A multi-kingdom genetic barcoding system for precise target clone isolation

Soh Ishiguro[1,15,19], Kana Ishida[2,19], Rina C. Sakata[1,19], Hideto Mori[3-5], Mamoru Takana[3], Samuel King[1], Omar Bashth[1], Minori Ichiraku[6], Nanami Masuyama[1,4,5], Ren Takimoto[1], Yusuke Kijima[1,7], Arman Adel[1], Hiromi Toyoshima[3], Motoaki Seki[3], Ju Hee Oh[8], Anne-Sophie Archambault[8], Keiji Nishida[9,10], Akihiko Kondo[9-11], Satoru Kuhara[12], Hiroyuki Aburatani[13], Ramon I. Klein Geltink[8], Yasuhiro Takashima[6], Nika Shakiba[1], and Nozomu Yachie[1,3,14,16-18]*

1. School of Biomedical Engineering, Faculty of Applied Science and Faculty of Medicine, The University of British Columbia, Vancouver, Canada.
2. Spiber Inc., Tsuruoka, Japan.
3. Synthetic Biology Division, Research Center for Advanced Science and Technology, The University of Tokyo, Tokyo, Japan.
4. Institute for Advanced Biosciences, Keio University, Tsuruoka, Japan.
5. Systems Biology Program, Graduate School of Media and Governance, Keio University, Fujisawa, Japan.
6. Center for iPS Cell Research and Application, Kyoto University, Kyoto, Japan.
7. Department of Aquatic Bioscience, Graduate School of Agricultural and Life Sciences, The University of Tokyo, Tokyo, Japan.
8. BC Children's Hospital Research Institute, Department of Pathology and Laboratory Medicine, The University of British Columbia, Vancouver, Canada.
9. Engineering Biology Research Center, Kobe University, Kobe, Japan.
10. Graduate School of Science, Technology and Innovation, Kobe University, Kobe, Japan.
11. Department of Chemical Science and Engineering, Graduate School of Engineering, Kobe University, Kobe, Japan.
12. Graduate School of Bioresource and Bioenvironmental Sciences, Faculty of Agriculture, Kyushu University, Fukuoka, Japan.
13. Genome Science Division, Research Center for Advanced Science and Technology, The University of Tokyo, Tokyo, Japan.
14. Premium Research Institute for Human Metaverse Medicine (WPI-PRIMe), Osaka University, Suita, Osaka, Japan
15. Twitter: @soh__i
16. Twitter: @nzmyachie
17. Twitter: @yachielab
18. Mastodon: sciencemastodon.com/@nzm
19. These authors contributed equally.

*Corresponding author: nozomu.yachie@ubc.ca.

Teaser: A multi-kingdom CRISPR-activatable barcoding system enables the precise isolation of target barcode-labeled clones from a complex cell population.

Clonal heterogeneity underlies diverse biological processes, including cancer progression, cell differentiation, and microbial evolution. Cell tagging strategies with DNA barcodes have recently enabled analysis of clone size dynamics and clone-restricted transcriptomic landscapes of heterogeneous populations. However, isolating a target clone that displays a specific phenotype from a complex population remains challenging. Here, we present a new multi-kingdom genetic barcoding system, CloneSelect, in which a target cell clone can be triggered to express a reporter gene for isolation through barcode-specific CRISPR base editing. In CloneSelect, cells are first barcoded and propagated so their subpopulation can be subjected to a given experiment. A clone that shows a phenotype or genotype of interest at a given time can then be isolated from the initial or subsequent cell pools stored throughout the experimental timecourse. This novel CRISPR-barcode genetics platform provides many new ways of analyzing and manipulating mammalian, yeast, and bacterial systems.

Cells are not homogeneous in any system. Although they proliferate and replicate the genome that encodes molecular regulatory programs into their progenies, they also change their statuses according to the dynamic response of gene expression patterns to environmental signals. As typically shown in multicellular organisms, cells self-organize through mutual molecular and mechanical communications and dynamically create complex structures. In these processes, spontaneous mutations in the genome sometimes impair the cellular program and cause cellular malfunction. Conversely, some other mutations confer growth or survival advantages to the cells, which can be beneficial or catastrophic to the system.

During cancer chemotherapy, resistant clones arise and expand with unique genetic, epigenetic, and cellular statuses, contributing to cancer recurrence and metastasis (*1-4*). In hematopoiesis, stem cells of an analytically indistinguishable group show fate-restricted differentiation patterns, in which some cells seem to be primed for specific lineages by incompletely understood factors (*5-8*). Similarly, *in vitro* stem cell differentiation and direct reprogramming experiments have demonstrated that “elite” clones reproducibly transform into target cell states (*9-11*). In laboratory microbial evolution experiments, different cells within the initial population dynamically expand and shrink their clone sizes through the acquisition of new mutations over multiple generations (*12-16*).

These views on clonal heterogeneity and cell lineage bias have been rapidly shaped by approaches that use DNA barcodes and deep sequencing for high-resolution clonal cell tagging. DNA barcoding involves integrating short and unique variable fragments of DNA—so-called “barcodes”—to individual cells by lentiviral transduction, transposon-mediated delivery, site-specific DNA recombination in the host genome, or plasmid transformation. In a given assay, the change in abundance of the barcoded clones is traced *en masse* by subsampling the cell populations and quantifying the DNA barcodes by deep sequencing read counts. Furthermore, the combination of barcode transcription and single-cell RNA sequencing (scRNA-seq) allows clonal lineages to be analyzed alongside cell states, revealing cell lineage-restricted state trajectories (*9, 17-23*). However, these measurements only provide snapshots of clone population sizes and gene expression statuses at the time of observation (*24*), limiting the analysis of the early causal molecular and/or environmental factors that derive specific fate outcomes of clones.

Did the chemotherapy-resistant clones exist in the initial cell population with the genetic mutations or altered cell state from the beginning? Did any molecular factor(s) underlie the observed stem cell differentiation fates? Was the progression of the specific clone conditional on the existence of any other clones? Flow cytometry cell sorting with immunostaining and emerging image cytometry cell sorting technologies enable the dissociation of heterogeneous cell populations into single cells with their observed phenotypes (*25-28*) but cannot do the same for a population of clones before they exhibit a phenotype of interest. Furthermore, any current high-content omic measurement requires the destruction of samples at the time of observation, which precludes the analysis of dynamic molecular and cellular behaviors within the same biological systems. The new concept of “retrospective clone isolation” has recently emerged to tackle the above-raised questions (*29-33*). In such a system, a barcoded cell population is first propagated, and its subpopulation is subjected to a given assay (Fig. 1A). After identifying a barcoded clone of interest, the same clone (or its close relative) is isolated in a

barcode-specific manner from the initial or any other subpopulation stored during the experiment. The isolated live clone can then be subjected to any following experiments, including omics measurements and reconstitution of a synthetic cell population with the isolates.

The idea of retrospective clone isolation has been implemented using CRISPR–Cas9 genome editing. In one form of the CRISPR activation (CRISPRa)-based approach (*29*), cells are tagged with DNA barcodes that are located upstream of a fluorescent reporter gene with a minimal promoter sequence. The barcoded cells are designed not to express the reporter gene before the guide RNA (gRNA)-induced transcription activation is triggered. Once a barcoded clone is identified for isolation, a gRNA targeting its barcode is introduced to the cell population with catalytically dead Cas9 (dCas9) fused to transcriptional activator(s) to trigger reporter expression of the target clone. This approach suffers from low specificity due to leaky expression of the reporter in the non-activated condition, presumably because of the effect of another constitutively active promoter neighboring the reporter. This effect can be mitigated by barcoding cells with a gRNA library pool and then labeling a target cell by delivering a plasmid that only encodes a reporter targeted by its corresponding gRNA, with no additional promoters present (*30, 31*). However, this strategy prohibits selection of the cells harboring the reporter plasmid using a secondary selectable marker. The CRISPRa-based approach requires continuous activation and maintenance of all components when a downstream experiment necessitates reporter expression in isolated cells. Furthermore, CRISPRa-based retrospective clone isolation has only been demonstrated in mammalian cell systems.

