2642–2654 *Nucleic Acids Research, 2021, Vol. 49, No. 5* *Published online 16 February 2021*
doi: 10.1093/nar/gkab035

Application of counter-selectable marker PIGA in engineering designer deletion cell lines and characterization of CRISPR deletion efficiency

Donghui Li[1,2], Xiaoji Sun[1], Fangzhou Yu[1], Mary Ann Perle[3], David Araten[4] and Jef D. Boeke[1,5,*]

[1]Institute for Systems Genetics and Department of Biochemistry & Molecular Pharmacology, NYU Langone Health, New York, NY 10016, USA, [2]McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA, [3]Department of Pathology, NYU Langone Health, New York, NY 10016, USA, [4]Division of Hematology, Laura and Isaac Perlmutter Cancer Center, NYU Langone Health, and the New York VA Medical Center, New York, NY 10016, USA and [5]Department of Biomedical Engineering, NYU Tandon School of Engineering, Brooklyn NY 11201, USA

Received September 08, 2020; Revised January 11, 2021; Editorial Decision January 11, 2021; Accepted February 15, 2021

ABSTRACT

The CRISPR/Cas9 system is a technology for genome engineering, which has been applied to indel mutations in genes as well as targeted gene deletion and replacement. Here, we describe paired gRNA deletions along the *PIGA* locus on the human X chromosome ranging from 17 kb to 2 Mb. We found no compelling linear correlation between deletion size and the deletion efficiency, and there is no substantial impact of topologically associating domains on deletion frequency. Using this precise deletion technique, we have engineered a series of designer deletion cell lines, including one with deletions of two X-chromosomal counterselectable (negative selection) markers, *PIGA* and *HPRT1*, and additional cell lines bearing each individual deletion. *PIGA* encodes a component of the glycosylphosphatidylinositol (GPI) anchor biosynthetic apparatus. The *PIGA* gene counterselectable marker has unique features, including existing single cell level assays for both function and loss of function of *PIGA* and the existence of a potent counterselectable agent, proaerolysin, which we use routinely for selection against cells expressing *PIGA*. These designer cell lines may serve as a general platform with multiple selection markers and may be particularly useful for large scale genome engineering projects such as Genome Project-Write (GP-write).

INTRODUCTION

Genome engineering has become an essential study tool for fundamental research and industry, including the targeted engineering of genes, biosynthetic pathways, and networked systems analysis. Such designer genomic manipulations rely on selectable markers to distinguish successfully targeted cells from a much larger number of wild-type cells. Since the number of useful selectable markers discovered thus far is limited, reusing markers that feature counterselection (also known as negative selection) is extremely helpful for multistep engineering. Counterselectable markers have been extensively applied in yeast genetics since the exploitation of 5-fluoroorotic acid (5-FOA) counterselection against the *URA3* gene, which confers uracil prototrophy and thus makes *URA3* both selectable and counterselectable (1). In addition to marker reuse and sequential delivery, counterselection contributes to removing auxiliary genes, cassettes and other engineered DNAs whenever this may be desirable.

Genome engineering and stem cell reprogramming may require transient transfection of auxiliary genes (like CRISPR) for permanent genomic modification (2) or modification of cell fate (3). Sometimes, for efficient reprogramming or for therapeutic purposes, these systems are also packaged on an episomal vector, allowing for sufficiently long-term expression. First, selection for cells that have undergone stable genome modification is valuable. Second, it may also be important to subsequently selectively lose the auxiliary genes or the whole episomal vector, to eliminate potential toxicity, e.g. ultimately, during clinical treatments.

*To whom correspondence should be addressed. Tel: 646 501 0504; Email: jef.boeke@nyulangone.org
Present addresses:
Donghui Li, Tessera Therapeutics, Cambridge, MA, USA.
Xiaoji Sun, Cellarity, Cambridge, MA, USA.
Fangzhou Yu, University of the Pacific Arthur A. Dugoni School of Dentistry, San Francisco, CA, USA.

For use in mammalian cells, a number of natural and fusion genes serve the purpose of counterselection, like the thymidine kinase (*Tk*) gene of HSV (4), the human hypoxanthine-guanine phosphoribosyl transferase gene (*HPRT1*) (5), and the fusion constructs of various positive selection markers with the *Tk* gene or the *codA* (cytosine deaminase) gene of *Escherichia coli* (6–9). The further development of counterselectable markers with potent selection is needed to facilitate the expanding needs of gene targeting and genomic rearrangement research.

The human *PIGA* gene is a suitable candidate for a counterselectable marker, which encodes phosphatidylinositol *N*-acetylglucosaminyltransferase subunit A (PIG-A), the catalytic subunit of the multiprotein complex that catalyzes the first step of glycosylphosphatidylinositol (GPI) anchor synthesis (10–12). The GPI anchor is attached post-translationally to various eukaryotic proteins to direct them to the cell membrane where it provides a high degree of membrane fluidity relative to anchoring the protein directly into the membrane and protecting the attached proteins from most extracellular proteases and lipases (13,14). Of the 22 genes encoding proteins that are involved in GPI anchor synthesis, only *PIGA* is located on the X chromosome in mammals. Thus 'single hit' mutations in *PIGA* disrupt GPI anchor synthesis, resulting in a general deficiency of GPI-anchored proteins, a phenotype which can be detected and quantified by flow cytometry. Flow sorting can also be used to collect live cells based on their expression of GPI-linked proteins (15). Flow cytometry is used to diagnose the disease paroxysmal nocturnal hemoglobinuria (PNH), characterized by the deficiency of GPI-linked proteins on the surface of blood cells caused by clonal expansion of hematopoietic cells with acquired somatic mutations in *PIGA* (16). A broad spectrum of somatic mutations in the *PIGA* gene from PNH patients has been reported (17), including small a 737 bp deletion, large deletions ∼4 kb (18,19) and chromosome aberrations ∼500 kb spanning *PIGA* (20). In addition to flow cytometry-based cell sorting, GPI-deficient cells can be selected for using proaerolysin, a protoxin form of aerolysin, a cytolytic pore-forming toxin produced by *Aeromonas hydrophila* that causes cell death by forming transmembrane pores and cell content release (21). Proaerolysin binds to the GPI anchor and is proteolytically cleaved into active aerolysin. All these features make *PIGA* a well-suited counterselectable marker. In addition to the diagnosis of PNH, published applications of *PIGA* include in vivo mutation assays (22,23), quantitative measurement of somatic mutation rates (15), and monitoring of gene targeting efficiency (24). Moreover, the mutation assays of *PIGA* have been extended to in vitro mammalian cells for chemically-induced mutagenesis and safety assessment in the pharmaceutical and consumer product industries (25).

Since *PIGA* is an endogenous human gene, a marker-deficient background is required in order to carry out negative selection. We conducted paired gRNA CRISPR deletion, to remove endogenous *PIGA*, in which two independent constructs of Cas9 coupled with gRNA were co-transfected to create two double strand breaks (DSBs) around *PIGA*, which are subsequently repaired by non-homologous end joining (NHEJ). While performing gene targeting, PIGA cDNA can be used as a positive selection marker (26). Using the paired CRISPR deletion tool, we engineered a series of *PIGA* deficient human HCT116+ch3 cell lines to establish a genetic background that prepares the cell line for large-scale genome rearrangements. HCT116 is a human male colon cancer cell line. Different from other massively rearranged cancer genomes, HCT116 is nearly diploid, making it well-suited for genome rearrangement targeting. The extra chromosome 3 (ch3) in HCT116+ch3 was previously introduced to complement the microsatellite instability and mismatch repair deficiency resulting from a homozygous mutation in the mismatch repair gene *MLH1* on chromosome 3 (27), resulting in a cell line that is predicted to be much more stable genomically than the parental HCT116 line.

In this study, we also employed a quantitative assay based on the counterselection associated with *PIGA* to measure paired gRNA based deletion efficiency in HCT116+ch3. This assay is distinct from single gRNA CRISPR/Cas9 based NHEJ, often used to inactivate genes through INDELs in the coding region. In the paired gRNA assay, CRISPR deletion can be used for a wide variety of purposes, such as making nonrevertable mutations, completely eliminating entire native genes with their noncoding elements, or eliminating noncoding RNA gene regions. Previous studies reported robust efficiency of paired gRNA CRISPR deletion in multiple mammalian cell lines, targeting different loci (28–30) but systematic studies are limited. We sought to understand whether there was a correlation of deletion efficiency with intended deletion size, as there was disagreement between prior studies (29,30). Our assay was designed to be performed at and flanking the *PIGA* locus in the nearly diploid HCT116+ch3 cell line, with one gRNA always anchored at a consistent location (the 'anchor' gRNA), and paired with various 'partner' gRNAs positioned at different distances, allowing for a systematic analysis of paired gRNA CRISPR deletion efficiency.

