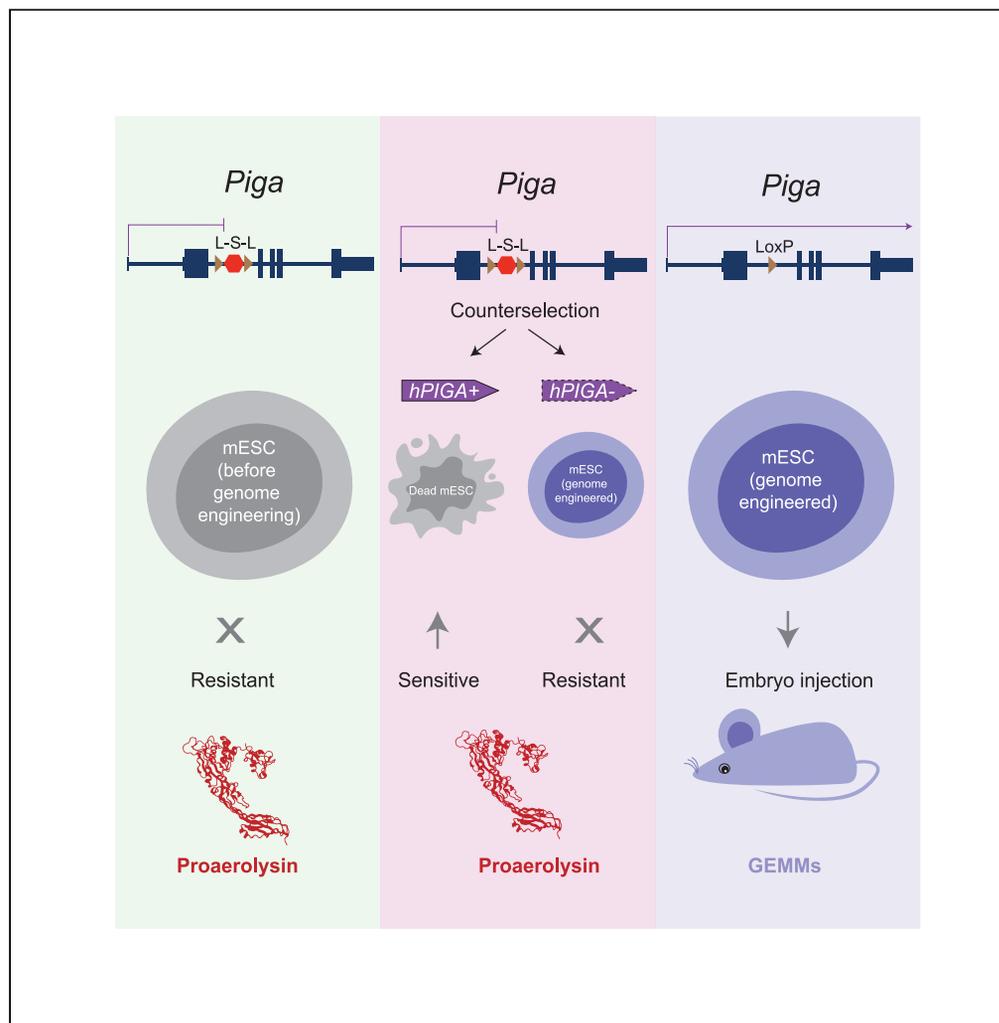


Article

A conditional counterselectable *Piga* knockout in mouse embryonic stem cells for advanced genome writing applications



Weimin Zhang,
Ran Brosh, Laura
H. McCulloch, ...,
Sang Yong Kim,
Matthew T.
Maurano, Jef D.
Boeke

jef.boeke@nyulangone.org

Highlights

Transcriptional
inactivation of *Piga*
renders mESCs full
resistance to proaerolysin

PIGA serves as
counterselectable marker
in endogenous *Piga*-
inactivated mESCs

Piga function is readily
restored by Cre-mediated
STOP cassette excision

Conditional knockout and
restoration strategy is also
applicable to *Hprt*

Zhang et al., iScience 25,
104438
June 17, 2022 © 2022 The
Authors.
[https://doi.org/10.1016/
j.isci.2022.104438](https://doi.org/10.1016/j.isci.2022.104438)



Article

A conditional counterselectable *Piga* knockout in mouse embryonic stem cells for advanced genome writing applications

Weimin Zhang,¹ Ran Brosh,¹ Laura H. McCulloch,¹ Yinan Zhu,¹ Hannah Ashe,¹ Gwen Ellis,¹ Brendan R. Camellato,¹ Sang Yong Kim,² Matthew T. Maurano,^{1,2} and Jef D. Boeke^{1,3,4,5,*}

SUMMARY

Overwriting counterselectable markers is an efficient strategy for removing wild-type DNA or replacing it with payload DNA of interest. Currently, one bottleneck of efficient genome engineering in mammals is the shortage of counterselectable (negative selection) markers that work robustly without affecting organismal developmental potential. Here, we report a conditional *Piga* knockout strategy that enables efficient proaerolysin-based counterselection in mouse embryonic stem cells. The conditional *Piga* knockout cells show similar proaerolysin resistance as full (non-conditional) *Piga* deletion cells, which enables the use of a *PIGA* transgene as a counterselectable marker for genome engineering purposes. Native *Piga* function is readily restored in conditional *Piga* knockout cells to facilitate subsequent mouse development. We also demonstrate the generality of our strategy by engineering a conditional knockout of endogenous *Hprt*. Taken together, our work provides a new tool for advanced mouse genome writing and mouse model establishment.

INTRODUCTION

Mouse models have long served as key tools for studying disease mechanisms. Advances in stem cell manipulation and genome engineering have facilitated the development of many useful genetically engineered mouse models (GEMMs). In recent decades, as biologists advanced the ability to synthesize and integrate megabase-sized DNA sequences of interest across systems from *Mycoplasma* subspecies to *Saccharomyces cerevisiae* (Gibson et al., 2008, 2010; Dymond et al., 2011; Annaluru et al., 2014; Richardson et al., 2017; Shen et al., 2017; Xie et al., 2017; Mitchell et al., 2017; Wu et al., 2017; Zhang et al., 2017, 2020), increased attention has turned to how such technologies could be applied to mouse genome writing. Large-scale mouse genome engineering on the order of hundreds of kilobases (kb) at a time makes it possible to design and integrate complex sequences that would be difficult to complete using one-edit-at-a-time approaches. Additionally, fully overwriting a few mouse genes with their human counterparts allows for better recapitulation of human gene functions in mice (Lee et al., 2014; Mitchell et al., 2021; Wallace et al., 2007). At present, challenges in mouse genome writing have mainly focused on delivering large pieces of DNA into mouse embryonic stem cells (mESCs) in a site-specific manner, which enables facile analysis of the resultant animals.

Existing big DNA delivery technologies in non-mammalian organisms provide useful insights for mouse genome writing. For example, the Synthetic Yeast Genome Project (Sc2.0) uses the Switch Auxotrophies Progressively for Integration (SwAP-In) method to integrate large sequences of interest in a stepwise manner. In more detail, SwAP-In alternates auxotrophic markers during each engineering step to select for insertion of synthetic DNA and against the preexisting wild-type (WT) sequences (Dymond and Boeke, 2012; Richardson et al., 2017). A similar approach relying on both selection and counterselection was deployed for the assembly of the 61-codon synthetic *Escherichia coli* genome (Wang et al., 2016; Fredens et al., 2019). In mESCs, many positive selection markers, including the puromycin-N-acetyltransferase gene (*pac*), blasticidin S-resistance gene (*bsr*), and hygromycin phosphotransferase gene (*hph*), are readily available. However, effective counterselectable markers are generally in short supply, limiting the development of efficient DNA delivery methods for mESCs.