Genetic circuits based on DNA code alteration generally show highly specific input responses (*34*). A wild-type Cas9-based retrospective clone isolation has also been prototyped as a DNA code alteration-based approach (*32*). In this system, an out-of-frame start codon is followed by in-frame stop codons and a DNA barcode is encoded upstream of a reporter gene, whose translation is inactivated. The reporter translation can be restored by Cas9-induced double-stranded DNA break (DSB) and deletion through non-homologous end-joining (NHEJ) DNA repair at the barcode region, which stochastically removes the stop codons and brings the start codon into the coding frame. While this does not usually show unexpected reporter activation for non-target clones and confers permanent labeling of a target clone, the system's sensitivity relies on stochastic DNA deletion, which can be a bottleneck to efficiency. Cas9-induced DSB is also cytotoxic and potentially impairs the target clone during the reporter activation procedure (*35, 36*).

Another approach has been proposed using RNA fluorescence *in situ* hybridization (FISH), where barcoded cells are first fixed, and a target transcribing barcoded clone can be labeled with a fluorophore probe and isolated by cell sorting (*37*). While RNA FISH is specific and sensitive, the isolated cells cannot be used for cell culture-based assays due to the fixation of the analyte.

Here, we report a new CRISPR base editing-based approach, CloneSelect, that overcomes the current technical constraints. CloneSelect is based on the restoration of reporter protein translation by base editing of an impaired start codon or an upstream stop codon in a barcode-specific manner. The new method is highly scalable, programmable, and compatible with scRNA-seq. Its specificity markedly surpasses other CRISPR-based systems. We also present the versatility of the method by implementing CloneSelect in human embryonic kidney (HEK) cells, mouse embryonic stem cells (mESCs), human pluripotent stem cells (hPSCs), yeast cells, and bacterial cells.

Results

Mammalian CloneSelect by C→T base editing

CRISPR base editing has widely been used to induce a single nucleotide substitution at a target genomic site without DSB (*38*). We first hypothesized that a C→T base editor (CBE)-based circuit (*39*) would enable highly sensitive, precise barcode-specific clone isolation with better performance than the previous CRISPR-based methods. In this CloneSelect C→T system, a barcode is encoded immediately upstream of a reporter gene whose start codon is mutated to GTG (Fig. 1B left). A constitutively active promoter transcribes the reporter-encoding region, but the impaired start codon makes its translation inactive. C→T base editing at the target barcode site enables the antisense strand of the first guanine (cytosine) to be substituted into thymine, restores ATG on the sense strand, and activates the reporter

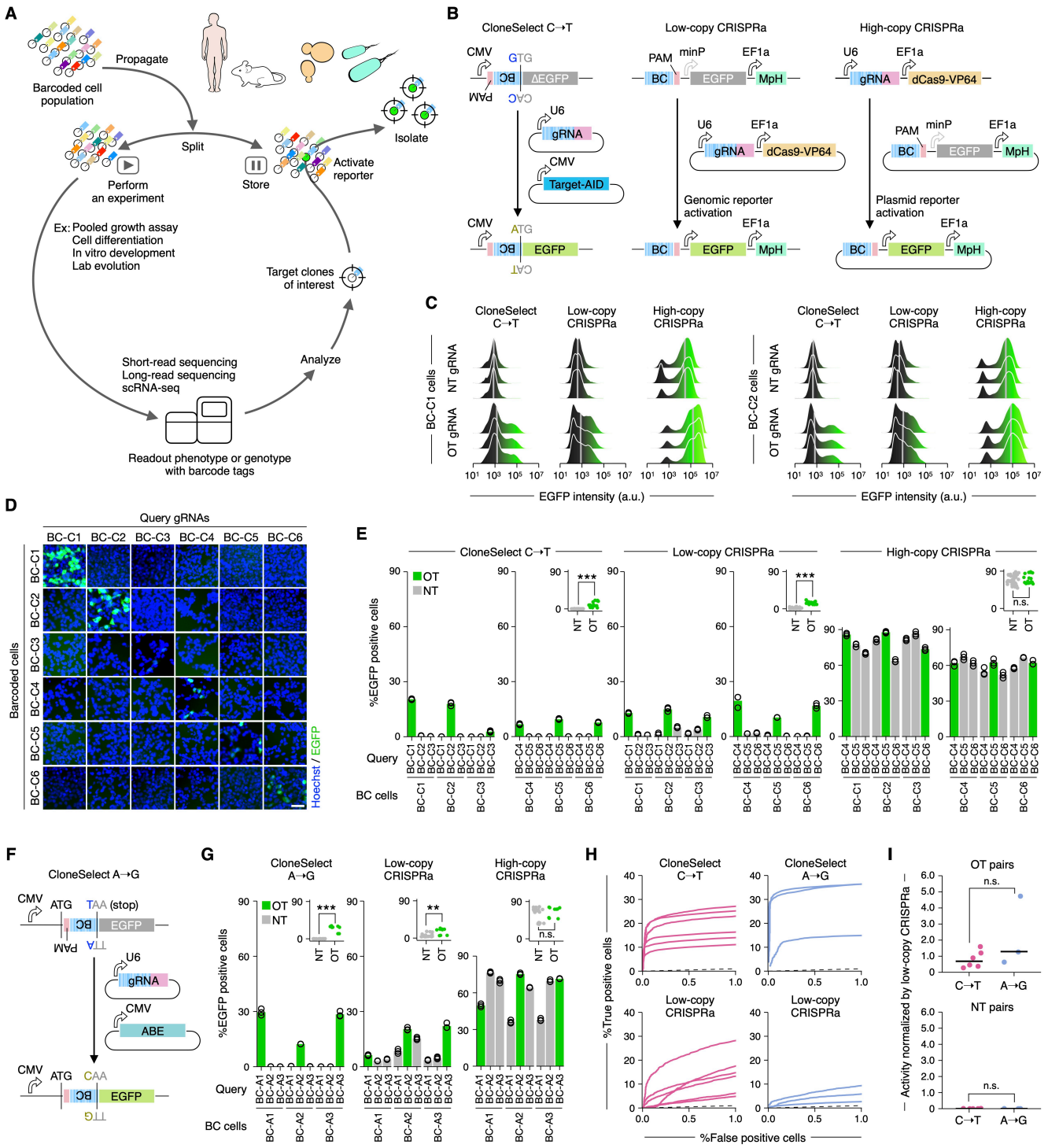

Fig. 1. The CloneSelect circuit. (**A**) Conceptual diagram of retrospective clone isolation. (**B**) Different barcode-specific gRNA-dependent reporter activation circuits. CloneSelect C→T, low-copy CRISPRa, and high-copy CRISPRa. (**C**) Flow cytometry analysis of single-cell EGFP activation levels. (**D**) Barcode-dependent reporter activation of six barcoded cell lines by CloneSelect C→T. Scale bar, 50 μm. (**E**) Comparison of CloneSelect C→T, low-copy CRISPRa, and high-copy CRISPRa across the same barcode-gRNA pairs (n=3). For each approach, Welch's t-test was performed to compare on-target (OT) and non-target (NT) activations. (**F**) CloneSelect A→G. (**G**) Comparison of CloneSelect A→G, low-copy CRISPRa, and high-copy CRISPRa across the same barcode-gRNA pairs (n=3). Welch's t-test was performed to compare OT and NT activations. (**H**) ROC curves along varying EGFP intensity thresholds for target barcoded cells. Left, CloneSelect C→T and low-copy CRISPRa by the same targeting gRNAs. Right, CloneSelect A→G and low-copy CRISPRa for the same set of targeting gRNAs. (**I**)

Performance comparison of CloneSelect C→T and CloneSelect A→G. Activated cell frequencies of OT and NT barcodes were normalized by activated cell frequencies of OT barcodes conferred by low-copy CRISPRa using the same targeting gRNA. The Mann-Whitney U test was performed to compare the two groups of datasets. *P < 0.05; **P < 0.01; ***P < 0.001.

translation. To achieve this, we employed Target-AID, one of the CBEs with a narrow C→T base editing window in the gRNA target site (*36, 38*).