We designed the correlation analysis of deletion efficiency not only by examining inter-gRNA distance but also by considering potential 3D topological intracellular associations between them. Recent technological advancements in Hi-C and related methods have enabled the study of the hierarchy and dynamics of chromatin domains in a genome manner at a high resolution. Hi-C has helped generate genome-wide DNA interaction maps for different cell types and organisms, defining discrete self-interacting genomic regions in the genome, termed topologically associated domains (TADs) (31). DNA sequences within a TAD physically interact with each other more frequently than with sequences outside TADs, and TADs are bordered by low interaction regions called TAD boundaries (32). Based on published Hi-C data for the closely related HCT116 cell line, we were able to generate a heat map of the target region to allow combined evaluation of paired CRISPR gRNA efficiencies and assess the impact of local TAD boundary analysis. Surprisingly, we conclude from these experiments that TAD boundaries do not interfere significantly with deletion formation.

MATERIALS AND METHODS

gRNA design and plasmid construction

The target gRNAs for deletion efficiency analysis were selected using the website http://crispor.tefor.net/crispor.py/ by feeding 2000 bp sequence of the targeted region to the website and selecting at least two gRNAs. The target gRNAs for engineering cell lines were selected using the online tool listed in the CRISPR/Cas9 protocol (33). Following this protocol, gRNAs were ordered as oligos and cloned into the human codon optimized Cas9 plasmid. gRNAs for cell line engineering and deletion efficiency analysis at *PIGA* locus are summarized in Supplementary Table S1. gRNAs for deletion efficiency analysis at *HPRT1* locus are summarized in Supplementary Table S2.

Cell culture and transfection for HCT116+ch3 cell line

HCT116+ch3 cells were maintained in DMEM/F12, HEPES (ThermoFisher Scientific, catalog # 11330032) supplemented with 10% fetal bovine serum (FBS) and 400 μg/ml active G418 (Geneticin). Cells were seeded in 6-well plates (30,000 cells/well) the day before transfection. The next day, paired gRNA plasmids (750μg of each) were co-transfected using 3 μl Fugene HD (Promega, Madison WI) according to the manufacturer's protocol into each well. The day after transfection, DMEM/F12 medium with 0.6 μg/ml puromycin was applied for 48 hours. Three biological repeats of each paired CRISPR deletion transfections were done. Cell lines used and generated in this study are summarized in Supplementary Table S3.

Deletion junction analysis

Genomic DNA (gDNA) was extracted 3 d after CRISPR transfection with QIAamp DNA mini Kit (Qiagen, Valencia, CA, USA). To detect genomic deletions, gDNA was subjected to PCR analysis using GoTaq Green Master Mix (Promega) and appropriate primers (Supplementary Table S4). PCR products were analyzed by agarose gel electrophoresis. For sequencing analysis, PCR product bands corresponding to genomic deletions were cut and purified using Zymoclean Gel DNA Recovery Kit (Zymo Research) and cloned into pCR™4-TOPO® TA vector with TOPO™ TA Cloning™ Kit for Sequencing (ThermoFisher Scientific). Cloned plasmids were sequenced using M13 primers.

GPI staining and Flow cytometry

Ten days post-transfection, cells were detached and stained in 24 μl DMEM/F12, HEPES (ThermoFisher Scientific, catalog # 11330032) with 3μl HLA-ABC PE antibody (ThermoFisher Scientific, Catalog # 12-9983-42) and 3 μl CD59-FITC antibody (BioRad, catalog # MCA1054F) for 30 min on ice, cells were washed with PBS twice. The frequency of GPI (−) cells from each deletion reaction was evaluated by FACS using BD Accuri™ C6 flow cytometer. HCT116+ch3 WT and HCT116+ch3 *piga*Δ cells were used for gating. The result of three repeated experiments was used to calculate the average frequency and standard deviation. The fcs files of the flow are available at https://flowrepository.org/id/FR-FCM-Z3BN.

Mini-PIGA expression plasmid construction

The human *PIGA* expression plasmids pDL365 (TK promoter) and pDL366 (native *PIGA* promoter) were first made by the use of Gibson assembly to combine the promoter and *PIGA* cDNA into the pCEP4 plasmid (XhoI/BamHI) which contains the SV40 polyadenylation signal. In order to exchange the human *PIGA* cDNA with other *PIGA* ortholog cDNAs, a BstZ17I site was introduced into pDL365 between the promoter and the CDS to make the parent plasmid, pDL371. Protein sequences of *Mus musculus* (mouse, AAI38759.1), *Macaca mulatta* (monkey, NP_001247532.1), *Drosophila melanogas*ter (fruit fly, XP_020815746.1) and *Plasmodium falciparum* (malaria parasite, PKC43453.1) *PIGA* genes were acquired from NCBI. Human codon-optimized cDNA sequences of each ortholog were synthesized by a commercial provider. The cDNA was then inserted into the parent plasmid by Gibson assembly after cutting host plasmid pDL371 with BstZ17I and BamHI. Plasmids are summarized in Supplementary Table S5.

Drug selection

In order to isolate the HCT116+ch3 *piga*Δ cell line, cells from paired CRISPR deletion with gRNAs 5′ USS and 3′ DSS were trypsinized and resuspended in culture medium with 5nM proaerolysin 7 days post transfection. After 24 h of treatment, dead cells were removed and fresh medium without proaerolysin was applied. In order to isolate *HPRT1* deletion cell lines, 72 h post CRISPR transfection, cells were treated with 30 μM 6-thioguanine (Sigma, catalog # A4660). For both selections, after at least 7 days of outgrowth, clones were picked into 96-well plates for genomic deletion verification. Proaerolysin was obtained from Aerohead Scientific Ltd, 1533 Cairns Avenue, Saskatoon SK, S7H 2H5, Canada. One noteworthy point is the potential toxicity of proaerolysin. The active form of proaerolysin is known to be quite toxic. Its LD_{50} in mice is 7 μg/kg. However, proaerolysin is inactive, and requires cleavage of a small fragment from its carboxyl terminal to be activated. Proaerolysin is least 250 times less active than aerolysin in mice (34). Further, proaerolysin is about 200 times less active on human erythrocytes than mouse erythrocytes (35), suggesting toxicity of proaerolysin to humans will be even lower. Nonetheless, appropriate protection of laboratory personnel and attention to standard lab safety precautions for procedures involving toxic compounds are important when using proaerolysin.

High throughput human genomic DNA extraction

After the clones were grown to confluence in 96-well plates, medium was removed and the attached cells were washed with 100 μl PBS twice. Then the plate was fast frozen at −80°C for at least 30 min. Cells were resuspended in 50 μl TE buffer with 0.3 μg/μl proteinase K and transferred to a 96-well PCR plate using the following program: 37°C for 60 min→ 99°C for 5 min→ 4°C.

Hi-C data analysis and visualization

Hi-C data (.hic files) of HCT116 cells were downloaded from NCBI GEO (accession number GSM2795535) (36). We extracted the interaction matrices from .hic format at 50 kb resolution using the straw software (37) and Knight-Ruiz Matrix Balancing (KR) (38) normalization was applied to the matrices. Interaction scores within the range of interest were plotted using in-house developed scripts in R. Heat maps in Figure 4 were generated using the 3D Genome Browser (39). All Hi-C analysis scripts are available at https://github.com/sunnysun515/hic_analysis.

RESULTS

Robust selection with proaerolysin in HCT116+ch3 cell line

To evaluate the robustness of the proaerolysin, we evaluated its killing curve and then examined the kinetics of appearance of proaerolysin-resistant cells after treatment with a single exon-targeted CRISPR guide sequence. The proaerolysin killing curve indicates nearly 100% cell death after 24 h treatment with as little as 0.1 nM proaerolysin (Figure 1A). Usually cell shrinkage (a consequence of leakage of cellular contents through transmembrane pores caused by aerolysin) can be observed within 1h after addition of proaerolysin to the medium (Figure 1B). Dead cells were floating within 24 h and were removed by changing the medium (Figure 1B). By employing a single gRNA targeting *PIGA* Exon 2, we showed that there is a clean background after proaerolysin treatment by comparing transfections with and without gRNA (Figure 1C). Importantly, we found that it was critical to incubate the cells for at least 7 days prior to initiating counterselection with proaerolysin. This time is presumably required to allow for turnover of both preexisting PIGA enzyme as well as preexisting GPI-anchored proteins in the membrane. Figure 1D indicates the single gRNA targeting efficiency reflected on different days of proaerolysin treatment post-CRISPR transfection. When proaerolysin was added less than 7 days after CRISPR, there were barely any survivors. As indicated in Figure 1E, the correlation of CRISPR efficiency (reflected as proaerolysin survival rate) is almost linearly correlated with the number of days between proaerolysin application and CRISPR transfection.

Designer deletion cell lines with counterselectable markers

In order to generate a 'designer deletion' cell line in which the entire *PIGA* transcription unit was deleted, we designed paired gRNA 3′DSS and 5′USS in the flanking region of *PIGA*. Even though INDELs caused by single gRNA CRISPR in an exon of *PIGA* are sufficient to disrupt its function, we preferred to engineer a marker-deficient cell line via elimination of the full gene, because the *PIGA* DNA sequence in the genome might otherwise allow for subsequent recombination with incoming DNA when introducing it as a selectable marker.