¹Institute for Systems Genetics, NYU Langone Health, New York, NY 10016, USA

²Department of Pathology, NYU Langone Health, New York, NY 10016, USA

³Department of Biochemistry and Molecular Pharmacology, NYU Langone Health, New York, NY 10016, USA

⁴Department of Biomedical Engineering, NYU Tandon School of Engineering, Brooklyn, NY 11201, USA

⁵Lead contact

*Correspondence:

jef.boeke@nyulangone.org
<https://doi.org/10.1016/j.isci.2022.104438>



Recombinase-mediated cassette exchange has proven promising for targeted DNA delivery to cells, and has been adapted for delivery of payloads larger than 100 kb (Iacovino et al., 2011; Lee et al., 2014; Wallace et al., 2007). Recently, the Big-IN method, which relies on overwriting an intermediate landing pad harboring a counterselectable marker, was shown to enable site-specific >100 kb payload delivery, leaving only recombinase binding sites as genomic scars (Brosh et al., 2021). Additionally, if iteratively delivering DNA to mESCs, counterselectable markers can be embedded near the end of the payload DNA and used for selecting against residual unwanted DNA in the next engineering step. However, many common counterselectable markers used in mESCs have significant limitations. One popular counterselectable marker, the herpes simplex virus type 1–thymidine kinase (HSV1-TK, or TK), produces a bystander effect by which toxic ganciclovir-triphosphate can diffuse from TK positive cells to TK negative cells via gap junctions, resulting in unwanted TK negative cell death following ganciclovir treatment and limiting TK's counterselection accuracy (Brosh et al., 2021; Elshami et al., 1996; Mesnil et al., 1996). Diphtheria toxin binds to human heparin-binding EGF-like growth factor (HB-EGF) and abolishes host protein synthesis. Introducing the diphtheria toxin receptor (DTR) in mice allows targeted cell ablation in transgenic mice (Saito et al., 2001), but it cannot be used as a counterselectable marker in human cells. Other counterselectable markers rely on knocking out important endogenous genes, which may require additional engineering steps to reconstitute the knocked-out gene using a minigene before phenotyping the engineered cells or converting mESCs to mice. Neither of these strategies for reintroducing genes is ideal; indeed, minigenes lack the natural regulatory sequences associated with their native loci of interest. Moreover, their reintroduction can be technically challenging and can lead to additional off-target genome modifications.

One such option for counterselection in mESCs is *Piga* (phosphatidylinositol glycan anchor biosynthesis class A), an essential gene in the glycosylphosphatidylinositol (GPI)-anchor biosynthesis pathway (Miyata et al., 1993). *Piga*-expressing cells are sensitive to proaerolysin, a bacterial prototoxin that binds cell-surface GPI anchors and perforates the cell membrane, causing rapid cell death (Abrami et al., 1998). Knocking out the single copy of the X-linked *Piga* gene in male genomes readily generates proaerolysin-resistant cells. This enables the use of a *Piga* minigene as a counterselectable marker on intermediate landing pads and consequently, efficient isolation of cells in *Piga*-overwritten cells following genome writing with no bystander effect (Brosh et al., 2021; Li et al., 2021). However, as PIGA is essential for murine development (Rosti et al., 1997), PIGA function must be restored before mESCs can give rise to mice.

Here, we demonstrate conditionally knocking out and restoring *Piga* using a floxed STOP cassette inserted into the second intron of *Piga*. Using this strategy, we induce a temporal endogenous *Piga* knockout state, enabling the use of a *Piga* transgene as a counterselectable marker for genome writing. We demonstrate highly efficient excision of the STOP cassette using Cre recombinase, which can be applied once mESC engineering is complete to enable mouse development. This easily deployable “counterselection switch” preserves the native regulatory structure (and thus native levels and splice isoforms of the gene in multiple tissues) of *Piga*. This approach provides an efficient counterselectable marker for mESCs engineering, while preserving the potential of developing GEMMs from engineered mESCs that contain large, targeted DNA modifications of interest.

RESULTS

Construction of a reversible *Piga* knockout mouse embryonic stem cell lines

To conditionally knock out *Piga* in a male mESC line, we inserted an STOP cassette consisting of a β -geo (β -galactosidase and neomycin fusion) marker and three tandemly repeated SV40 polyadenylation signals (pA) into the second intron of the *Piga* gene to prematurely terminate its transcription (Figure 1A). The STOP cassette is flanked by two forward-facing loxP sites to enable Cre-mediated excision. Integration of the STOP cassette was achieved by co-transfecting mESCs with a plasmid co-expressing Cas9 and a guide RNA (gRNA) targeting the second intron of *Piga* (chrX:164,425,912, mm10) along with a plasmid carrying the STOP cassette flanked by homology arm (HA) sequences (Figure S1A). Successful STOP cassette insertion (henceforth referred to as *Piga*-STOP) was initially verified by identifying the novel genomic junctions by PCR in G418-resistant clones (Figure 1B). Two candidate clones that passed the junction PCR screening were subjected to targeted capture sequencing for further verification of engineering precision (Brosh et al., 2021). We used a modular mapping approach, in which the sequencing reads were separately mapped to three references: the STOP cassette, the plasmid backbone of the STOP cassette, and the mouse genome (mm10). We found that both *Piga*-STOP clones had reads mapping to the STOP cassette (Figure 1C), indicating that the STOP cassette was successfully inserted into the genome. We then applied *bamintersect* integration site analysis (Brosh et al., 2021), which identifies reads mapping to two different

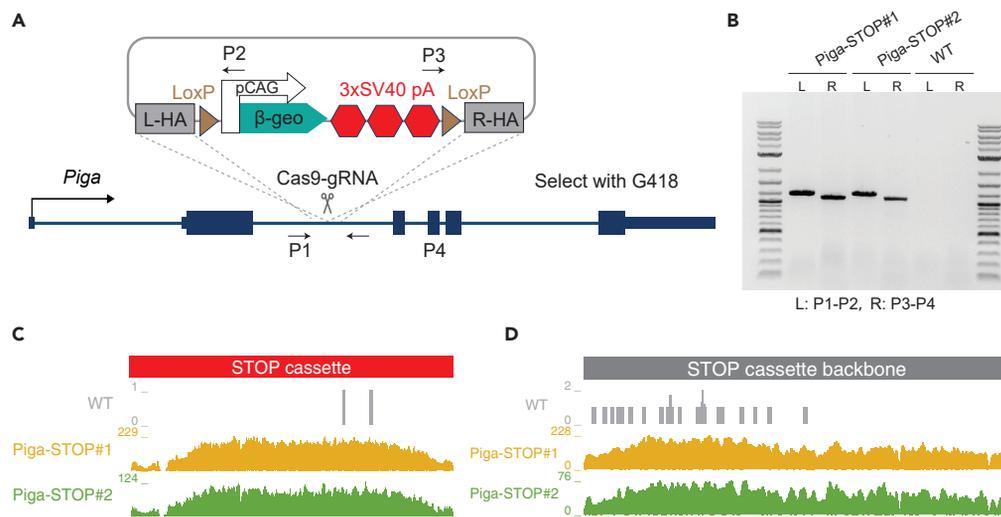


Figure 1. *Pigta* transcriptional knockout by a floxed STOP cassette

(A) Schematic for *Pigta* conditional knockout. Dark blue lines, *Pigta* introns, dark blue boxes, coding exons or untranslated region (UTR). L-HA and R-HA, left and right homology arms (~900 bp), respectively. pCAG, synthetic promoter consists of CMV enhancer, chicken beta-Actin promoter and rabbit beta-Globin splice acceptor site. Arrows indicate the position of genotyping primers. Scissors indicate the Cas9-gRNA cutting site.

