A versatile platform for locus-scale genome rewriting and verification

Ran Brosha1,1, Jon M. Laurent1,2, Raquel Ordoñez3, Emily Huang4, Megan S. Hogan5, Angela M. Hitchcock6, Leslie A. Mitchell3, Sudarshan Pinglay7, John A. Cadley8, Raven D. Luther9, David M. Truong4,4, Jef D. Boeke1,2,3,4, and Matthew T. Maurano2,4

Institute for Systems Genetics, NYU Langone Health, New York, NY 10016; Department of Biochemistry and Molecular Pharmacology, NYU Langone Health, New York, NY 10016; Department of Biomedical Engineering, NYU Tandon School of Engineering, Brooklyn 11201, NY; and Department of Pathology, NYU Langone Health, New York, NY 10016

Contributed by Jef D. Boeke, January 14, 2021 (sent for review November 19, 2020; reviewed by Barak A. Cohen and Evgeny Kvon)

Routine rewriting of loci associated with human traits and diseases would facilitate their functional analysis. However, existing DNA integration approaches are limited in terms of scalability and portability across genomic loci and cellular contexts. We describe Big-IN, a versatile platform for targeted integration of large DNAs into mammalian cells. CRISPR/Cas9-mediated targeting of a landing pad enables subsequent recombinase-mediated delivery of variant payloads and efficient positive/negative selection for correct clones in mammalian stem cells. We demonstrate integration of constructs up to 143 kb, and an approach for one-step scarless delivery. We developed a staged pipeline combining PCR genotyping and targeted capture sequencing for economical and comprehensive verification of engineered stem cells. Our approach should enable combinatorial interrogation of genomic functional elements and systematic locus-scale analysis of genome function.

Significance

Functional analysis of noncoding genomic regulatory elements, which harbor the majority of common human disease and trait associations, is complicated by their cellular and genomic context sensitivity. We developed Big-IN, a method for rewriting large segments of mammalian genomes, including full genes and their surrounding regulatory elements. We demonstrate a flexible genomic verification pipeline to identify correctly engineered cells. We expect Big-IN will enable technologies for synthesis and assembly of large DNAs to catalyze a synthetic approach to regulatory genomics.

Author contributions: R.B., J.M.L., S.P., J.D.B., and M.T.M. designed research; R.B., J.M.L., R.O., E.H., M.S.H., A.M.H., L.A.M., and R.D.L. performed research; R.B., J.M.L., R.O., L.A.M., J.A.C., D.M.T., and J.D.B. contributed new reagents/analytic tools; R.B., J.M.L., R.O., J.A.C., J.D.B., and M.T.M. analyzed data; and R.B., J.D.B., and M.T.M. wrote the paper.

Reviewers: B.A.C., Washington University in St. Louis School of Medicine; and E.K., University of California, Irvine.

Competing interest statement: R.B., J.M.L., J.D.B., and M.T.M. are listed as inventors on a patent application describing Big-IN. L.A.M. is a founder of Neochromosome, Inc. and is on the Scientific Advisory Board of ReOpen Diagnostics. D.M.T. is an employee of Neochromosome, Inc. J.D.B. is a founder and director of CDI Labs, Inc., a founder of Neochromosome, Inc., a founder of and consultant to ReOpen Diagnostics, and serves or served on the Scientific Advisory Board of Sangamo, Inc., Modern Meadow, Inc., and Sampili, Inc.

This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

1Present address: Department of Research and Development, Pandemic Response Lab NYC, New York, NY 10016.
2Present address: Department of Research and Development, Neochromosome, New York, NY 10016.
3To whom correspondence may be addressed. Email: jef.boeke@nyulangone.org.

This article contains supporting information online at https://www.pnas.orglookup/suppl/doi:10.1073/pnas.2023952118/-/DCSupplemental.

Published March 1, 2021.
Results

Engineering the HPRT1 Locus in Human ESCs. To enable repeated, precise, and efficient delivery of large DNAs to a given locus, we employed a two-stage approach that first targets a short LP to replace a genomic locus of interest using CRISPR/Cas9-mediated homology directed repair (HDR) (Fig. 1A). A plasmid (pLP-TK) was engineered to include the human EF1α promoter (pEF1α) to drive ubiquitous expression of a single open reading frame (ORF) comprising a puromycin-resistance gene (PuroR) fused to a truncated Herpes simplex virus thymidine kinase (HSV1-ATK) gene (28) and a CreERT2 gene (29), separated by a P2A peptide (30). Interposed between the LP ORF and the vector backbone are heterotypic loxM (lox 2272) and loxP sites to permit subsequent RMCE. The lox sites are flanked by homology arms (HAs) corresponding to the genomic sequences flanking target sites at the targeted genomic locus. To facilitate clearance of the transiently transfected plasmid by inducing its linearization in vivo, the same gRNA target sequences and protospacer adjacent motifs (PAM) were cloned into the vector backbone just outside the HAs.

We targeted the X-linked HPRT1 locus for LP integration to permit counterselection with the cytotoxic antimetabolite 6-Thioguanine (6-TG) (31). H1 male hESCs, which harbor a single copy of HPRT1, were cotransfected with pLP-TK and pCas9 plasmids (32) expressing gRNAs targeting a 42-kb region, including the HPRT1 gene for replacement. Cells were sequentially treated with 6-TG and puromycin to select for HPRT1 loss and LP-TK gain, followed by clonal isolation. Correct LP-TK integration was verified by PCR genotyping using primers targeting the novel junctions between LP-TK and the genomic sequences beyond the HAs (Fig. 1B). A candidate clone (58I) was selected for further validation. Junction PCR amplicons were subjected to Sanger sequencing, to verify correct LP-TK integration at base pair resolution (SI Appendix, Fig. S1A). Quantitative real-time PCR (qRT-PCR) confirmed loss of HPRT1 mRNA expression and gain of CreERT2 expression (SI Appendix, Fig. S1B). Robust cytotoxic activity of HSV1-ΔTK following ganciclovir (GCV) treatment was validated in a kill curve (SI Appendix, Fig. S1C). We also developed a lentiviral reporter assay for Cre activity, which indicated that CreERT2 is rapidly and efficiently activated by tamoxifen (SI Appendix, Fig. S1D). Thus, the function of all three components of the LP ORF was verified.

To facilitate comprehensive genomic verification of multistep cellular engineering with these complex constructs, we developed a modular next-generation sequencing analysis approach, which independently maps short reads to both reference genomes (hg38 and mm10) and custom references for each engineering construct. We further applied hybridization capture sequencing (Capture-seq) approach to efficiently verify correct engineering of screened clones (Fig. 1C). We employed nick translation to generate bait in a rapid, flexible, and cost-effective fashion. Using this mapping pipeline, whole-genome sequencing (WGS) of clone 58I verified loss of the targeted HPRT1 locus, gain of LP-TK, and absence of LP-TK backbone and pCas9 (Fig. 1D–F).