Single clones were picked after CRISPR–Cas9 transfection, followed by outgrowth and proaerolysin selection (Figure 2A). Of 15 clones picked, 13 gave rise to positive products from deletion junction PCR. The deletion junction fragments of six clones were gel purified and revealed by Sanger sequencing. Three were precise deletions, and the other three were deletions with insertions (Figure 2B). After blasting the two insertion sequences, #4 showed no genomic similarities, whereas #6 had a 99% identity to a sequence on chromosome X 15336146–15336415, which lies within the deletion region between gRNAs 3′DSS and 5′USS (Supplementary Table S6). In addition to the *piga*Δ cell line, we used similar methods to obtain precise *hprt1*Δ cell lines in the backgrounds of HCT116+ch3 WT and HCT116+ch3 *piga*Δ (Figure 2C). Sequencing revealed four clones of HCT116+ch3 *piga*Δ *hprt1*Δ of which two were precise deletions and the other two had INDELs. For the two clones of HCT116+ch3 *hprt1*Δ analyzed, one was a precise deletion, and the other one had a total 74 bp deletion beyond the two expected Cas9 cut sites (Figure 2D). After blasting the 74 bp insertion against the human genome, no similarities were found.

In addition to these modest sized designer deletion cell lines, we recovered cell lines with ∼2 Mb mega-deletions centered on *PIGA* by using CRISPR deletion with the outermost pair of gRNAs 1 Mb and 1 Mb' (Supplementary Figure S1A). Eleven clones were recovered from proaerolysin selected populations corresponding to the 2 Mb deletion, and six of these were positive for deletion junction PCR. Sequencing indicated all six had INDELs which fell into two groups, seemingly representing sibling clones derived from two original deletion events (Supplementary Figure S1B, Supplementary Table S6). This could be due to the fact that the 2 Mb deletion is a relatively rare event and the proaerolysin treatment was only applied 16 days post-transfection, in order to recover the largest number of survivors. Remarkably, although this 2 Mb deletion removes 15 human genes in addition to *PIGA*, namely *GLRA2*, *FANCB*, *MOSPD2*, *ASB9*, *ASB11*, *VEGFD*, *PIR*, *BMX*, *ACE2*, *TMEM27*, *CA5B*, *ZRSR2*, *AP1S2*, *GRPR* and *MAGEB17*, the cells grow normally.

Complementing the loss of *PIGA* with mini*PIGA* cDNA

In order to complement the loss of *PIGA* in the deletion cell line, we cloned the cDNA of *PIGA* gene into the episomal (oriP-containing) pCEP4 plasmid under either the TK (pDL365) or the native promoter (pDL366) (Supplementary Figure S2A). After transfection and 7 days of puromycin selection, we treated transfected cells with proaerolysin and employed PrestoBlue cell viability assay to evaluate drug resistance. Unexpectedly, we only observed ∼40% complementation (Figure 3A). To further analyze whether this result was due to problems in initial transfection or gradual loss/silencing of transfected plasmids, previously described flow cytometry methods employing staining with fluorescent CD59 antibody (40) were used to sort the fully complemented population, and proaerolysin treatment was initiated after outgrowth. We conducted the proaerolysin treatment right after sorting, 5 and 9 days after sorting and outgrowth (Supplementary Figure S3). The PrestoBlue assay after outgrowth demonstrated a gradual loss of complementation (Supplementary Figure S3A–C),

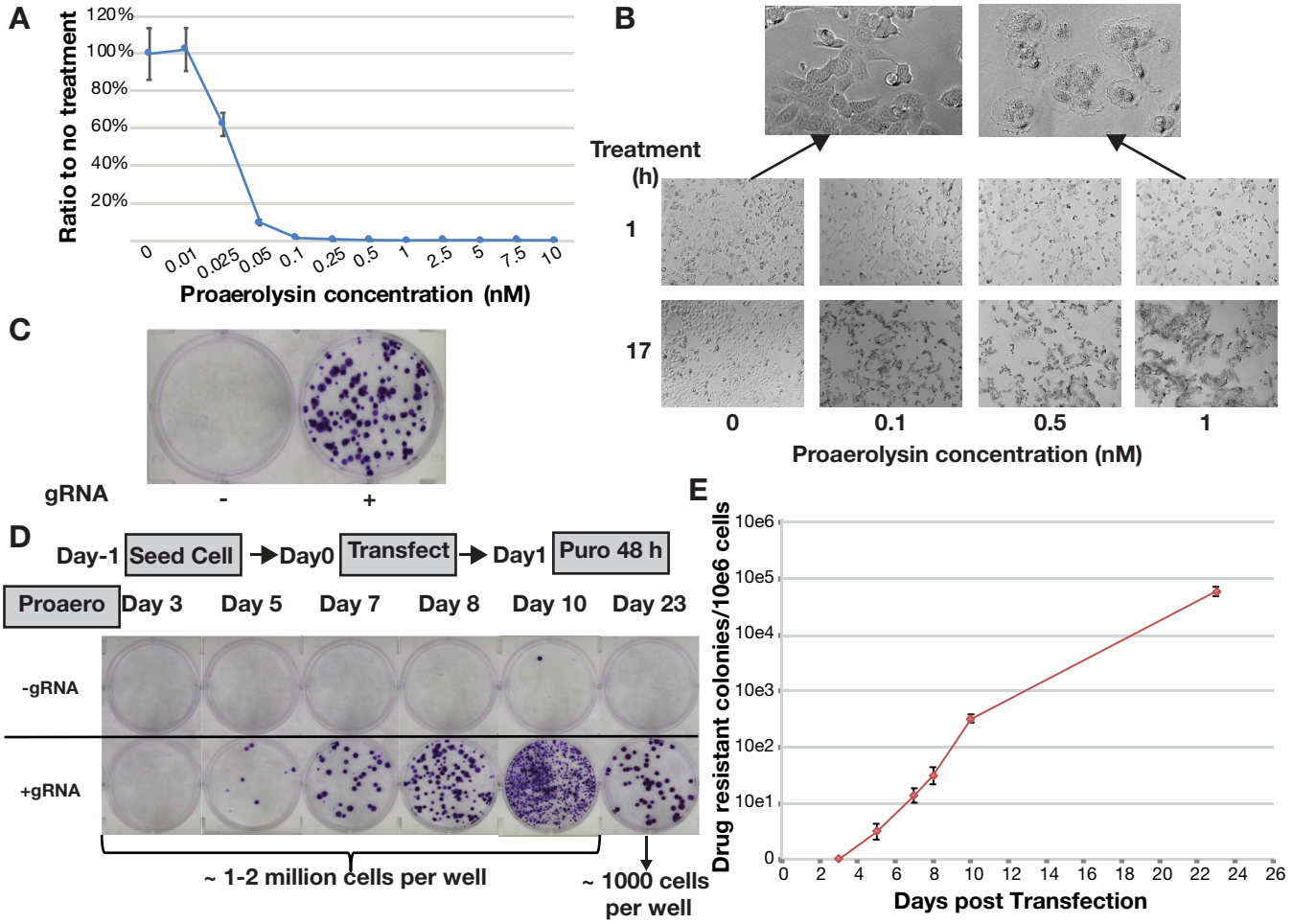

Figure 1. Proaerolysin treatment in HCT116+ch3 cell line. (**A**) Proaerolysin killing curve conducted in HCT116+ch3 cell line using the PrestoBlue assay. (**B**) Cell images for HCT116+ch3 cells after 1 and 17 h showing treatment with different concentrations of proaerolysin. (**C**) Crystal violet stain for CRISPR Cas9 targeting *PIGA* Exon 2 with proaerolysin treatment. (**D**) Crystal violet stain for CRISPR Cas9 targeting *PIGA* Exon 2 with proaerolysin treatment on different days post-transfection. (**E**) CRISPR efficiency from (D) calculated on a $\log_{10}$ scale, plotting proaerolysin resistant colony number per million treated cells.

which may be due to episomal plasmid loss. pCEP4 plasmid was used in this analysis because of its episomal expression. However, our result suggests that there is still a chance to lose this episomal vector even under the proper selection.

GPI anchor is a highly conserved post-translational modification. PIGA is essential for the first step of GPI synthesis and has the active site of the enzyme, but it is only one of multiple subunits involved in this step, so we planned to investigate the degree of conservation of this biosynthetic process. We cloned cDNA of *PIGA* orthologs from *M. musculus* (mouse), *M. mulatta* (monkey), *D. melanogaster* (fruit fly) and *P. falciparum* (malaria parasite) into pCEP4 (Supplementary Figure S2B and C). After transfecting these plasmids into the *piga*Δ cell line, we used proaerolysin resistance analysis to evaluate the complementation efficiency (Figure 3B). The orthologs of mouse and monkey complemented as efficiently as the human cDNA, but those of Drosophila and plasmodium did not demonstrate any effect. This result agrees that PIGA participates in a multisubunit step.

What factors affect paired gRNA deletion efficiency?