(B) Genotyping PCR of STOP cassette insertion clones. L, left junction (primers P1 and P2); R, right junction (primers P3 and P4). Ladder, 1 kb plus DNA ladder (NEB).

(C) Capture sequencing analysis of *Pigta*-STOP mESC clones. Coverage plots of reads mapping to STOP cassette.

(D) Capture sequencing analysis of *Pigta*-STOP mESC clones. Coverage plots of reads mapping to the STOP cassette backbone.

reference genomes, enabling unbiased detection of the STOP cassette's integration site. This analysis detected no off-target junction for the two sequenced clones (Table S1). We also found both *Pigta*-STOP clones had reads mapping to the plasmid backbone of the STOP cassette (Figure 1D). Also, a 2-fold increase in sequencing coverage in the *Pigta* homology arm regions relative to the surrounding genome coverage was observed (Figure S1B), indicating that two copies of the STOP cassette and a copy of the plasmid backbone might have integrated. Genotyping PCR spanning the backbone and STOP cassette confirmed this duplicated configuration (Figure S1C). Nonetheless, the duplicated STOP cassette should still fully retain the conditional knockout's reversibility as the entire insertion is still flanked by two loxP sites.

Evaluation of *Pigta*-STOP mESCs

To evaluate whether transcription of *Pigta* can be terminated by the STOP cassette, we examined the mRNA level of *Pigta* using primers targeting exons downstream of the STOP cassette integration site. As expected, *Pigta* transcription was significantly decreased in *Pigta*-STOP cells compared to wild-type cells (Figure 2A). Consistent with this, PIGA protein was significantly depleted in *Pigta*-STOP cells as determined by western blotting (Figure 2B).

Proaerolysin-mediated cell death is initiated when the active form of the proaerolysin monomer binds to the glycosylphosphatidylinositol (GPI) anchors on the cell membrane. The activation of proaerolysin leads to a conformational change followed by the formation of a heptameric oligomer. The heptamer inserts into cell membrane, forming a membrane pore that leads to rapid cell death (Iacovache et al., 2011, 2016) (Figure 2C). Blocking *Pigta* transcriptionally should abolish GPI-anchor biosynthesis, thus preventing proaerolysin from binding GPI-anchored proteins and rendering *Pigta*-STOP mESCs resistant to proaerolysin. As a control, we used a complete *Pigta* deletion (Δ *Pigta*) mESC line (Brosh et al., 2021). A proaerolysin kill curve assay was first performed to determine the lowest effective working concentration. We found that 0.5 nM and higher concentrations of proaerolysin eliminated almost all the sensitive mESCs (Figure S2A). To make sure no sensitive cells escaped proaerolysin-mediated cell death, we withdrew proaerolysin after 24 h and continued to culture any residual cells for three days. Consistent with our initial observations, concentrations of 0.5 nM proaerolysin and above produced no escapees (Figure S2B). Next, we tested the resistance of *Pigta*-STOP mESCs to proaerolysin. Severe cell shrinkage was observed in wild-type mESCs during

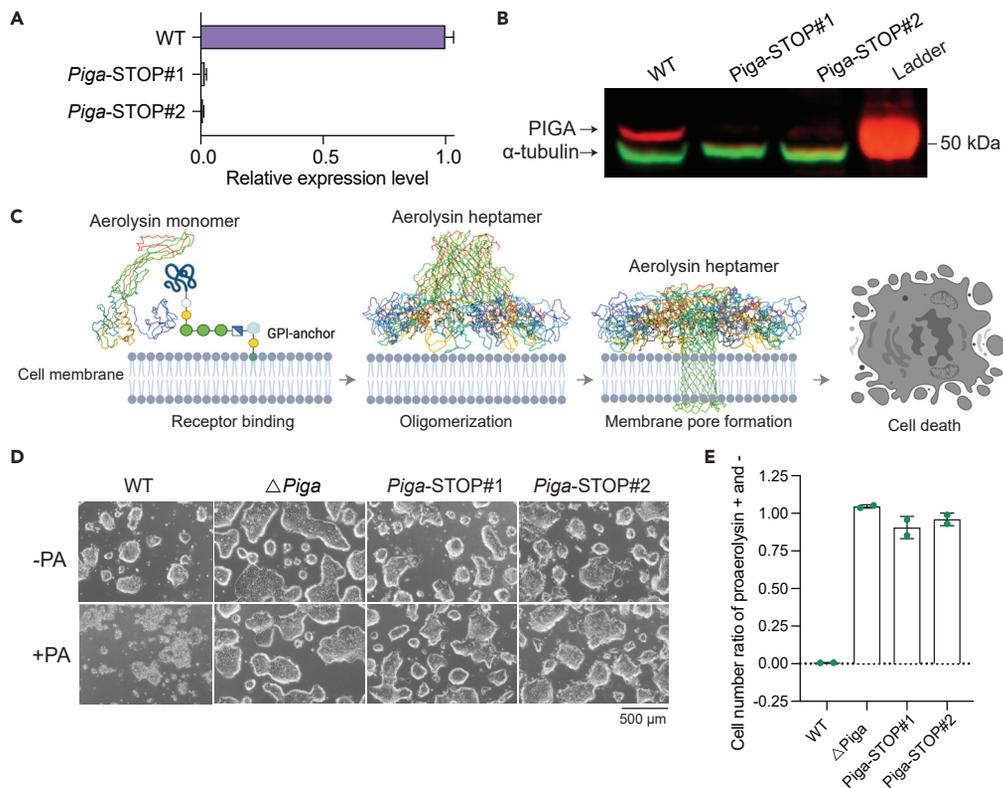


Figure 2. Evaluation of proaerolysin resistance in *Piga*-STOP mESCs

(A) RT-qPCR analysis of *Piga* mRNA expression in WT and *Piga*-STOP mESC lines. Values are normalized to *Actb* mRNA. Bars represent mean \pm SD of four technical replicates.

(B) Western blot analysis of PIGA in wild-type and *Piga*-STOP mESCs. α -tubulin serves as a loading control.

(C) Illustration of proaerolysin-mediated cell death; images were generated with bioRender.

(D) Bright-field images of mESCs after 24 h of 5 nM proaerolysin treatment.

(E) Cell viability comparison between proaerolysin-treated and untreated mESCs (5 nM proaerolysin for 24 h). Cell viability was quantified using PrestoBlue (see STAR Methods). Bars represent mean \pm SD of two replicates.

proaerolysin treatment, suggesting loss of cell content due to cell membrane perforation. By contrast, both $\Delta Piga$ and *Piga*-STOP mESCs showed no obvious morphological changes following proaerolysin treatment (Figure 2D). We then evaluated cell viability in proaerolysin-treated and untreated cells. The viability ratio between proaerolysin-treated and untreated cells reflects the degree of resistance to proaerolysin (with zero indicating full proaerolysin sensitivity and one indicating complete resistance). The *Piga*-STOP lines displayed nearly complete resistance to proaerolysin in comparison to the $\Delta Piga$ mESCs (Figure 2E), consistent with functional inactivation of PIGA by the STOP cassette.

