

Linking a cell-division gene and a suicide gene to define and improve cell therapy safety

Qin Liang^{1,2,8}, Claudio Monetti^{1,8}, Maria V. Shutova¹, Eric J. Neely^{1,2}, Sabiha Hacibekiroglu^{1,2}, Huijuan Yang^{1,3}, Christopher Kim^{1,2}, Puzheng Zhang¹, Chengjin Li¹, Kristina Nagy^{1,3}, Maria Mileikovsky¹, Istvan Gyongy⁴, Hoon-Ki Sung^{1,2,6,7*} & Andras Nagy^{1*}

Human pluripotent cell lines hold enormous promise for the development of cell-based therapies. Safety, however, is a crucial prerequisite condition for clinical applications. Numerous groups have attempted to eliminate potentially harmful cells through the use of suicide genes¹, but none has quantitatively defined the safety level of transplant therapies. Here, using genome-engineering strategies, we demonstrate the protection of a suicide system from inactivation in dividing cells. We created a transcriptional link between the suicide gene herpes simplex virus thymidine kinase (HSV-TK) and a cell-division gene (CDK1); this combination is designated the safe-cell system. Furthermore, we used a mathematical model to quantify the safety level of the cell therapy as a function of the number of cells that is needed for the therapy and the type of genome editing that is performed. Even with the highly conservative estimates described here, we anticipate that our solution will rapidly accelerate the entry of cell-based medicine into the clinic.

Most randomly integrated transgenes show variegated expression^{2–4}. To achieve reliable expression, we generated a transcriptional link between a cell-division locus (CDL), which is a gene that is essential for a cell to divide or to survive, and a drug-inducible suicide system (SU), which resulted in a CDL–SU allele (Fig. 1a). Therefore, the expression of the CDL and the drug-inducible suicide system are tightly linked and, if required, either only the dividing cells or all of the cells can be arrested or eliminated by treatment with the drug that induces the suicide system.

We consider a cell population a safe-cell batch when all cells within that batch contain a functional suicide system. The safe-cell level (SCL) is the number of therapeutic batches in which there is expected to be only one non-safe batch (Fig. 1b), such that one in a thousand gives a SCL of 1,000 and one in a million has a SCL of 1,000,000. We used various *in vitro* and *in vivo* experiments as well as mathematical modelling to define the SCL as the function of the number of cells needed for therapy.

From a list of CDL candidates (Supplementary Table 1), we chose *CDK1* and *HSV-TK.007*⁵ (*TK*) as prototypes. The absence of *CDK1* causes a block in the G2 to M transition of the cell cycle, and other CDKs are not able to rescue this deficiency^{6–8}. *TK* has been extensively used for cell ablation⁹ and its mechanism of action in the presence of its clinically approved prodrug ganciclovir (GCV), is well-characterized¹⁰. *CDK1* is not expressed in non-dividing cells⁸. Therefore, *TK* is not expressed from the *CDK1-TK* allele, eliminating the potential effect of its immunogenicity¹¹.

To generate a transcriptional link between *CDK1* and *TK* (Fig. 1c), we inserted *TK* into the 3' untranslated region of *Cdk1* in mouse C2 embryonic stem (ES) cells¹² (Extended Data Fig. 1a–c) and of *CDK1* in human H1¹³ (Extended Data Fig. 2a–f) and human CA1¹⁴ (Extended Data Fig. 3a–e) ES cells. We determined the optimal GCV concentration for the heterozygous *Cdk1-TK*-expressing mouse and heterozygous *CDK1-TK*-expressing human ES cells *in vitro* (Extended Data

Figs. 1d, 2h) as well as for controlling the growth of the teratomas that were generated using these lines. Human ES cells implanted in immunodeficient NOD/SCID/IL2R γ (NSG) mice and mouse ES cells implanted in isogenic C57BL/6N recipient mice resulted in teratoma formation with the expected efficiency (Extended Data Figs. 1e, 2g). At a volume of 500 mm³ (day 0), we administered GCV daily by intraperitoneal injection for up to four weeks. GCV rendered the C2 ES-cell-derived teratomas dormant, without growth rebound following the treatment (Fig. 1d and Extended Data Fig. 4a, b). H1-derived teratomas responded similarly although occasionally, repeated GCV treatments were required to stabilize the teratoma size (Fig. 1f and Extended Data Fig. 4d). The decrease in teratoma size after GCV readministration indicates regained proliferation of quiescent or slow-dividing cells following drug withdrawal. Consequently, *TK* is expressed in dividing cells and induces subsequent GCV sensitivity. The volume of the human teratomas frequently increased in the later phases; however, in agreement with previous reports^{15,16}, this was the result of cyst formation (Extended Data Fig. 4e) and not solid tissue growth.

The induced long-term dormancy of teratomas was encouraging, but was also unexpected given that such a large tissue (approximately 10⁹ cells¹⁷) could contain numerous cells that might be capable of escaping the suicide system through different types of mutations. Within the well-encapsulated teratoma, however, these presumably resistant cells (escapees) could have been eliminated by the bystander-killing effect of the *TK-GCV* system¹⁸.

To further investigate the capacity of the safe-cell system to control cell proliferation, we performed a breast cancer transplantation assay using safe-cell mammary epithelial tumour cells¹⁹. Upon isogenic transplantation, we observed that after a delayed period (approximately 100 days), heterozygous safe-cell tumours became resistant to GCV and they continued to grow in the presence of the drug (Extended Data Fig. 5a, b).

To identify escapees that appeared during the expansion of heterozygous safe-cell ES cells, we designed an *in vitro* experiment that mitigated the bystander-killing effect (Extended Data Fig. 6a, b) and characterized the mechanisms by which resistance occurred in eight independent clonal escapee lines obtained from 120 million cells. To determine whether the mechanism by which resistance occurred was caused by large genomic changes or *Cdk1-TK* locus-specific mutations, we analysed the copy number of *Cdk1*, the *TK* transgene and six endogenous genes that are found on chromosome 10 (Extended Data Fig. 6d, f). Only one escapee (E3 in Extended Data Fig. 6f) contained the *TK* gene (Extended Data Fig. 6c). We did not detect mutations in the coding region of either *Cdk1* or *TK* (data not shown); however, the expression level of this allele was reduced and rendered cells GCV-resistant (Extended Data Fig. 6e). Another escapee, E5, was the result of a regional deletion that included the *Cdk1-TK* locus, which led to a more than 18.5-Mb hemizygous region in the wild-type chromosome (Extended Data Fig. 6d, f). Diploidy (copy number of 2) in the ten genes

¹Lunenfeld-Tanenbaum Research Institute, Sinai Health System, Toronto, Ontario, Canada. ²Institute of Medical Science, University of Toronto, Toronto, Ontario, Canada. ³Department of Physiology, University of Toronto, Toronto, Ontario, Canada. ⁴School of Mathematics and Maxwell Institute, The University of Edinburgh, Edinburgh, UK. ⁵The Hospital for Sick Children Research Institute, Toronto, Ontario, Canada. ⁶Australian Regenerative Medicine Institute, Monash University, Melbourne, Victoria, Australia. ⁷Department of Obstetrics & Gynaecology, University of Toronto, Toronto, Ontario, Canada. ⁸These authors contributed equally: Qin Liang, Claudio Monetti. *e-mail: nagy@lunenfeld.ca

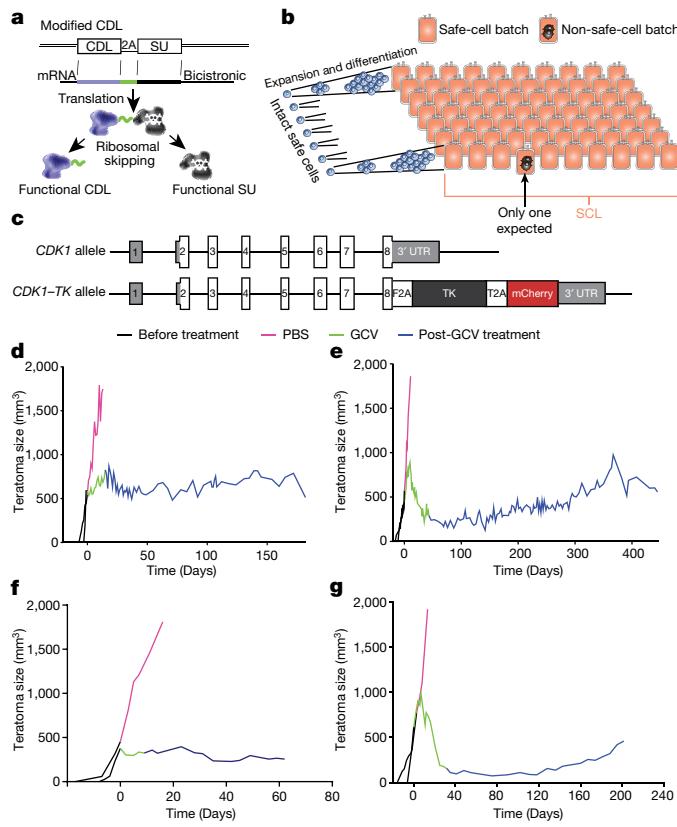


Fig. 1 | The concept, the definition, realization and properties of the safe-cell system. **a**, The suicide gene is placed into a cell-division essential locus (CDL), resulting in a bicistronic mRNA that is translated into two proteins; a cell division essential factor and a drug-inducible suicide factor. **b**, Visual representation of the SCL defined by one non-safe-cell batch out of many batches. **c**, The link between the prototype drug-inducible suicide system (HSV-TK) and the prototype CDL (*CDK1*). 3' UTR, 3' untranslated region. **d–g**, Representative growth of teratomas formed by mouse *Cdk1-TK/Cdk1* cells, when the recipient mice were treated with PBS or GCV. **e**, Representative growth of teratomas formed by human *CDK1-TK/CDK1* ES cells, when the recipient mice were treated with PBS or GCV. **f**, Representative growth of teratomas in mice formed by mouse *Cdk1-TK/Cdk1-TK* ES cells. **g**, Representative growth of teratomas in mice formed by human *CDK1-TK/CDK1-TK* ES cells. **d–g**, Experiments were repeated multiple times (see Extended Data Fig. 4) with similar results.

and the lack of TK in the remaining six resistant clones suggested that these were formed by diploid loss of heterozygosity (dLOH), which was probably caused by mitotic recombination or chromosomal non-disjunction, leading to homozygosity of the wild-type *Cdk1* allele in a diploid form. These data indicate that dLOH is the dominant mechanism by which the *Cdk1-TK* allele is lost in heterozygous ES cells, consistent with a study of mouse *Aprt^{+/−}* heterozygous cells, in which dLOH accounted for 78% of the loss of gene function events²⁰.

To mitigate the generation of escapees by dLOH, we established both mouse and human ES cell lines that were homozygous for the *Cdk1-TK* and *CDK1-TK* alleles, respectively (Extended Data Figs. 1a–c, 2a–h, 3a–e). As expected, we were unable to identify any escapees (Extended Data Fig. 6b). Homozygous ES-cell-derived teratomas behaved similarly to teratomas derived from heterozygous ES cells; a brief GCV treatment was sufficient to render the teratomas dormant (Fig. 1f, g and Extended Data Fig. 4c, f). As in human heterozygous ES-cell-derived lines, cyst formation also occurred in human homozygous ES-cell-derived teratomas (Extended Data Fig. 4f).

To test the limits of our system, we generated homozygous safe-cell mouse mammary tumour cells and transplanted the cells into wild-type isogenic C57BL/6N recipient mice. We found that the size of tumours was reduced following transplantation and their growth was

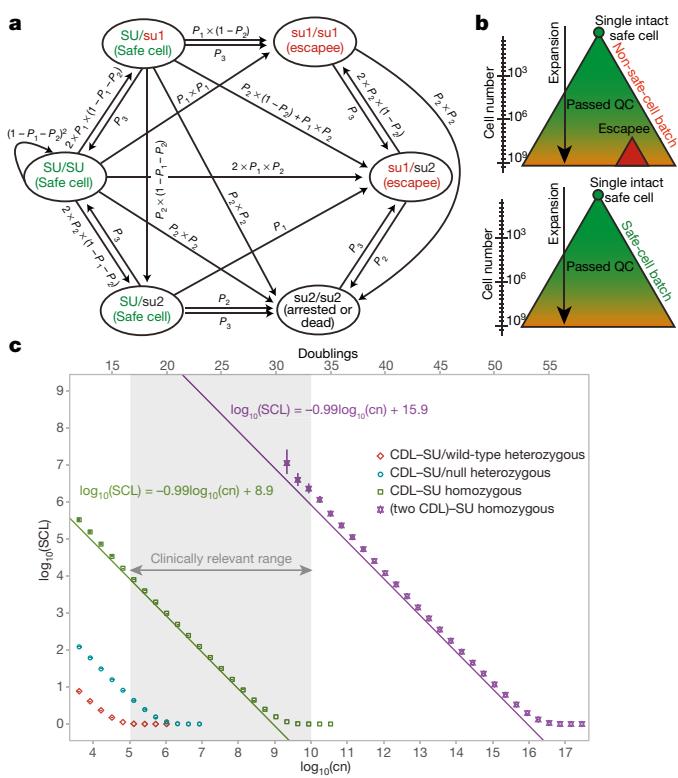


Fig. 2 | Modelling of the safe-cell system and calculating the SCL.

a, Genotype transition matrix considered in the modelling and Monte Carlo simulations. **b**, Visual illustration of safe-cell and non-safe-cell batch formation during cell expansion. QC, quality control. **c**, The function between therapeutic cell number and SCL was determined by Monte Carlo simulation (data points) for different initial cell genotypes. cn, cell number. Solid lines show the approximated linear regression on the ‘close to linear’ parts of the functions. Monte Carlo simulations were run n times for each genotype ($n = 80 \times 10^6$ for CDL-SU/wild-type; $n = 10^8$ for CDL-SU homozygous; $n = 10^7$ for CDL-SU/null heterozygous; $n = 13.5 \times 10^6 - 90.5 \times 10^6$ for (two CDL)-SU homozygous). The bars on certain data points represent the 95% confidence intervals of the SCL estimates.

restrained after GCV administration (Extended Data Fig. 5c). The observed growth rebound following GCV withdrawal was not surprising as slow-dividing and quiescent tumour-prone cells survive GCV administration and can start proliferating in the absence of the drug. Nevertheless, even in this non-clinical, extreme situation in which a tumour cell line is used for cell transplantation, the homozygous safe-cell system is capable of controlling tumour growth.