Since *PIGA* is essential for the synthesis of GPI anchors, measuring the loss of GPI anchors using flow cytometry in a cell population can help provide an estimate of deletion efficiency. In order to investigate the associated factors of the deletion efficiency, series of partner gRNAs were designed 50 kb, 100 kb, 250 kb, 500 kb, 750 kb and 1 Mb upstream and downstream of the *PIGA* gene (Figure 4C, Supplementary Table S1). gRNAs were chosen to minimize the likelihood of off-target effects based on publicly available online tools (33). At each position, at least two gRNAs were designed. gRNA IN2-3 was used as the 'anchor gRNA' and paired with each of the partner gRNAs to systematically analyze deletion efficiency over a wide deletion size range from 50 kb to 1 Mb. All these deletions contain the loss of at least one *PIGA* exon, and all are able to disrupt the synthesis of GPI anchors through deletion of at least one critical segment(s) of *PIGA*. Since *PIGA* was located at the boundary between two well-defined TADs and the anchor gRNA IN2–3 belongs to the TAD upstream of *PIGA* (Fig-

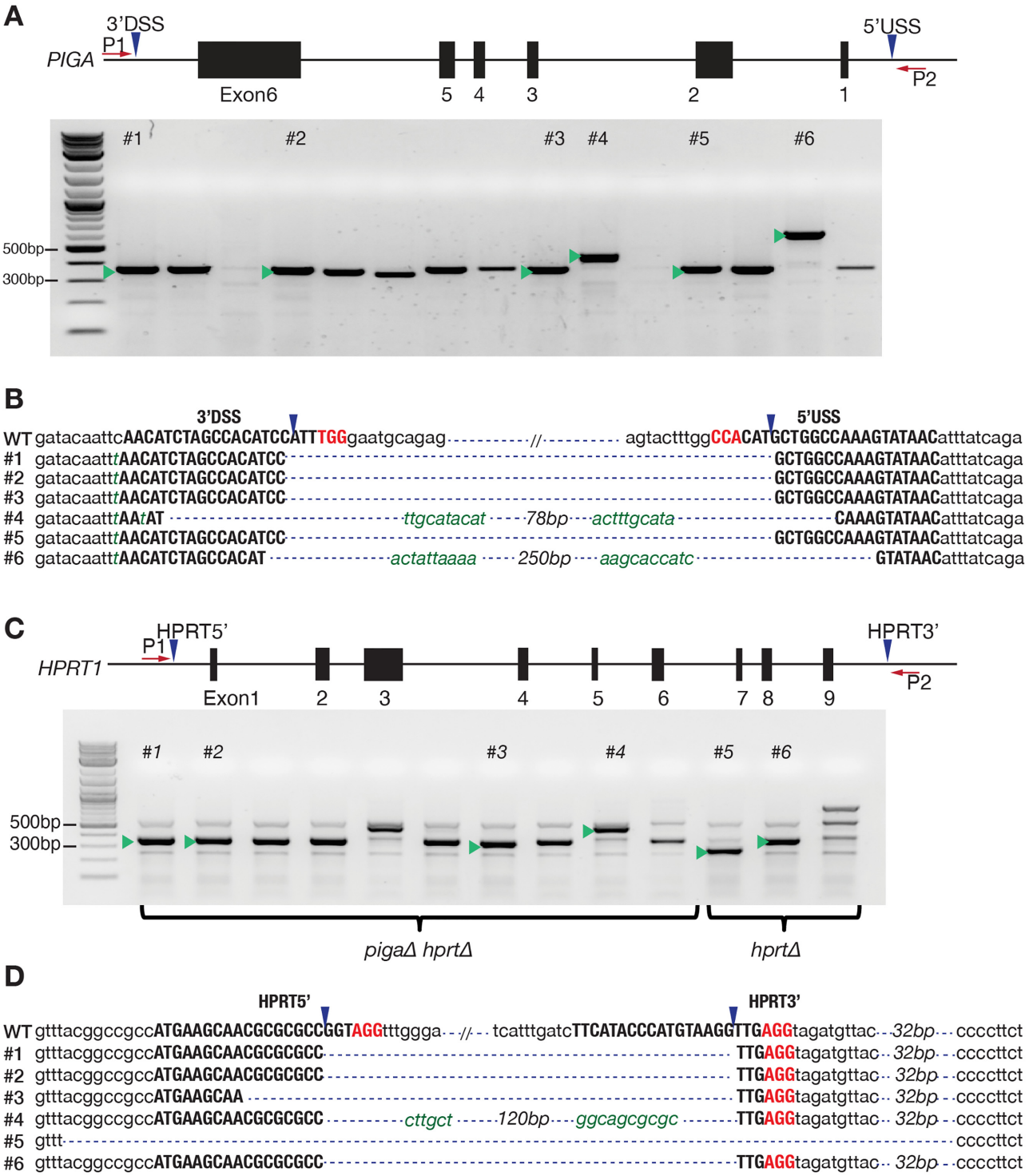

Figure 2. Designer deletion cell lines. (**A**) gRNA designed to delete the full sequence of *PIGA*. The gel image indicates the deletion product amplified with P1 and P2 for the 15 clones picked after proaerolysin treatment. Green triangles indicate the band cut and purified for sequencing. (**B**) Sequencing results for the bands cut from the gel image in (A). (**C**) gRNA designed to delete the full sequence of *HPRT1* in HCT116+ch3 WT and *pigaΔ* cell lines. The gel image shows the deletion product for the numbered clones which were picked. Green triangles indicate band cut and purified for sequencing. (**D**) Sequencing results for bands cut from gel image in (C). gRNA sequences are in bold uppercase letters. The PAM sequence is in red. Blue triangles indicate predicted cleavage sites. WT refers to the wild type sequence from ensemble.org. Green italic lowercase letters represent mismatch sequences or insertions.

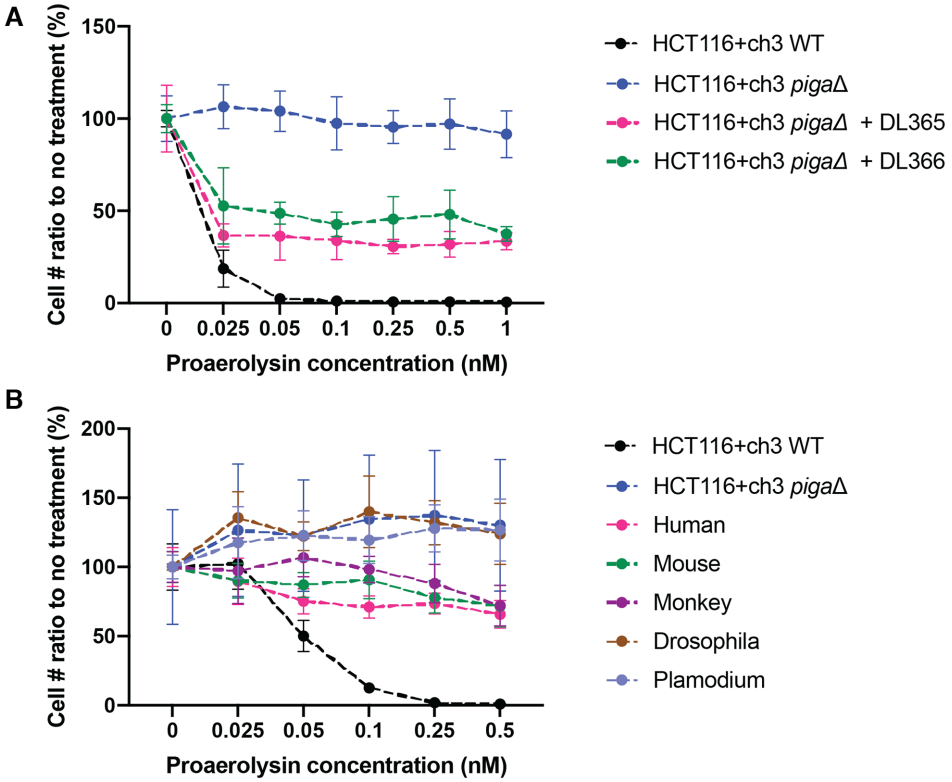

Figure 3. Proaerolysin resistance analysis for mini-*PIGA* cDNA complementation. (**A**) pDL365 (TK promoter) and pDL366 (native *PIGA* promoter) were transfected to *pigaΔ* cell line, after 7 days of puromycin selection, transfected cells were treated with proaerolysin for 24 h and then assayed with PrestoBlue. (**B**) Plasmids expressing cDNA copies of the orthologs were transfected into the *pigaΔ* cell line, and after 7 days of puromycin selection, the transfected cells were treated with proaerolysin for 24 h and assayed with PrestoBlue.

ure 4A), we used this design to test the effect of TAD on deletion efficiency. Even though the mechanism of paired gRNA CRISPR deletion is not clear, it is likely that NHEJ is required to join the ends generated by CRISPR/Cas9 cutting, we hypothesized that deletion efficiency might be affected by Topologically Associated Domains (TADs) as two targeted regions might be more likely to be ligated when they in close proximity. The gRNAs *PIGA* 50 kb_D, 100 kb_D, 250 kb_D, 500 kb_D, 750 kb_D and 1 Mb_D, mapping downstream of *PIGA*, all lie within the TAD downstream of the anchor gRNA IN2–3. On the other hand, all gRNAs 50 kb_U, 100 kb_U, 250 kb_U, 500 kb_U, 750 kb_U and 1 Mb_U are within the same TAD in which gRNA IN2–3 is located (Figure 4A, C). Interaction scores between the anchor guide region and the regions of the upstream partner guides within the same TAD were much higher relative to downstream partner guides that lie in the downstream TAD (Figure 4B)

Using flow cytometry, deletion efficiency was measured as percentage of GPI (−) cells out of a total 20,000 cells, which varied between 30% and 0.5%. Using this assay, we observed no clear correlation between deletion efficiency and distance between paired gRNAs (Figure 5A). The deletion efficiency of the anchor gRNA paired with the 1 Mb_D gRNA was as high as 20%. (Figure 5B). Moreover, by comparing deletion efficiency of paired gRNAs within the same TAD or crossing the TAD boundary, we noticed that the paired gRNAs that cross the TAD boundary actually resulted in overall slightly higher deletion efficiencies (Figure 5B and Supplementary Figure S4A), the opposite of what one would expect if TAD boundaries interfered with deletion formation. We also tested the correlation of TAD boundaries with paired gRNA mediated deletion at the *HPRT1* locus. Similar to *PIGA*, *HPRT1* is located at the boundary of two TADs. gRNAs at 50, 100, 250 and 500 kb upstream or downstream of *HPRT1* were designed to pair with the anchor gRNA inside *HPRT1*(Supplementary Table S2). The deletion efficiency measured by G418 selection and crystal violet staining and it again indicated no significant difference between deletion efficiency and whether or not the paired gRNAs reside in the same or the neighboring TAD (Supplementary Figure S4B and C). These observations are inconsistent with the hypothesis that TAD boundaries interfere with deletion formation.