The human *PIGA* minigene serves as a counterselectable marker in *Piga*-STOP mESCs

Next, we asked whether *Piga*-STOP mESCs can serve as a clean genetic background for proaerolysin-based counterselection. We first restored proaerolysin sensitivity by integrating a human *PIGA* transgene (*hPIGA*)-containing marker cassette into the mouse *Hprt* locus of *Piga*-STOP mESCs (Figure 3A). Cells were selected for the presence of the marker cassette with puromycin and for the loss of HPRT activity using 6-thioguanine (6-TG). The integration was validated by the presence of novel junctions between the integrated *hPIGA* marker cassette and *Hprt* flanking regions (Figure S2C, left). The presence of the STOP cassette in *Piga* was confirmed in these clones (Figure S2C, right). Capture sequencing of two *Piga*-STOP-*hPIGA* clones confirmed on-target cassette integration and no plasmid backbone integration (Figures 3B, 3C, S2D, and Table S1). To test whether we can select for *hPIGA* loss in the *Piga*-STOP background, we subsequently disrupted *hPIGA* by introducing two Cas9-gRNAs into *Piga*-STOP-*hPIGA* mESCs. We found that following “no DNA control” transfections, both *Piga*-STOP-*hPIGA* lines produced zero colonies after proaerolysin selection, suggesting the *hPIGA* transgene renders mESCs fully sensitive to

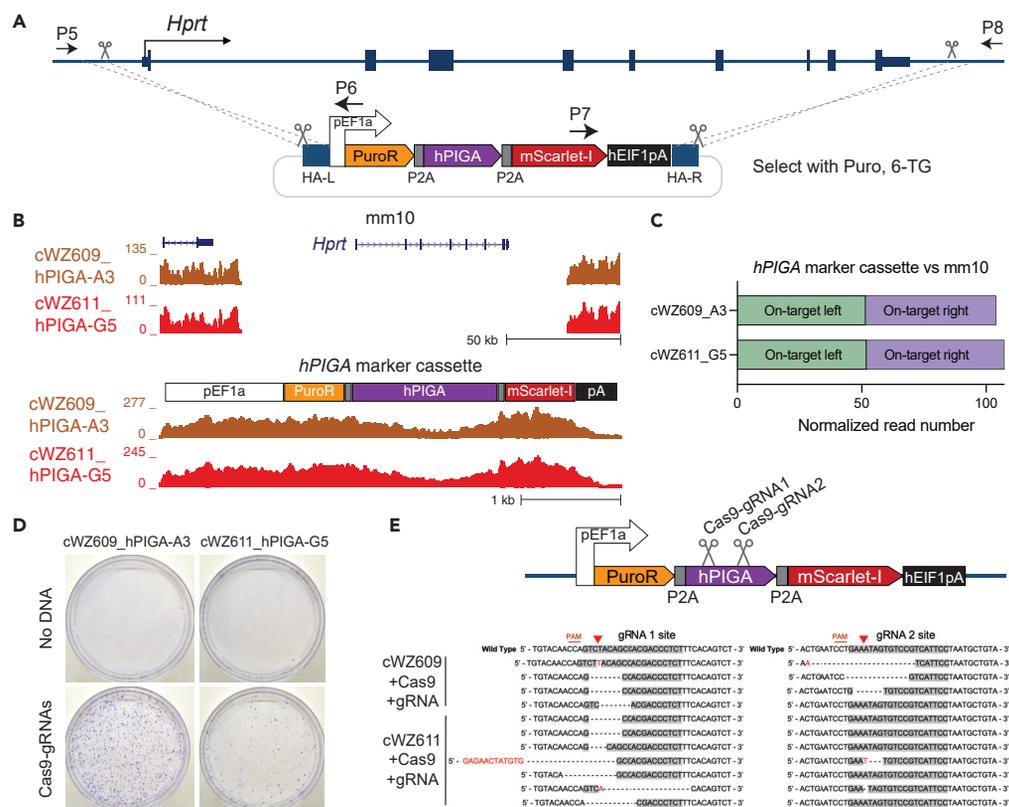


Figure 3. *hPIGA* integration and deletion

- (A) Schematic for integration of the *hPIGA* marker cassette into the *Hprt* locus.
 (B) Capture sequencing of two *Piga*-STOP-*hPIGA* clones. Reads were mapped to mm10 (top) or the marker cassette (bottom).
 (C) *Bamintersect* integration site analysis of junctions spanning *hPIGA* marker cassette and mm10.
 (D) Crystal violet staining of *Piga*-STOP-*hPIGA* mESC colonies with or without *hPIGA* targeting Cas9-gRNAs.
 (E) Sanger sequencing of *hPIGA* gRNA sites in Cas9-gRNA-transfected clones.

proaerolysin. In contrast, following the targeting of the *hPIGA* marker cassette by Cas9-gRNAs, hundreds of colonies were resistant to proaerolysin (Figure 3D). Furthermore, Sanger sequencing revealed that at least one of the two gRNA sites had indels (Figure 3E).

Piga restoration by efficient removal of STOP cassette

Given the essential role of *PIGA* in murine development, *Piga* expression necessitates restoration before generating viable mice from *Piga*-STOP mESCs. To facilitate this, we transiently expressed Cre recombinase in *Piga*-STOP mESCs to excise the STOP cassette (Figure 4A). mESCs were sparsely plated after transfecting Cre-expressing plasmid, allowing single cell-derived colony formation. We randomly selected 94 mESC colonies for PCR genotyping, and 58.5% of the colonies showed deletion of the STOP cassette (Figures 4B, S3A, and S3B). These results demonstrate that the removal of the STOP cassette is highly efficient, despite the extra integrated copy of the STOP cassette and its plasmid backbone (Figure 1D). Capture sequencing of *Piga*-restored clones revealed that the duplicated homology arm regions were excised together with the STOP cassette, as expected (Figure S3C), and Sanger sequencing of the excision-spanning PCR amplicons confirmed the expected 76 bp scar (Figure S3D). Robust *Piga* transcription was detected in STOP cassette removal lines by RT-qPCR (Figure 4C). Importantly, we found reacquisition of proaerolysin sensitivity in STOP cassette-excised mESCs (Figures 4D and 4E).

Applying the STOP cassette to *Hprt*

To demonstrate the general applicability of the conditional knockout strategy to other counterselectable markers in mESCs, we next sought to introduce the STOP cassette into the mouse hypoxanthine