To examine the performance of CloneSelect C→T employing EGFP (enhanced green fluorescent protein) as a reporter, we compared it with two CRISPRa-based approaches. In the low-copy CRISPRa approach, cells are lentivirally labeled by barcoded CRISPRa reporters, each consisting of a barcode followed by a minimal promoter and EGFP with an MCP-p65-HSF1 (MpH) transcriptional activator expression unit encoded on the same vector (Fig. 1B middle). The reporter can be activated by introducing catalytically dead Cas9 (dCas9) fused to VP64 and a targeting gRNA fused to MS2 hairpins, interacting with the MCP (MS2 coat protein) domain of MpH to recruit transcriptional activators (*40*). In the high-copy CRISPRa approach, cells are labeled by gRNAs, in which gRNA spacers serve as barcodes (Fig. 1B right). Upon transfection of the high-copy CRISPRa reporter-encoding plasmids, the reporter expression can be induced in the cell encoding the targeting gRNA.

CloneSelect C→T was more sensitive than low-copy CRISPRa and more specific than high-copy CRISPRa in HEK293T cells. For each of the three approaches, we prepared two cell lines with the same set of barcodes (BC-C1 and BC-C2). CloneSelect C→T with on-target gRNAs activated the reporter expression in 19.17% (BC-C1) and 17.89% (BC-C2) of target cells while minimizing the activation by non-targeting gRNAs to less than 0.1% (Fig. 1C and Fig. S1A). Low-copy CRISPRa showed a lower reporter activation rate of 5.24–6.26% (Fig. 1C and Fig. S1B), and high-copy CRISPRa showed a very high background activation rate of 52.12–62.15% by non-targeting gRNAs (Fig. 1C and Fig. S1C). The reporter intensity of activated cells for CloneSelect C→T was slightly but significantly higher (1.76-fold) than that of low-copy CRISPRa (Fig. S1D). The sensitivity of CloneSelect C→T using Target-AID constructed based on a nickase Cas9 (nCas9) was higher than that constructed based on dCas9 (Fig. S1E). CloneSelect C→T was also demonstrated to activate the reporter expression in HeLa cells in a barcode-specific manner (Fig. S1F).

To systematically examine the orthogonality between barcodes and gRNAs across the three approaches, we expanded the analysis to six barcodes (Fig. 1D and Fig. S2A and B) and demonstrated that CloneSelect C→T and low-copy CRISPRa both specifically activated the reporter expression of target barcoded cells, where 2.81–20.46% and 10.39–19.53% of target cells were EGFP positive, respectively, with minimal non-target activation (Fig. 1E). Consistent with the preceding result, CloneSelect C→T exhibited significantly higher overall reporter expression level than low-copy CRISPRa (Fig. S2C) and the least false positive rate (Fig. S2D). High-copy CRISPRa showed substantially lower orthogonality (Fig. 1E and Fig. S2B and D), indicating that this approach is not practical for retrospective clone isolation. For CloneSelect C→T, we also optimized the input DNA amount without elevating the minimal false positive rate (Fig. S2E and F).

We also tested a wild-type CRISPR-based system in which the reporter expression interrupted by an upstream stop codon is removed by the barcode-specific DSB, followed by its stochastic deletion with NHEJ (Fig. S3A). The reporter activation rate of this approach remained at 2.14–2.30% for target barcoded cells with a relatively high-false positive rate of 0.51–0.56 % for non-target barcoded cells (Fig. S3B and C). Additionally, as an alternative to the EGFP reporter, we attempted to develop an mCherry reporter for CloneSelect C→T (Fig. S4A). However, the mutation of the start codon did not inactivate the fluorescent expression of the wild-type, presumably because the second methionine served as a start codon (Fig. S4B and C). We found that a stringent barcode-specific gRNA-dependent reporter activation can be established with the M1V (GTG)+M9A mutant without losing the fluorescence intensity level (Fig. S4D).

Mammalian CloneSelect by A→G base editing

While exhibiting high efficiency and specificity in human cells, CloneSelect C→T is not applicable for clone isolation of bacterial species as they use GTG to initiate protein translation (*41*). We thus designed

another system, CloneSelect A→G, using an adenine base editor ABE-7.10 that induces A→G base substitution at the gRNA target sequence (*42*). In CloneSelect A→G, following a constitutively active promoter and a start codon, a barcode encoding a TAA stop codon prevents downstream reporter translation (Fig. 1F). The stop codon can be removed in a gRNA-dependent manner by mutating the antisense strand of the thymine (adenine) to guanine, converting the stop codon into CAA (Proline). We compared the performance of CloneSelect A→G, low-copy CRISPRa, and high-copy CRISPRa (Fig. 1G and Fig. S5A–F) for three arbitrarily designed barcodes in HEK293T cells. Similar to CloneSelect C→T, CloneSelect A→G activated the reporter expression for 12.27–31.47% of on-target cells, whereas the non-target activation was maintained below 0.5% with high orthogonality of barcodes and gRNAs. Low-copy CRISPRa also exhibited barcode-specific reporter activation, but despite a lower reporter expression level, its overall false positive rate was higher than CloneSelect A→G (Fig. S5H). High-copy CRISPRa again showed a high background activation level.

To quantitatively compare the performances of CloneSelect C→T, CloneSelect A→G, low-copy CRISPRa, and high-copy CRISPRa, we analyzed reporter activation frequencies of on-target barcoded cells when a stringent false positive rate of 0.5% was permitted. CloneSelect C→T and CloneSelect A→G exhibited 10.05–24.88% and 14.12–35.19%, respectively (Fig. 1H). Low-copy CRISPRa examined with the same barcode sets also used for CloneSelect C→T and CloneSelect A→G exhibited 2.71–22.37% and 1.92–6.60%, respectively (Fig. 1H). In contrast, high-copy CRISPRa did not show reporter-activated cells at this threshold. The efficacy of the gRNA to recruit the effector Cas9, in general, has been known to depend highly on its targeting sequence (*43, 44*). To compare CloneSelect C→T and CloneSelect A→G, their cell activation frequencies for different target barcodes were normalized by those of low-copy CRISPRa for the same barcodes. As a result, we did not observe a marked difference in normalized performance between CloneSelect C→T and CloneSelect A→G (Fig. 1I).

Isolation of barcoded human cells from a heterogeneous population

To examine if CloneSelect C→T can isolate target barcoded cells from a complex population, we next generated a barcoded lentiviral library by pooled ligation of barcoded EGFP fragments (Fig. S6A). The barcoded EGFP fragments were amplified by PCR using an upstream forward primer pool encoding semi-random barcodes and a common reverse primer (Fig. S6A and B). The barcode sequences were designed to be WSNS repeats (W=A or T; S=G or C) to avoid additional start codons from appearing. The ligation product was used to transform *Escherichia coli* (*E. coli*) cells, and transformant colonies on an agar plate were scraped to purify a lentiviral barcode library pool. We estimated that our protocol normally confers a barcode

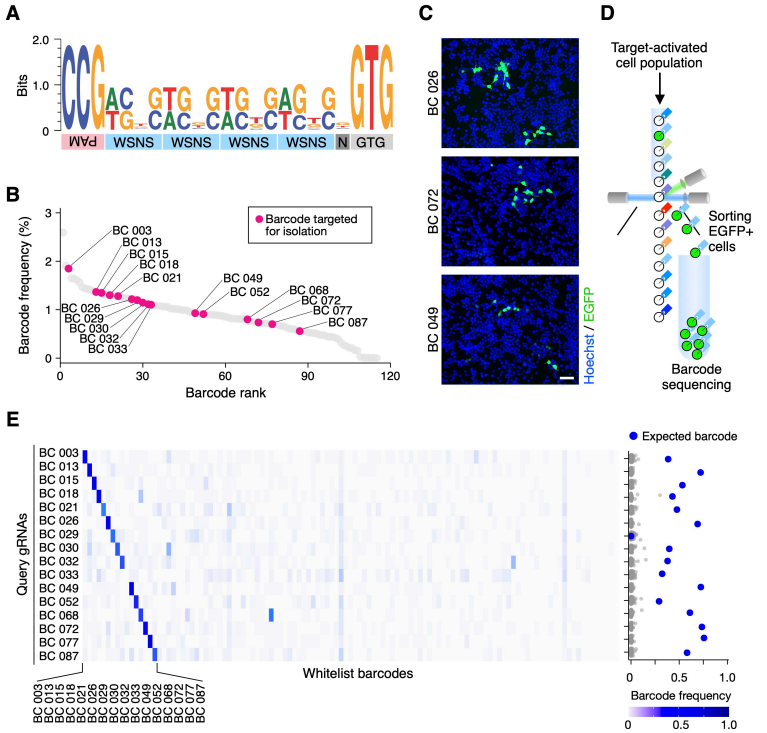

Fig. 2. Barcode-dependent cell isolation from a HEK293T cell population. (**A**) Nucleotide compositions of barcodes in the mammalian CloneSelect C→T plasmid mini-pool. Five barcodes that had unexpected lengths were excluded from this visualization. The full barcode sequence list can be found in Table S1. (**B**) Barcode abundances in the cell population labeled by the mini-lentiviral barcode pool of CloneSelect C→T. (**C**) gRNA-dependent labeling of target barcoded cells in a population. (**D**) Flow cytometry cell sorting of reporter-activated cells. (**E**) Barcode enrichment analysis after cell sorting of the reporter-activated cells. Each row represents the barcode enrichment profile for each target isolation assay.

complexity of multiple hundreds of thousands (Fig. S6C) and confirmed most of them to have barcodes by DNA restriction digestion (Fig. S6D).