Analysis of a heterozygous mutation in an autosomal gene, *PIGL*, suggests high frequency loss of heterozyogosity

In addition to *PIGA*, a number of other 'PIG' genes are required for the formation of GPI anchors, and the other *PIG* genes are all autosomal (10). We were interested in using the GPI system to study the stability of heterozygous CRISPR induced mutations in autosomal '*PIG*' genes. Because homozygous missense mutations of *PIGU* result in reduced cell-surface expression of fluorescent aerolysin (FLAER) flow cytometry profile (41), we anticipated that

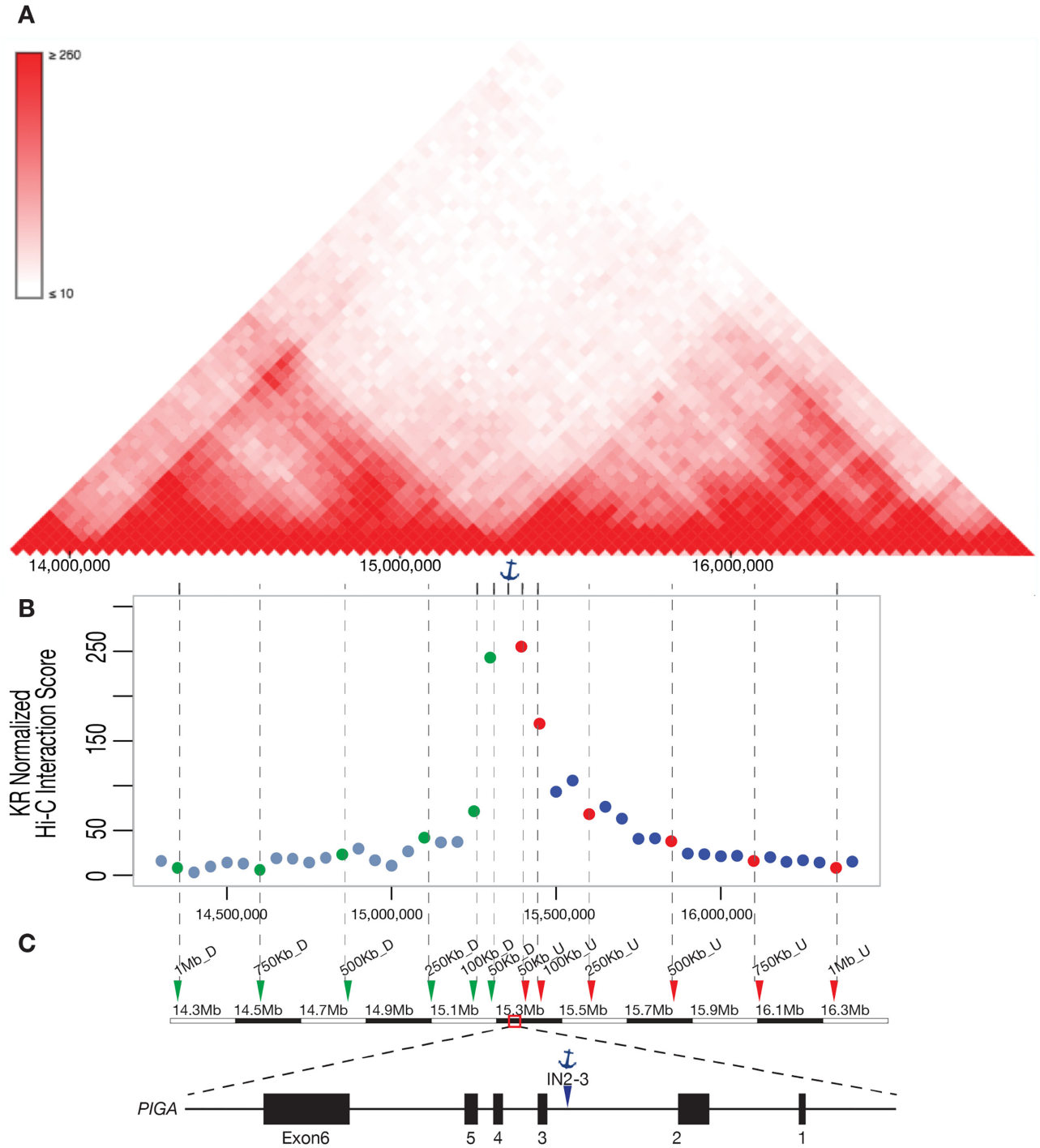

Figure 4. Paired gRNA deletion design. (**A**) Interaction frequencies from HI-C data were shown as a heat map at the *PIGA* region. Positions of gRNAs located 50 kb–1 Mb from the anchored gRNA IN2–3 were indicated by gray dashed lines. Anchor icon indicates position of gRNA IN2–3. (**B**) Hi-C interaction scores of partner gRNAs regions and anchor gRNA (IN2–3) region were plotted according to the partner gRNA positions around *PIGA*. (**C**) gRNA target sites. Red dots and triangles represent gRNAs 50 kb_U, 100 kb_U, 250 kb_U, 500 kb_U, 750 kb_U and 1 Mb_U. Green dots and triangles represent gRNAs 50 kb_D, 100 kb_D, 250 kb_D, 500 kb_D, 750 kb_D and 1 Mb_D.