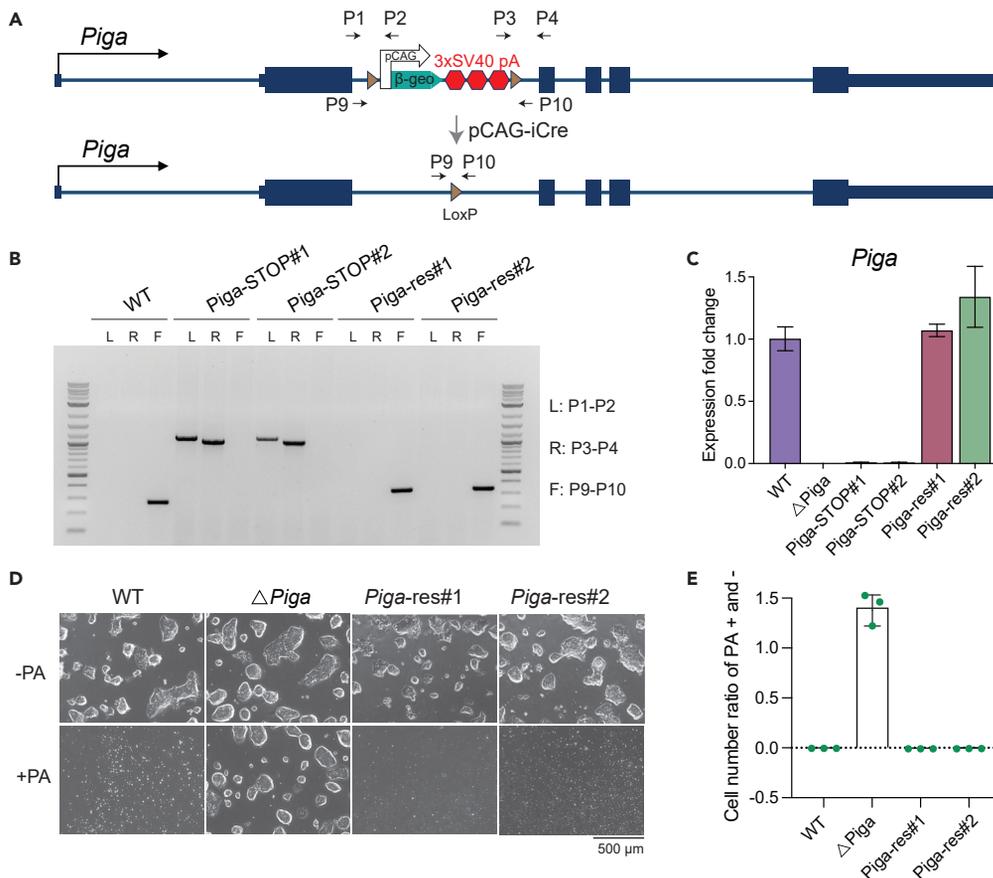


Figure 4. *Piga* restoration by Cre-mediated STOP cassette excision

(A) Schematic diagram showing *Piga* gene before and after STOP cassette excision.

(B) PCR genotyping of WT, *Piga*-STOP, and *Piga*-restoration clones. L, left junction of the STOP cassette; R, right junction of the STOP cassette; F, STOP excision. The expected size of the F product in pre-excision cells is 6.4 kb, which is challenging to amplify in this PCR assay. Ladder, 1 kb plus DNA ladder (NEB).

(C) RT-qPCR analysis of *Piga* expression in *Piga*-restored clones. Fold change was calculated by normalizing to WT. Bars represent mean \pm SD of four technical replicates.

(D) Bright field view of mESCs 24 h after proaerolysin (5 nM) treatment.

(E) Cell viability comparison (measured with PrestoBlue) between proaerolysin-treated and untreated mESCs. Bars represent mean \pm SD of three replicates.

phosphoribosyltransferase (*Hprt*) gene. The deletion of *Hprt* renders cells resistant to 6-thioguanine (6-TG), a purine analog that is converted by HPRT to the toxin 6-thioguanosine monophosphate. We determined the 6-TG concentration that eliminates all escapees (Figure S4A). Similar to the approach we used for *Piga*-STOP engineering, we precisely inserted the STOP cassette into the first intron of *Hprt* (chrX:52,992,027, mm10) (Figures 5A, S4B, and Table S1). A control 6-TG-resistant mESC line ($\Delta Hprt$) was generated by a complete *Hprt* knockout using paired CRISPR-Cas9 deletion (Figure S4C). We then treated WT, *Hprt*-STOP, and $\Delta Hprt$ mESCs with 6-TG, and found the *Hprt*-STOP mESCs displayed the same level of 6-TG resistance as $\Delta Hprt$ cells (Figure 5B), suggesting that the STOP cassette-mediated gene inactivation effectively eliminated *Hprt* expression.

DISCUSSION

Advances in genome writing demand the development of new tools to bridge the gap between DNA delivery to mESCs and eventual mouse model generation. Here, we have created a conditional *Piga* knockout in mESCs to enable the utilization of *Piga* as a counterselectable marker during genetic engineering. *Piga*-STOP mESCs demonstrate abrogated *Piga* mRNA levels and acquired proaerolysin resistance, similar to full *Piga* deletion lines. Following transient expression of Cre recombinase, we observed restoration of

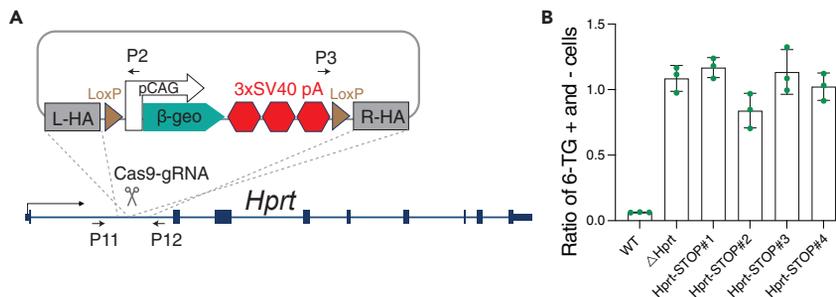


Figure 5. Engineering counterselectable *Hprt* knockout by STOP cassette

(A) Schematic of *Hprt*-STOP engineering.

(B) Cell viability comparison (measured with PrestoBlue) between 6-TG treated and untreated mESCs. Bars represent mean \pm SD of three replicates.

mRNA levels and proaerolysin sensitivity, consistent with normal PIGA function. Together, our findings suggest that this conditional counterselectable marker can be readily introduced to and removed from mESC genomes during the course of mESC engineering.

This approach represents a proof of concept for increasing the flexibility of mESC engineering. While our system does leave behind a modest-sized intronic scar (76 bp), theory and prior work suggest that small modifications to introns are unlikely to affect gene function (Lee and Rio, 2015; Yoshimatsu and Nagawa, 1989). Existing delivery technologies, such as inducible cassette exchange (ICE), routinely leave scars following DNA integration, and the resulting cell lines are able to generate mice (Iacovino et al., 2011; Wallace et al., 2007). Thus, we expect the restored gene to behave normally and support organismal development.

More broadly, we hope to adapt this approach to work with a variety of DNA delivery approaches. LoxP sites could be replaced with FRT sites or other loxP variants to improve compatibility with the Big-IN delivery system (Brosh et al., 2021), or with other DNA integration strategies that already use loxP sites for Cre-mediated recombination. We could further increase system flexibility by adding an inducible Cre segment within the STOP cassette, such as Cre-ERT2 (Indra et al., 1999). Subsequent tamoxifen treatment would trigger cassette removal and restoration of gene function, without an additional Cre recombinase introduction step. This approach is also feasible for studying noncoding RNA genes to achieve temporal control.

Conditional counterselectable markers may also prove valuable for genome engineering efforts in other species, including rats, flies, and other commonly used model organisms (Murata et al., 2012). As in mice, such tools may help strike a balance between ease of genome editing and downstream organismal viability. Additionally, conditional counterselectable markers could be used in other cell lines, such as human-induced pluripotent stem cells (iPSCs). While such cell lines may not give rise to full organisms, there may be benefits to restoring the natural regulatory environment surrounding counterselectable markers following cellular engineering, particularly when studying genes or pathways related to the counterselectable marker of interest.

As mammalian genome writing efforts expand, demand for tools to accelerate the pipeline from design and DNA synthesis to downstream applications will likely continue to grow. Ultimately, conditional counterselection systems like the one piloted here may become key components of the genome writing toolbox, accelerating our ability to study basic biology and model diseases.