Because we were unable to identify any escapees that appeared in homozygous safe-cell ES cells, we used Monte Carlo simulation to estimate the odds of escapees in this scenario. The model considers three types of mutations that could potentially influence the function of the CDL-SU link (Extended Data Fig. 7a). Type-1 mutations (su1) render the suicide system non-functional while keeping the linked CDL operational. Type-2 mutations (su2) eliminate both the suicide system and CDL functionality through epigenetic or genetic changes to the entire locus, including a hemizygous LOH-dependent mechanism. Type-3 mutations (su3) remove a functional CDL-SU allele by dLOH.

To estimate the probabilities of type-1, -2 and -3 mutations (P_1 , P_2 and P_3 , respectively) per cell generation (Extended Data Fig. 7a), we designed our own experiment (Supplementary Table 4) and also used published data^{21–24}. We used the values $P_1 = P_2 = 10^{-6}$ and $P_3 = 2 \times 10^{-5}$ per cell per division, all of which are intentional overestimates. Consequently, our calculated SCLs represent underestimates, being equal to or lower than the actual SCL.

In silico, we subsequently generated a sufficient number of cell batches that were expanded from a single cell with an intact suicide

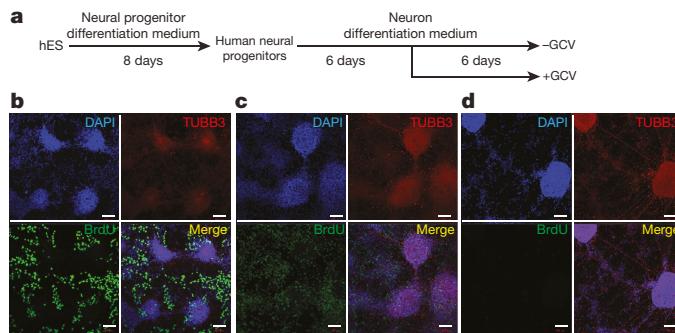


Fig. 3 | The safe-cell system eliminates only proliferating cells.

a, Schematic. b–d, Human *CDK1-TK/CDK1-TK* ES cells are differentiated into neural epithelial progenitors (b) and subsequently into neurons without (c) or with (d) GCV treatment. b, Neural epithelial progenitors consist of a mixed culture of proliferating neural epithelial progenitors and non-proliferating differentiated neurons that were treated with GCV. Scale bars, 100 μ m. This experiment was repeated three times with similar results.

system. During each doubling, the model permits allele transitions (Extended Data Fig. 7b) that determine the transition graph (Fig. 2a), reflecting the genotype change that could occur during cell expansion. For the homozygous *CDK1-TK/CDK1-TK* simulation, we initiated the batch production from an SU/SU cell with two intact suicide-system alleles, whereas for the heterozygous *CDK1-TK/CDK1^{WT}* simulation, the initiating cell was SU/su1, because the su1 allele is functionally equivalent to the *CDK1^{WT}* allele. For the compound heterozygous *CDK1-TK/CDK1^{null}* simulation, the initiating cell was SU/su2, as the su2 allele is the same as a *CDK1^{null}* allele (Fig. 2a).

A batch of cells is considered to be a safe-cell batch if it does not contain any escapees (Fig. 2b). On the basis of the frequency of getting a non-safe-cell batch of cells determined by the Monte Carlo simulation, we calculated the SCL as the function of the number of cells that is needed for a therapeutic cell batch. Figure 2c shows these functions for the different initiating cell genotypes described above.

The number of cells that is required for cell therapy is disease-specific and is estimated to range between 10^5 (for example, the eye^{25,26}) and 10^{10} (for example, the heart²⁷) cells. The genotype scenarios presented in Fig. 2c show that a single TK insertion (*CDK1-TK/CDK1^{WT}* or *CDK1-TK/CDK1^{null}*) gives a low SCL. By contrast, a homozygous TK insertion into *CDK1* significantly increases the SCL and brings the safety level into a clinically relevant range. However, for diseases that

require a larger number of therapeutic cells (10^8 to 10^{10} cells), the SCL provided even by the homozygous CDL-SU is insufficient (SCL < 10). For this disease category, we propose the use of a homozygous modification of two different CDLs. In this scenario, our Monte Carlo simulation showed a strong increase in SCL (SCL > 10^6 for all clinically relevant batch volumes, Fig. 2c).

We observed that the logarithm of SCL, as a function of the logarithm of the cell number, is very close to a linear function when the SCL is above 10 and the cell number is in the clinically relevant range. Therefore, for an estimation of SCL, we applied linear regression to these segments (Fig. 2c). Using these approximates, the calculation of SCLs becomes very simple, while retaining the desired underestimates (Fig. 2c): for one CDL, this equates to $SCL = 10^9/cn$, and for two CDLs this equates to $SCL = 10^{16}/cn$, where 'cn' is the number of cells needed for a therapeutic cell batch.

During the production of therapeutic cells, some are lost during differentiation or expansion. Therefore, the efficiency of cell production should be accurately estimated and the cell numbers that are needed to generate a therapeutic batch should be corrected accordingly.

In future cell therapies, if allogeneic cells are desired or HLA haplobanks²⁸ of pluripotent cell lines are available, generation of off-the-shelf cell batches would be advantageous. This would require the production of large pools of cells that, following quality control, would be aliquoted into therapeutic batches. We calculated the effect of aliquoting on SCL using both mathematical and Monte Carlo modelling approaches (equation (1), Supplementary Information and Extended Data Fig. 8a). Notably, aliquoting resulted in an approximately fivefold decrease in SCL using both approaches.

Quality control should be performed on every cell pool to ensure that the originating cell was a safe-cell and consequently, the SCL calculation is correct. To this end, we grew several batches from a single, homozygous safe-cell ES cell. At the early phase of expansion, we verified that both *CDK1-TK* alleles were expressed and intact using flow cytometry, allele-specific PCR (Extended Data Fig. 8b–e) and sequencing of the TK coding region.

Both mouse and human ES cells with homozygous modifications of *CDK1* have a normal ES cell morphology, self-renewing capacity and ability to differentiate (Extended Data Fig. 9). Additionally, using in vitro neural differentiation, we demonstrate the selective killing of dividing cells by the safe-cell system. Following a brief GCV treatment, all mitotically active cells were eliminated whereas non-dividing cells were spared (Fig. 3). This ability could represent a valuable safety measure before transplantation of cells into a patient.

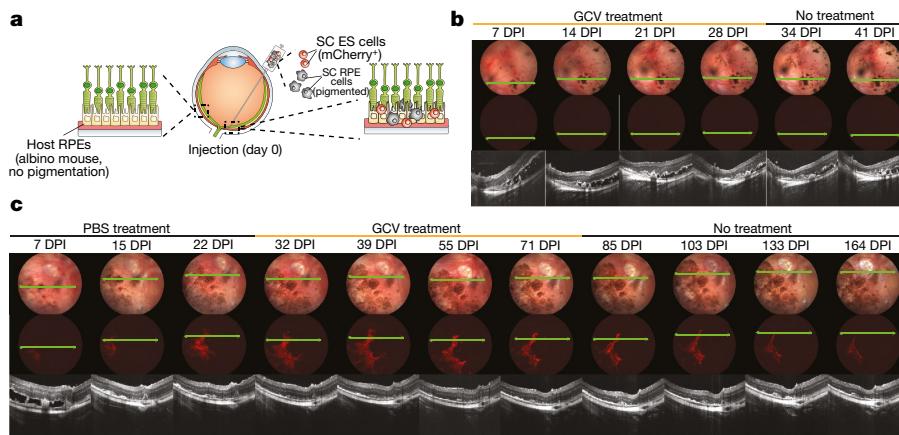


Fig. 4 | An in vivo proof of principle study shows the safe-cell system in action. a, A 3:1 mixture of human homozygous safe-cell ES-cell-derived RPE cells and homozygous safe-cell ES cells were subretinally injected into NSG mice and imaged using fundoscopy and optical coherence tomography throughout GCV or PBS treatment. b, Fundoscopy, optical coherence tomography and fluorescence imaging of the eyes of mice that received GCV treatment (4 weeks). c, Fundoscopy, optical coherence tomography and fluorescence imaging of the eyes of a mouse that received PBS treatment (3 weeks) and developed an actively growing ES-cell-derived lesion (mCherry⁺ cells). d.p.i., days post-injection.

b, c, Experiments were repeated multiple times with similar results (Extended Data Fig. 10).

received GCV treatment (4 weeks). c, Fundoscopy, optical coherence tomography and fluorescence imaging of the eyes of a mouse that received PBS treatment (3 weeks) and developed an actively growing ES-cell-derived lesion (mCherry⁺ cells). d.p.i., days post-injection.

To simulate a clinical cell transplantation scenario gone awry, we injected a 3:1 mixture of human homozygous safe-cell retinal pigment epithelium (RPE) cells and human homozygous safe-cell ES cells (that were tagged with mCherry) into the subretinal space in the eyes of NSG mice. Among four injected eyes, we did not observe any ES cell derivatives when GCV was administered 24 h post-injection and for 28 days, as a preventative measure (Fig. 4a, b and Extended Data Fig. 10a, b). However, cell growth was detectable in six eyes that received PBS as an initial treatment (Fig. 4c and Extended Data Fig. 10d–f). Notably, even when GCV administration was delayed three weeks post-injection and cell growth was present, the homozygous safe-cell system efficiently arrested the ES-cell-derived component of the graft; only non-dividing cells remained (Fig. 4c and Extended Data Fig. 10e, f). This experiment illustrates the ability of the safe-cell system to selectively eliminate proliferating cells after cell transplantation. Neither the initial nor the delayed GCV treatment affected the RPE graft or the integrity of the surrounding retinal tissue (Fig. 4b, c and Extended Data Fig. 10).

No therapy is without risk. Our safe-cell system (concept and genome-editing approach) provides a definition of risk and a quantification of the safety level as a function of the number of cells that is needed for any given cell therapy. We contend that the risks associated with the safe-cell system are sufficiently low to provide an indispensable component of prospective cell therapies. Our approach to assessing and quantifying the safety of cell-based therapies will be critical for informed decision-making by regulators, clinicians and patients while advancing modern medicine.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at <https://doi.org/10.1038/s41586-018-0733-7>.

Received: 13 October 2017; Accepted: 27 September 2018;

Published online: 14 November 2018

- Li, W. & Xiang, A. P. Safeguarding clinical translation of pluripotent stem cells with suicide genes. *Organogenesis* **9**, 34–39 (2013).
- Dobie, K., Mehtali, M., McClenaghan, M. & Lathe, R. Variegated gene expression in mice. *Trends Genet.* **13**, 127–130 (1997).
- Yagyu, S., Hoyos, V., Del Bufalo, F. & Brenner, M. K. Multiple mechanisms determine the sensitivity of human-induced pluripotent stem cells to the inducible caspase-9 safety switch. *Mol. Ther. Methods Clin. Dev.* **3**, 16003–16008 (2016).
- Kotini, A. G., de Stanchina, E., Themeli, M., Sadelain, M. & Papapetrou, E. P. Escape mutations, ganciclovir resistance, and teratoma formation in human iPSCs expressing an HSVtk suicide gene. *Mol. Ther. Nucleic Acids* **5**, e284 (2016).
- Preuß, E. et al. TK.007: A novel, codon-optimized HSVtk(A168H) mutant for suicide gene therapy. *Hum. Gene Ther.* **21**, 929–941 (2010).
- Diril, M. K. et al. Cyclin-dependent kinase 1 (Cdk1) is essential for cell division and suppression of DNA re-replication but not for liver regeneration. *Proc. Natl Acad. Sci. USA* **109**, 3826–3831 (2012).
- Santamaría, D. et al. Cdk1 is sufficient to drive the mammalian cell cycle. *Nature* **448**, 811–815 (2007).
- Satyaranayana, A. et al. Genetic substitution of Cdk1 by Cdk2 leads to embryonic lethality and loss of meiotic function of Cdk2. *Development* **135**, 3389–3400 (2008).
- Fillat, C., Carrió, M., Cascante, A. & Sangro, B. Suicide gene therapy mediated by the herpes simplex virus thymidine kinase gene/ganciclovir system: fifteen years of application. *Curr. Gene Ther.* **3**, 13–26 (2003).
- Tomicic, M. T., Thust, R. & Kaina, B. Ganciclovir-induced apoptosis in HSV-1 thymidine kinase expressing cells: critical role of DNA breaks, Bcl-2 decline and caspase-9 activation. *Oncogene* **21**, 2141–2153 (2002).
- Berger, C., Flowers, M. E., Warren, E. H. & Riddell, S. R. Analysis of transgene-specific immune responses that limit the *in vivo* persistence of adoptively transferred HSV-TK-modified donor T cells after allogeneic hematopoietic cell transplantation. *Blood* **107**, 2294–2302 (2006).
- Gertsenstein, M. et al. Efficient generation of germ line transmitting chimeras from C57BL/6N ES cells by aggregation with outbred host embryos. *PLoS ONE* **5**, e11260 (2010).
- Thomson, J. A. et al. Embryonic stem cell lines derived from human blastocysts. *Science* **282**, 1145–1147 (1998).
- The International Stem Cell Initiative. Characterization of human embryonic stem cell lines by the International Stem Cell Initiative. *Nat. Biotechnol.* **25**, 803–816 (2007).