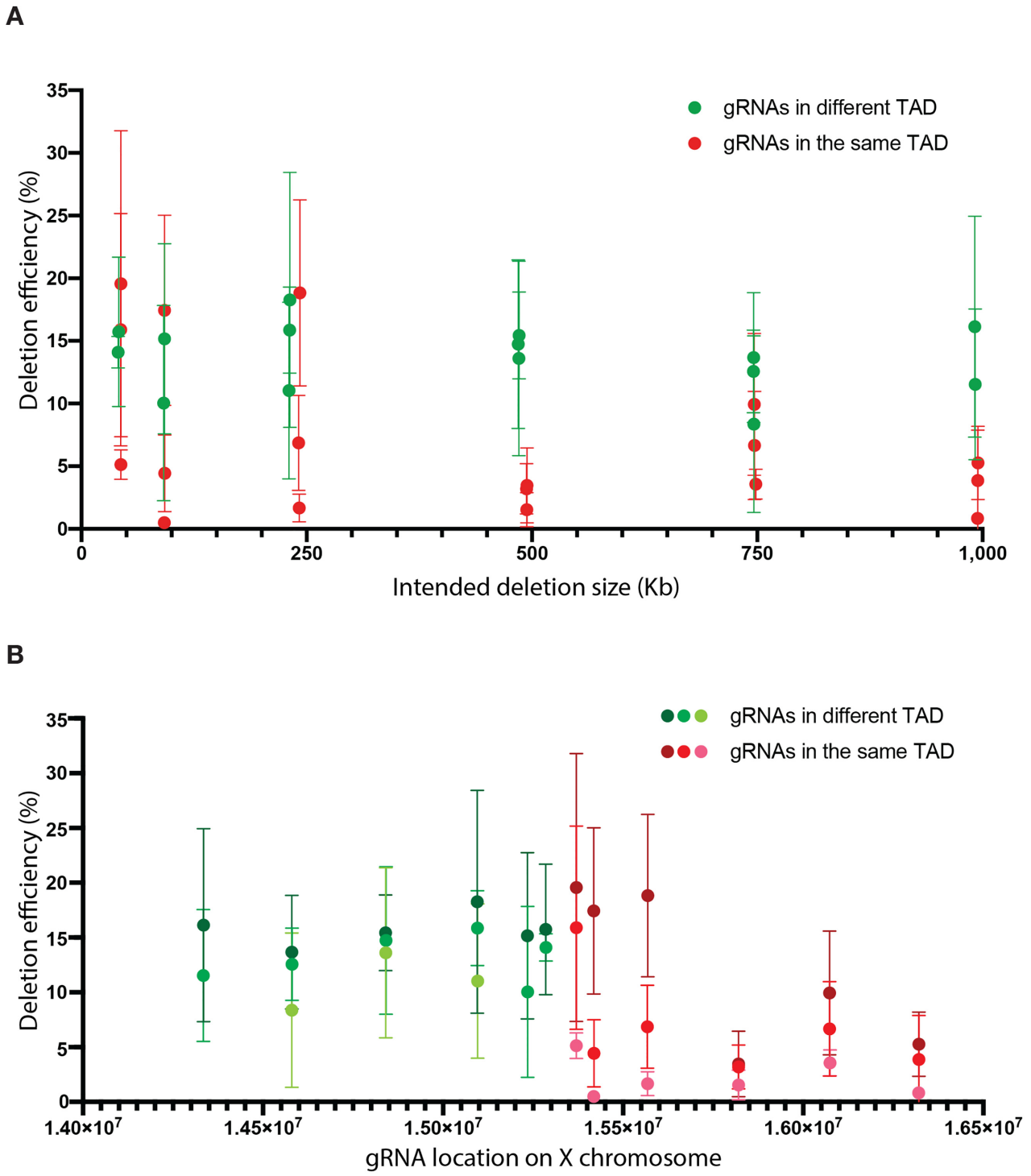

Figure 5. Paired gRNA deletion efficiency is not significantly affected by deletion size nor TAD boundaries. (**A**) Deletion efficiency versus deletion size. Red dots represent gRNAs in the TAD upstream of *PIGA*, green dots represent gRNAs in the TAD downstream of *PIGA*. (**B**) Deletion efficiencies versus partner gRNA position. Red dots represent gRNAs in the TAD upstream of *PIGA*, green dots represent gRNAs in the TAD downstream of *PIGA*. The shading from dark to light represent gRNAs with deletion efficiency from high to low at the same position. Error bars are standard deviations calculated from three biological repeats.

homozygous mutations in all other autosomal *PIG* genes would also confer proaerolysin resistance. Thus, we used CRISPR–Cas9 to generate a heterozygous NHEJ-induced mutation at a single site in the *PIGL* gene (Supplementary Figure S5A). These studies revealed that a homozygous mutation in *PIGL* indeed led to proaerolysin resistance, as expected, and also that cells with a heterozygous mutation in *PIGL* remained sensitive to proaerolysin.

We then studied the heterozygous cell line, HCT116+ch3 *PIGL* (+/−) by selecting derivatives that were proaerolysin resistant, expecting to find new spontaneous mutations in the wild-type *PIGL* allele at a high frequency. However, when using PCR amplification of the gRNA target site to verify heterozygosity in single clones from this cell line after proaerolysin treatment and outgrowth, we unexpectedly observed that the majority of the survivors (21/24) had lost heterozygosity for the original CRISPR/NHEJ-induced mutation (Supplementary Figure S5B, primers in Supplementary Table S4). The wild type allele was maintained in only three of the lines. Further, cytogenetic analyses showed that the loss of heterozygosity (LoH) was not accompanied by loss of chromosome 17p, on which *PIGL* resides, in four of seven cases evaluated. It may be that the observed LoH occurred through gene conversion, although the experiments done thus far cannot be considered definitive on this point.

DISCUSSION

Canver *et al.* reported an inverse relationship between deletion efficiency and deletion size based on their CRISPR/Cas9-mediated genomic deletions in mouse erythroleukemia (MEL) cells (29). However, another analysis conducted along the *HPRT1* locus in HEK293FT cell line did not report any apparent correlation between inter-gRNA distance and deletion efficiency (30). Both groups achieved deletions up to 1 Mb. For the paired gRNA CRISPR deletion study conducted in MEL cells, a range of deletions with different inter-gRNA distances were designed and constructed at different loci throughout the genome. This design introduced many confounding variables beyond inter-gRNA distance, and thus is not an optimal experimental design for directly evaluating the effect of length on deletion efficiency. For the analysis conducted by He *et al.* (30) at the *HPRT1* locus, the cell line chosen was HEK 293FT, which is hypotriploid and contains at least three X chromosomes. Such a complex genome confounds direct genotyping to detect deletions, but did benefit from being a single locus study, with an anchored guide sequence to make results potentially more comparable.

By employing the nearly diploid HCT116+ch3 cell line, the *PIGA* locus on the X chromosome and the quantitative measurement of GPI(–) cells, we performed a systematic analysis for the efficiency of paired gRNA CRISPR deletion. Our results reveal no significant impact of gRNA positioning on deletion efficiency associated with the distance between the two gRNAs nor their positions relative to nearby TAD boundaries. This finding was confirmed by the same design in the *HPRT1* locus, providing a valuable extension to the observations of Canver *et al.* (29) and He *et al.* (30). Even though TAD boundaries are mostly defined based on studies of asynchronous cells, and the highly compartmentalized organization is restricted to interphase (42), the consideration of the TAD effect is practically relevant to those wishing to perform paired gRNA CRISPR deletion. It is reasonable to ask whether the TAD boundaries are active in cells actually undergoing CRISPR-mediated deletion. Based on single cell Hi-C studies, the locations of TAD borders are generally unchanged in G1, early-S and late-S/G2 cells and are absent during mitosis (43). The duration of mitosis is short, constant and independent of varied cell cycle length, and lasts about 1 hour in rapidly proliferating human cells (44). Thus, more than 90% of the cells assayed in this study were in interphase and presumably the boundaries at the *PIGA* and *HPRT1* loci were intact. Furthermore, the NHEJ repair pathway utilized by paired gRNA CRISPR deletion is active throughout interphase and predominates in the G1 phase when sister chromatids are available for homologous recombination based repair (45). Since G1 phase is the longest phase of the cell cycle, the population engineered by paired gRNA CRISPR deletion is presumed to have active TAD boundaries. The lack of impact of nearby TAD boundaries on paired gRNA CRISPR deletion provides confidence when using paired gRNA CRISPR to engineer large deletions.

As a genome engineering technology, CRISPR is a promising tool and has been applied extensively in research and clinical fields due to its high efficiency. INDELs associated with NHEJ repair of CRISPR/Cas9-mediated DSBs that occur within an exon can disrupt protein-coding open reading frames. However, INDELS placed in noncoding DNA are usually insufficient to disrupt the function of noncoding transcripts, gene clusters or regulatory sequences, where large genomic deletions would be preferable. Also, CRISPR deletions could play roles in genetic diseases driven by genome amplification or copy number abnormalities for animal model generation and therapeutics with large genomic rearrangements. During our study of cell line engineering via CRISPR, we also observed the value of paired gRNA CRISPR deletion in the generation of heterozygous mutations. The loss of heterozygosity which occurred in the HCT116+ch3 *PIGL* (+/−) cell line indicates the high frequency of LoH of point alterations like the small INDELs produced by single gRNA targeting. It might be possible to harness such LoH, particularly if it is produced by gene conversion, to make homozygous designer alterations to diploid mammalian chromosomes. Since gene-conversion tracts are usually short in mammalian cells, and rarely exceed 1 kb in length (46), it is interesting to consider ways to limit the impact of such gene conversion, e.g. in order to develop an assay for mutation in a large autosomal gene like *PIGL*. By making a heterozygous deletions very large, or even extending them into an adjacent essential gene, it might be possible to minimize or entirely eliminate the recovery of this type of LoH event.

The era of editing and writing mammalian genomes using synthetic genomics techniques is well underway. Launched in 2016, the GP-write project aims to reduce genomic engineering and testing costs to 1/1000th of the previous within 10 years and address a number of human health challenges (47). A recently announced plan by GP-write, making virus-resistant human cells, would require >400 000 changes to

the genome (48). All this editing will be associated with extensive recoding, DNA fragment delivery, and sequential delivery with marker swapping. Paired gRNA CRISPR deletion may be important in endogenous genomic content removal when introducing synthetic fragments. Designer deletion cell lines described in this study may be useful for DNA delivery, and cell lines in which both *HPRT1* and *PIGA* are deleted could be used as a platform for delivering DNAs independently or perhaps sequentially. The wide use and adoption of the 5-Foa selection in *Saccharomyces cerevisiae* (1), and similarly, the wide use of *TK* negative selection using gancyclovir in mammalian cells (4) suggests that the proaerolysin selection, in conjunction with deployment of CRISPR represents a highly versatile mammalian genome engineering tool that will enjoy wide usage.

One potential limitation of our study is that it was performed in male cell lines, and thus it remains to be practically tested how well these counterselections work in female cells. The 6-thioguanine selection for *hprt1* mutants works equally well in male and female cells (49) and furthermore, the gene lies well outside the pseudoautosomal region, and is not a gene known to 'escape' X inactivation (50). Nonetheless some caution might be exercised in female embryonic stem cells, as these are known to lose X inactivation. The availability of lines with deletions in both *PIGA* and *HPRT1*, located on the left and right arms of the X chromosome, respectively opens up some interesting opportunities to look at aspects of X inactivation and beyond.