To perform a proof-of-principle demonstration, we transformed *E. coli* cells again with the constructed library pool, isolated barcoded plasmid clones into a 96-well plate, and pooled 93 that were confirmed to have single barcodes by Sanger sequencing (Fig. S6E). The plasmid mini-pool was used to transduce HEK293T cells with a low multiplicity of infection (MOI) of <0.1, ensuring a single barcode was introduced to each cell. We amplified the barcode region from the plasmid mini-pool and the transduced cells by PCR and analyzed them by high-throughput sequencing. We identified 115 barcodes (Fig. 2A) and found that the variation in barcode abundance was replicable in two independent sequencing library preparations (Fig. S6F) and largely inherited from that in the plasmid pool (Fig. S6G and H), suggesting no substantial barcode-dependent bias in lentiviral packaging and transduction.

We then tested if we could enrich cells with 16 arbitrarily selected barcodes of different abundances in the cell population (Fig. 2B). For each target barcode, the cell population was co-transfected with the gRNA and Target-AID plasmids. After four days, we observed EGFP-positive cells in each assay (Fig. 2C) and sorted them by flow cytometry cell sorting (Fig. 2D). For each target barcode, its enrichment in the sorted cell population was analyzed by PCR and high-throughput sequencing (Fig. 2E). In sum, we succeeded in enriching the target barcoded cells in 15 out of 16 experiments (93.75%) with a relative barcode abundance of 29.18–75.75% after sorting. For the 15 successful targets, the mutated start codon was restored to ATG with an efficiency of 91.63–99.85% (Fig. S6I). A fraction of cells with the expected barcodes did not demonstrate the GTG→ATG mutation, suggesting that the cytidine deamination by Target-AID on the antisense strand might be sufficient to express the codon-repaired reporter transcripts.

Isolation of clones identified in a scRNA-seq platform

To extend the utility of CloneSelect to isolate living clones identified according to their high-dimensional transcriptome profiles from a cell population, we made CloneSelect C→T compatible with 3′ capture scRNA-seq platforms and established scCloneSelect. In scCloneSelect, the barcode located upstream of the reporter with the mutated start codon (hereafter referred to as “uptag”) is paired with another barcode (“dntag”) downstream of the reporter followed by a hard-coded 30-nt poly(A) sequence (Fig. 3A). The dntag is captured as part of the short-read scRNA-seq reads through the common poly(A)-tailed 3′ enrichment strategy (*45, 46*) and used to refer its corresponding uptag for the reporter start codon restoration. This change in the circuit design did not affect the reporter activation performance of CloneSelect C→T in HEK293T cells (Fig. 3B and C and Fig. S7A and B) as well as the high orthogonality between barcodes and gRNAs (Fig. S7C and D). We also confirmed that dntag barcodes were transcribed (Fig. S7E) and efficiently captured by a scRNA-seq platform (Fig. S7F).

One intriguing application of scCloneSelect would be to study the fate-determining factors of stem cell differentiation and cell reprogramming. scCloneSelect can be used to retrospectively isolate, from the initial population, cell clones whose states have been identified using scRNA-seq after differentiation. These clones could then be subjected to further analyses. We therefore tested if the scCloneSelect system is functional in selectively labeling target barcoded cells in mouse embryonic stem cell (mESC) and human pluripotent stem cell (hPSC) populations. In one of our approaches, a constitutively active Target-AID expression cassette was first stably integrated into a cell population by piggyBac transposon-based genomic integration, followed by lentiviral transduction of the cells with a scCloneSelect barcode library (Fig. S7G). The reporter activation was then triggered by introducing a targeting gRNA by lentiviral transduction or lipofectamine transfection. In mESCs, both of the gRNA reagent delivery methods activated the reporter expression in a barcode-dependent manner with no marked false positive activation, but lentiviral delivery of the gRNA showed an overall higher cell activation of 4.57–33.76% in comparison to 2.91–15.77% by the lipofectamine-based plasmid delivery method (Fig. 3D and E and Fig. S7H). The barcode-specific reporter activation was also successful in hPSCs, where targeting gRNA reagents were delivered by transfection (Fig. S7I). Additionally, we tested an approach that required a minimal number of steps, where hPSCs were first lentivirally barcoded, and the reporter was activated by electroporation with the targeting gRNA and Target-AID delivered together (Fig. S7J). This also enabled barcode-dependent reporter activation in hPSCs (Fig. S7K).