- Damjanov, I. & Andrews, P. W. Teratomas produced from human pluripotent stem cells xenografted into immunodeficient mice - a histopathology atlas. *Int. J. Dev. Biol.* **60**, 337–419 (2016).
- Wu, S. M. & Hochedlinger, K. Harnessing the potential of induced pluripotent stem cells for regenerative medicine. *Nat. Cell Biol.* **13**, 497–505 (2011).
- Bianconi, E. et al. An estimation of the number of cells in the human body. *Ann. Hum. Biol.* **40**, 463–471 (2013).
- Mesnil, M. & Yamasaki, H. Bystander effect in herpes simplex virus-thymidine kinase/ganciclovir cancer gene therapy: role of gap-junctional intercellular communication. *Cancer Res.* **60**, 3989–3999 (2000).
- Guy, C. T., Cardiff, R. D. & Muller, W. J. Induction of mammary tumors by expression of polyomavirus middle T oncogene: a transgenic mouse model for metastatic disease. *Mol. Cell. Biol.* **12**, 954–961 (1992).
- Cervantes, R. B., Stringer, J. R., Shao, C., Tischfield, J. A. & Stambrook, P. J. Embryonic stem cells and somatic cells differ in mutation frequency and type. *Proc. Natl Acad. Sci. USA* **99**, 3586–3590 (2002).
- Larson, J. S., Yin, M., Fischer, J. M., Stringer, S. L. & Stringer, J. R. Expression and loss of alleles in cultured mouse embryonic fibroblasts and stem cells carrying allelic fluorescent protein genes. *CBM Mol. Biol.* **7**, 36 (2006).
- Mortensen, R. M., Conner, D. A., Chao, S., Geisterer-Lowrance, A. A. & Seidman, J. G. Production of homozygous mutant ES cells with a single targeting construct. *Mol. Cell. Biol.* **12**, 2391–2395 (1992).
- Koike, H. et al. Efficient biallelic mutagenesis with Cre/loxP-mediated inter-chromosomal recombination. *EMBO Rep.* **3**, 433–437 (2002).
- Yusa, K. et al. Genome-wide phenotype analysis in ES cells by regulated disruption of Bloom's syndrome gene. *Nature* **429**, 896–899 (2004).
- Schwartz, S. D. et al. Human embryonic stem cell-derived retinal pigment epithelium in patients with age-related macular degeneration and Stargardt's macular dystrophy: follow-up of two open-label phase 1/2 studies. *Lancet* **385**, 509–516 (2015).
- Mandal, M. et al. Autologous induced stem-cell-derived retinal cells for macular degeneration. *N. Engl. J. Med.* **376**, 1038–1046 (2017).
- Chong, J. J. H. et al. Human embryonic-stem-cell-derived cardiomyocytes regenerate non-human primate hearts. *Nature* **510**, 273–277 (2014).
- Barry, J., Hyllner, J., Stacey, G., Taylor, C. J. & Turner, M. Setting up a haplobank: issues and solutions. *Curr. Stem Cell Rep.* **1**, 110–117 (2015).

Acknowledgements We thank the TCP Transgenic Core and M. Gertsenstein for mouse line derivation; A. Bang for flow cytometry; the TCP Pathology Core and K. Harpal for histology analysis; M. Kownacka for providing MEFs; M. S. Shoichet for HAMC; N. Mitrousis for qPCR primers; S. Nurk for advice on Monte Carlo simulation; I. P. Michael, P. D. Tonge, B. V. Varga, C. He and R. El-Rass for experimental advice and J. S. Harding and K. C. Davidson for their proofreading of the manuscript; to J. S. Harding for the artwork in Fig. 4. This work was supported by CIHR Foundation Grant, Foundation Fighting Blindness, Canadian Research Chair and Medicine by Design (University of Toronto) funding to A.N.

Reviewer information *Nature* thanks A. Hewitt, T. J. Kieffer and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions Q.L. designed and conducted most of the experiments, including targeting of mouse and human ES cells, teratoma analysis, mammary gland tumour analysis, neural *in vitro* differentiation, mutation rate calculation, as well as analysing the data and writing the manuscript. C.M. designed and conducted experiments, designed and constructed the CDK1-TK targeting cassette, targeted mouse and human ES cells, analysed safe-cell escapees, performed *in vitro* differentiation assays, analysed data, composed figures and wrote the manuscript. M.V.S. conducted the Monte Carlo simulations and analysed the data. I.G. performed the mathematical modelling. E.J.N. targeted human ES cells, differentiated RPEs and wrote the manuscript. S.H. performed the eye experiment and analysed the data. H.Y. performed *in vitro* differentiation into endoderm and analysed the data. C.K. performed *in vitro* differentiation into mesenchymal stem cells and analysed the data. P.Z. performed Southern blots. C.L. performed animal teratoma experiments. K.N. performed animal teratoma experiments, analysed the data and edited the manuscript. M.M. targeted human ES cells. H.-K.S. analysed teratoma histology. A.N. conceived and supervised the study, designed experiments, analysed the data and wrote the manuscript.

Competing interests A.N., C.M. and Q.L. are inventors on a patent application covering the SC technology (PCT/CA2016/050256). A.N. is a co-founder and shareholder of panCELLa Inc. C.M. is a senior scientist at panCELLa Inc. The other authors declare no competing interests.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41586-018-0733-7>.

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-018-0733-7>.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

Correspondence and requests for materials should be addressed to A.N.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

Generation of targeting vectors. Targeting vectors were generated by DNA synthesis, molecular cloning, recombineering and the NEBuilder HiFi DNA Assembly Cloning Kit (New England Biolabs).

Generation of CRISPR-Cas9 vectors. pX330-U6-Chimeric_BB-CBh-hSpCas9 was a gift from F. Zhang (Addgene plasmid 42230)²⁹. Guide sequences for CRISPR-Cas9 were analysed using the online CRISPR design tool (<http://crispr.mit.edu>). Guide sequence for mouse *Cdk1* targeting: TAAGAAGATGTAGCCCTC. Guide sequence for human *CDK1* targeting: CTATCTGTTGACATAACATA.

Mouse ES cell culture. C57BL/6N C2 ES cells were grown at 37 °C in 95% air, 5% CO₂ on mouse embryonic fibroblasts (MEFs) obtained from TgN(DR1)1Jae/J mice (The Jackson Laboratory, 003208) at all times except for one passage on gelatinized tissue-culture plates before aggregation³⁰. Two types of medium were used. The first medium, FBS-DMEM ES cell medium, was used for gene targeting. This medium consisted of high-glucose DMEM supplemented with 15% FBS (previously shown to support germline chimera generation), 2 mM GlutaMAX, 1 mM Na pyruvate, 0.1 mM non-essential amino acids (NEAA), 50 U ml⁻¹ penicillin-streptomycin (all Thermo Fisher Scientific), 0.1 mM 2-mercaptoethanol (Sigma-Aldrich) and 1,000 U ml⁻¹ LIF prepared with LIF-producing plasmid³¹. 2). The second medium, KSR+2i medium was used for 2–4 passages before the generation of ES cell chimaeras¹². KSR+2i medium consisted of high-glucose DMEM medium supplemented with 15% knockout serum replacement (KSR) (Thermo Fisher Scientific), 1 mM Na pyruvate, 0.1 mM NEAA, 0.1 mM 2-mercaptoethanol, 2 mM GlutaMAX, 50 U ml⁻¹ penicillin-streptomycin, 500 U ml⁻¹ LIF, 5 mg ml⁻¹ insulin (Thermo Fisher Scientific), 1 μM of the mitogen-activated protein kinase inhibitor PD0325901 (StemGent) and 3 μM of the glycogen synthase kinase-3 inhibitor CHIR99021 (StemGent). Cells were fed daily and passaged when they reached a confluence of 70–80%. Then, 0.05% trypsin-EDTA (Thermo Fisher Scientific) was used for the passaging of cells grown in FBS-DMEM and accutase (STEMCELL Technologies) was used for cells grown in KSR+2i medium. Cells were tested (negative) for mycoplasma contamination but were not authenticated.

Mouse ES cell targeting. In brief, 50,000 mouse C57BL/6N C2 ES cells were transfected with 2 μg DNA (mouse target vector I or II, 1.5 μg; CRISPR vector, 0.5 μg) using JetPrime for transfection (Polyplus). The cells were selected for G418 resistance (160 μg ml⁻¹) starting 48 h after transfection. Resistant clones were picked independently and replicated in 96-well plates for freezing and genotyping using PCR. PCR-positive clones were expanded, frozen in multiple vials and genotyped by Southern blotting.

Selection cassette excision in mouse ES cells. Correctly targeted ES cell clones were transfected with episomal-hyPBase (for mouse target vector I) or pCAGGS-NLS-Cre-IRES-puromycin (for mouse target vector II). Then, 2–3 days after transfection, cells were trypsinized and plated clonally (1,000–2,000 cells per 10-cm plate). mCherry⁺ clones were picked and transferred to 96-well plates into independent wells and genotyped by PCR and Southern blotting to confirm the excision event. The junctions of the removal region were PCR-amplified, sequenced and confirmed to be intact and without any frameshift mutations. GCV (Sigma-Aldrich) to test for TK activity was used at a final concentration of 1 μM.

Human ES cell culture. Human CA1 and H1 ES cells were cultured on Geltrex (Thermo Fisher Scientific) using mTeSR1 medium (STEMCELL Technologies) containing 50 U ml⁻¹ penicillin-streptomycin (Thermo Fisher Scientific). Cells were passaged using TrypLE Express (Thermo Fisher Scientific) and were subsequently plated in mTeSR medium containing 10 μM ROCK inhibitor (Selleckchem) for 24 h. Cells were tested (negative) for mycoplasma contamination but were not authenticated.

Human ES cell targeting. For human targeting vectors I and II, six million human ES cells were electroporated using a Neon Transfection System (Thermo Fisher Scientific) with protocol 14 (pulse voltage, 1,200 mV; pulse width, 20 ms; pulse number, 2) with 24 μg DNA (target vector, 18 μg; CRISPR vector, 6 μg). After transfection, cells were plated on four 10-cm plates. G418 selection at 30 μg ml⁻¹ or puromycin selection at 0.75 μg ml⁻¹ was initiated 48 h after transfection. Independent colonies were picked to 96-well plates and each plate was duplicated for further growth and genotyping with PCR. PCR-positive clones were expanded, frozen in multiple vials and genotyped with Southern blotting. For human target vector III targeting, 10 million human ES cells were electroporated using Neon protocol 14 with 40 μg DNA (human target vector III, 30 μg; CRISPR vector, 10 μg) and plated in four 10-cm plates. After 3–4 days of transfection, cells that were double-positive for mCherry and eGFP were sorted into one well of a 96-well plate. After recovery from fluorescence-activated cell sorting (FACS), cells were dissociated and plated clonally (1,000–2,000 cells per 10-cm plate). Next, clones were picked independently, replicated and transferred to 96-well plates for freezing and genotyping with PCR. PCR-positive clones were expanded, frozen in multiple vials and genotyped by Southern blotting.

Selection cassette excision in human ES cells. In brief, one million correctly targeted ES cell clones were electroporated with 2 μg episomal-hyPBase-IRES-puro

(for human target vector I) or 2 μg episomal-Cre-IRES-puro (for human target vector II) using Neon protocol 14. Once the cells were confluent in six-well plates, mCherry⁺ cells were sorted into one well of a 96-well plate by FACS. After recovery, cells were dissociated and plated clonally (1,000–2,000 cells per 10-cm plate). Clones were picked and transferred to 96-well plates into independent wells and genotyped by PCR and Southern blotting to confirm the excision event. The junctions of the removal region were PCR-amplified, sequenced and confirmed to be intact and without frameshift mutations.

PCR genotyping. For all PCR reactions, 2× Taq PCR master mix (Biomart) was used. Genomic DNA from human cell pellets was extracted using the DNeasy Blood & Tissue Kit (Qiagen). The primer pairs and conditions used for each reaction are listed in Supplementary Table 2.

Southern blotting. In brief, 10 μg of genomic DNA was extracted from PCR-positive clones, digested with ScaI-HF overnight, resolved by 0.6–0.7% gel electrophoresis, and transferred to Hybond N+ (GE Healthcare). The following probes were labelled with ³²P and used to hybridize with the membrane (around 25 ng probe per ml hybridization solution). Human *CDK1* genomic probe: PCR-amplified with primers h*CDK1*-Probe6-F and h*CDK1*-Probe6-R. Mouse *Cdk1* genomic probe: PCR-amplified with primers 647302FWD and 647302REV. mCherry probe: the entire length of mCherry. eGFP probe: the entire length of eGFP. TK-mCherry probe: cut from h*CDK1*-PB-neo-TK-mCherry with Bsu36I and SgrAI: 1,092 bp, gel-purified.

Mice. The CD-1 (ICR) (Charles River) outbred albino mouse stock was used as embryo donors for aggregation with ES cells and as pseudopregnant recipients. Six-to-ten-week-old C57BL/6NCrl mice (Charles River) were used as the host for teratoma assays with mouse C2 ES cells. Six-to-ten-week-old C57BL/6NCrl or B6N-Tyr_c N4/Crl#493 (Charles River) mice were used as the host for mammary fat pad transplantation of mammary epithelial cells. Six-to-ten-week-old NSG/J#5557 mice (Jackson Laboratories) were used as the host for teratoma assays with human H1 or CA1 ES cells. FVB/N-Tg(MMTV-PyVT)634Mul/J mice were a gift from members of W. Muller's laboratory¹⁹, and the backcross to B6J background was done by members of A. Pawson's laboratory. Animals were maintained on a 12-h light/dark cycle and provided with food and water ad libitum in individually ventilated units (Techniplast) in the specific-pathogen free facility at The Centre for Phenogenomics (TCP). All procedures involving animals were performed in compliance with the Animals for Research Act of Ontario and the Guidelines of the Canadian Council on Animal Care. Animal protocols performed in this study were approved by the Toronto Centre for Phenogenomics Animal Care Committee, chaired by A. Jurisicova and L. Phaneuf. The number of animals used in different experiments was determined in accordance with similar studies in the field; owing to the nature of most experiments, blinding was impossible, because the results are visible at the time of analysis. Animals were allocated randomly when possible.

Generation of chimaeras and mouse lines. Morula aggregations were performed as previously described³⁰. Chimaeras were identified at birth by the presence of black eyes and later by coat pigmentation. Male chimaeras with more than 50% ES cell contribution to coat colour were bred with CD-1 females to identify germline transmitters. The transmitter was then bred with C57BL/6NCrl females and pups were confirmed by genotyping to obtain *Cdk1-TK/Cdk1* mice. *Cdk1-TK/Cdk1* MMTV-PyMT males were generated by breeding MMTV-PyMT (B6) males and *Cdk1-TK/Cdk1* females. *Cdk1-TK/Cdk1* MMTV-PyMT and *Cdk1-TK/Cdk1-TK* MMTV-PyMT female mice were generated by breeding *Cdk1-TK/Cdk1* MMTV-PyMT males and *Cdk1-TK/Cdk1* females.

