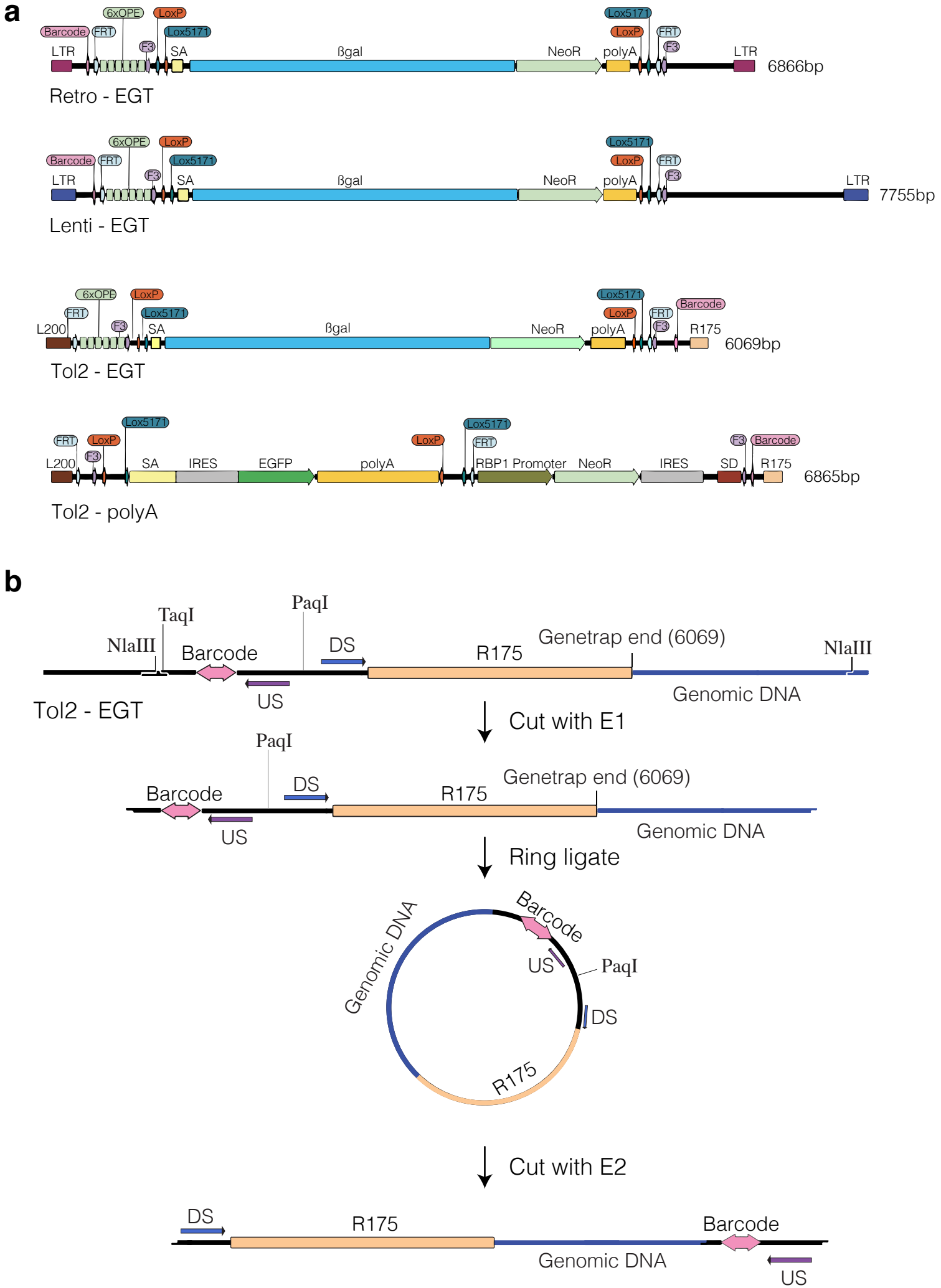
**Supplementary Data file 2.** Sequence file for the Lentiviral-EGT construct (Lenti-EGT) presented in Supplementary Figure 1a.

**Supplementary Data file 3.** Sequence file for the transposon construct (Tol2-EGT) presented in Supplementary Figure 1a.

**Supplementary Data file 4.** Sequence file for the transposon construct (Tol2-polyA-EGT) presented in Supplementary Figure 1a.

## References

1. Elling, U. *et al.* A reversible haploid mouse embryonic stem cell biobank resource for functional genomics. *Nature* **550,** 114 (2017).


a
6xOPE
LoxP
Barcode
FRT
Lox5171
F3
LTR
SA
ßgal
NeoR
polyA
LTR
6866bp
Retro - EGT
7755bp
Lenti - EGT
L200
R175
6069bp
Tol2 - EGT
IRES
EGFP
RBP1 Promoter
SD
6865bp
Tol2 - polyA
b
TaqI
NlaIII
PaqI
DS
US
R175
Genetrap end (6069)
Genomic DNA
Tol2 - EGT
Cut with E1
Ring ligate
Cut with E2