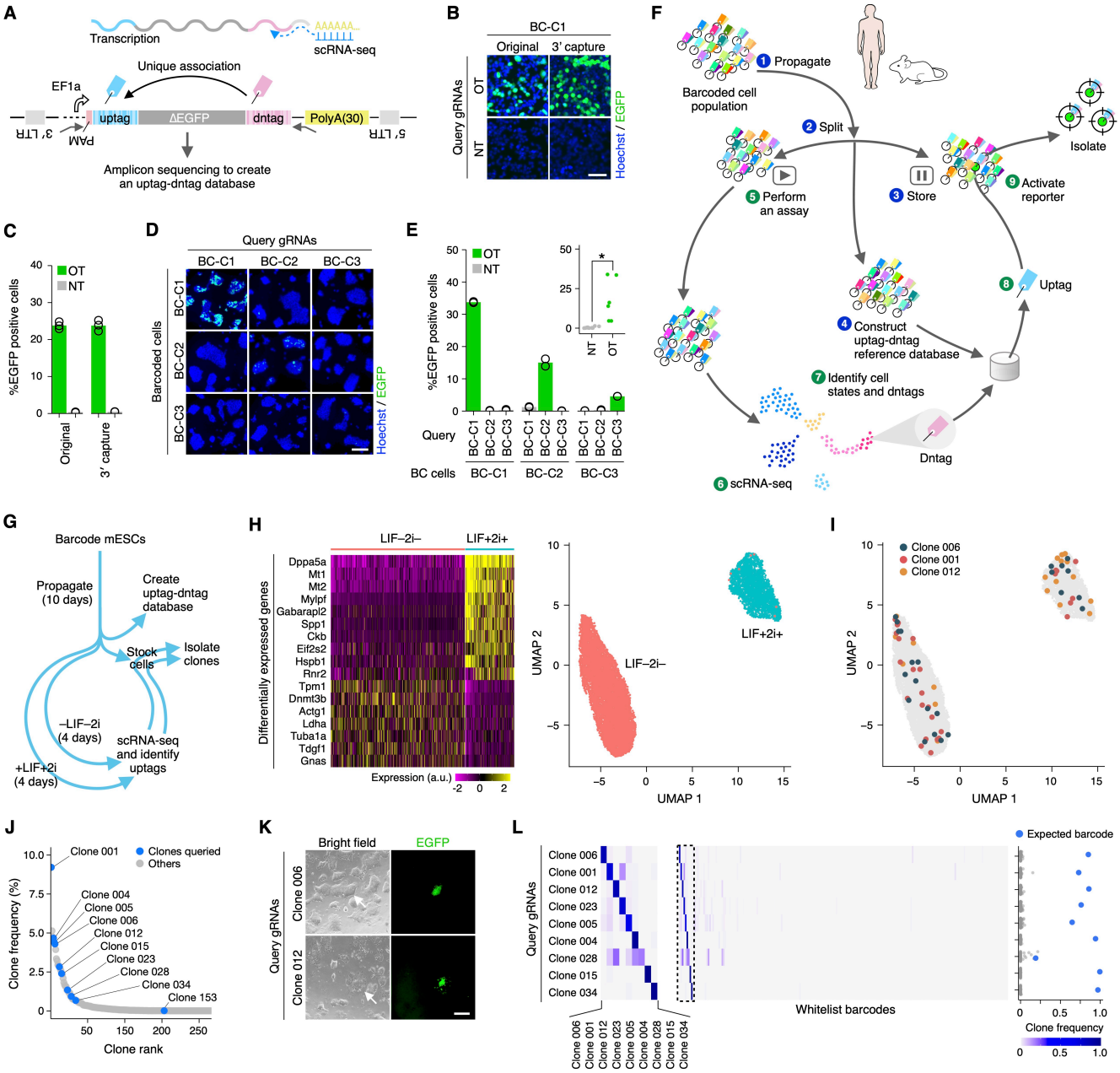

Fig. 3. Isolation of live stem cell clones characterized by scRNA-seq. (**A**) scCloneSelect. (**B** and **C**) Barcode-specific gRNA-dependent reporter activation of the original CloneSelect C→T and scCloneSelect in HEK293T cells (n=3). Scale bar, 50 μm. (**D** and **E**) Barcode-specific gRNA-dependent reporter activation of three barcoded mESC lines by scCloneSelect. Target-AID was stably integrated prior to the barcoding. gRNAs were delivered by lentiviral transduction. Scale bar, 100 μm. Welch's t-test was performed to compare on-target (OT) and non-target (NT) activations. *P < 0.05; **P < 0.01; ***P < 0.001. (**F**) Schematic diagram of a scCloneSelect workflow to retrospectively isolate a cell clone demonstrating a gene expression profile of interest from a cell population stored before they demonstrate the target gene expression pattern. (**G**) mESC cell culture assays and clone isolation performed in this work. (**H**) scRNA-seq of mESC populations treated with LIF and 2i and those without LIF or 2i. (**I**) Distribution of cells for arbitrarily selected clones in the two-dimensional embedding of high-dimensional gene expression space by UMAP (uniform manifold approximation and projection). (**J**) Abundance of barcoded cell clones in the mESC population. The data was generated based on dntags identified by reamplifying the dntag reads from the original scRNA-seq libraries. (**K**) gRNA-specific activation of target barcoded clones in the mESC population. Scale bar, 50 μm. (**L**) Barcode enrichment analysis after cell sorting of the reporter-activated cells. Each row represents the barcode enrichment profile for each target isolation assay. The left heatmap was expanded from the dashed box area of the right heatmap.

To demonstrate that barcoded cell clones identified with single-cell transcriptomes can be isolated from a barcoded cell pool sub-populated in parallel with the one used in scRNA-seq, we set up the following experiments using mESCs (Fig. 3F and Fig. S8A). In this experimental pipeline, a ΔEGFP (no start codon) fragment reporter is first amplified with forward primers encoding semi-random uptags of WSNS repeats accompanying a mutated start codon and reverse primers encoding random dntags. They are cloned into a common backbone plasmid *en masse* (Fig. S8B and C), and the constructed barcode plasmid pool is used to lentivirally barcode cells. Barcoded cells are then cultured to propagate cell clones (Step 1) and separated into three groups of sub-pools (Step 2). The first group is stored for later clone isolation (Step 3). The second group is used to identify the uptag-dntag combination reference database by PCR amplification and high-throughput sequencing (Step 4). The last group is subjected to a given assay, during which intermediate subpopulations can be stored at any point (Step 5). A cell population obtained during the assay is then analyzed by scRNA-seq to identify high-content gene expression profiles of single cells (Step 6). For a single cell demonstrating a cell state of interest, its dntags can be identified (Step 7), and the corresponding uptag can be retrieved from the uptag-dntag combination reference database (Step 8). Finally, using a gRNA targeting the identified uptag, the reporter expression of the clone containing the dntag is activated and isolated by cell sorting (Step 9). Because retroviral transduction, in general, is prone to having chimeric products of input vector sequences integrated into the genome through its high recombination activity (*47, 48*), the uptag-dntag database of the analyte population needs to be determined every time after the pooled transduction. We optimized the PCR protocol for the amplicon sequencing that identifies this database by minimizing chimeric PCR artifact products (Fig. S8D and E).

Using a plasmid pool with a barcode complexity of ~150,000, we lentivirally transduced mESCs with Target-AID at an MOI of <0.1. Following the creation of a clone variation bottleneck by sparse sampling and 10 days of expansion, a subpopulation of the barcoded cells was used to construct the uptag-dntag database, in which 216 unique barcode pairs were identified (Fig. 3G). After preserving another subpopulation, the remaining cells were cultured with serum with LIF and 2i (LIF+2i+) or serum without LIF or 2i (LIF–2i–), to maintain or to lose pluripotency, respectively. Four days after, scRNA-seq was performed independently for the two conditions. While the RNA capture rates per cell of the two datasets were similar (Fig. S9A), the gene expression profiles of single cells were clustered into two distinct groups along with the culture conditions (Fig. 3H and Fig. S9B). Pluripotency marker genes, such as *Dppa5a*, were highly expressed in the pluripotency maintenance condition, whereas differentiation marker genes, such as *Tpm1*, were enriched in the non-pluripotent condition. Although the barcoded clones did not show a significantly biased distribution between the two conditions, we attempted to isolate the top 10 abundant clones in the scRNA-seq datasets (Fig. 3I and Fig. S9C and D) from the initial barcoded population. The abundances of these clones varied from 0.0133% to 9.21% in the initial population according to the analysis determined by the uptag-dntag database (Fig. 3J). Except for one experiment targeting Clone 153, we obtained a sufficient number of EGFP-positive cells after introducing the targeting gRNA using lentivirus followed by flow cytometry cell sorting (Fig. 3K and Fig. S9E). For each of the remaining nine clone isolation attempts, eight showed target clone enrichment with relative target abundances of 64.8–99.4%, whereas one (Clone 028) showed an enrichment degree of 18.9% (Fig. 3L). Furthermore, we isolated single cells from the EGFP-positive cell samples obtained for Clone 006 and Clone 012, expanded them (Fig. S9F), and confirmed the purification of the isolated clones (Fig. S9G). Accordingly, scCloneSelect was demonstrated to greatly enrich a target clone identified by scRNA-seq from a complex population with a success rate of 80% (eight out of 10).

Yeast CloneSelect

Clonal barcoding approaches have also been used in microorganisms, such as yeast *Saccharomyces cerevisiae* (*S. cerevisiae*) and *E. coli*, to study their laboratory evolution and the genomic mutations accompanying clonal expansions of cells (*14, 16*). However, the analysis methods have been limited to time-course tracing of clone size dynamics. No technology has been developed to explore genetic, epigenetic, and other molecular factors contributing to laboratory evolution.