Limitations of study

Here, we present a conditional knockout of two X-linked endogenous genes, *Piga* and *Hprt*, to enable proaerolysin and 6-TG-based counterselections in a male mESC line. We have discussed some limitations of this study, including the compatibility to recombinase-based big DNA delivery methods. The insertion and excision of the STOP cassette is simplified due to the single copy nature of X chromosome in male mESC. Biallelic insertion of the STOP cassette is required when implementing this strategy to autosomal genes, thus complicating the engineering process. Further optimization of the STOP cassette by introduction of a positive selection marker other from NeoR (for example PuroR), co-delivering both STOP cassettes with distinct selection markers will facilitate the selection of biallelic insertions.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- **KEY RESOURCES TABLE**
- **RESOURCE AVAILABILITY**
 - Lead contact
 - Materials availability
 - Data and code availability
- **EXPERIMENTAL MODELS AND SUBJECT DETAILS**
 - mESCs
- **METHOD DETAILS**
 - Plasmid construction
 - Nucleofection
 - Cell viability assay
 - RT-qPCR
 - Immunoblotting
 - Targeted capture sequencing
- **QUANTIFICATION AND STATISTICAL ANALYSIS**

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.isci.2022.104438>.

ACKNOWLEDGMENTS

We thank the NIH/NHGRI for supporting this work via CEGS grant 1RM1HG009491.

AUTHOR CONTRIBUTIONS

W.Z., R.B., and J.D.B. conceptualized the study. J.D.B., M.T.M. and S.Y.K. supervised the study. W.Z., Y.Z., H.A., G.E., and B.R.C. performed the experiments. H.A., G.E., and M.T.M. performed capture sequencing and data analysis. Manuscript was drafted by W.Z. and L.H.M., and reviewed and edited by R.B. and J.D.B.

DECLARATION OF INTERESTS

Jef Boeke is a Founder and Director of CDI Labs, Inc., a Founder of and consultant to Neochromosome, Inc, a Founder, SAB member of and consultant to Re-Open Diagnostics, LLC and serves or served on the Scientific Advisory Board of the following: Sangamo, Inc., Modern Meadow, Inc., Rome Therapeutics, Inc., Sample6, Inc., Tessera Therapeutics, Inc., and the Wyss Institute.

Received: September 4, 2021

Revised: March 18, 2022

Accepted: May 17, 2022

Published: June 17, 2022

REFERENCES

- Abrami, L., Fivaz, M., Decroly, E., Seidah, N.G., Jean, F., Thomas, G., Leppla, S.H., Buckley, J.T., and van der Goot, F.G. (1998). The pore-forming toxin proaerolysin is activated by furin. *J. Biol. Chem.* 273, 32656–32661. <https://doi.org/10.1074/jbc.273.49.32656>.
- Annaluru, N., Muller, H., Mitchell, L.A., Ramalingam, S., Stracquadanio, G., Richardson, S.M., Dymond, J.S., Kuang, Z., Scheifele, L.Z., Cooper, E.M., et al. (2014). Total synthesis of a functional designer eukaryotic chromosome. *Science* 344, 55–58. <https://doi.org/10.1126/science.1249252>.
- Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>.
- Brosh, R., Laurent, J.M., Ordoñez, R., Huang, E., Hogan, M.S., Hitchcock, A.M., Mitchell, L.A., Pinglay, S., Cadley, J.A., Luther, R.D., et al. (2021). A versatile platform for locus-scale genome rewriting and verification. *Proc. Natl. Acad. Sci. U S A* 118. <https://doi.org/10.1073/pnas.2023952118>.
- Concordet, J.-P., and Haeussler, M. (2018). CRISPOR: intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens. *Nucleic Acids Res.* 46, W242–W245. <https://doi.org/10.1093/nar/gky354>.
- Dymond, J., and Boeke, J. (2012). The *Saccharomyces cerevisiae* SCRaMbLE system and genome minimization. *Bioeng. Bugs* 3, 170–173. <https://doi.org/10.4161/bbug.19543>.
- Dymond, J.S., Richardson, S.M., Coombes, C.E., Babatz, T., Muller, H., Annaluru, N., Blake, W.J., Schwerzmann, J.W., Dai, J., Lindstrom, D.L., et al. (2011). Synthetic chromosome arms function in yeast and generate phenotypic diversity by design. *Nature* 477, 471–476. <https://doi.org/10.1038/nature10403>.
- Elshami, A.A., Saavedra, A., Zhang, H., Kucharczuk, J.C., Spray, D.C., Fishman, G.I., Amin, K.M., Kaiser, L.R., and Albelda, S.M. (1996). Gap junctions play a role in the “bystander effect”

of the herpes simplex virus thymidine kinase/ganciclovir system in vitro. *Gene Ther.* 3, 85–92.

Faust, G.G., and Hall, I.M. (2014). SAMBLASTER: fast duplicate marking and structural variant read extraction. *Bioinformatics* 30, 2503–2505. <https://doi.org/10.1093/bioinformatics/btu314>.

Fredens, J., Wang, K., de la Torre, D., Funke, L.F.H., Robertson, W.E., Christova, Y., Chia, T., Schmied, W.H., Dunkelmann, D.L., Beránek, V., et al. (2019). Total synthesis of *Escherichia coli* with a recoded genome. *Nature* 569, 514–518. <https://doi.org/10.1038/s41586-019-1192-5>.

Gibson, D.G., Benders, G.A., Andrews-Pfannkoch, C., Denisova, E.A., Baden-Tillson, H., Zaveri, J., Stockwell, T.B., Brownley, A., Thomas, D.W., Algire, M.A., et al. (2008). Complete chemical synthesis, assembly, and cloning of a *Mycoplasma genitalium* genome. *Science* 319, 1215–1220. <https://doi.org/10.1126/science.1151721>.

Gibson, D.G., Glass, J.I., Lartigue, C., Noskov, V.N., Chuang, R.-Y., Algire, M.A., Benders, G.A., Montague, M.G., Ma, L., Moodie, M.M., et al. (2010). Creation of a bacterial cell controlled by a chemically synthesized genome. *Science* 329, 52–56. <https://doi.org/10.1126/science.1190719>.

Han, Y., Lin, Y.B., An, W., Xu, J., Yang, H.-C., O'Connell, K., Dordai, D., Boeke, J.D., Siliciano, J.D., and Siliciano, R.F. (2008). Orientation-dependent regulation of integrated HIV-1 expression by host gene transcriptional read-through. *Cell Host Microbe* 4, 134–146. <https://doi.org/10.1016/j.chom.2008.06.008>.

Iacovache, I., Degiacomi, M.T., Pernot, L., Ho, S., Schiltz, M., Dal Peraro, M., and van der Goot, F.G. (2011). Dual chaperone role of the C-terminal propeptide in folding and oligomerization of the pore-forming toxin aerolysin. *PLoS Pathog.* 7, e1002135. <https://doi.org/10.1371/journal.ppat.1002135>.

Iacovache, I., De Carlo, S., Cirauqui, N., Dal Peraro, M., van der Goot, F.G., and Zuber, B. (2016). Cryo-EM structure of aerolysin variants reveals a novel protein fold and the pore-formation process. *Nat. Commun.* 7, 12062. <https://doi.org/10.1038/ncomms12062>.

Iacovino, M., Bosnakovski, D., Fey, H., Rux, D., Bajwa, G., Mahen, E., Mitanoska, A., Xu, Z., and Kyba, M. (2011). Inducible cassette exchange: a rapid and efficient system enabling conditional gene expression in embryonic stem and primary cells. *Stem Cell.* 29, 1580–1588. <https://doi.org/10.1002/stem.715>.

Indra, A.K., Warot, X., Brocard, J., Bornert, J.-M., Xiao, J.-H., Chambon, P., and Metzger, D. (1999). Temporally-controlled site-specific mutagenesis in the basal layer of the epidermis: comparison of the recombinase activity of the tamoxifen-inducible Cre-ERT and Cre-ERT2 recombinases. *Nucleic Acids Res.* 27, 4324–4327. <https://doi.org/10.1093/nar/27.22.4324>.