Integration relied on 1-kb HAs to correctly target the LP, but HA length reduces the efficiency of PCR genotyping from genomic DNA (Fig. 1B) and impedes the mapping of short sequencing reads that definitively span the LP-HA/genome junctions. Therefore, we measured relative integration efficiency with shorter HAs. We integrated a series of pLP-TK plasmids with varying HA lengths and estimated on-target integration as the relative number of cells surviving puromycin and 6-TG selection, revealing that efficient integration could be performed with HAs as short as 100 bp (SI Appendix, Fig. S1E), facilitating subsequent sequence-based mapping of integration sites.

We also assessed the efficacy of our in vivo linearization strategy to reduce off-target integration of transiently transfected plasmids. We designed two pLP-TK plasmids differing only in the presence of the LP-flanking gRNA sites required for in vivo linearization, targeted them to HPRT1, selected for correct integrants with puromycin and 6-TG, and subjected the pool of cells to Capture-seq. We found that the relative coverage depth of the LP backbone was lower for the in vivo-linearized pLP-TK (SI Appendix, Fig. S1F), possibly due to enhanced HDR efficiency (33) and reduced plasmid half-life (which was evident from shortened transient puromycin resistance of the transfected cells).

Delivery of large DNA through cassette exchange is an infrequent event, requiring selection to obtain practical efficiency. The HSV1-ATK gene encoded by LP-TK is a widely used counterselectable marker that renders cells sensitive to GCV by converting it to the toxic metabolite GCV-triphosphate (GCV-TP), which inhibits DNA synthesis and leads to cell death (34). To demonstrate a counterselection-based approach to isolation of successful RMCE events, we designed a minimal 27-kb playout (PL1), comprising a pEF1α-driven GFP-T2A-BSD (blasticidin S deaminase) ORF flanked by loxM and loxP sites (Fig. 1G). H1 LP-TK cells were transfected with a PL1-harboring plasmid (pPL1) and LP-derived CreERT2 activity was induced with tamoxifen. Cells were selected with blasticidin to enrich for PL1-expressing cells, following by GCV counterselection of TK-expressing cells. PCR genotyping of isolated clones showed a 100% rate of replacement of LP-TK with PL1 (Fig. 1H). Capture-seq analysis of four selected clones confirmed the presence of PL1, the absence of any plasmid backbone, and the loss of LP-TK (Fig. 1I and J). The integrated PL1 was transcriptionally active, as evident from GFP expression (SI Appendix, Fig. S1G).

Efficient Counterselection for Delivery. To quantify the efficacy of TK/GCV counterselection in H1 hESCs, we mixed TK− and TK+ (LP-TK) cells at different ratios and treated these co-cultures with GCV. More than 80% of the TK− cells died when mixed at a 1:1 ratio with TK+ cells, and all died when mixed at a 1:10 ratio (Fig. 2A). Indeed, it is known that GCV-TP can diffuse from TK+ cells to TK− cells via gap junctions (35, 36). The resulting bystander cell death in TK− cells limits the ability to recover rare events (Fig. 2B).

Therefore, we tested an alternative counterselection strategy (Fig. 2C) that relies on the X-linked PIGA (phosphatidylinositol glycan anchor biosynthesis class A) gene, which encodes an enzyme crucial for the biosynthesis of glycosylphosphatidylinositol (GPI) anchors (37) and renders cells sensitive to proaerolysin, a bacterial protoxin. Proaerolysin perforates the plasma membrane upon binding to GPI anchors on the cell surface, resulting in rapid cell death (38). Furthermore, PIGA activity can be quantitatively monitored by measuring levels of CD59, a broadly expressed membrane-linked GPI-anchored protein (39). Deletion of PIGA can be selected for with proaerolysin after a short period to allow for loss of PIGA protein and subsequent loss of GPI-anchored proteins from the cell surface (40).

While proaerolysin efficiently killed parental H1 hESCs, ΔPIGA cells, in which the PIGA gene was deleted using CRISPR/Cas9 (Materials and Methods), were entirely resistant (SI Appendix, Fig. S2). Integration of an LP expressing a human mini PIGA gene (hminiPIGA) to the HPRT1 locus resensitized H1 ΔPIGA hESCs to proaerolysin and restored CD59 expression (SI Appendix, Fig. S2 B and F). Importantly, rare ΔPIGA H1 hESCs were efficiently isolated when cocultured with parental H1 cells by applying proaerolysin selection (Fig. 2D). This suggested that
LP-expressed hmPIGA permits negative selection of LP-PIGA cells to effectively enrich for correct delivery events. Recovery of rare events where a payload replaces the LP requires that expression of hmPIGA is stably maintained following puromycin withdrawal. However, while nearly all H1 LP-PIGA cells maintained high CD59 levels in the presence of puromycin, a substantial proportion of cells spontaneously lost expression of hmPIGA following puromycin withdrawal (41). To develop an approach for allele-specific engineering of diploid loci, we employed C57BL/6j × CAST/EiJ (BL6CAST or BL6xC) F1 hybrid mESC cells (41), the genome of which harbors heterozygous point variants every 140 bp on average (42). We targeted the Sox2 locus, which encodes a master transcription factor essential for regulation of pluripotency and differentiation (43, 44). We designed gRNAs targeting the flanks of a 143-kb genomic region that includes the Sox2 coding sequence, promoter, long-distance regulatory regions, and several noncoding genes (44, 45). These gRNAs target BL6-specific PAMs to facilitate allele-specific engineering. We constructed pLP-PIGA to support allele-specific engineering of the Murine Sox2 Locus. To develop an approach for allele-specific engineering of diploid loci, we employed C57BL/6j × CAST/EiJ (BL6CAST or BL6xC) F1 hybrid mESC cells (41), the genome of which harbors heterozygous point variants every 140 bp on average (42). We targeted the Sox2 locus, which encodes a master transcription factor essential for regulation of pluripotency and differentiation (43, 44). We designed gRNAs targeting the flanks of a 143-kb genomic region that includes the Sox2 coding sequence, promoter, long-distance regulatory regions, and several noncoding genes (44, 45). These gRNAs target BL6-specific PAMs to facilitate allele-specific engineering. We constructed pLP-PIGA to support
Proaerolysin resistant generated from a BAC covering the specific primer harboring 4 mismatched base pairs relative to the Sox2 LP-PIGA backbone or pCas9. We confirmed the allele-specific (50%), likely resulting from retention or off-target integration of neighboring cells and induces bystander cell death in TK toxic membrane-impermeable compound GCV-TP, which diffuses into multiple vector backbones), which eliminated eight Ori+ clones BL6xCAST gRNA target sites. Isolated clones. Of 40 clones screened using PCR genotyping, 16 proaerolysin for 1 d and stained with Crystal violet 3 d later. (A) Parental (TK−) and LP-TK (TK+) H1 hESCs were cocultured at the indicated ratios, treated with 1 μM GCV for 4 d, and assayed for the number of live cells using PrestoBlue. Cell counts are shown relative to unmixed parental cells. Bars show mean ± SD (n = 2). (B) GCV enters TK+ cells and is metabolized into the toxic membrane-impermeable compound GCV-TP, which diffuses into neighboring cells and induces bystander cell death in TK− cells. (C) Big-IN counterselection strategy using PIGA/proaerolysin. (D) Parental and ΔPIGA H1 hESCs cocultured at the indicated ratios for 3 d were treated with 1 nM proaerolysin for 1 d and stained with Crystal violet 3 d later.