We next extended the CloneSelect C→T system for the yeast *S. cerevisiae* using mCherry as a fluorescent reporter (Fig. 4A). Like the mCherry reporter in mammalian cells, we also needed to truncate

the first nine amino acids of mCherry to establish the reporter system (Fig. S10A). CBEs, including Target-AID, developed for mammalian species normally accompany uracil glycosylase inhibitor (UGI) to inhibit the base excision repair pathway, enhancing both the efficacy and purity of C→T substitution at the target site (*38*). However, Target-AID has been tested in yeast without UGI and was demonstrated to confer C→D (non-C) substitution at the target sequence at a high rate (*36*). We, therefore, constructed

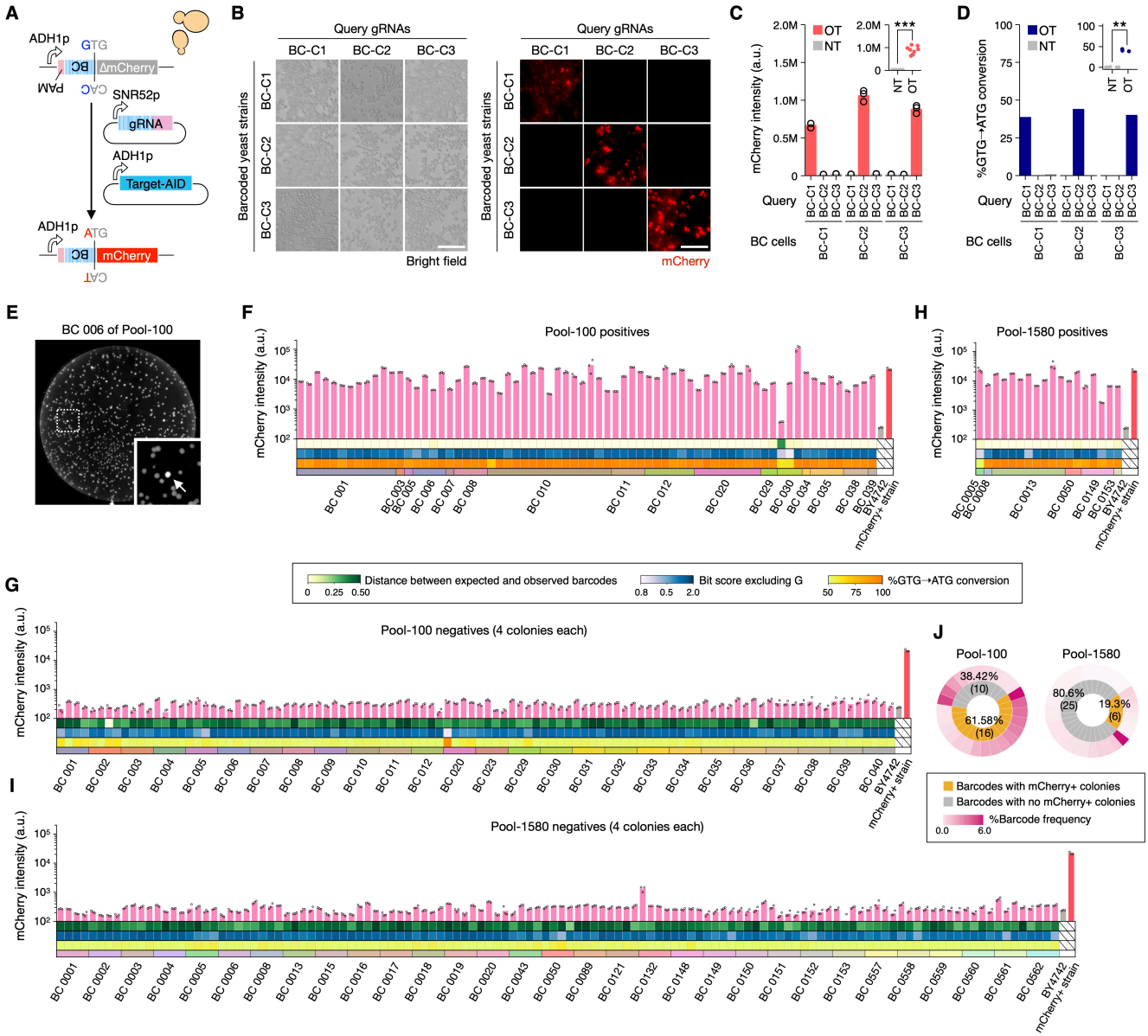

Fig. 4. Yeast CloneSelect. (**A**) Yeast CloneSelect C→T circuit. (**B** and **C**) Barcode-specific gRNA-dependent reporter activation. Scale bar, 25 μm. Mean mCherry intensity measured by a plate reader was normalized by $OD_{595\ nm}$ (n=3). Welch's t-test was performed to compare on-target (OT) and non-target (NT) activities. (**D**) GTG→ATG editing frequencies observed by high-throughput sequencing. Welch's t-test was performed to compare OT and NT datasets. (**E**) Yeast colonies formed on a 10-cm agar plate after performing a target clone labeling in the yeast cell population of Pool-100. (**F–J**) Analysis of colonies isolated after clone labeling using each targeting gRNA. (**F**) mCherry positive isolates from Pool-100. (**G**) mCherry negative isolates from Pool-100. (**H**) mCherry positive isolates from Pool-1580. (**I**) mCherry negative isolates from Pool-1580. (**J**) Summary of the analysis results. *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$.

411 a yeast Target-AID with UGI and found that it did not largely impair the base editing activity (Fig. S10B–
412 D) but as expected, greatly enhanced the frequency of resulting thymine at the target sequence (Fig.
413 S10E). The efficient yeast CloneSelect C→T reporter activation was only possible by the addition of
414 UGI (Fig. S10F). Similar to mammalian CloneSelect systems, yeast CloneSelect C→T was also
415 demonstrated to activate barcoded cells in a highly target-specific manner with a sensitivity of 38.80–
416 44.13% by co-transformation of Target-AID and gRNA plasmids (Fig. 4B–D and Fig. S10G). Unlike
417 mammalian cells, the labeled clones could be isolated by picking fluorescent colonies formed on a solid
418 agar plate (Fig. 4E).
419 To test the sensitivity of the clone isolation strategy, we generated a barcode plasmid pool by a
420 pooled ligation of semi-random barcode fragments to a backbone vector (Fig. S11A and B). Following
421 a quality control of the pooled cloning reaction (Fig. S11C), *E. coli* cells were then transformed with the
422 plasmid pool, and either 100 or ~1,580 colonies were pooled to prepare barcode plasmid pools of
423 defined complexities (hereafter referred to as Pool-100 and Pool-1580, respectively). After establishing
424 the barcoded yeast cell populations of Pool-100 and Pool-1580 by yeast plasmid transformation, we
425 arbitrarily targeted 26 and 31 barcodes of a range of abundances (0.83–5.79% for Pool-100 and 0.12–
426 0.37% for Pool-1580) to be isolated (Fig. S11D). For each barcoded clone isolation, its corresponding
427 gRNA plasmid and Target-AID plasmid was co-transformed into the barcoded yeast cells, and
428 fluorescent colonies were isolated, if any, together with four non-fluorescent colonies. The colony
429 isolates were then cultured in liquid selective media in a microwell plate to measure the fluorescence
430 intensities by a plate reader and barcode sequences were examined by PCR followed by Sanger
431 sequencing (Fig. 4F–I). For Pool-100, 16 out of the 26 attempts (61.58%) conferred positive colonies,
432 all of which, except for one of the three positive colonies obtained for BC 030, had the expected
433 barcodes with a GTG→ATG conversion rate of 48.92–97.41% (Fig. 4J). For Pool-1580, six out of the
434 31 attempts (19.35%) conferred positive colonies, and all of them had the expected barcodes with a
435 GTG→ATG conversion rate of 81.51–97.20%. The abundance of these successful barcodes in the Pool-
436 100 and Pool-1580 yeast pools ranged from 0.120% and 5.78%, respectively, demonstrating that yeast
437 CloneSelect can also isolate rare clones.
438
439 **Bacterial CloneSelect**
440 Lastly, we also established a bacterial CloneSelect system for *E. coli* clone isolation. We first
441 implemented CloneSelect A→G using an arabinose-inducible promoter for the EGFP reporter and an
442 IPTG (Isopropyl β-d-1-thiogalactopyranoside)-inducible promoter for both the gRNA and ABE (Fig. 5A).
443 The inducible promoter circuits showed an overall higher reporter expression in a barcode-specific
444 manner (Fig. 5B). At the same time, the EGFP expression level by a targeting gRNA was higher than
445 that by a non-targeting gRNA regardless of IPTG induction (Fig. 5C). The expected A→G substitution
446 was also observed in a target barcode-dependent manner even for the no IPTG condition (Fig. 5D).
447 These results collectively suggest that the background expression levels of the genome editing reagents
448 with no inducer were sufficient to confer the base editing. Furthermore, when the EGFP positive control
449 reporter was used, the arabinose-induced reporter intensity was suppressed by the targeting gRNA (Fig.
450 5E and Fig. S12A and B), suggesting a transcriptional silencing effect by ABE recruited at the target
451 site.
452 Because the fold change in EGFP reporter intensity for the targeting gRNA compared to the non-
453 targeting gRNA was not high (0.901–2.52 fold), we then sought to establish a drug-selectable system
454 for bacterial CloneSelect. We first realized that when a Zeocin resistance gene (*Sh ble*) (*49*) was used
455 for the barcode-specific reporter system (Fig. 5F), a constitutively active J23119 promoter conferred cell
456 growth under the selective condition regardless of the upstream TAA stop codon (Fig. S12C). In contrast,
457 the arabinose-inducible promoter was selective for the stop codon removal even without the arabinose
458 supplement, suggesting that a small gene expression level is sufficient for drug resistance. We also
459 found that targeting gRNA expression from the T7 promoter substantially dropped the number of colony-
460 forming units (Fig. S12D), presumably because nickase Cas9 has known to be toxic to the bacterial
461 cells (*50, 51*). Finally, we found that using the same setup as EGFP for the Zeocin resistance reporter
462 with no inducer condition demonstrated cell growth in a gRNA-dependent manner (Fig. S12E). A