Lee, Y., and Rio, D.C. (2015). Mechanisms and regulation of alternative pre-mRNA splicing.

Annu. Rev. Biochem. 84, 291–323. <https://doi.org/10.1146/annurev-biochem-060614-034316>.

Lee, E.-C., Liang, Q., Ali, H., Bayliss, L., Beasley, A., Bloomfield-Gerdes, T., Bonoli, L., Brown, R., Campbell, J., Carpenter, A., et al. (2014). Complete humanization of the mouse immunoglobulin loci enables efficient therapeutic antibody discovery. *Nat. Biotechnol.* 32, 356–363. <https://doi.org/10.1038/nbt.2825>.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>.

Li, D., Sun, X., Yu, F., Perle, M.A., Araten, D., and Boeke, J.D. (2021). Application of counter-selectable marker PIGA in engineering designer deletion cell lines and characterization of CRISPR deletion efficiency. *Nucleic Acids Res.* 49, 2642–2654. <https://doi.org/10.1093/nar/gkab035>.

Mesnil, M., Piccoli, C., Tiraby, G., Willecke, K., and Yamasaki, H. (1996). Bystander killing of cancer cells by herpes simplex virus thymidine kinase gene is mediated by connexins. *Proc. Natl. Acad. Sci. U S A* 93, 1831–1835. <https://doi.org/10.1073/pnas.93.5.1831>.

Mitchell, L.A., Wang, A., Stracquadanio, G., Kuang, Z., Wang, X., Yang, K., Richardson, S., Martin, J.A., Zhao, Y., Walker, R., et al. (2017). Synthesis, debugging, and effects of synthetic chromosome consolidation: synVI and beyond. *Science* 355, eaaf4831. <https://doi.org/10.1126/science.aaf4831>.

Mitchell, L.A., McCulloch, L.H., Pinglay, S., Berger, H., Bosco, N., Brosh, R., Bulajić, M., Huang, E., Hogan, M.S., Martin, J.A., et al. (2021). De novo assembly and delivery to mouse cells of a 101 kb functional human gene. *Genetics* 218. <https://doi.org/10.1093/genetics/iyab038>.

Miyata, T., Takeda, J., Iida, Y., Yamada, N., Inoue, N., Takahashi, M., Maeda, K., Kitani, T., and Kinoshita, T. (1993). The cloning of PIG-A, a component in the early step of GPI-anchor biosynthesis. *Science* 259, 1318–1320. <https://doi.org/10.1126/science.7680492>.

Murata, D., Nomura, K.H., Dejima, K., Mizuguchi, S., Kawasaki, N., Matsuishi-Nakajima, Y., Ito, S., Gengyo-Ando, K., Kage-Nakadai, E., Mitani, S., et al. (2012). GPI-anchor synthesis is indispensable for the germline development of the nematode *Caenorhabditis elegans*. *Mol. Biol. Cell* 23, 982–995. <https://doi.org/10.1091/mbc.e10-10-0855>.

Neph, S., Kuehn, M.S., Reynolds, A.P., Haugen, E., Thurman, R.E., Johnson, A.K., Rynes, E., Maurano, M.T., Vierstra, J., Thomas, S., et al. (2012). BEDOPS: high-performance genomic feature operations. *Bioinformatics* 28, 1919–1920. <https://doi.org/10.1093/bioinformatics/bts277>.

Ran, F.A., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A., and Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* 8, 2281–2308. <https://doi.org/10.1038/nprot.2013.143>.

Richardson, S.M., Mitchell, L.A., Stracquadanio, G., Yang, K., Dymond, J.S., DiCarlo, J.E., Lee, D., Huang, C.L.V., Chandrasegaran, S., Cai, Y., et al. (2017). Design of a synthetic yeast genome. *Science* 355, 1040–1044. <https://doi.org/10.1126/science.aaf4557>.

Rosti, V., Tremml, G., Soares, V., Pandolfi, P.P., Luzzatto, L., and Bessler, M. (1997). Murine embryonic stem cells without pig-a gene activity are competent for hematopoiesis with the PNH phenotype but not for clonal expansion. *J. Clin. Invest.* 100, 1028–1036. <https://doi.org/10.1172/JCI119613>.

Saito, M., Iwakaki, T., Taya, C., Yonekawa, H., Noda, M., Inui, Y., Mekada, E., Kimata, Y., Tsuru, A., and Kohno, K. (2001). Diphtheria toxin receptor-mediated conditional and targeted cell ablation in transgenic mice. *Nat. Biotechnol.* 19, 746–750. <https://doi.org/10.1038/90795>.

Shen, Y., Wang, Y., Chen, T., Gao, F., Gong, J., Abramczyk, D., Walker, R., Zhao, H., Chen, S., Liu, W., et al. (2017). Deep functional analysis of synII, a 770-kilobase synthetic yeast chromosome. *Science* 355, eaaf4791. <https://doi.org/10.1126/science.aaf4791>.

Wallace, H.A.C., Marques-Kranc, F., Richardson, M., Luna-Crespo, F., Sharpe, J.A., Hughes, J., Wood, W.G., Higgs, D.R., and Smith, A.J.H. (2007). Manipulating the mouse genome to engineer precise functional syntenic replacements with human sequence. *Cell* 128, 197–209. <https://doi.org/10.1016/j.cell.2006.11.044>.

Wang, K., Fredens, J., Brunner, S.F., Kim, S.H., Chia, T., and Chin, J.W. (2016). Defining synonymous codon compression schemes by genome recoding. *Nature* 539, 59–64. <https://doi.org/10.1038/nature20124>.

Wu, Y., Li, B.-Z., Zhao, M., Mitchell, L.A., Xie, Z.-X., Lin, Q.-H., Wang, X., Xiao, W.-H., Wang, Y., Zhou, X., et al. (2017). Bug mapping and fitness testing of chemically synthesized chromosome X. *Science* 355, eaaf4706. <https://doi.org/10.1126/science.aaf4706>.

Xie, Z.-X., Li, B.-Z., Mitchell, L.A., Wu, Y., Qi, X., Jin, Z., Jia, B., Wang, X., Zeng, B.-X., Liu, H.-M., et al. (2017). Perfect[™] designer chromosome V and behavior of a ring derivative. *Science* 355, eaaf4704. <https://doi.org/10.1126/science.aaf4704>.

Yoshimatsu, T., and Nagawa, F. (1989). Control of gene expression by artificial introns in *Saccharomyces cerevisiae*. *Science* 244, 1346–1348. <https://doi.org/10.1126/science.2544026>.

Zhang, W., Zhao, G., Luo, Z., Lin, Y., Wang, L., Guo, Y., Wang, A., Jiang, S., Jiang, Q., Gong, J., et al. (2017). Engineering the ribosomal DNA in a megabase synthetic chromosome. *Science* 355. <https://doi.org/10.1126/science.aaf3981>.