Teratoma assay. Matrigel Matrix High Concentration (Corning) was diluted 1:3 with cold DMEM medium on ice. Then, 1–5 million mouse ES cells or 5–10 million human ES cells were suspended in 100 μl of Matrigel-DMEM and injected subcutaneously into one or both dorsal flanks of C57BL/6NCrl mice (for mouse C2 ES cells) and NSG/J#5557 mice (for human H1 and CA1 ES cells). Teratomas formed 2–4 weeks after injection. Teratoma size was measured using callipers and volume was calculated using the formula $V = (L \times W \times H)/\pi/6$. GCV or PBS treatment was performed through daily intraperitoneal injections (50 mg kg⁻¹) with varying treatment durations. At the end of treatment, mice were euthanized and tumours were dissected and fixed in 4% paraformaldehyde for histological analysis.

Breast cancer transplantation assay. *Cdk1-TK/Cdk1* MMTV-PyMT and *Cdk1-TK/Cdk1-TK* MMTV-PyMT female mice developed mammary gland tumours between three and six months of age. Mammary epithelial tumorigenic cells were isolated from developed tumours by digestion in 10× collagenase-hyaluronidase (STEMCELL Technologies), and dilution to 1× with medium consisting of DMEM/F12 (Thermo Fisher Scientific), 10% FBS and 50 U ml⁻¹ penicillin-streptomycin for 1 h in 37 °C. The digested cells were washed and pelleted with DMEM/F12 and 10% FBS four times, and plated in CnT-PRIME epithelium culture medium (CELLnTEC Advanced Cell Systems) on plates coated with 0.1% gelatin (Sigma-Aldrich). Without passaging, primary mammary epithelial cells were dissociated and resuspended in PBS at 10,000 cells per μl, and 50 μl (500,000 cells)

was transplanted into each mouse by intraductal injection after making a small abdominal skin incision as previously described³². Tumour measurement and PBS or GCV treatment were the same as described in the teratoma assay.

Differentiation of human ES cells into RPE cells. RPE differentiation was performed as previously described³³ with minor changes. Human ES cells were plated on Geltrex-coated six-well plates and cultured in feeder-free conditions with mTeSR medium until confluence was reached and the colonies lost their tight borders (7–10 days). Next, the medium was replaced with differentiation medium (basal media with 13% KSR) and changed every 2–3 days. The basal medium consisted of KO-DMEM supplemented with 50 U ml⁻¹ penicillin-streptomycin, 1 mM Na pyruvate, 0.1 mM NEAA, 2 mM GlutaMAX and 0.1 mM 2-mercaptoethanol. Initial pigmentation was observed approximately three weeks after the switch to differentiation medium. Clusters of RPE cells were manually picked and transferred to a Geltrex-coated 24-well plate (three clusters per well) when they were large enough (around 1 mm in diameter) for enrichment and the medium was changed to RPE medium, which consisted of basal medium with 5% FBS, 7% KSR and 10 ng ml⁻¹ bFGF (Peprotech).

Differentiation of human ES cells into definitive endoderm. Definitive endoderm differentiation was performed using the STEMdiff Trilineage Differentiation Kit (StemCell Technologies) and characterized by immunostaining for SOX17 and FOXA2 (Supplementary Table 3).

Differentiation of human ES cells into pharyngeal pouch endoderm. Differentiation into pharyngeal pouch endoderm was performed as previously described³⁴ with the only modification being that the induction from ES cells to definitive endoderm is one day shorter than reported.

Differentiation of human ES cells into mesenchymal stem cells and subsequent adipogenic, osteogenic and chondrogenic differentiation. ES cells were cultured in mTeSR medium for two days. Next, cells were induced into early mesoderm progenitor cells with STEMdiff Mesenchymal Induction Medium (STEMCELL Technologies) for four days and then maintained in MesenCult-ACF Medium (STEMCELL Technologies). Cells were continually passaged into six-well plates precoated with MesenCult-ACF Attachment Substrate (STEMCELL Technologies) to derive early mesenchymal progenitor cells. At day 21, the mesenchymal stem cells (MSCs) showed a fibroblast-like morphology and the culture medium was changed every three days. For adipogenic differentiation, MSCs at a density of 20,000 cells per well were plated with MesenCult-ACF Attachment Substrate and cultured with MesenCult-ACF Medium for two days. Adipogenesis was induced using the StemPro Adipogenesis Differentiation Kit (Thermo Fisher Scientific). After 21 days, lipid droplets were visualized using Oil Red O (Sigma). For osteogenic differentiation, ES-derived MSCs at a density of 50,000 cells per well were plated with MesenCult-ACF Attachment Substrate and cultured with MesenCult-ACF Medium for two days. Osteogenesis was induced using the StemPro Osteogenesis Differentiation Kit (Thermo Fisher Scientific). After 21 days, calcium deposition was visualized using Alizarin Red (Sigma-Aldrich). For induction of chondrogenic differentiation, ES-derived MSCs were centrifuged in 15-ml conical tubes at 500g for 5 min to create cell pellets with 5,000,000 cells per pellet. Chondrogenesis was induced using the StemPro Chondrogenesis Differentiation Kit (Thermo Fisher Scientific). After 21 days, cartilage was visualized using Alcian Blue (Sigma-Aldrich). Differentiation medium was changed every three days.

Differentiation of human ES cells into beating cardiomyocytes. Cardiomyocyte differentiation was performed using the STEMdiff Cardiomyocyte Differentiation Kit (STEMCELL Technologies).

Differentiation of human ES cells into neuronal progenitors and neurons. To differentiate human ES cells into neuronal progenitors, human ES cells were plated at 50–100,000 cells per cm² in 1:1 DMEM/F12: Neurobasal (Thermo Fisher Scientific), 0.5× N2 supplement (made in-house, 1.92 mg ml⁻¹ putrescine, 2.376 µg ml⁻¹ progesterone, 3.6 µM selenium, 10 mg ml⁻¹ apo-transferrin, 0.75% BSA, 20 g ml⁻¹ insulin), 0.5× B27 supplement with vitamin A (Thermo Fisher Scientific), 2 mM GlutaMAX, 0.1 mM β-mercaptoethanol, 50 U ml⁻¹ penicillin-streptomycin, 10 µM SB431542 (Selleckchem), 100 nM LDN193189 (Selleckchem) (10 µM ROCK inhibitor was added overnight only and was subsequently removed). Cells were maintained in this condition for eight days and the medium was changed every other day. Next, neuronal progenitors were dissociated with accutase and plated at a density of 5 × 10⁴ cells per cm² on laminin (Sigma-Aldrich, 1 µl for 1 cm², diluted in 250 µl PBS without Ca and Mg) in fast neuron differentiation medium, 1:1 DMEM-F12:Neurobasal, 1× B27 supplement with vitamin A, 5 µM DAPT (Selleckchem), 2 mM GlutaMAX, 0.1 mM β-mercaptoethanol, 50 U ml⁻¹ penicillin-streptomycin. Medium was changed every three days. Then, 10 µM GCV was added six days after neuron differentiation and kept for six days. After five days of GCV treatment, 10 µM BrdU (Sigma-Aldrich) was added and the cultures were fixed after six days of GCV treatment and were immunostained for BrdU and β-tubulin III (Supplementary Table 3).

Flow cytometry analysis and FACS. Flow cytometry and FACS experiments were both performed and analysed by the Lunenfeld-Tanenbaum Research Institute

flow cytometry facility. FACS was performed using the ASTRIOS EQ cell sorter. Flow cytometry was performed using the GALLIOS flow cytometer and evaluated using Kaluza Analysis Software (Beckman Coulter). Samples were gated for live single cells using forward scatter, side scatter and DAPI staining. Wild-type and single-colour samples of the same cell type as the experimental samples were used as negative controls and compensation calculations. Human ES cell samples were single-cell sorted using StemFlex Medium (Thermo Fisher Scientific) and 10 µM ROCK inhibitor.

Immunostaining. Cells fixed in 4% PFA were blocked and permeabilized with 5% goat serum, 1 M glycine and 1% Triton X-100 (all Sigma-Aldrich) in PBS without Ca and Mg or animal-free blocker (Vector Laboratories) with 1% Triton X-100 in milliQ water. All of the primary antibody information can be found in Supplementary Table 3. Staining was visualized using a Zeiss LSM780 confocal microscope.

Histology analysis. Paraffin embedding, paraffin block sectioning, and haematoxylin and eosin staining were performed by the Pathology Core of The Centre for Phenogenomics.

qPCR. Gene expression analyses were completed as follows: RNA extraction by GenElute Mammalian Total RNA Miniprep Kit (Sigma), reverse transcription with QuantiTect Reverse Transcription Kit (Qiagen), qPCR with SensiFAST SYBR No-Rox Kit (Bioline) on Bio-Rad CFX Real-Time Systems (Bio-Rad) and analysis with Bio-Rad CFX Manager 3.1. All information regarding primers and probes for the TaqMan qPCR analyses can be found in Supplementary Table 2. For copy number analysis, the reactions were performed using TaqMan Genotyping Master Mix (Thermo Fisher Scientific) and CFX Real-Time Systems, and were analysed by CopyCaller Software v.2.1 (Thermo Fisher Scientific).

Luria and Delbrück assay. The Luria and Delbrück assay was performed as previously described³⁵. CDK1-TK homozygous 3C cells were single-cell plated in a 96-well plate using FACS. Subsequently, 21 single-cell-derived cultures were grown to an average of five million cells per culture, and the numbers of single-positive cells in each culture were analysed using flow cytometry. The mutation rate was calculated using the previously described equation³⁶ available at <https://www.wolframalpha.com/>.

In vivo transplantation of RPEs. CDK1-TK homozygous 3C ES cells were transfected with the PB-CAGGs-mCherry-pA plasmid and sorted for highly expressing cells. Then, 40,000 3C-derived RPEs only or 30,000 3C-RPEs and 10,000 mCherry-tagged 3C ES cells were injected subretinally with 0.5%/0.5% (wt/vol) hydrogel blend of hyaluronan and methylcellulose (HAMC) in HBSS. PBS or GCV (50 mg kg⁻¹) treatments were started the day after cell injection or as stated in the Figures, and were given every other day through intraperitoneal injections. Monitoring by fundoscopy and optical coherence tomography was performed on the day after transplantation and then once a week.

SCL calculations. To establish the rate of mutation in the CDK1-TK allele, we used our targeted human H1 CDK1-TK-mCherry/CDK1-TK-eGFP dichromatic cell line, as most mutations in either CDK1-TK allele result in monochromatic cells. We grew 21 parallel cultures from a single dichromatic cell to an average of 5 × 10⁶ cells per culture (>22 consecutive doublings) and determined the number of monochromatic cells in the culture using flow cytometry. Next, we applied Luria-Delbrück fluctuation analysis^{35,36} to calculate the sum of the $P_1 + P_2 + P_3$ probabilities in the two CDK1-TK alleles. We found that the mutation rate of losing mCherry was 9.05 × 10⁻⁶ per cell per division while the mutation rate of losing eGFP was similar at 7.68 × 10⁻⁶ per cell per division (Extended Data Fig. 8a, b). To further validate the probabilities of these various mutations, we also analysed published studies that focused on these events. In mouse ES cells, the mutation rate ($P_1 + P_2 + P_3$) of changing from a dichromatic to a monochromatic phenotype in the Rosa26 locus (mouse chromosome 6) was 1 × 10⁻⁵ per cell, per division²¹. Similarly, another study calculated the mutation rate of gene function loss in the Gdf9 locus (mouse chromosome 11) to be 2.3 × 10⁻⁵ events per cell per division³⁷. Furthermore, the probability of the type 3 mutation, P_3 alone, has been calculated as 1 × 10⁻⁵, 7.2 × 10⁻⁶ and 8.5 × 10⁻⁶ in three different studies^{22–24} by performing high-G418 selection in mouse ES cells. The $P_1 + P_2$ mutation rate has also been estimated in the human HPRT locus on the X chromosome to be 1.7–6 × 10⁻⁷ by Luria-Delbrück fluctuation analysis³⁸, and 5 × 10⁻⁶ through mutation frequency analysis in population datasets³⁹. Next, we performed Monte Carlo simulations to establish the SCL of cell batches derived from different SU genotypes. On the basis of both published data and our own, we used the values $P_1 = P_2 = 10^{-6}$ and $P_3 = 2 \times 10^{-5}$ per cell per division; all of which are intentional overestimates. An ES cell population was considered to be a mix of mutant and non-mutant cells with reference to the CDL-SU locus (or loci). All possible mutations were categorized into three different types: type 1, when only the SU part of the locus becomes non-functional (su1 allele); type 2, when both the CDL and SU become non-functional (su2 allele); type 3, when any of the above occurs as a result of LOH (Fig. 2a, b). Back mutations, such as su1 to SU, su2 to SU or su2 to su1, were not considered, because of their extremely low probabilities (Fig. 2c). Back mutations,

such as SU/su1 to SU/SU, SU/su2 to SU/SU and su1/su2 to su1/su1, were considered as a part of the more frequent LOH process. P_1 , P_2 and P_3 were designated the probabilities of each mutation type, respectively. We distinguished between two types of P_3 : $P_{3\text{mr}}$ (probability LOH occurred through mitotic recombination), where both daughter cells survive; and $P_{3\text{cnd}}$ (probability LOH occurred through chromosomal non-disjunction), where one of the daughter cells with the single remaining copy of the chromosome is likely to die⁴⁰. On the basis of these probabilities, matrices of transitions between all possible genotypes within one or two CDL systems were constructed (Fig. 2c, <https://github.com/mashutova/failsafe>). With each division cycle (d), all cells within the population except cells with the su2/su2 genotype, were allowed to divide. Genotypes with su1/su2 and su1/su1 were considered escapees and the simulation initiated from one non-mutant cell. $n(g_1, d)$ was the number of cells of genotype g_1 at doubling d , and $P(g_1, g_2)$ was the probability of transition from genotype g_1 to g_2 . In each doubling, the number of cells changing genotype from g_1 to g_2 was determined through random sampling from a binomial distribution with parameters $2n(g_1, d - 1)$ and $P(g_1, g_2)$. We used a Poisson approximation of a binomial distribution to work with ultra-low $P(g_1, g_2)$ values. For each division, the number of cells of each genotype was assessed, and the simulation proceeded until the first escapee was detected. For each starting genotype, we performed more than 10 million simulations and obtained a distribution of the number of doublings (d) from the detection of the first escapee. On the basis of these data, we generated a function of SCL (overall number of trials divided by number of trials with escapees) over cell population size (2^d) (Fig. 2d). Because all graphs contain almost linear regions, we used linear models to extrapolate them to high SCL values. To obtain linear regression lines, we used only simulated points from the linear-like part of the graph ($R^2 > 0.999$) with 95% confidence intervals less than 1,000. To obtain a conservative boundary for the SCL, we used only the lowest confidence interval values to build linear regressions. To analyse the outcome from the aliquoting of the pool of safe-cell cells possibly containing escapees, using probability modelling we developed the following formula to calculate the drop of SCL (for details see equation (1))

$$\frac{\text{SCL}_{ap}}{\text{SCL}_a} \approx \frac{1}{A \left[1 - \sum_{k=0}^{m-1} 2^{-k-1} \left(\frac{A-1}{A} \right)^{2^k} \right]} \quad (1)$$

To reduce the complexity of the model, we considered only one escapee event in the pool, as the possibility of two independent escapees occurring in a pool is low in the quasi-linear phase of SCL. Nevertheless, we tested the effect of this omission on the drop of SCL due to aliquoting using Monte Carlo simulations. We performed 10 million independent trials for a doubling of 20 and a doubling of 27, and obtained a distribution of the number of escapees for each of them. Through

randomly sampling a number of escapees from each trial to the A aliquots, we calculated the number of 'bad' aliquots containing one or more escapees (A_b). To calculate a new SCL of the population after aliquoting, where SCL_p is the SCL of the original population, we used the formula $A \times \text{SCL}_p / \text{mean}(A_b)$. The drop in SCL was measured in silico and was compared with the value that we obtained from the equation obtained from the probability model.