Counterselection markers might be useful in selective loss of DNA templates delivered on episomal vectors after proper genome editing has occurred. The instability of episomal vectors discovered in this study indicates that the selected loss of auxiliary gene on an episomal vector via proaerolysin treatment seems more efficient than expected. It is worth noting that an intronless *PIGA* processed pseudogene (Ψ*PIGA*) resides on chromosome 12 (51). Ψ*PIGA* should not cause issues when using the designer deletion cells in the study for DNA delivery or other applications via proaerolysin selection or other GPI anchor based assays, for it is not functionally expressed as a consequence of several nonsense mutations. However, Ψ*PIGA* should be taken into account due to its homology to *PIGA* cDNA, which might in theory lead to recombination with *PIGA* cDNA constructs.

In addition to the use of designer deletion cell lines in DNA delivery, it would be interesting to investigate the TAD architecture of the cell line HCT116+ch3 *piga* 2Mb deletion, since the deleted region harbors extensive CTCF (a master regulator of TAD structure) binding sites and includes a TAD boundary (Figure 4A). It has been reported that TAD boundary disruption can occur after deletions of minimal or very large genomic regions with CTCF binding sites (52–54). However, a recent study focusing on the CTCF-rich *Firre* locus on X chromosome revealed that its deletion preserved the surrounding TAD even though CTCF binding was depleted (55). It would be interesting to evaluate another X chromosome locus like *PIGA* to investigate whether such TAD preservation is a sex chromosome associated phenomena. The series of paired gRNAs designed for deletion efficiency measurements in this study can also be utilized to screen the effects on TAD reformation by length variation and inclusion/exclusion TAD boundary deletions. Since *PIGA* has the potential to become a widely used counterselection marker, its endogenous locus may well serve as a future target of DNA delivery or other types of analysis, more extensive studies of TADs in our cell lines (Supplementary Table S3) may be of interest.

In summary, we have engineered a series of designer deletion cell lines featureing counterselectable markers of *PIGA* and *HPRT1*. These cell lines can facilitate large DNA fragment delivery and other types of study involved with *PIGA* and *HPRT1* biology. In the process of building this resource, we acquired systematic data on a controlled set of deletions with a common anchor point, and used this to provide evidence supporting the hypothesis that TAD boundaries do not interfere in a practical way with deletion formation using paired CRISPR guides.

DATA AVAILABILITY

All the Hi-C analysis scripts are available at https://github.com/sunnysun515/hic_analysis. Flow cytometry experiments were deposited in FlowRepository under Repository ID FR-FCM-Z3BN.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Drs Jeff Corden, Maria Jasin, Todd Macfarlan and David Truong, and anonymous reviewers for valuable discussions and suggestions. We thank Dr Thomas A. Kunkel for sharing the HCT116+ch3 cell line. Portions of this work were previously published in a PhD thesis by D.L. and supported in part by NIH grants P50-GM107632 and RM1-HG009491 to JDB. JDB is a founder and Director of CDI Labs, Inc., a founder of Neochromosome, Inc., a Founder of and Consultant to the ReOpen Diagnostics, and serves or served on the Scientific Advisory Board of the following: Sangamo, Inc., Modern Meadow, Inc. and Sample6, Inc.

FUNDING

National Human Genome Research Institute [RM1-HG009491]; National Institute of General Medical Sciences [P50-GM107632]. Funding for open access charge: National Human Genome Research Institute [RM1-HG009491].
Conflict of interest statement. Jef Boeke is a Founder and Director of CDI Labs, Inc., a Founder of Neochromosome, Inc., a Founder of and Consultant to the ReOpen Diagnostics, and serves or served on the Scientific Advisory Board of the following: Sangamo, Inc., Modern Meadow, Inc. and Sample6, Inc.

REFERENCES

1. Boeke,J.D., Trueheart,J., Natsoulis,G. and Fink,G.R. (1987) 5-Fluoroorotic acid as a selective agent in yeast molecular genetics. *Methods Enzymol.*, **154**, 164–175.

2. Steyer,B., Bu,Q., Cory,E., Jiang,K., Duong,S., Sinha,D., Steltzer,S., Gamm,D., Chang,Q. and Saha,K. (2018) Scarless genome editing of human pluripotent stem cells via transient puromycin selection. *Stem Cell Rep.*, **10**, 642–654.
3. Yu,J., Vodyanik,M.A., Smuga-Otto,K., Antosiewicz-Bourget,J., Frane,J.L., Tian,S., Nie,J., Jonsdottir,G.A., Ruotti,V., Stewart,R. *et al.* (2007) Induced pluripotent stem cell lines derived from human somatic cells. *Science*, **318**, 1917–1920.
4. Wigler,M., Silverstein,S., Lee,L.S., Pellicer,A., Cheng,Y. and Axel,R. (1977) Transfer of purified herpes virus thymidine kinase gene to cultured mouse cells. *Cell*, **11**, 223–232.
5. Lester,S.C., LeVan,S.K., Steglich,C. and DeMars,R. (1980) Expression of human genes for adenine phosphoribosyltransferase and hypoxanthine-guanine phosphoribosyltransferase after genetic transformation of mouse cells with purified human DNA. *Somatic Cell Genet.*, **6**, 241–259.
6. Lupton,S.D., Brunton,L.L., Kalberg,V.A. and Overell,R.W. (1991) Dominant positive and negative selection using a hygromycin phosphotransferase-thymidine kinase fusion gene. *Mol. Cell. Biol.*, **11**, 3374–3378.
7. Schwartz,F., Maeda,N., Smithies,O., Hickey,R., Edelmann,W., Skoultchi,A. and Kucherlapati,R. (1991) A dominant positive and negative selectable gene for use in mammalian cells. *Proc. Natl. Acad. Sci. U.S.A.*, **88**, 10416–10420.
8. Karreman,C. (1998) New positive/negative selectable markers for mammalian cells on the basis of Blasticidin deaminase-thymidine kinase fusions. *Nucleic Acids Res.*, **26**, 2508–2510.
9. Karreman,C. (1998) A new set of positive/negative selectable markers for mammalian cells. *Gene*, **218**, 57–61.
10. Kawagoe,K., Takeda,J., Endo,Y. and Kinoshita,T. (1994) Molecular cloning of murine pig-a, a gene for GPI-anchor biosynthesis, and demonstration of interspecies conservation of its structure, function, and genetic locus. *Genomics*, **23**, 566–574.
11. Watanabe,R., Kinoshita,T., Masaki,R., Yamamoto,A., Takeda,J. and Inoue,N. (1996) PIG-A and PIG-H, which participate in glycosylphosphatidylinositol anchor biosynthesis, form a protein complex in the endoplasmic reticulum. *J. Biol. Chem.*, **271**, 26868–26875.
12. Watanabe,R., Inoue,N., Westfall,B., Taron,C.H., Orlean,P., Takeda,J. and Kinoshita,T. (1998) The first step of glycosylphosphatidylinositol biosynthesis is mediated by a complex of PIG-A, PIG-H, PIG-C and GPI1. *EMBO J.*, **17**, 877–885.
13. Low,M.G. and Saltiel,A.R. (1988) Structural and functional roles of glycosyl-phosphatidylinositol in membranes. *Science*, **239**, 268–275.
14. Paulick,M.G. and Bertozzi,C.R. (2008) The glycosylphosphatidylinositol anchor: a complex membrane-anchoring structure for proteins. *Biochemistry*, **47**, 6991–7000.
15. Araten,D.J., Golde,D.W., Zhang,R.H., Thaler,H.T., Gargiulo,L., Notaro,R. and Luzzatto,L. (2005) A quantitative measurement of the human somatic mutation rate. *Cancer Res.*, **65**, 8111–8117.
16. Takeda,J., Miyata,T., Kawagoe,K., Iida,Y., Endo,Y., Fujita,T., Takahashi,M., Kitani,T. and Kinoshita,T. (1993) Deficiency of the GPI anchor caused by a somatic mutation of the PIG-A gene in paroxysmal nocturnal hemoglobinuria. *Cell*, **73**, 703–711.
17. Young,N.S. and Moss,J. (2000) In: *Paroxysmal Nocturnal Hemoglobinuria and the Glycosylphosphatidylinositol-linked Proteins*. Academic Press, San Diego.
18. Bessler,M., Mason,P., Hillmen,P. and Luzzatto,L. (1994) Somatic mutations and cellular selection in paroxysmal nocturnal haemoglobinuria. *Lancet*, **343**, 951–953.
19. Endo,M., Ware,R.E., Vreeke,T.M., Singh,S.P., Howard,T.A., Tomita,A., Holguin,M.H. and Parker,C.J. (1996) Molecular basis of the heterogeneity of expression of glycosyl phosphatidylinositol anchored proteins in paroxysmal nocturnal hemoglobinuria. *Blood*, **87**, 2546–2557.
20. O'Keefe,C.L., Sugimori,C., Afable,M., Clemente,M., Shain,K., Araten,D.J., List,A., Epling-Burnette,P.K. and Maciejewski,J.P. (2011) Deletions of Xp22.2 including PIG-A locus lead to paroxysmal nocturnal hemoglobinuria. *Leukemia*, **25**, 379–382.
21. Buckley,J.T. and Howard,S.P. (1988) Aerolysin from *Aeromonas hydrophila*. *Methods Enzymol.*, **165**, 193–199.
22. Miura,D., Dobrovolsky,V.N., Kasahara,Y., Katsuura,Y. and Heflich,R.H. (2008) Development of an in vivo gene mutation assay using the endogenous Pig-A gene: I. Flow cytometric detection of CD59-negative peripheral red blood cells and CD48-negative spleen T-cells from the rat. *Environ. Mol. Mutagen.*, **49**, 614–621.
23. Bryce,S.M., Bemis,J.C. and Dertinger,S.D. (2008) In vivo mutation assay based on the endogenous Pig-a locus. *Environ. Mol. Mutagen.*, **49**, 256–264.
24. Karnan,S., Konishi,Y., Ota,A., Takahashi,M., Damdindorj,L., Hosokawa,Y. and Konishi,H. (2012) Simple monitoring of gene targeting efficiency in human somatic cell lines using the PIGA gene. *PLoS One*, **7**, e47389.
25. Bemis,J.C. and Heflich,R.H. (2019) In vitro mammalian cell mutation assays based on the Pig-a gene: a report of the 7th International Workshop on Genotoxicity Testing (IWGT) Workgroup. *Mutat. Res.*, **847**, 403028.
26. Zou,J., Maeder,M.L., Mali,P., Pruett-Miller,S.M., Thibodeau-Beganny,S., Chou,B.K., Chen,G., Ye,Z., Park,I.H., Daley,G.Q. *et al.* (2009) Gene targeting of a disease-related gene in human induced pluripotent stem and embryonic stem cells. *Cell Stem Cell*, **5**, 97–110.
27. Koi,M., Umar,A., Chauhan,D.P., Cherian,S.P., Carethers,J.M., Kunkel,T.A. and Boland,C.R. (1994) Human chromosome 3 corrects mismatch repair deficiency and microsatellite instability and reduces N-methyl-N'-nitro-N-nitrosoguanidine tolerance in colon tumor cells with homozygous hMLH1 mutation. *Cancer Res.*, **54**, 4308–4312.
28. Zheng,Q., Cai,X., Tan,M.H., Schaffert,S., Arnold,C.P., Gong,X., Chen,C.Z. and Huang,S. (2014) Precise gene deletion and replacement using the CRISPR/Cas9 system in human cells. *BioTechniques*, **57**, 115–124.
29. Canver,M.C., Bauer,D.E., Dass,A., Yien,Y.Y., Chung,J., Masuda,T., Maeda,T., Paw,B.H. and Orkin,S.H. (2014) Characterization of genomic deletion efficiency mediated by clustered regularly interspaced short palindromic repeats (CRISPR)/Cas9 nuclease system in mammalian cells. *J. Biol. Chem.*, **289**, 21312–21324.
30. He,Z., Proudfoot,C., Mileham,A.J., McLaren,D.G., Whitelaw,C.B. and Lillico,S.G. (2015) Highly efficient targeted chromosome deletions using CRISPR/Cas9. *Biotechnol. Bioeng.*, **112**, 1060–1064.
31. Nora,E.P., Lajoie,B.R., Schulz,E.G., Giorgetti,L., Okamoto,I., Servant,N., Piolot,T., van Berkum,N.L., Meisig,J., Sedat,J. *et al.* (2012) Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature*, **485**, 381–385.
32. Pombo,A. and Dillon,N. (2015) Three-dimensional genome architecture: players and mechanisms. *Nat. Rev. Mol. Cell Biol.*, **16**, 245–257.
33. 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.
34. Howard,S.P. and Buckley,J.T. (1985) Activation of the hole-forming toxin aerolysin by extracellular processing. *J. Bacteriol.*, **163**, 336–340.
35. Bernheimer,A.W., Avigad,L.S. and Avigad,G. (1975) Interactions between aerolysin, erythrocytes, and erythrocyte membranes. *Infect. Immun.*, **11**, 1312–1319.
36. Rao,S.S.P., Huang,S.C., Glenn St Hilaire,B., Engreitz,J.M., Perez,E.M., Kieffer-Kwon,K.R., Sanborn,A.L., Johnstone,S.E., Bascom,G.D., Bochkov,I.D. *et al.* (2017) Cohesin loss eliminates all loop domains. *Cell*, **171**, 305–320.
37. Durand,N.C., Robinson,J.T., Shamim,M.S., Machol,I., Mesirov,J.P., Lander,E.S. and Aiden,E.L. (2016) Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. *Cell Syst.*, **3**, 99–101.
38. Philip,A. and Knight,D.R. (2013) A fast algorithm for matrix balancing. *IMA J. Numer. Anal.*, **33**, 1029–1047.
39. Wang,Y., Song,F., Zhang,B., Zhang,L., Xu,J., Kuang,D., Li,D., Choudhary,M.N.K., Li,Y., Hu,M. *et al.* (2018) The 3D Genome Browser: a web-based browser for visualizing 3D genome organization and long-range chromatin interactions. *Genome Biol.*, **19**, 151.
40. Araten,D.J., Martinez-Climent,J.A., Perle,M.A., Holm,E., Zamechek,L., DiTata,K. and Sanders,K.J. (2010) A quantitative analysis of genomic instability in lymphoid and plasma cell neoplasms based on the PIG-A gene. *Mutat. Res.*, **686**, 1–8.
41. Knaus,A., Kortum,F., Kleefstra,T., Stray-Pedersen,A., Dukic,D., Murakami,Y., Gerstner,T., van Bokhoven,H., Iqbal,Z., Horn,D. *et al.* (2019) Mutations in PIGU impair the function of the GPI