Zhang, W., Mitchell, L.A., Bader, J.S., and Boeke, J.D. (2020). Synthetic genomes. *Annu. Rev. Biochem.* 89, 77–101. <https://doi.org/10.1146/annurev-biochem-013118-110704>.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Anti-PIGA antibody	Abcam	Cat# ab69768, RRID: AB_1566593
Mouse monoclonal anti-alpha-Tubulin antibody	Sigma-Aldrich	Cat# T5168, RRID: AB_477579
Chemicals, peptides, and recombinant proteins		
Crystal violet solution	Sigma-Aldrich	V5265-500ML
LightCycler 480 SYBR green I master	Roche	04887352001
Geneticin (G418 sulfate)	Gibco	10131027
Critical commercial assays		
PrestoBlue cell viability assay	Invitrogen	A13261
Mouse ES cell nucleofection kit	Lonza	VPH-1001
NEBNext Ultra II FS DNA library prep kit	NEB	E7645S
SuperScript IV reverse transcriptase	Invitrogen	18090200
NextSeq High-output 75-cycle V2.5 Kit	Illumina	20024906
Experimental models: Cell lines		
C57BL/6J (MK6) mouse embryonic stem cells	NYU Langone Health Rodent Genetic Engineering Laboratory	N/A
Recombinant DNA		
Plasmid: STOP cassette	This study	N/A
Plasmid: pX330-U6-Chimeric_BB-CBh-hSpCas9	Addgene	42230
pCAG-iCre	Addgene	89573
Plasmid: Puro-hPIGA-mScarlet	This study	N/A
Software and algorithms		
Gen5	BioTek	N/A
CRISPOR	Concordet and Haeussler (2018)	http://crispor.tefor.net/
Image Studio Lite	LI-COR	https://www.licor.com/bio/image-studio/
Trimmomatic v0.39	Bolger et al. (2014)	https://github.com/usadellab/Trimmomatic
BWA v0.7.17	Li and Durbin (2009)	https://github.com/sghignone/bwa
samblaster v0.1.24	Faust and Hall (2014)	https://github.com/GregoryFaust/samblaster
BEDOPS v2.4.35	Neph et al. (2012)	https://github.com/bedops/bedops/releases/tag/v2.4.35
Bamintersect	Brosh et al. (2021)	https://github.com/mauranolab/mapping/tree/master/dnase/bamintersect

RESOURCE AVAILABILITY

Lead contact

Further information should be directed to and will be fulfilled by the lead contact Jef D. Boeke (jef.boeke@nyulangone.org).

Materials availability

All unique/stable reagents generated in this study are available from the [Lead contact](#) with a completed Materials Transfer Agreement.

Data and code availability

Data: All data reported in this paper will be shared by the [Lead contact](#) upon request.

Code: This work did not generate any code.

Any additional information required to reanalyze the data reported in this paper is available from the [Lead contact](#) upon request.

EXPERIMENTAL MODELS AND SUBJECT DETAILS

mESCs

C57BL/6J male mESCs (MK6) were provided by NYU Langone Health, Rodent Genetic Engineering Laboratory. mESCs were cultured in 80/20 medium [80% of 2i medium (2i basal medium supplement with 3 μ M CHIR99021 and 1 μ M PD0325901) mixed 20% of ES medium [Knockout DMEM, 15% fetal bovine serum, 1% MEM Non-Essential Amino Acids, 10³ U/mL Leukemia Inhibitory Factor, 0.1 mM 2-mercaptoethanol, 1% Pen-Strep]) on plates coated with 0.1% gelatin (EMD Millipore, ES-006-B). mESCs were grown in a humidified tissue culture incubator at 37°C, 5% CO₂. Medium was exchanged daily.

METHOD DETAILS

Plasmid construction

The STOP cassette plasmid is derived from pWA203 used in a previous study (Han et al., 2008). The STOP cassette fragment was released from pWA203 by *AscI* and *KpnI* digestion. *Piga* and *Hprt* homology arms were amplified from mouse BACs RP23-32H22 and RP23-412J16 using primers shown in Table S2. Gibson assembly was used to construct the STOP cassette with flanking homology arms into the multiple cloning site of pRS413 vector. Puro-hPIGA-mScarlet plasmid was constructed by stitching all the components together using Gibson assembly enzyme mix. Guide RNAs (gRNAs, listed in Table S3) were chosen using the CRISPOR (crispor.tefor.net/) online tool, and subsequently cloned into pX330 (Addgene, 42230) using Golden Gate assembly (Ran et al., 2013). The pCAG-iCre plasmid was purchased from Addgene (89573).

Nucleofection

The mouse embryonic stem cell nucleofector kit (Lonza, VPH-1001) was used for the nucleofection. mESCs were harvested by trypsinization followed by a PBS wash and were counted using a hemocytometer. Approximately 2 \times 10⁶ cells were used for each reaction. A total of 10 μ g DNA was mixed with mESCs prior to the nucleofection. The nucleofection was conducted using the Lonza 2b Nucleofector, program A-023. mESCs were resuspended in culture medium after the nucleofection and plated immediately onto gelatin-coated 10 cm plates.

Cell viability assay

Approximately 1 \times 10⁵ mESCs were seeded in a 24 well plate. After growing for 24 h, mESCs were treated either with 5 nM of proaerolysin for 24 h or 5 μ M of 6-thioguanine for 3 days. Both treated and untreated cells were harvested by trypsinization. mESCs were resuspended in 1 mL of 80/20 medium, and 180 μ L cell suspension was mixed with 20 μ L of PrestoBlue reagent (Invitrogen, A13261) in a 96 well plate for 30 min at 37°C, 5% CO₂. Absorbance was read using a plate reader following manufacturer's instructions (Biotek, synergy H1).

RT-qPCR

mESCs were harvested by trypsinization. Total RNA was isolated using the RNeasy kit (QIAGEN, 74106) following the manufacturer's instructions. 1 μ g of total RNA was used for reverse transcription (Invitrogen, 18090200). 1 μ L of 10-fold diluted cDNA was used as template in a 10 μ L SYBR (Roche, 04887352001) qPCR reaction on a LightCycler 480 instrument. Relative expression was calculated using the $\Delta\Delta$ CT method.

Immunoblotting

mESCs were harvested by trypsinization and pellets were resuspended in 200 μ L RIPA buffer supplemented with protease inhibitor cocktail (Roche, 11873580001). Cell suspensions were then incubated on ice for 30 min. After the on-ice incubation, 1 μ L Benzonase (Sigma-Aldrich, E1014) was added to the cell lysate for a 10 min 37°C digestion of genomic DNA. Protein lysates were mixed with LDS loading buffer (Invitrogen, NP0007), and boiled at 70°C for 10 min. Proteins were separated using a NuPAGE 4–12% Bis-Tris gel (Life Technologies, WG1402BOX) and transferred to PVDF membranes (Millipore, IPFL00010), which were blotted with PIGA (Abcam, ab69768) and Tubulin (Sigma-Aldrich, T5168) primary antibodies, followed by

goat anti-mouse (LI-COR Biosciences, 926-32210) and goat anti-rabbit (LI-COR Biosciences, 926-68071) secondary antibodies. Blots were visualized using the LI-COR system.

Targeted capture sequencing

Targeted capture sequencing was performed as previously described (Brosh et al., 2021). In brief, mESC genomic DNA was isolated with QIAamp gDNA extraction kit (QIAGEN, 51306) following the manufacturer's instructions. 1 μ g of genomic DNA was used as input for the large fragment size (>550 bp) DNA library preparation (NEBNext Ultra II FS). Mouse BACs RP23-32H22 and RP23-412J16 (BACPAC Resources Center), STOP cassette plasmid, Puro-hPIGA-mScarlet and Cas9-gRNA plasmids were used for nick translation to generate the biotinylated bait set. Captured DNA library was sequenced using a NextSeq 500 75 cycles high output kit (Illumina, 20024906).

QUANTIFICATION AND STATISTICAL ANALYSIS

For RT-qPCR assay, data were collected from four technical replicates. $\Delta\Delta$ CT method was used for computing the relative expression level. For cell viability assays, two (Figure 2E) or three (Figures 4E and 5B) wells of cells were used as replicates. All graphs were created in GraphPad Prism 9.