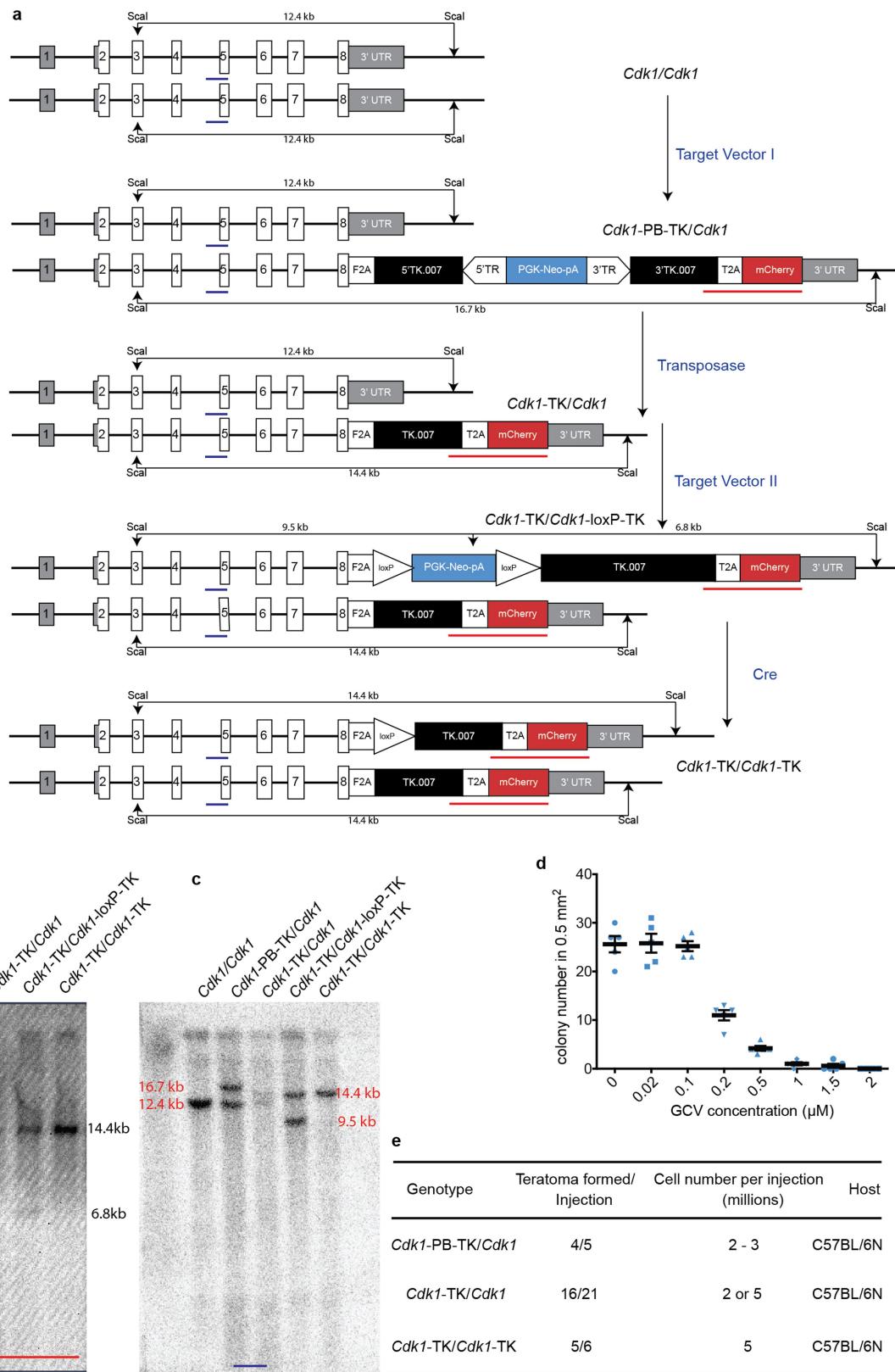
Code availability. The code used in this manuscript is publicly available at <https://github.com/mashutova/failsafe>.

Reporting summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

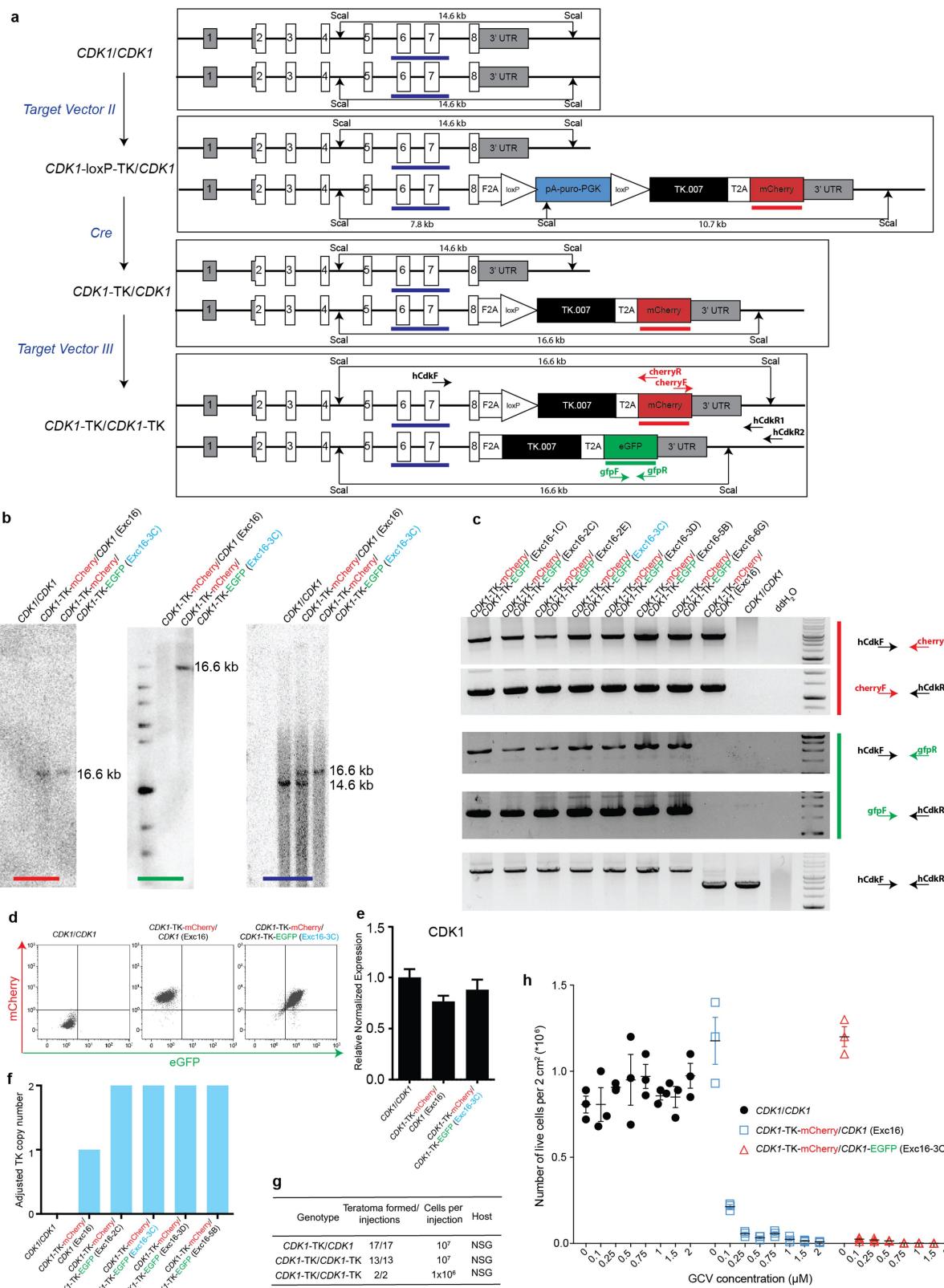
The data that support the findings of this study are available from the corresponding author upon reasonable request.

29. Cong, L. et al. Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).
30. Behringher, R., Gertsenstein, M., Nagy, K. V. & Nagy, A. *Manipulating the Mouse Embryo: A Laboratory Manual* 4th edn (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 2013).
31. Mereau, A., Grey, L., Piquet-Pellorce, C. & Heath, J. K. Characterization of a binding protein for leukemia inhibitory factor localized in extracellular matrix. *J. Cell Biol.* **122**, 713–719 (1993).
32. Park, J., Kusminski, C. M., Chua, S. C. & Scherer, P. E. Leptin receptor signaling supports cancer cell metabolism through suppression of mitochondrial respiration in vivo. *Am. J. Pathol.* **177**, 3133–3144 (2010).
33. Klimanskaya, I. et al. Derivation and comparative assessment of retinal pigment epithelium from human embryonic stem cells using transcriptomics. *Cloning Stem Cells* **6**, 217–245 (2004).
34. Sun, X. et al. Directed differentiation of human embryonic stem cells into thymic epithelial progenitor-like cells reconstitutes the thymic microenvironment in vivo. *Cell Stem Cell* **13**, 230–236 (2013).
35. Luria, S. E. & Delbrück, M. Mutations of bacteria from virus sensitivity to virus resistance. *Genetics* **28**, 491–511 (1943).
36. Capizzi, R. L. & Jameson, J. W. A table for the estimation of the spontaneous mutation rate of cells in culture. *Mutat. Res.* **17**, 147–148 (1973).
37. Luo, G. et al. Cancer predisposition caused by elevated mitotic recombination in Bloom mice. *Nat. Genet.* **26**, 424–429 (2000).
38. Monnat, R. J. Jr. Molecular analysis of spontaneous hypoxanthine phosphoribosyltransferase mutations in thioguanine-resistant HL-60 human leukemia cells. *Cancer Res.* **49**, 81–87 (1989).
39. Green, M. H. L., O'Neill, J. P. & Cole, J. Suggestions concerning the relationship between mutant frequency and mutation rate at the *hprt* locus in human peripheral T-lymphocytes. *Mutat. Res.* **334**, 323–339 (1995).
40. Biancotti, J.-C. et al. The in vitro survival of human monosomies and trisomies as embryonic stem cells. *Stem Cell Res.* **9**, 218–224 (2012).



Extended Data Fig. 1 | Generation, genotyping and characterization of mouse C57BL/6N C2 Cdk1-TK/Cdk1 and Cdk1-TK/Cdk1-TK ES cells. **a**, Summary of the targeting steps used to generate mouse C2 Cdk1-TK/Cdk1 and Cdk1-TK/Cdk1-TK ES cells. **b**, Southern blot genotyping with internal TK-mCherry probe. **c**, Southern blot genotyping with

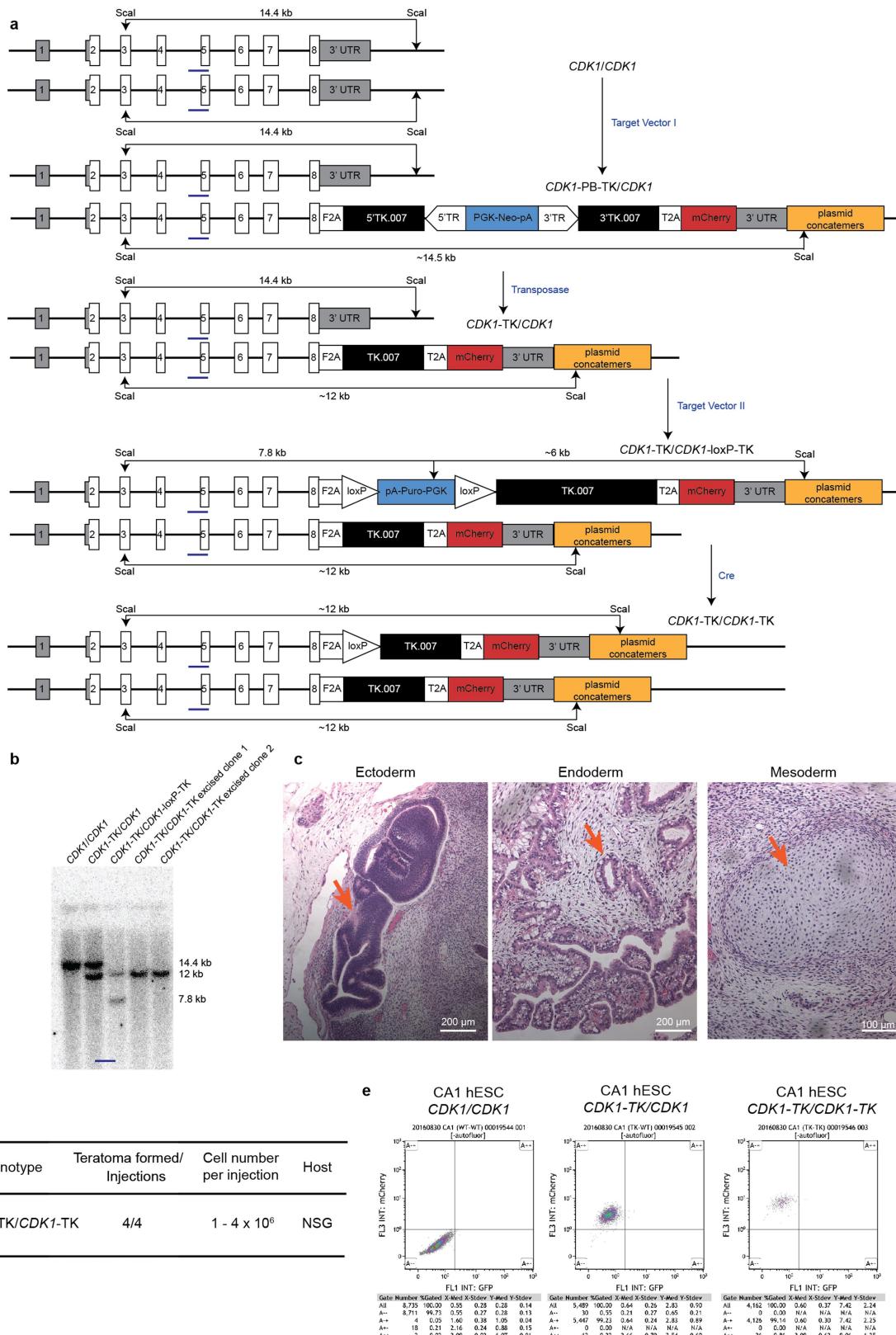
mouse Cdk1 genomic probe. **d**, In vitro GCV dose-response killing curve of mouse C2 CDK1-TK/CDK1 ES cells. Data are mean \pm s.d., $n = 3$. **e**, Teratoma formation efficiency of mouse C2 Cdk1-TK-PB/Cdk1, Cdk1-TK/Cdk1 and Cdk1-TK/Cdk1-TK ES cells.