Blasticidin S-resistance gene (*Bsr*) (*52*) also showed utility for gRNA-dependent cell growth with a similar experimental setup (Fig. S12F and G).

The Zeocin reporter system enabled the isolation of different barcoded strains with high specificity (Fig. 5G). To demonstrate barcoded cell isolation from a complex population, we constructed a pooled plasmid library for semi-random barcodes of VNN repeats (V=non-T), preventing the appearance of stop codons (Fig. S13A), and prepared sub-pool libraries by pooling 100 and ~1,550 colonies, respectively (Fig. S13B; hereafter referred as to Pool-100 and Pool-1550). After establishing barcoded *E. coli* cell

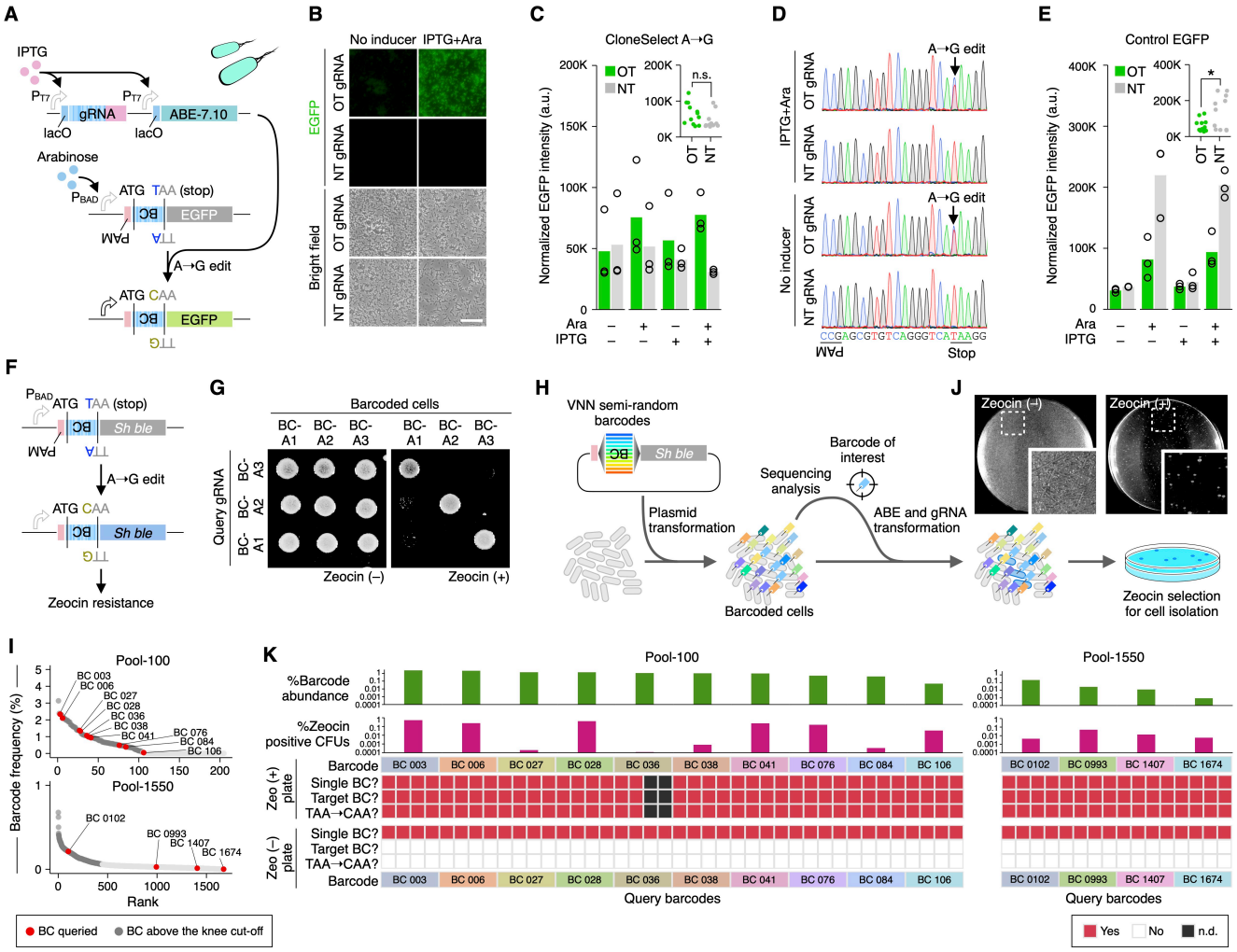

Fig. 5. Bacterial CloneSelect. (**A**) Bacterial CloneSelect A→G circuit. ABE and gRNA expressions were controlled by IPTG-inducible promoters, and the EGFP reporter expression was controlled by an arabinose-inducible promoter. (**B** and **C**) EGFP reporter activation of *E. coli* cells under different inducer conditions. Scale bar, 25 μm. Mean EGFP intensity measured by a plate reader was normalized by $OD_{595\ nm}$ (n=3). Welch's t-test was performed to compare on-target (OT) and non-target (NT) activities. (**D**) Base editing outcomes analyzed by Sanger sequencing. (**E**) Activities of the positive control EGFP reporter under the same conditions tested for (C) (n=3). Welch's t-test was performed to compare OT and NT activities. (**F**) Zeocin resistance marker-based circuit. (**G**) Barcode-specific gRNA-dependent Zeocin resistance reporter activation. (**H**) Schematic diagram of a bacterial CloneSelect workflow using a drug selective condition for the target barcoded cell isolation. (**I**) Abundance of barcoded cells in Pool-100 and Pool-1550. (**J**) Colonies formed on Zeocin-selective and non-selective solid agar plates after performing the reporter activation of Clone 106 in the *E. coli* cell population of Pool-100. (**K**) Analysis of colonies isolated from Zeocin selective and non-selective plates obtained after clone labeling using each targeting gRNA. *P < 0.05; **P < 0.01; ***P < 0.001.

populations of Pool-100 and Pool-1550, we arbitrarily selected target barcodes and attempted to select cells having each of those barcodes by co-transforming the targeting gRNA and ABE, followed by Zeocin selection (Fig. 5H). We targeted 10 and four barcodes of a range of abundances in Pool-100 (0.047–2.33%) and Pool-1550 (0.00089–0.211%), respectively (Fig. 5I). In every isolation experiment, the Zeocin selective conditions showed a substantially lower number of colonies than the non-selective conditions (Fig. 5J and K). For each of the successful target barcodes, except for BC 036 of Pool-100 in which we obtained only two colonies in the selective condition, four and four colonies were isolated from the selective and non-selective conditions, respectively, and their barcodes and base editing patterns were analyzed by Sanger sequencing. All isolates from the selective conditions had the expected barcodes and all of the isolates from the non-selective conditions had non-targeted barcodes. Accordingly, bacterial CloneSelect with the drug resistance marker system demonstrated great sensitivity and was able to isolate clones that were extremely low in frequency.