counterselection-based delivery to cell lines lacking a functional Piga gene (Fig. 2C). The LP ORF includes four components, each separated by three mutually recoded P2A peptides: mScarlet (46), CreERT2, PuroR, and hmPIGA (Fig. 3A). The ORF is flanked by heterotypic loxM/loxP sites, short HAs, and gRNA target sites.

We transfected pLP-PIGA and pCas9 plasmids into BL6xCAST ΔPIGA mESCs, selected cells with puromycin, and isolated clones. Of 40 clones screened using PCR genotyping, 16 (40%) contained both novel junctions (Fig. 3B). Passing clones were further screened with primers to detect Ori (common to multiple vector backbones), which eliminated eight Ori+ clones (50%), likely resulting from retention or off-target integration of LP-PIGA backbone or pCas9. We confirmed the allele-specific loss of Sox2 in 15 (94%) of the 16 clones using a BL6 allele-specific primer harboring 4 mismatched base pairs relative to the CAST allele (SI Appendix, Table S1).

A successful LP-PIGA integration (clone A1) and a clone that failed PCR genotyping were subjected to Capture-seq using bait generated from a BAC covering the Sox2 region, and the pLP-PIGA and pCas9 plasmids. Inspection of coverage depth at the 143-kb Sox2 genomic locus revealed a 50% reduction for clone A1 compared with parental mESCs or the failed clones (Fig. 3C), as expected for complete loss of the targeted BL6 allele. Clone A1 also showed specific gain of LP-PIGA with no coverage of the LP-PIGA backbone or pCas9, whereas the failed clone showed clear presence of the LP-PIGA backbone (Fig. 3 D and E). Expression of LP-PIGA components and BL6 allele-specific loss of Sox2 expression in clone A1 was verified through qRT-PCR analysis (SI Appendix, Fig. S3A), which was chosen for future payload deliveries. We confirmed efficient isolation of rare mESCs using the Piga/proaerolysin counterselection strategy (SI Appendix, Fig. S3 B and C), and observed a similar silencing effect to LP-TK in the absence of positive selection (SI Appendix, Fig. S3D). In summary, we have demonstrated an efficient strategy for allele-specific LP integration and a comprehensive pipeline for verification of correctly engineered cells.

Efficient Delivery to mESCs. We attempted to deliver payloads to LP-PIGA mESCs using a positive/negative selection strategy. However, all clones that survived blasticidin and proaerolysin selection manifested multiple-copy payload gain, including its vector backbone, and without LP-PIGA loss (SI Appendix, Fig. S4A). We transiently augmented Cre activity through cotransfection of a Cre expression plasmid (pCAG-Cre). Additionally, we cloned a ΔTK expression cassette (BBTK) into the payload backbone to permit GCV-based counterselection against surviving colonies harboring off-target integrants. Cotransfection of pPL1-BBTK and pCAG-Cre readily resulted in efficient PL1 integration (Fig. 4B). To assess efficiency of larger payloads, pSox2346kb-MC-BBTK was constructed including a 46-kb region of the Sox2 locus and containing a marker cassette to enable positive selection (Fig. 4A). Upon delivery and selection, PCR genotyping verified that 99% of clones harbored correct payload integration (Fig. 4B). Six PCR-validated clones of each payload type were then chosen for Capture-seq analysis. Mapping sequencing reads to the PL1 sequence or mouse genome revealed that all clones had complete coverage of the delivered payload (Fig. 4C). In Sox2346kb-MC clones, coverage depth was restored to parental levels over the genomic region corresponding to Sox2346kb while the remaining 97 kb of the Sox2 deletion was unaffected (Fig. 4D and SI Appendix, Fig. S4B). Analysis of known CAST single nucleotide variants (SNVs) further confirmed reintroduction of BL6 alleles. There was no evidence for the gain of the payload backbone in any of the clones analyzed (SI Appendix, Fig. S4C), and all 79 clones lost LP-PIGA (SI Appendix, Fig. S4D). Selected PL1 and Sox2346kb-MC cells both expressed the payload-derived BSD, while Sox2346kb-MC clones also partially restored the expression of the BL6 allele of Sox2 (SI Appendix, Fig. S4E). In addition, both cell types showed expression of payload-derived GFP (SI Appendix, Fig. S4F).

This approach leaves a BSD-GFP transcriptional unit (TU) integrated with the payload, which might affect the activity of nearby genes or regulatory elements. To develop an alternate architecture and selection strategy for scarless delivery, we constructed pSox2143kb, which harbors the entire 143-kb Sox2 locus and contains a marker cassette to enable positive selection (Fig. 4A). We delivered pSox2143kb to LP-PIGA mESCs together with pCAG-Cre, which encodes a codon-optimized Cre recombinase, and selected cells transiently with blasticidin to enrich for payload-transfected cells, followed by proaerolysin selection to eliminate unrecombined LP-PIGA mESCs. PCR genotyping identified four clones that lost LP-PIGA, one of which (G11) was positive for the newly formed BL6 allele genomic junctions (Fig. 4E). Capture-seq analysis verified the restoration of the entire 143-kb BL6 allele in clone G11, without gain of the payload backbone (Fig. 4F). Finally, qRT-PCR analysis confirmed
that the expression of the BL6 allele of Sox2 was completely restored, and expression of hmPIGA and BSD was undetectable (Fig. 4G).