Extended Data Fig. 2 | Generation, genotyping and characterization of human H1 CDK1-TK/CDK1 and CDK1-TK/CDK1-TK ES cells.

a, Generation of human H1 CDK1-TK/CDK1 and CDK1-TK/CDK1-TK ES cells. **b**, Southern blot genotyping of CDK1-TK/CDK1 clone Exc16, which was used in teratoma assays (Fig. 1c and Extended Data Fig. 7) and CDK1-TK/CDK1-TK clone Exc16-3C, which was used in the differentiation assays in Fig. 3. **c**, PCR genotyping of all the correct clones. **d**, Flow cytometry analysis of the CDK1-TK/CDK1 clone Exc16 and the CDK1-TK/CDK1-TK clone Exc16-3C. **e**, SybrGreen qPCR of human

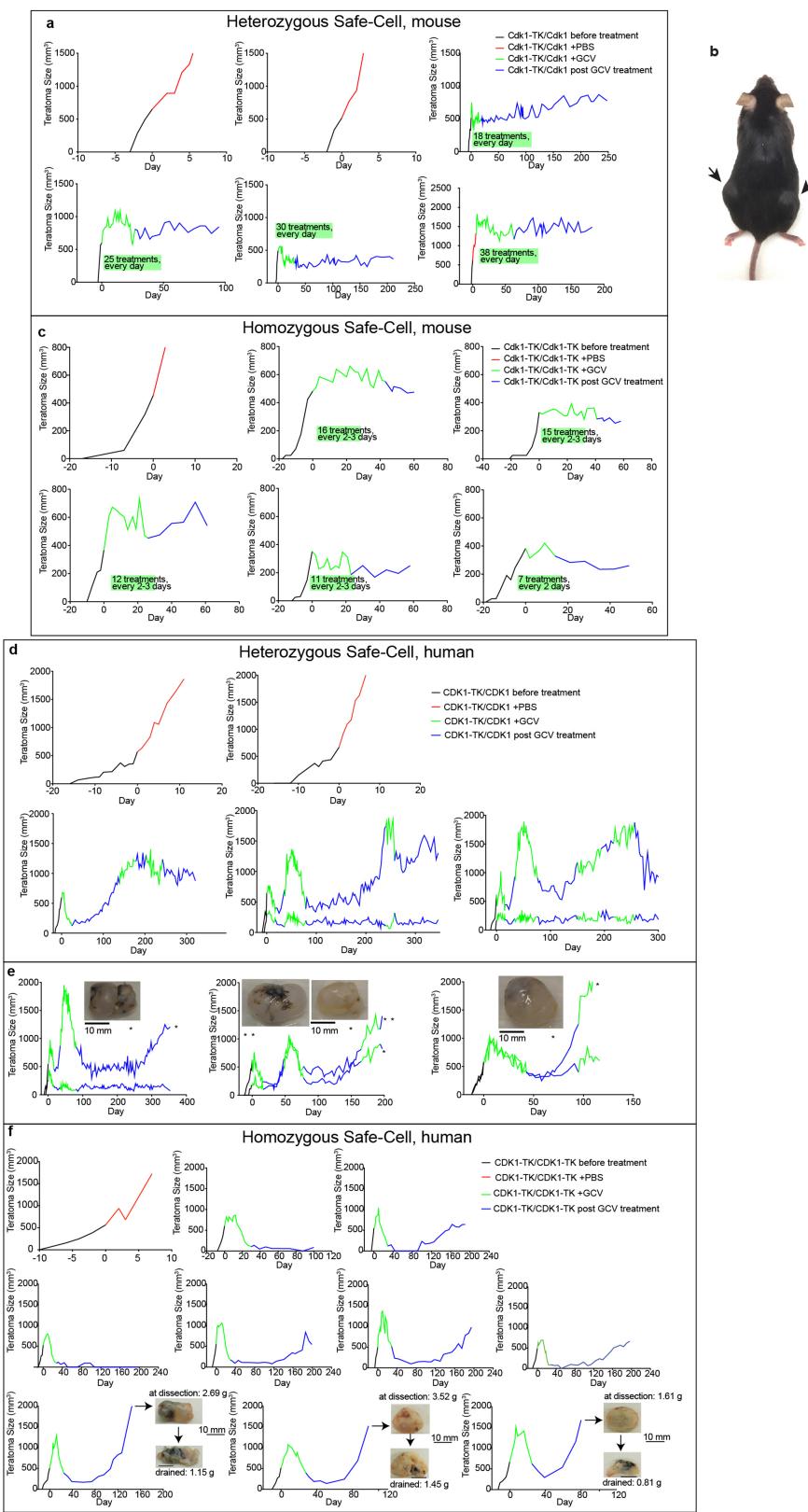
CDK1 expression in H1 wild-type cells, and cells expressing the CDK1-TK/CDK1 clone Exc16 and the CDK1-TK/CDK1-TK clone Exc16-3C. Data are mean \pm s.d., $n = 3$. No significant difference was found between groups. **f**, TaqMan qPCR copy-number analysis of TKs of all clones with the correct genotype. **g**, The efficiency of teratoma formation in NSG mice using human H1 ES cells. **h**, Dose-response analysis of wild-type, CDK1-TK/CDK1 and CDK1-TK/CDK1-TK human H1 ES cells. Cells were treated with different GCV concentrations, dissociated and counted after seven days. Data are mean \pm s.d., $n = 3$.



Extended Data Fig. 3 | Generation, genotyping and characterization of human CA1 CDK1-TK/CDK1 and CDK1-TK/CDK1-TK ES cells.

a, Generation of human CA1 CDK1-TK/CDK1 and CDK1-TK/CDK1-TK ES cells. **b**, Southern blot genotyping of human CA1 CDK1-TK/CDK1 and CDK1-TK/CDK1-TK ES cells. The plasmid concatemers are multiple copies of plasmid integration (including backbone). The ampicillin gene in

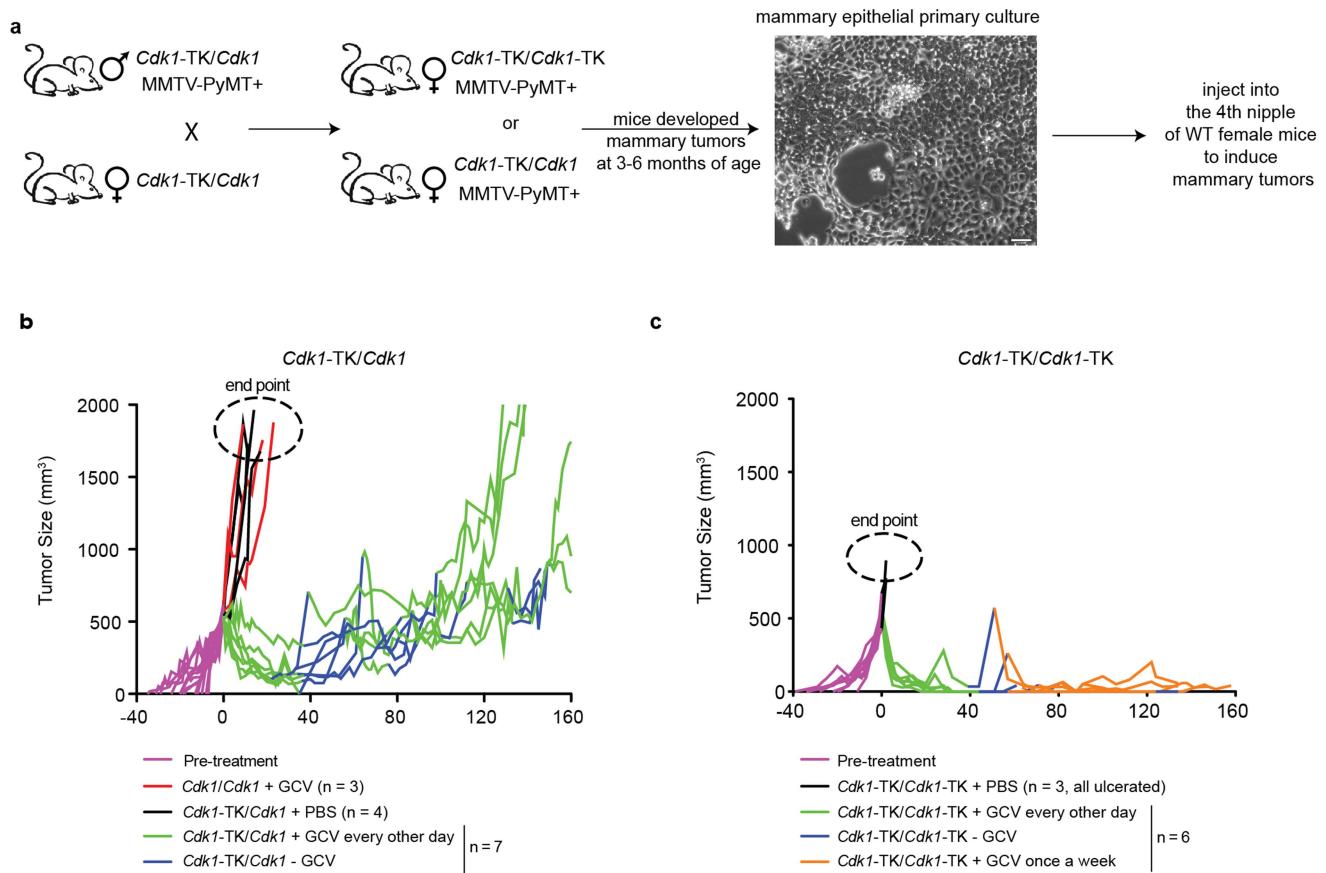
the backbone contains a ScalI restriction enzyme site, which is consistent with the sizes of the band in Southern blots. **c**, Haematoxylin and eosin staining of a CDK1-TK/CDK1-TK CA1 ES-cell-derived teratoma. **d**, The efficiency of teratoma formation in NSG mice using human CA1 ES cells. **e**, Flow cytometry analysis shows a direct correlation between the number of CDK1-TK alleles and mCherry fluorescence levels.



Extended Data Fig. 4 | See next page for caption.