transamidase complex, causing severe intellectual disability, epilepsy, and brain anomalies. *Am. J. Hum. Genet.*, **105**, 395–402.
42. Naumova,N., Imakaev,M., Fudenberg,G., Zhan,Y., Lajoie,B.R., Mirny,L.A. and Dekker,J. (2013) Organization of the mitotic chromosome. *Science*, **342**, 948–953.
43. Nagano,T., Lubling,Y., Varnai,C., Dudley,C., Leung,W., Baran,Y., Mendelson Cohen,N., Wingett,S., Fraser,P. and Tanay,A. (2017) Cell-cycle dynamics of chromosomal organization at single-cell resolution. *Nature*, **547**, 61–67.
44. Araujo,A.R., Gelens,L., Sheriff,R.S. and Santos,S.D. (2016) Positive feedback keeps duration of mitosis temporally insulated from upstream cell-cycle events. *Mol. Cell*, **64**, 362–375.
45. Aylon,Y., Liefshitz,B. and Kupiec,M. (2004) The CDK regulates repair of double-strand breaks by homologous recombination during the cell cycle. *EMBO J.*, **23**, 4868–4875.
46. Chen,J.M., Cooper,D.N., Chuzhanova,N., Ferec,C. and Patrinos,G.P. (2007) Gene conversion: mechanisms, evolution and human disease. *Nat. Rev. Genet.*, **8**, 762–775.
47. Boeke,J.D., Church,G., Hessel,A., Kelley,N.J., Arkin,A., Cai,Y., Carlson,R., Chakravarti,A., Cornish,V.W., Holt,L. *et al.* (2016) GENOME ENGINEERING. The Genome Project-Write. *Science*, **353**, 126–127.
48. Dolgin,E. (2018) Scientists downsize bold plan to make human genome from scratch. *Nature*, **557**, 16–17.
49. Fukuchi,K., Martin,G.M. and Monnat,R.J. Jr. (1989) Mutator phenotype of Werner syndrome is characterized by extensive deletions. *Proc. Natl. Acad. Sci. U.S.A.*, **86**, 5893–5897.
50. Carrel,L. and Willard,H.F. (2005) X-inactivation profile reveals extensive variability in X-linked gene expression in females. *Nature*, **434**, 400–404.
51. Bessler,M., Hillmen,P., Longo,L., Luzzatto,L. and Mason,P.J. (1994) Genomic organization of the X-linked gene (PIG-A) that is mutated in paroxysmal nocturnal haemoglobinuria and of a related autosomal pseudogene mapped to 12q21. *Hum. Mol. Genet.*, **3**, 751–757.
52. Lupianez,D.G., Kraft,K., Heinrich,V., Krawitz,P., Brancati,F., Klopocki,E., Horn,D., Kayserili,H., Opitz,J.M., Laxova,R. *et al.* (2015) Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell*, **161**, 1012–1025.
53. Franke,M., Ibrahim,D.M., Andrey,G., Schwarzer,W., Heinrich,V., Schopflin,R., Kraft,K., Kempfer,R., Jerkovic,I., Chan,W.L. *et al.* (2016) Formation of new chromatin domains determines pathogenicity of genomic duplications. *Nature*, **538**, 265–269.
54. Rodriguez-Carballo,E., Lopez-Delisle,L., Zhan,Y., Fabre,P.J., Beccari,L., El-Idrissi,I., Huynh,T.H.N., Ozadam,H., Dekker,J. and Duboule,D. (2017) The HoxD cluster is a dynamic and resilient TAD boundary controlling the segregation of antagonistic regulatory landscapes. *Genes Dev.*, **31**, 2264–2281.
55. Barutcu,A.R., Maass,P.G., Lewandowski,J.P., Weiner,C.L. and Rinn,J.L. (2018) A TAD boundary is preserved upon deletion of the CTCF-rich Firre locus. *Nat. Commun.*, **9**, 1444.