To demonstrate the flexibility of Big-IN for delivery of payloads to additional loci, LP-PIGA2 was integrated into chromosome 7 of BL6xCAST ΔPiga mESCs, replacing a 157-kb region of the Isg2/H19 locus (SI Appendix, Fig. S5A). We transduced these cells with pCAG-iCre and either the nonscarless payload pSox246kb-MC-BBTK or the scarless pSox246kb payload. Following stable positive selection with blasticidin and negative selection with proaerolysin and GCV, 95 of 96 (99%) of the presence of the right payload junction for 24 of 25 clones and similarly verified. Further verification of selected clones confirmed the presence of the right payload junction for 24 of 25 clones and the absence of pCAG-iCre in all clones. Capture-seq analysis of parental BL6xCAST mESCs junctions are depleted of BL6 reads, both junctions in-identical to the CAST allele, the high rate of endogenous variability, efficiency, and precision at three loci in mouse and human integrations in BL6xCAST mESCs permitted specific detection of the correctly integrated payload. Analysis of read pairs overlapping informative BL6xCAST variants revealed that, while LP-PIGA mESCs junctions are depleted of BL6 reads, both junctions including the BL6 allele are restored in Sox2Δ143kb clone A1, and an example failed clone from an independent LP-PIGA delivery. Reads were mapped to the references indicated above. Cross-mapping sequences are shaded gray.

Genomic Screening of On- and Off-Target Integrations. To screen genomic data for on- and off-target integration events, we developed bamintersect, which leverages a modular mapping approach where sequencing reads are mapped separately to two reference genomes. Bamintersect then jointly analyzes both mappings to detect read pairs indicative of a junction (Fig. 5A). Nearby reads in each reference are clustered and masked for uninformative regions (Materials and Methods). We applied bamintersect to confirm LP integration and payload delivery for the genomic engineering events described herein, the majority of which were verified by identifying multiple reads supporting the novel junctions between the integrated sequence and its flanks (Figs. 5 and Dataset S2).

For LP-PIGA integration at Sox2 in BL6xCAST mESCs, two of the four analyzed clones (A1 and C5) were validated for the presence of both correct junctions, whereas one clone (C2) was validated only for the left junction, and an additional clone (G2) demonstrated off-target LP integration at chromosome 1 (Fig. 5C). Bamintersect also detected an unexpected junction between the right and left HAs for clones A1 and C5 (Dataset S2). PCR confirmed a tandem head-to-tail multimeric LP integration (SI Appendix, Fig. S7 A and B). All payloads delivered to clone A1 were verified as correctly targeted (Fig. 5 B and D–H) and lacked tandem LP junctions (SI Appendix, Fig. S7C and Dataset S2), suggesting the tandem LP supported productive recombination upon Cre expression.

Several junctions were impossible to confirm using bamintersect for technical reasons. For LP-TK integration at Hprt1, the 1-kb HAs precluded mapping reads spanning the junction between LP-TK and hg38. For PL1 deliveries to both Hprt1 and Sox2, the left junction was nearly identical to that of the replaced LP. Although Sox2Δ143kb delivery results in junctions nearly identical to the CAST allele, the high rate of endogenous variation in BL6xCAST mESCs permitted specific detection of the correctly integrated payload. Analysis of read pairs overlapping informative BL6xCAST variants revealed that, while LP-PIGA mESCs junctions are depleted of BL6 reads, both junctions including the BL6 allele are restored in Sox2Δ143kb clone G11 mESCs (Fig. 5F). These results support the utility of bamintersect as a sensitive, scalable, and unbiased tool for detection of on and off-target integration events.

Discussion

We have described Big-IN, a platform for scalable targeted integration into mammalian genomes, and demonstrated its flexibility, efficiency, and precision at three loci in mouse and human
ESCs. Big-IN first targets an LP to a locus of interest using CRISPR/Cas9-mediated HDR, which permits single-step payload integration through Cre-mediated RMCE (Fig. 6). Single-step payload integration minimizes confounding technical factors by permitting repeated deliveries to the same allele, and is thus ideal for in-depth interrogation of a given locus (47). LP cell lines can be

Fig. 4. Efficient delivery to mESCs. (A) Delivery of three payloads to BL6xCAST ΔPiga LP-PIGA mESCs. (B) PCR genotyping of PL1 (Upper) and Sox246kb-MC mESCs. (Lower) mESC clones for novel junctions illustrated in A. E, empty well; L, ladder. (C and D) Capture-seq analysis of chosen PL1 and Sox246kb-MC mESC clones, with Parental and LP-PIGA mESCs as controls. (C) Sequencing coverage mapped to PL1. pEF1α (shaded gray) is present in both LP-PIGA and PL1. (D) Gain of coverage in Sox246kb-MC mESCs at the 46-kb payload region. Black ticks under each coverage track indicate detection of BL6 alleles at known SNVs. Internal payload duplication marked in Clone C9 (SI Appendix, Fig. S6). (E) PCR genotyping of Sox2143kb clones for BL6-specific junctions and loss of LP-PIGA, as illustrated in A. (F) Sox2143kb mESCs show restored coverage of the full 143-kb genomic region corresponding to the payload. Black ticks under each coverage track indicate detection of BL6 alleles at known SNVs. Coverage at right shows no retention of payload backbone. Cross-mapping sequences are shaded gray. (G) qRT-PCR expression analysis of Sox2143kb clone G11 and LP-PIGA mESCs for mRNAs from BL6 and CAST Sox2 alleles, payload-derived BSD, and LP-harbored hmPIGA. Bars represent mean + SD for technical replicates (n = 3).
intensively verified following CRISPR/Cas9 expression to ensure the absence of undesired rearrangements or other off-target events, while subsequent Cre expression for payload delivery is expected to be less mutagenic (12).

Our cell-engineering approach is designed to scale rapidly across multiple loci and cell lines. While we have demonstrated Big-IN in both mouse and human ESCs, it is possible that engineering other mammalian cell lines with LPs may require optimization. Indeed, we note that despite the success of the LP-expressed CreERT2 strategy in H1 hESCs, exogenous Cre was required in mESCs. We have shown that the selection and delivery methods described herein can be redeployed in a modular
fashion to overcome challenges associated with different cell types and loci. For example, the LP can employ either HSV1-ΔTK or hmPIGA as a counterselectable marker, with the former suffering from a bystander effect, and the latter requiring prior engineering to inactivate the endogenous PIGA/Piga gene. While loss of GPI-anchored proteins has no detectable phenotype in culture, mice completely lacking Piga function are inviable (48).

A reversible Piga knockout using an excisable intronic transcription terminator as previously engineered for HPRT1 (49) would enable efficient recovery of Piga-expressing cells by sorting for a GPI-anchored membrane protein. A similar trade-off relates to the inclusion of a positive selection marker on the payload, which augments delivery efficiency, while its placement in the payload backbone enables scarless integration (SI Appendix, Fig. S5B). Quantitative comparison of the efficiency of the Big-IN deliveries described herein is confined by technical differences and the need to replate rapidly growing ESCs, but we expect that future improvements will enhance overall efficiency and its application to diverse cellular contexts.