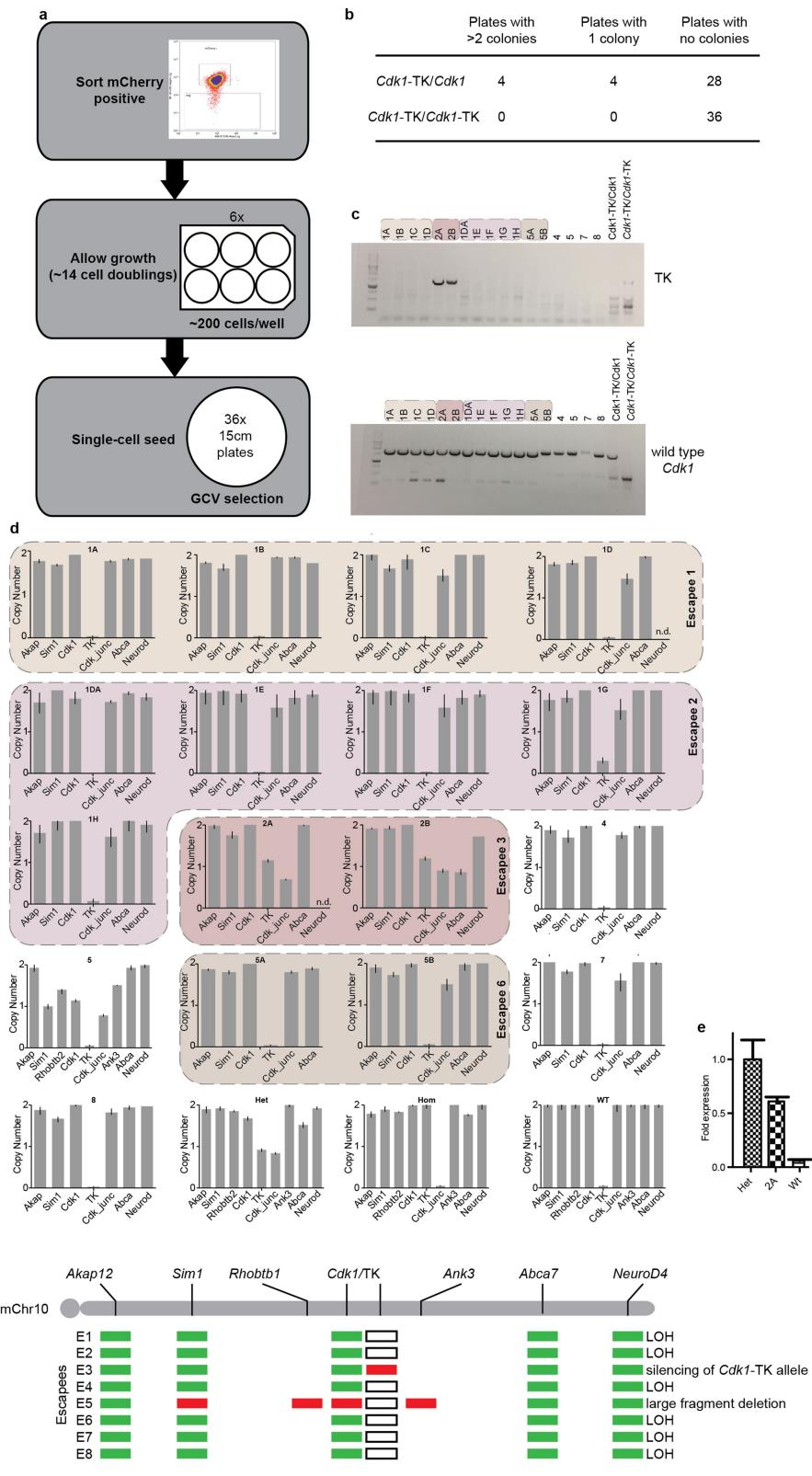
Extended Data Fig. 4 | Growth graphs of mouse and human ES-cell-derived teratomas. **a**, Growth of teratomas derived from mouse heterozygous safe-cell ES cells (*C2 Cdk1-TK/Cdk1*) **b**, Adult mouse with stabilized subcutaneous tissue (safe-cell ES-cell-derived dormant teratoma), 2.5 months after GCV treatment. **c**, Growth of teratomas derived from mouse homozygous safe-cell ES cells (*C2 Cdk1-TK/Cdk1-TK*). **d**, Growth of teratomas derived from human heterozygous safe-cell ES cells (*H1 CDK1-TK/CDK1*, clone Exc16); daily GCV treatment. **e**, Examples of teratomas from human heterozygous safe-cell ES cells showing cyst formation; images of cystic teratomas at dissection are shown next to the corresponding growth line; daily GCV treatment. The graphs

with two lines represent mice that had cells injected into both flanks. The graphs with one line represent mice that had cells injected into one flank. The GCV treatment regime varies among mice because each teratoma behaves differently; we started GCV when the teratoma size started to increase. **f**, Growth of teratomas derived from human homozygous safe-cell ES cells (*H1 CDK1-TK/CDK1-TK*), GCV treatment was every other day. Images of cystic teratomas are shown next to the corresponding growth line; cysts were drained after dissection to show the difference in tumour weight due to the fluid present in the tissue. Each graph represents one mouse. **a, c, d, f**, All replicates of these experiments are shown.



Extended Data Fig. 5 | Breast cancer transplantation assay using heterozygous and homozygous safe-cell mammary tumour cells.
a, Generation of mouse lines and experimental design. **b**, Growth of mammary gland tumours derived from mouse *Cdk1/Cdk1* and

Cdk1-TK/Cdk1 mammary epithelial cells with PBS or GCV treatment. **c**, Growth of mammary gland tumours derived from mouse *CDK1-TK/CDK1-TK* mammary epithelial cells with PBS or GCV treatment. **b, c**, The sample sizes of each group are indicated at the bottom of the graphs.

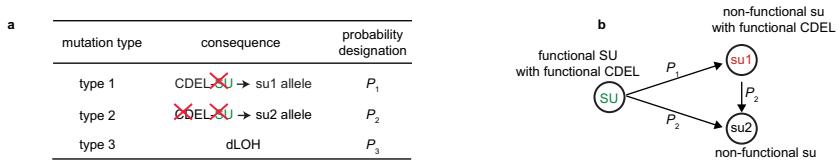


Extended Data Fig. 6 | See next page for caption.

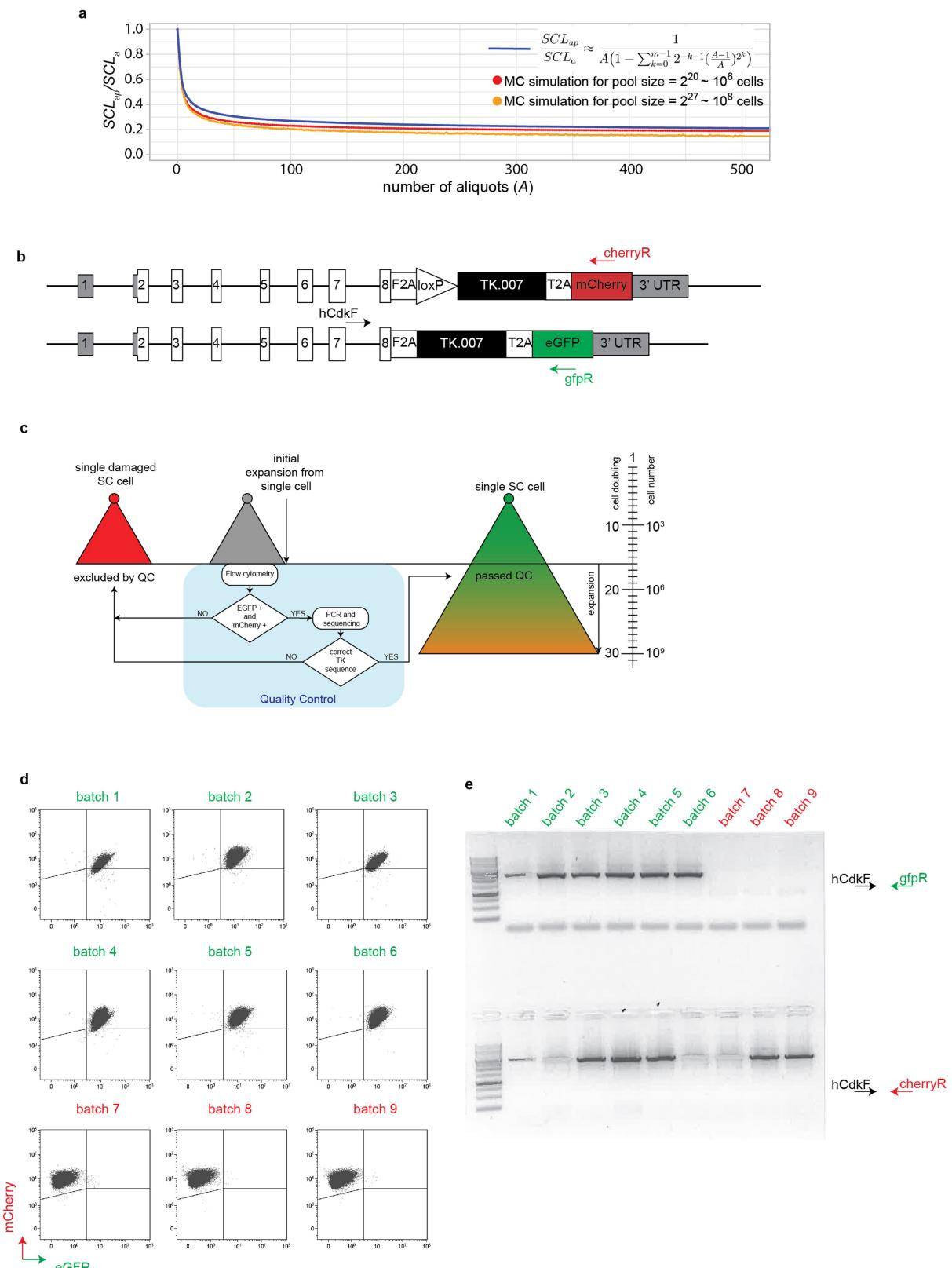
Extended Data Fig. 6 | In vitro experiments with mouse C2

Cdk1-TK/Cdk1 and Cdk1-TK/Cdk1-TK ES cells and subsequent characterization of escapees. **a**, Experimental design: mCherry⁺ cells were selected by sorting to ensure that the starting cell population did not contain escapees. These cells were plated on six-well plates (200 cells per well, in a total of 36 wells) and allowed to grow to 14 cell doublings (this was estimated by counting cells in sample wells). The 36 cultures were then resuspended to a single-cell suspension and each was plated in a 15-cm plate (4×10^6 cells). One day after plating, selection with GCV was started and maintained until escapee colonies appeared. **b**, Escapee numbers obtained in 36 independent cultures growing from *Cdk1-TK/Cdk1* and *Cdk1-TK/Cdk1-TK* ES cells. **c**, PCR to determine the presence of TK.

d, TaqMan copy number qPCR analysis of *Akap7*, *Sim1* and *Cdk1* junction of exon 8 and 3' UTR, *Neurod*, *Cdk1*, TK transgene and *Abca* on mouse chromosome 10. Data are the copy number calculated by CopyCaller Software v.2.1 and the error bars indicate the range from the minimum to the maximum number. $n = 3$. The same colour in the background of **c** and **d** indicates that they are from the same independent culture. n.d., not determined. **e**, qPCR to compare TK expression level in *Cdk1-TK/Cdk1* escapee clone 2A and C2 wild-type ES cells. Data are mean \pm s.e.m., $n = 3$. **f**, Summary of the copy number analysis of mouse *Cdk1-TK/Cdk1* escapees. **a**, **b**, Experiments were repeated twice on a smaller scale but with similar results. **d**, Experiments were repeated twice with similar results.

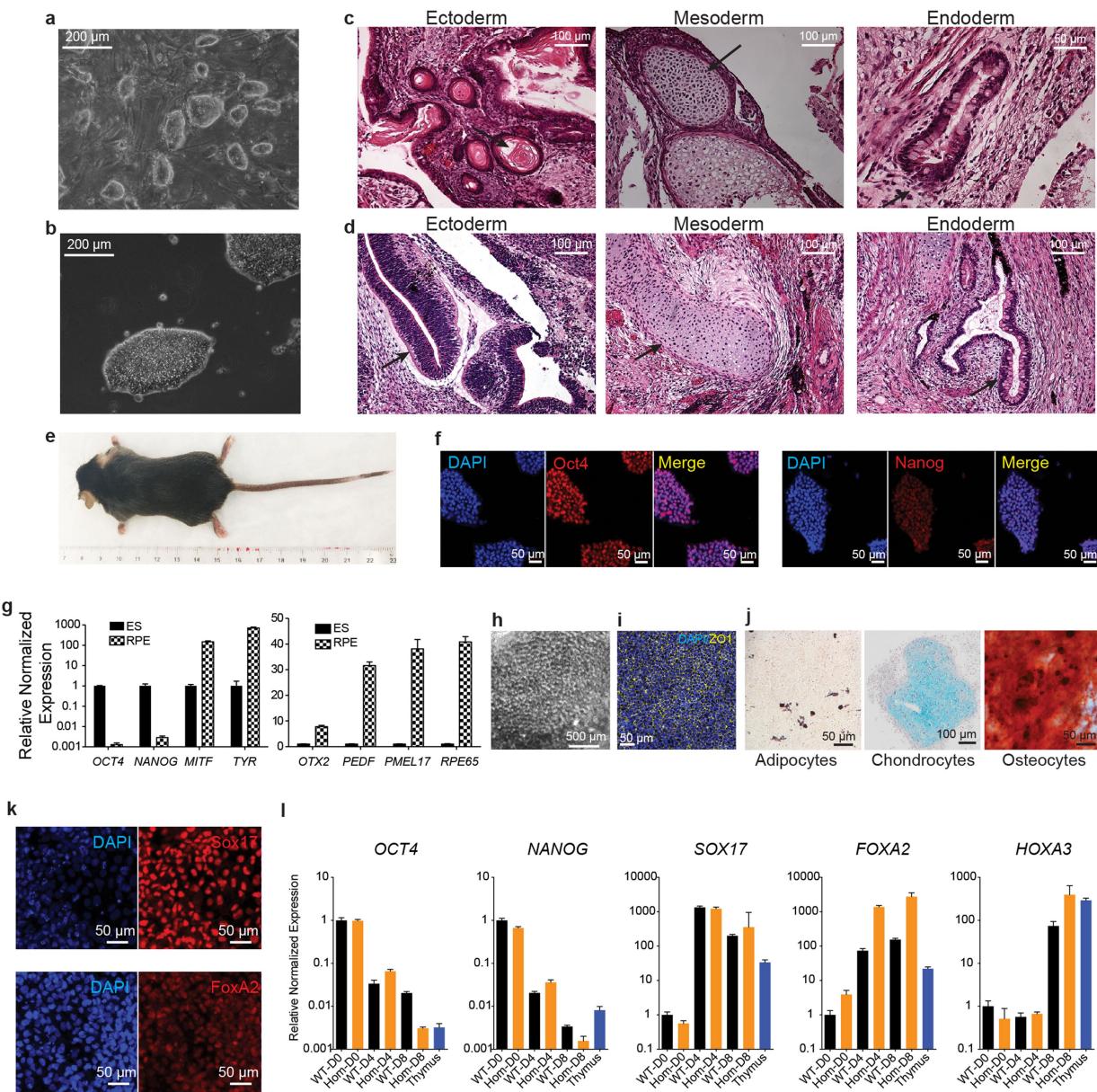


Extended Data Fig. 7 | Schematics of the possible mutation types affecting the CDL-SU allele. **a**, Three types of mutations that could affect the CDL-SU allele. **b**, Safe-cell allele transition considered in the modelling and Monte Carlo simulations.