Our verification strategy is tailored to enable early verification of engineering outcomes. For example, the use of locally generated Capture-seq bait circumvents the cost and delay of complex assembly to diverse cellular contexts. For a GPI-anchored membrane protein. A similar trade-off relates to the inclusion of a positive selection marker on the payload, which augments delivery efficiency, while its placement in the payload backbone enables scarless integration (SI Appendix, Fig. S5B). Quantitative comparison of the efficiency of the Big-IN deliveries described herein is confined by technical differences and the need to replate rapidly growing ESCs, but we expect that future improvements will enhance overall efficiency and its application to diverse cellular contexts.

The effectiveness of Big-IN for integration of large DNA constructs (SI Appendix, Fig. S5B) suggests that it might also be optimized to support integration of complex libraries for saturation mutagenesis of shorter elements (50–52), and eventually, analysis of large constructs in a pooled library format. When combined with the rapidly evolving big DNA synthesis field (14, 25), we envision that Big-IN will enable designer-like control over mammalian genomes and facilitate a synthetic approach to genome biology.

Materials and Methods

Additional information is available in SI Appendix.

Cell Culture. WA01 (H1) hESCs were purchased from WiCell. The use of H1 hESCs was approved by the New York University School of Medicine Embryonic Stem Cell Research Oversight Committee. H1 hESCs were initially grown for 2 wk on plates coated with Matrigel (Corning 354277) in mTeSR medium (Stem Cell Technologies B8580) and subsequently transferred to plates coated with Geltrix ( Gibco A1413302) and StemFlex medium (ThermoFisher A3349401) supplemented with 1% Pen-Strep (ThermoFisher 15140122). For routine passaging, cells were dissociated into clumps with Versene ( Gibco 15-040-066) and gentle trituration. Wide-orifice pipette tips were used when handling small volumes of cell suspension.

C57BL6/J x CAST/EiJ (BL6xCAST) clone 4 mESCs (41) were kindly provided by David Spector, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY. mESCs were cultured on plates coated with 0.1% gelatin (EMD Millipore ES-006-B) in 80/20 medium comprising 80% 2i medium and 20% mESC medium. 2i medium contained a 1:1 mixture of Advanced DMEM/F12 (ThermoFisher 12634010) and Neurobasal-A (ThermoFisher 10398016) supplemented with 10% FBS (BenchMark 100-106), 0.1 mM 2-mercaptoethanol (Sigma M3148), 1,250 U/mL LIF (ESGRO ESG11070), 3 μM CHIR99021 (R&D Systems 4423), and 1 μM PD0325901 (Sigma P20162), mESC medium contained knockout DMEM (ThermoFisher 10082018) supplemented with 15% FBS ( Benchmark 100-106), 0.1 mM 2-mercaptoethanol, 1% glutamax, 1% MEM nonessential amino acids (ThermoFisher 11400050), 1% nucleosides (EMD Millipore ES-008-D), 1% Pen-Strep, and 1,250 U/mL LIF. HEK-293T cells were cultured in DMEM supplemented with 10% FBS, 1 mM sodium pyruvate (ThermoFisher 11360070), 1% glutamax, and 1% Pen-Strep. All cells were grown at 37 °C in a humidified atmosphere of 5% CO2 and passaged on average twice per week.

Genome Engineering. Relevant genomic coordinates are listed in SI Appendix, Table S3.

H1 hESCs were transfected using the Neon Transfection System (ThermoFisher). Cells were treated several hours prior to transfection with StemFlex medium supplemented with 1% RevitaCell Supplement (ThermoFisher A2644501). Cells were washed with PBS, disassociated into a single-cell suspension using TrypLE-Select (ThermoFisher 12536011), which was neutralized with StemFlex medium, spun down at 200 relative centrifugal force

Materials and Methods

Additional information is available in SI Appendix.

Fig. 6. Targeted locus-scale genome rewriting using Big-IN. An allele of interest is replaced by a LP using CRISPR/Cas9-mediated HDR. A pair of gRNAs target the replaced allele and the LP, and short HAs mediate precise LP integration. Puromycin selects for LP-harboring cells. Next, Cre-mediated recombination of two pairs of heterotypic loxM and loxP sites results in LP/payload cassette exchange and resistance to either GCV for HSV1-ΔTK LPS, or proarolysin for hmPIGA LPS in cells where endogenous PIGA is inactivated. Positioning the blasticidin cassette (BSD) within the payload permits selection for high-efficiency integration; positioning BSD on the payload backbone permits transient selection for scarless delivery. Additionally, backbone HSV-ΔTK (Left) can be counterselected with GCV to limit off-target integration. Each engineering step is comprehensively verified by PCR genotyping, WGS or Capture-seq, and functional assays.
(rcf) for 3 min, supernatant aspirated, and cells resuspended in PBS. Next, 1 × 10^7 cells per transfection were spun down at 200 rcf for 3 min and resuspended in 200 μL Neon Buffer R at a final concentration of 2 × 10^7 cells/mL; 50 μL of cell suspension were mixed with 50 μL Neon Buffer R containing 10 μg of total DNA per transfection. Nucleofection used Neon 100 μL Tips with two 20-ms pulses at 1,100 V. Transfected cells were transferred into plates coated with rhLaminin-521 (Gibco A29249) prefilled with StemFlect medium supplemented with 1% RevitaCell. PIGA deletion was performed with 5 μg of each pCas9 plasmid expressing gRNAs hPIGA-g1 and hPIGA-g2 and cells were selected with 200 μM puromycin for 1 to 2 wk posttransfection. These ΔPIGA cells were used for subsequent LP-PIGA integrations. All LP integrations were performed using the Neon Transfection System. Cells were washed with PBS, dissociated into a single-cell suspension using TrypLE-Select (Gibco), which was neutralized with mESC medium, spun down at 200 rcf for 3 min, supernatant aspirated, and cells resuspended in PBS. Next 1 × 10^7 cells per transfection were spun down at 200 rcf for 3 min and resuspended in 100 μL Neon Buffer R at a final concentration of 2 × 10^7 cells/mL. Per transfection, 50 μL of cell suspension were mixed with 50 μL Neon Buffer R containing 10 μg of total DNA and nucleofected using Neon 100 μL Tips with two 20-ms pulses at 1,200 V. Transfected cells were transferred into gelatin-coated plates pre-filled with 80/20 medium. Piga deletion was performed with 5 μg of each pCas9 plasmid expressing gRNAs mPiga-g1 and mPiga-g2, and cells were selected with 2 nM pyr-G418. Integration events were selected using a combination of 1 μg/mL puromycin and 6-TG, as indicated. H1 PL1 integrations were performed using 5 μg PPL1. Cells were treated with 2 μM 4-hydroxytamoxifen (Tam) the day following transfection for 3 h, selected with 5 μg/mL blasticidin S for 8 d, followed by 4 of 4 selection with 100 nM GCV to eliminate TK-expressing cells.