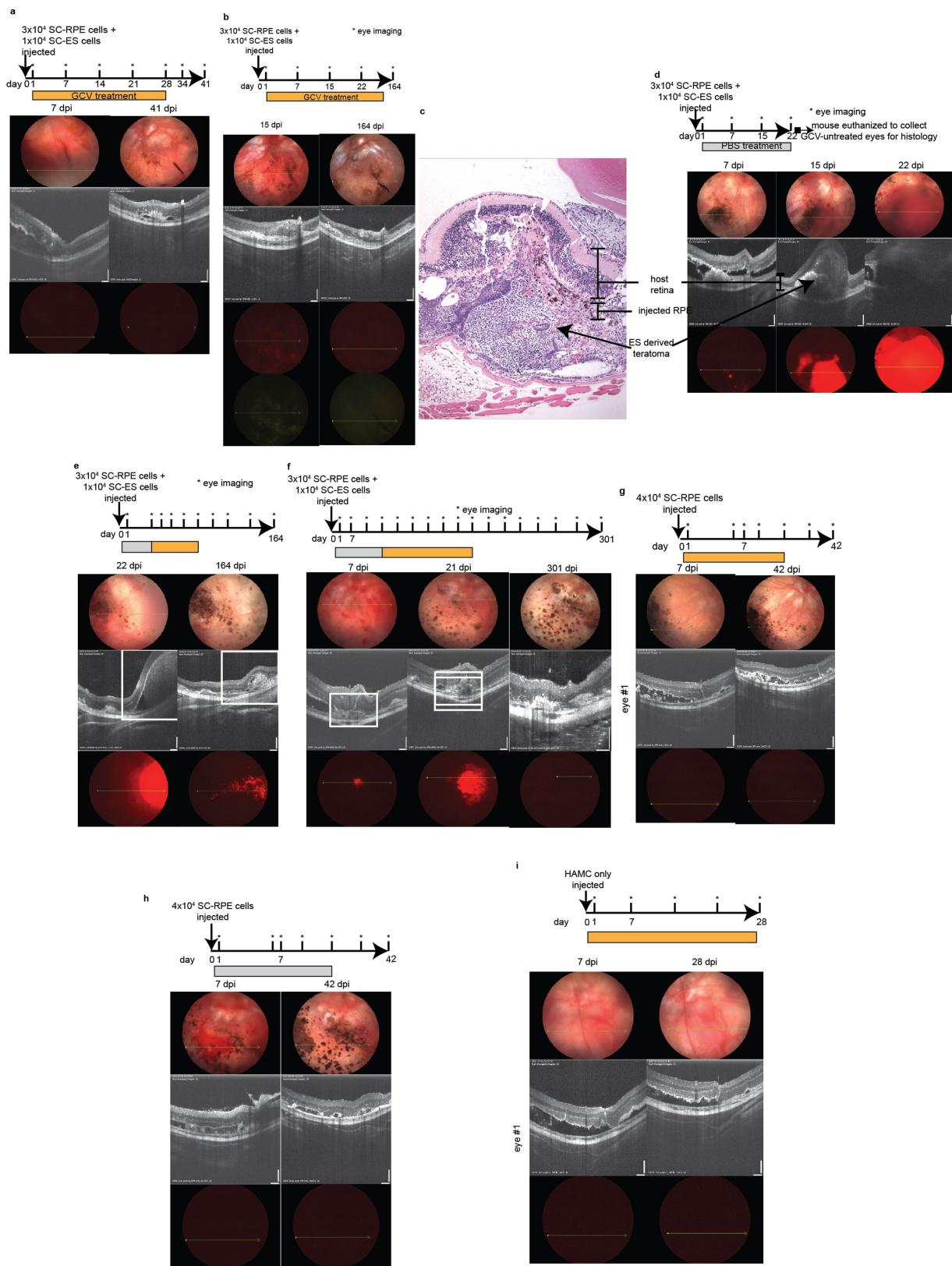
Extended Data Fig. 8 | Quality control of batches generated from single human H1 CDK1-TK/CDK1-TK ES cells. **a**, The drop of SCL due to aliquoting from a pool of cells relative to non-aliquoted batches of the same size. **b**, Schematics of the alleles in the CDK1-TK/CDK1-TK human

ES cells used in the quality control. **c**, Workflow schematic of performing quality control (QC) on several ES cell batches. **d**, An example of the flow cytometry for the quality control of nine clonally derived batches. **e**, An example of PCR for the quality control of nine clonally derived batches.



Extended Data Fig. 9 | Mouse and human safe-cell homozygous CDK1-TK/CDK1-TK cells demonstrate pluripotency. All experiments were performed using the same clone of mouse C2 or human H1 (Exc16-3C) CDK1-TK/CDK1-TK cells. **a**, Bright-field photograph showing mouse homozygous *Cdk1-TK/Cdk1-TK* ES cell morphology. **b**, Bright-field photograph showing human homozygous *CDK1-TK/CDK1-TK* ES cell morphology. **c**, Haematoxylin and eosin staining of a mouse *Cdk1-TK/Cdk1-TK* ES-cell-derived teratoma. **d**, Haematoxylin and eosin staining of a human *CDK1-TK/CDK1-TK* ES-cell-derived teratoma. **e**, An adult *Cdk1-TK/Cdk1-TK* mouse. **f**, OCT4 and NANOG staining of human *CDK1-TK/CDK1-TK* ES cells. **g**, qPCR characterization of human *CDK1-TK/CDK1-TK* ES cells. Data are mean \pm s.d., $n = 3$. **h–j**, Bright-field picture of human *CDK1-TK/CDK1-TK* ES-cell-derived RPE cells. **i**, ZO1 staining of human *CDK1-TK/CDK1-TK* ES-cell-derived RPE cells. **j**, Human *CDK1-TK/CDK1-TK* ES-cell-derived adipocytes, chondrocytes and osteocytes. **k**, SOX17 and FOXA2 staining of human *CDK1-TK/CDK1-TK* ES-cell-derived definitive endoderm. **l**, OCT4 (also known as *POU5F1*), NANOG, SOX17, FOXA2 and HOXA3 qPCR characterization of ES cell (day 0) differentiation into definitive endoderm (day 4) and pharyngeal pouch endoderm (day 8). Data are mean \pm s.d., $n = 3$. **a–d**, **f–i**, Experiments were repeated three times with similar results. **j–l**, Experiments were repeated twice with similar results.

CDK1-TK/CDK1-TK ES cell differentiation into RPE cells. Data are mean \pm s.d., $n = 3$. **h**, Bright-field picture of human *CDK1-TK/CDK1-TK* ES-cell-derived RPE cells. **i**, ZO1 staining of human *CDK1-TK/CDK1-TK* ES-cell-derived RPE cells. **j**, Human *CDK1-TK/CDK1-TK* ES-cell-derived adipocytes, chondrocytes and osteocytes. **k**, SOX17 and FOXA2 staining of human *CDK1-TK/CDK1-TK* ES-cell-derived definitive endoderm. **l**, OCT4 (also known as *POU5F1*), NANOG, SOX17, FOXA2 and HOXA3 qPCR characterization of ES cell (day 0) differentiation into definitive endoderm (day 4) and pharyngeal pouch endoderm (day 8). Data are mean \pm s.d., $n = 3$. **a–d**, **f–i**, Experiments were repeated three times with similar results. **j–l**, Experiments were repeated twice with similar results.



Extended Data Fig. 10 | See next page for caption.

Extended Data Fig. 10 | Representative images of eyes transplanted with both safe-cell RPE and safe-cell ES cells, and only with safe-cell RPE cells. **a, b**, Fundoscopy, optical coherence tomography and fluorescence imaging of eyes transplanted with safe-cell RPE and safe-cell ES cells (four-week GCV treatment). The absence of a mCherry signal indicates that ES cell growth has not occurred. **b**, Bottom, images of the green fluorescence channel are included to illustrate that the observed signal in the red fluorescence channel is actually autofluorescence. This experiment was repeated four times in four mice with similar results. **c**, Histological analysis of the eye presented in **d**. **d**, Fundoscopy, optical coherence tomography and fluorescence imaging of eyes transplanted with safe-cell RPE and safe-cell ES cells (PBS treatment). This experiment was repeated twice in two mice with similar results. **e, f**, Fundoscopy, optical coherence tomography and fluorescence imaging of eyes transplanted with safe-cell

RPE and safe-cell ES cells, mCherry signal is detectable and indicates ES cell growth. GCV treatment began three weeks post-injection following an initial PBS treatment. This experiment was repeated four times in three mice with similar results. **g**, Fundoscopy, optical coherence tomography and fluorescence imaging of eyes receiving only safe-cell RPE cells (four-week GCV treatment). This demonstrates that GCV treatment did not affect the RPE cells. This experiment was repeated five times in three mice with similar results. **h**, Fundoscopy, optical coherence tomography and fluorescence imaging of eyes receiving only safe-cell RPE cells (four-week PBS treatment). This experiment was repeated six times in three mice with similar results. **i**, Fundoscopy, optical coherence tomography and fluorescence imaging of eyes receiving only HAMC (four-week GCV treatment). This experiment was repeated twice in one mouse with similar results.

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated
- Clearly defined error bars
State explicitly what error bars represent (e.g. SD, SE, CI)

Our web collection on [statistics for biologists](#) may be useful.

Software and code

Policy information about [availability of computer code](#)

Data collection

Zen 2011 (black edition) for the confocal images taken for immuno-staining.
OpenLab 5.5.2 for bright-field image.
Leica IM50 Image Manager Version 1.20 for histology imaging.
Bio-Rad Quality One Version 4.6.9 for southern blot imaging.
Bio-rad Image Lab Version 5.2 for agarose gel imaging.
InSight software for eye imaging.

Data analysis

Bio-Rad CFX Manager 3.1, for qPCR analysis
Thermofisher CopyCaller Software v2.1, for copy number analysis.
R, version 3.3.3, for Monte Carlo simulation.
GraphPad Prism software v.5.0a for all the data processing.
Beckman Coulter Kaluza 1.5 for Flow cytometry analysis
The code is available here: <https://github.com/mashutova/failsafe>

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about availability of data

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size The number of samples used in different experiments was determined in accordance to similar studies in the field.

Data exclusions No data was intentionally excluded.

Replication Findings were reproduced as necessary and replications attempts are reported in the manuscript.

Randomization Samples were allocated randomly when possible.

Blinding Due to the nature of most experiments, blinding was impossible because the results are visible at the time of analysis.

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> Unique biological materials
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input type="checkbox"/>	<input checked="" type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology
<input type="checkbox"/>	<input checked="" type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input type="checkbox"/>	<input checked="" type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Unique biological materials

Policy information about availability of materials

Obtaining unique materials All the plasmids reported in the study will be available through Addgene upon publication.

Antibodies

Antibodies used

Anti-BrdU, CAT#M0744, DAKO, LOT#87555, diluted 1:200
 Anti-FoxA2, CAT#ab40874, Abcam, LOT#GR141178-1, diluted 1:1000
 Anti-Nanog (human), CAT#AF1997, R&D systems, LOT#KKJ0308031, , diluted 10 µg/ml
 Anti-Oct3/4 (human), CAT#AF1759, R&D systems, LOT#JTWO015111, diluted 10 µg/ml
 Anti-Sox17, CAT#AF1924, R&D systems, LOT#KGA0613041, diluted 10 µg/ml
 Anti-Tubb3, CAT#AB9354, Millipore LOT#LV1462900, diluted 1:1000
 Anti-ZO-1, CAT#40-2200, Thermo Fisher, LOT#SG246825, diluted 1:2000

Validation

Validations were available from manufacturers and confirmed in the lab using positive and negative controls.
 Anti-BrdU, <https://www.agilent.com/en/products/immunohistochemistry/antibodies-controls/primary-antibodies/bromodeoxyuridine>
 Anti-FoxA2, CAT#ab40874, Abcam, LOT#GR141178-1, <https://www.citeab.com/antibodies/731184-ab40874-anti-foxa2-antibody>
 Anti-Nanog (human), https://www.rndsystems.com/products/human-nanog-antibody_af1997
 Anti-Oct3/4 (human), https://www.rndsystems.com/products/human-mouse-oct-3-4-antibody_af1759
 Anti-Sox17, https://www.rndsystems.com/products/human-sox17-antibody_af1924
 Anti-Tubb3, http://www.emdmillipore.com/CA/en/product/Anti-Beta-III-Tubulin-Antibody,MM_NF-AB9354
 Anti-ZO-1, <https://www.thermofisher.com/order/genome-database/generatePdf?productName=ZO-1&assayType=PRANT&detailed=true&productId=40-2200>

Eukaryotic cell lines

Policy information about [cell lines](#)

Cell line source(s)	Mouse C67BL6/N C2 embryonic stem cells were derived in Nagy lab and The Centre for Phenogenomics. Human H1 embryonic stem cells were imported from WiCell. Human CA1 embryonic stem cells were derived in Nagy lab.
Authentication	Cell lines were not authenticated
Mycoplasma contamination	All the cell lines were tested negative for mycoplasma contamination.
Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell lines were used.

Animals and other organisms

Policy information about [studies involving animals; ARRIVE guidelines](#) recommended for reporting animal research

Laboratory animals	The female CD-1 (ICR) (Charles River) outbred albino stock was used as embryo donors for aggregation with ES cells and as pseudopregnant recipients. C57BL/6NCrl (Charles River), 6-20 weeks old, was used as the host for teratoma assays with mouse C2 ES cells. NOD scid gamma /J#5557 (Jackson Laboratories), 6-20 weeks old, was used as the host for teratoma assays with human H1 or CA1 ES cells and eye experiments.
Wild animals	This study did not involve wild animals
Field-collected samples	This study did not involved field-collected samples

Flow Cytometry

Plots

Confirm that:

- The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
- All plots are contour plots with outliers or pseudocolor plots.
- A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	Mouse or human pluripotent stem cells were dissociated, washed with PBS once, and filtered to Falcon tubes with cell-strainer cap (REF352235).
Instrument	FACS: MoFlo Astrios EQ cell sorter (Beckman Coulter, Miami, FL, USA), equipped with 305nm, 405nm, 488nm, 561nm and 640nm lasers. Flow cytometry: GALLIOS flow cytometer, Beckman Coulter, equipped with 405nm, 488nm, 561nm and 640nm lasers.
Software	Sorter collection software: Summit v6.2 (Beckman Coulter); Analysis software (for Gallios data): Kaluza v1.5 (Beckman Coulter).
Cell population abundance	The sorting in Extended Data Figure 8 and 9 were collected as single cell, one cell per well of 96-well plate, and further expanded for different experimental purposes. All of the survived clones belonged to the desired population verified by further analyses. The sorting in human ES targeting was collected as bulk, and then we derived clonal lines from the bulk for genotyping. The genotyping confirmed that we had sorted the desired population.
Gating strategy	Samples were gated for live single cells using forward and side scatter pulse height and area, and DAPI viability staining. Wildtype

Gating strategy

and single-color samples of the same cell type as the experimental samples were used for negative controls and compensation calculations.

Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.