LP integrations and genomic deletions in BL6xCAST mESCs were performed using the Neon Transfection System. Cells were washed with PBS, dissociated into a single-cell suspension using TrypLE-Select (Gibco), which was neutralized with mESC medium, spun down at 200 rcf for 3 min, supernatant aspirated, and cells resuspended in ice-cold PBS, counted, and 5 × 10^5 cells was transfection was spun down at 200 rcf for 3 min and resuspended in a room temperature mixture of 82 μL nucleofector solution and 18 μL nucleofector supplement from the Mouse ES Cell Nucleofector kit (Lonza VPH-1001). Per transfection, 100 μL of cell suspension was mixed with 10 μL TE containing 2.25 to 5 μg of total DNA, and nucleofected using program A-23. PL1 deliveries were performed with 1.5 μg pPL1-BBTK and 0.75 μg pCas9-Cre (Addgene plasmid #13775). pSox2Plan-MC deliveries (failed deliveries) were performed with 35 μg pSox2Plan-MC. Payloads were treated with 0.1% SDS and 3 times at 65 °C for 15 min, and then 24 h after transfection. Cells were selected with blastidicin constitutively starting day 1 posttransfection and with 2 μM pyr-G418 for 2 d starting day 14 posttransfection. pSox2Plan-MC-BBTK deliveries were performed with 3 μg pSox2Plan-MC-BBTK and 1 μg pCAS-Cre. Payload-transfected mESCs were treated with 200 μM Tam for 24 h before and after transfection. mESCs were selected with 10 μg/mL blasticidin S for 2 d starting day 7 posttransfection. Payload deliveries (failed deliveries) were performed with 35 μg pSox2Plan-MC-BBTK and 2 μg pCAS-Cre. Cells were selected with blastidicin constitutively starting day 1 posttransfection and with 2 μM pyr-G418 for 2 d starting day 14 posttransfection. pSox2Plan-MC-BBTK deliveries were performed with 3 μg pSox2Plan-MC-BBTK and 1 μg pCAS-Cre. Payload-transfected mESCs were treated with 200 μM Tam for 24 h before and after transfection. mESCs were selected with 10 μg/mL blasticidin S for 2 d starting day 7 posttransfection. Payload deliveries to BL6xCAST Ig2DH19 were performed with 5 μg pSox2Plan-MC-BBTK or pSox2Plan-MC and 2 μg pCAS-Cre. Cells were selected with blastidicin either transiently during days 1 and 2 posttransfection (pSox2Plan-MC) or constitutively (pSox2Plan-MC-BBTK), followed by 2 nM pyr-G418 selection during days 7 and 8 posttransfection. pSox2Plan-MC-BBTK transfected cells were further selected with 1 μM GCV during days 9 and 10 posttransfection.

Preparation of Illumina Double-Stranded DNA Libraries. Genomic DNA was isolated from cells using the DNeasy Blood and Tissue Kit (Qiagen) according to the manufacturer's protocol. For the generation of per-base coverage depth tracks and quantification was performed using BEDOPS v2.4.35 (58). Data were visualized using the University of California, Santa Cruz Genome Browser. The sequencing processing pipeline is available at https://github.com/mauranlab/macmapping.

Brosh et al. A versatile platform for locus-scale genome rewriting and verification
PNAS | 9 of 11
https://doi.org/10.1073/pnas.2023952118

Targeted Resequencing Using Capture-Seq. Baits for sequence capture were prepared from BAC or plasmid DNA containing the sequence of interest. BAC coordinates are listed in SI Appendix, Table S3. Biotin-16-DUTP (Roche) was incorporated into bait DNA using a Nick Translation kit (Roche). The reaction (total volume 20 μL) was set-up in a 200-μL PCR tube on ice as follows: 2 μg of BAC DNA, 10 μL of 0.1 M Mm Biotin-DUTP/dITP mixture (1 volume Biotin-16-DUTP, 2 volumes dITP, 3 volumes dATP, 3 volumes dCTP, and 3 volumes dGTP), 2 μL of 10× nick translation buffer, and 2 μL of enzyme mixture. Nick translation was carried out at 15 °C for 16 or 8 h (for BAC or plasmid DNA, respectively) in a thermal cycler. The reaction was stopped by addition of 1 μL 0.5 M EDTA and heating at 65 °C for 10 min or cooling at 4 °C overnight. Biotinylated baits were purified by ethanol precipitation, resuspended in 50 μL H2O, and the concentration was measured on a Nanodrop instrument.

Targeted sequencing using in-solution hybridization capture (Capture-seq) was performed as described previously (SA), with modifications. One microgram biotinylated DNA bait and 10 μg Cot-1 human or mouse DNA (Invtrogen) were combined with universal and sample-specific blocking oligos and lyophilized using a SpeedVac. Lyophilized DNA was resuspended in 12 μL TE (pH 7.5) and overlaid with mineral oil. In a thermal cycler, the DNA mixture was denatured at 96 °C for 5 min, incubated at 65 °C for an additional 15 min, and then 12 μl of 2× hybridization buffer (1.5 M NaCl, 40 μM sodium phosphate buffer [pH 7.2], 10 μM EDTA [pH 8], 10× Denhardt's, and 0.2% SDS) was added to the DNA, and the mixture was prehybridized for 6 h at 65 °C.

A total of 1 μg from up to two to eight libraries were pooled into a single 200-μL PCR tube for a single-capture reaction. Library DNA was diluted in H2O to a final volume of 12 μL and overlaid with mineral oil. Library DNA was denatured at 96 °C for 5 min, incubated at 65 °C for an additional 15 min, and then 12 μL of 2× hybridization buffer was added to the tube of prehybridized bait DNA, and the mixture was incubated at 65 °C for 16 to 22 h. For each capture reaction, 50 μL of MyOne streptavidin-coated magnetic beads (Invitrogen) were washed with 1× B&W buffer (5 mM Tris-HCl pH 7.5, 0.5 mM EDTA, 1 M NaCl) three times, and resuspended in 150 μL 1× B&W buffer in a low-retention microcentrifuge tube. The hybridization mix (12 μL) plus 48 μL 2× B&W buffer (10 mM Tris-HCl pH 7.5, 1 mM EDTA, 2 M NaCl) were then combined with the prewashed magnetic beads, and incubated at room temperature for 30 min with rotation. The magnetic beads were washed once at 25 °C for 15 min in 1× SSC with 0.1% SDS and three times at 65 °C for 15 min in 0.1× SSC with 0.1% SDS. To denature the captured library DNA, the beads were resuspended in 100 μL 1× SSC at 65 °C for 10 min, allowing the beads to separate on a magnetic rack, the supernatant (containing enriched library DNA) was transferred to a new tube, neutralized with 100 μL 1× Tris-HCl pH 7.5, and purified using the DNA Clean and Concentrate-5 Kit (Zymo Research). Four microliters of the captured library DNA were evaluated using qPCR to determine the optimal number of final PCR amplification cycles. Captured libraries were then amplified with KAPA Hi-Fi Hotstart ReadyMix (Roche). Bait sets and sequencing statistics are listed in Dataset S1.

Sequencing Data Processing. Illumina libraries were sequenced in paired-end mode on an Illumina NextSeq 500 operated at the Institute for Systems Genetics or a NovaSeq 6000 operated by the New York University Langone Health Genome Technology Center. Reads were demultiplexed with Illumina bcl2fastq v2.20 requiring a perfect match to indexing BC sequences. All WGS and Capture-seq data were processed using a uniform mapping and peak calling pipeline. Illumina sequencing adapters were trimmed with Trimmomatic v0.39 (55). Sequencing reads were aligned using BWA v0.7.17 (56) to a reference genome (GRCh38/hg38 or GRCm38/mm10), including unscaffolded contigs and alternate references, as well as independently to custom references for relevant vectors. PCR duplicates were marked using samblaster v0.1.24 (57). Generation of per base coverage depth tracks and quantification was performed using BEDOPS v2.4.35 (58). Data were visualized using the University of California, Santa Cruz Genome Browser. The sequencing processing pipeline is available at https://github.com/mauranlab/macmapping.
Genotype Analysis. Variant calling was performed on sequenced BL6xCAST samples to verify correct allele-specific engineering using a standard pipeline based on bctools v1.9:

bctools mpileup--redu-BAQ-adjust-MQ 50-gap-fac 0.05-max-depth 10000-max-idpex 20000 --DP,AD-output-type u
bctools call-keep-alts --ploidy 1-multiallelic-calculator -f GQ-output-type u

Raw pileups were filtered using:
bctools norm-check-ref w-output-type u
bctools filter -i "INFO/DP>=10 & QUAL>=10 & GQ>=90 & FORMAT/DP>=10" --SnipGap 3--IndelGap 10-set-GTs --output-type u
bctools view -i "GT="alt-"--trim-alt-alleles-output-type z

SNVs called in each sample were intersected with expected BL6xCAST heterozygous sites based on known variants called for CAST/EU (42).

Analysis of Integration Junctions Using Bamintersect. Bamintersect enables efficient filtering paired-end genomic sequencing based on dual independent mapping to different references, typically a mammalian reference genome (hg38 or mm10) and an engineered reference of interest (LP or payload). Bamintersect identifies junctions through analysis of read pairs which map to a different reference. For LP/payload genomes, the read’s mate is required to be unmapped to that genome. Reads must be fully mapped with ≤1 mismatched base and no clipping, insertions, or deletions, or duplicate or supplementary alignments are excluded.

1. M. T. Maurano et al., Systematic localization of common disease-associated variation in regulatory DNA. Science 337, 1190–1195 (2012).
2. R. D. Palmer, R. L. Brinster, Germ-line transformation of mice. Annu. Rev. Genet. 20, 465–489 (1986).
3. A. Battle, C. D. Brown, B. E. Engelhardt, S. B. Montgomery; GTEx Consortium; Laboratory Data Analysis &Coordinating Center (LDACC)—Analysis Working Group; Statistical Methods groups—Analysis Working Group; Enhancing GETx (GETx) groups; NIH Common Fund; NICHHD; NIH/NIMHD; NIAID; Bioinformatic Collection Site—NIH; Biopsy Collection Source Site—NIH; Biopsy Core Resource—VARS; Brain Bank Repository—University of Miami Brain Endowment Bank; Leidos Biomedical—Project Management; ELSI Study; Genome Browser Data Integration &Visualization—EBI; Genome Browser Data Integration &Visualization—UCSC Genomics Institute, University of California Santa Cruz; Lead analysts; Laboratory Data Analysis &Coordinating Center (LDACC); NIH program management; Bioinformatics; collection; Pathology; eQTL manuscript working group; Genetic effects on gene expression across human tissues. Nature 550, 204–213 (2017).
4. M. T. Maurano et al., Identification of cellular context sensitive regulatory variation in mouse genomes. bioRxiv 2020.06.27.175422 (28 June 2020).
5. O. Smithies, R. G. Gregg, S. S. Bogg, M. A. Koralevich, R. S. Kucherlapati, Insertion of DNA sequences into the human chromosomal beta-globin locus by homologous recombination. Nature 317, 230–234 (1985).
6. K. R. Thomas, K. R. Folger, M. R. Capcchi, High frequency targeting of genes to specific sites in the mammalian genome. Cell 44, 419–428 (1986).
7. F. D. Urnov, B. C. Genome Editing, Genome editing B.C. (Before CRISPR): Lasting基因表达 across human tissues. Nature 465, 1049–1054 (2009).
8. J. Vierstra et al, Functional dissection of the Sox9-Kcnj2 locus identifies nonessential gene expression in embryonic stem cells. Blood 90, 3332–3342 (1997).
9. M. Iacovino et al., Inducible cassette exchange: A rapid and efficient system enabling conditional gene expression in embryonic stem and primary cells. Stem Cells 29, 1580–1588 (2011).
10. F. Zhu et al., DICE, an efficient system for iterative genomic editing in human pluripotent stem cells. Nucleic Acids Res. 42, e34 (2014).
11. K. A. Matreyek, J. J. Stepney, D. M. Fowler, A platform for functional assessment of large variant libraries in mammalian cells. Nucleic Acids Res. 45, e102 (2017).
12. H. A. Wallace et al., Manipulating the mouse genome to engineer precise functional syntenic replacements with human sequence. Cell 128, 197–209 (2007).
13. L. A. Mitchell et al., De novo assembly, delivery and expression of a 101 kb human gene in mouse cells. bioRxiv:10.1101/423426 (26 January 2019).
14. L. A. Pennacchio et al., In vivo enhancer analysis of human conserved non-coding sequences. Nature 444, 499–504 (2006).
15. M. E. Dickinson et al, International Mouse Phenotyping Consortium; Jackson Laboratory; Infrastructure Nationalo PHENOMIN, Institut Clinique de la Souris (ICS); Central River Laboratories; MRC Harwell; Toronto Centre for Phenogenomics; Wellcome Trust Sanger Institute; RIKEN BioResource Center, High-throughput discovery of novel developmental phenotypes. Nature 537, 509–514 (2016).
16. M. H. St Clair, C. U. Lambe, P. A. Furman, Inhibition by ganciclovir of cell growth and DNA synthesis of cells biochemically transformed with herpesvirus genetic information. Antimicrob. Agents Chemother. 31, 844–849 (1987).
17. R. H. Friedel, W. Wurst, B. Wefers, R. Kühn, Generating conditional knockout mice. Methods Mol. Biol. 693, 205–231 (2013).
18. M. D. Ryan, A. M. King, G. P. Thomas, Cleavage of foot-and-mouth disease virus polyprotein is mediated by residues located within a 19 amino acid sequence. J. Gen. Virol. 72, 2727–2732 (1991).
19. C. T. Caskey, G. D. Kruh, The HPRT locus. Cell 16, 1–9 (1979).
20. T. A. Timp, D. L. Cole, A. B. Rogers, A. A. Horsfall, A. C. Brick, T. C. Slattery, W. L. Perfecto, C. J. M. Scott, S. J. Goldstone, L. F. Martin, D. J. Gifford, S. L. J. M. Parry, The immune system response to the smallpox virus infected host: Determinants of viral pathogenesis. Virology 24, 369–381 (1966).
21. J. Seibler, D. Schöberle, S. Fiering, M. Grouincke, J. Bode, DNA cassette exchange in ES cells mediated by Flp recombinase: An efficient system for repeated modification of tagged loci by marker-free constructs. Biochemistry 37, 6229–6234 (1998).
22. E. E. Bouahissia, K. Westerman, P. Leboulch, Transcriptional behavior of LCR enhancer elements integrated at the human U6 pro-motor (pCas9); and 3) pEF1 (payload deliveries). Satellite repeats were similarly masked. Reported reads spanning references were required to include a minimum of 20-bp mapping outside additional filtered regions: 1) HAS (LP integrations), 2) lox sites and genome sequence corresponding to the payload (payload deliveries), and 3) the intervening deleted region (HA analyses for payload deliveries).
23. For all analyses, reads with the same strand and mapping to within 500 bp of each other were clustered for reporting. Regions below 75 bp or with fewer than 1 read/10M reads sequenced were excluded. A distance 1 kb and greater was required between regions mapping to the same chromosome.

Data Availability. The sequencing processing pipeline and genome browser visualization hub source code are available at https://github.com/mauranolab/mapping. Sequencing data have been deposited in the Gene Expression Omnibus (GEO) database, https://www.ncbi.nlm.nih.gov/geo (accession no. GSE194988).

ACKNOWLEDGMENTS. We thank Dr. David J. Araten, Dr. Dongui Li, Dr. Paolo Mita, Brendan Carmelotto, and Julie Tolle for providing reagents and advice. This work was partially funded by NIH Grants RM1HG009491 (to J.D.B.) and R35GM119703 (to M.T.M.), and by the Colton Center for Autoimmunity.
39. D. J. Araten, K. Nafa, K. Pakdeesuwan, L. Luzzatto, Clonal populations of hematopoietic cells with paroxysmal nocturnal hemoglobinuria genotype and phenotype are present in normal individuals. Proc. Natl. Acad. Sci. U.S.A. 96, 5209–5214 (1999).

40. D. Li et al., Application of counter-selectable marker PIGA in engineering designer deletion cell lines and characterization of CRISPR deletion efficiency. Nucleic Acids Res. 10.1093/nar/gkab035 (2021).

41. M. A. Eckersley-Maslin et al., Random monoallelic gene expression increases upon embryonic stem cell differentiation. Dev. Cell 28, 351–365 (2014).

42. T. M. Keane et al., Mouse genomic variation and its effect on phenotypes and gene regulation. Nature 477, 289–294 (2011).

43. A. A. Avilion et al., Multipotent cell lineages in early mouse development depend on SOX2 function. Genes Dev. 17, 126–140 (2003).

44. H. Y. Zhou et al., A Sox2 distal enhancer cluster regulates embryonic stem cell differentiation potential. Genes Dev. 28, 2699–2711 (2014).

45. Y. Li et al., CRISPR reveals a distal super-enhancer required for Sox2 expression in mouse embryonic stem cells. PLoS One 9, e114485 (2014).

46. D. S. Bindels et al., mScarlet: A bright monomeric red fluorescent protein for cellular imaging. Nat. Methods 14, 53–56 (2017).

47. J. M. Laurent et al., Big DNA as a tool to dissect an age-related macular degeneration-associated haplotype. Precis. Clin. Med. 2, 1–7 (2019).

48. K. Kawagoe et al., Glycosylphosphatidylinositol-anchor-deficient mice: Implications for clonal dominance of mutant cells in paroxysmal nocturnal hemoglobinuria. Blood 87, 3600–3606 (1996).

49. Y. Han et al., Orientation-dependent regulation of integrated HIV-1 expression by host gene transcriptional readthrough. Cell Host Microbe 4, 134–146 (2008).

50. D. M. Fowler, S. Fields, Deep mutational scanning: A new style of protein science. Nat. Methods 11, 801–807 (2014).

51. E. Z. Kvon et al., Comprehensive in vivo interrogation reveals phenotypic impact of human enhancer variants. Cell 180, 1262–1271.e15 (2020).

52. B. B. Maricique, H. G. Chaudhari, B. A. Cohen, A massively parallel reporter assay dissects the influence of chromatin structure on cis-regulatory activity. Nat. Biotechnol. 37, 90–95 (2018).

53. A. R. Perez et al., GuideScan software for improved single and paired CRISPR guide RNA design. Nat. Biotechnol. 35, 347–349 (2017).

54. E. Yigit et al., High-resolution nucleosome mapping of targeted regions using BAC-based enrichment. Nucleic Acids Res. 41, e87 (2013).

55. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120 (2014).

56. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754–1760 (2009).

57. G. G. Faust, I. M. Hall, SAMBLASTER: Fast duplicate marking and structural variant read extraction. Bioinformatics 30, 2503–2509 (2014).

58. S. Nephe et al., BEDOPS: High-performance genomic feature operations. Bioinformatics 28, 1919–1920 (2012).

59. R. Feil, J. Wagner, D. Metzger, P. Chambon, Regulation of Cre recombinase activity by mutated estrogen receptor ligand-binding domains. Biochem. Biophys. Res. Commun. 237, 752–757 (1997).