Discussion

We describe a method, CloneSelect, which enables the isolation of target barcoded cells from a complex population using CRISPR base editing. Compared to the CRISPRa and wild-type Cas9-based systems tested in this study, CloneSelect demonstrated an ability to isolate cells with a higher sensitivity when a certain degree of false positive isolation was permitted. The low-copy CRISPRa experiments confirmed the reporter expression leak in similar existing methods (*29*). High background activity observed in the high-copy CRISPRa experiments supports the utility of a strategy implemented in the approaches COLBERT (*30*) and ClonMapper (*31*), whereby cells are labeled with gRNAs as barcodes, and target barcoded clones are activated by provision of a plasmid that encodes only the target gRNA-corresponding reporter with no constitutively active promoters. However, this approach of barcoding cells using gRNAs does not allow selected cells to maintain reporter activation for subsequent experiments, when necessary. CloneSelect largely benefits from the precision of base editing and the simplicity of altering the genetic code, while a CRISPRa-based system involves more additional endogenous factors to be coordinated for the target barcode-dependent reporter activation. Furthermore, engineering evolutionarily prevalent rules of the genetic code enabled us to implement the same idea across diverse species. We demonstrated the retrospective isolation of barcoded cells from complex yeast populations for the first time. While the isolation of barcoded *E. coli* cells has recently been demonstrated using barcode-specific CRISPR interference of a counter selection marker (*53*), we have shown that bacterial CloneSelect achieves the isolation of target barcoded cells at unprecedented sensitivity and selectivity.

When practically examining its ability to isolate target barcoded cells from complex populations, CloneSelect was able to retrieve rare, barcoded cells whose estimated abundances were as small as 0.558%, 0.013%, 0.120%, and 0.00089% in the human HEK293T, mouse ESC, yeast, and *E. coli* populations, respectively. The theoretical complexity of semi-random barcodes is limited to 4.19×10^6 and 2.54×10^8 for CloneSelect C→T and CloneSelect A→G, respectively. The sensitivities to obtain target barcoded cells that align with the same possible minimal abundance thresholds above are 93.75%,90.00%, 61.56%, and 100.00%, respectively. The limited sensitivity per isolation attempt could be explained by the general gRNA-dependent genome editing efficacy (*43*) since isolation success did not correlate well with the abundance of the target in a population. We suggest that the current sensitivity range of CloneSelect is sufficient in most prospective assays, with the expectation that multiple independent clones show a phenotype of interest that allows the user to have multiple attempts. More enhanced genetic circuits can be considered to improve sensitivity and specificity. We tested an OR-gate with C→T base editing of tandemly arrayed barcodes (Fig. S14A) and an AND-gate with A→G base editing of tandemly arrayed barcodes (Fig. S14B). The OR-gate did not function well and enabled an efficient reporter activation for only one of the three barcodes. The AND-gate enabled the three-input-dependent reporter activation, but exhibited a tight trade-off with the sensitivity, as expected. Given the high specificity of the system, introducing multiple independent CloneSelect barcodes per cell might be a solution to increase the sensitivity if necessary.

With CloneSelect, new wide-ranging experiments can be conceived in broad fields of biology. A CloneSelect-barcoded cell population can be subjected to any assay. Existing time-course scRNA-seq measurement strategies enable interrogation of different clonal lineages in a barcoded population alongside the dynamic changes in their gene expression landscapes, if the clone population sizes are not too small (*9*). In contrast, CloneSelect would allow the clones isolated from different time points from the progressing population to be analyzed by diverse approaches (Fig. 6A). Such non-transcriptomic analyses could include morphological analyses under a microscope and molecular analyses available for small amounts of input cells, but any currently available methods should be applicable for the downstream analysis if the isolated clones can be propagated for the given hypothesis.

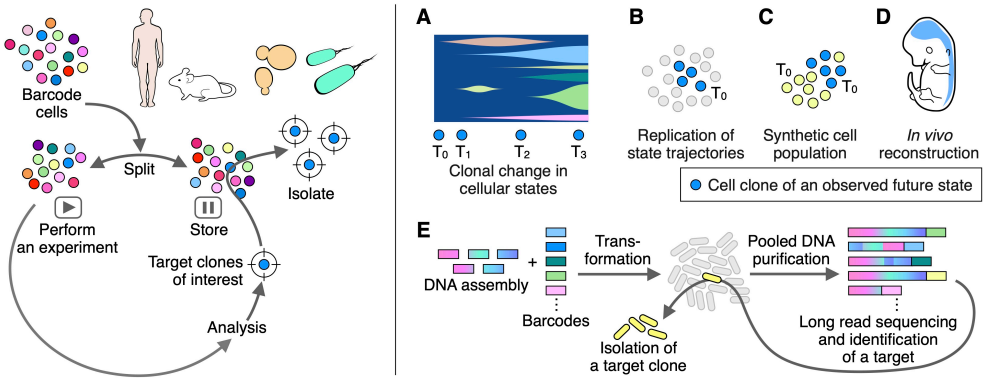

Fig. 6. New biology directions made possible by CloneSelect. (**A**) Clonal analysis of molecular profiles in a complex population. (**B**) Replication of cell state trajectories. (**C**) Reconstitution of synthetic cell populations. (**D**) Transplantation of a fate-mapped clone. (**E**) New DNA assembly strategy.

Cells isolated by CloneSelect are alive. The clones isolated from the initial stage of a once-performed assay can be tested to see if they take the same developmental trajectories (Fig. 6B) or used to reconstitute a synthetic population with another cell population or other isolated clones (Fig. 6C). For example, a variety of hPSC lines have been reported to be favorable for different cell differentiation and organoid models (*54-58*), suggesting that there could also be fate priming of stem cell clones due to undiscovered intrinsic factors. CloneSelect enables mapping of cell states after differentiation and their subsequent isolation from the initial population. The fate-mapped stem cell clones could be used to engineer new stem cell-based models or high-quality stem cell therapeutics. Cell clones isolated from diverse systems can also be transplanted into animal models (Fig. 6D). Examples include xenotransplantation of a cancer stem cell clone and aggregation of a fate-mapped stem cell clone with an early embryo. As opposed to the CRISPRa-based system, the barcode-dependent reporter activation is irreversible. When a fluorescent reporter gene is used, the spatial distribution of the targeted clone and their interaction with others in a complex biological system can be traced.

In alignment with having DNA sequencing as a readout, CloneSelect would also promote the genetic engineering of cells and DNA assembly (*59-63*). Any cell engineering cannot be perfect—only a fraction of cells that go through an engineering process will harbor the target genetic product. Therefore, obtaining successful cells becomes difficult when their fractions in the reaction pool are very small. In such a task, through the barcoding and sequencing of the product cells, CloneSelect could retrieve target cells with the desired genetic engineering outcome. We also envision CloneSelect to largely improve DNA assembly in general (Fig. 6E). Currently, it is common practice to transform a DNA assembly reaction sample into *E. coli* cells in order to clone each assembly product, followed by colony isolation and sequencing-based confirmation. A pool of CloneSelect barcodes can be used to molecularly tag DNA assembly products. In this framework, the assembly products are used to transform *E. coli*, followed by the pooling of transformants and extraction of the pooled plasmid products from its

subpopulation. The pooled plasmid product is then sequenced by long-read DNA sequencing to identify barcodes assigned to the desired product. Finally, the clone harboring the target product can be isolated using CloneSelect in a target barcode-specific manner. This strategy would enable the utilization of even inefficient DNA assembly protocols.

Accordingly, CloneSelect is a new way to precisely isolate target cell clones from complex populations. Its performance demonstrated in multi-kingdom species opens a vast array of possibilities to answer complex questions in diverse areas of biology.

Acknowledgments

We thank the members of the Yachie lab at the University of British Columbia and the University of Tokyo, especially Sanchit Chopra, Sean Okawa, Sofia Romero, Yin Liu, Yeganeh Dorri, Madina Kagieva, and Rina Yogo, for their constructive feedback. We also thank Kaori Shiina, Shiro Fukuda, and Tara Stach for their technical support in high-throughput sequencing and Andy Johnson for the flow cytometry cell sorting. This study was funded by the New Energy and Industrial Technology Development Organization (NEDO), the Japan Agency for Medical Research and Development (AMED), the Japan Science and Technology Agency (JST), the Takeda Science Foundation, the SECOM Science and Research Foundation, the Canada Foundation for Innovation (CFI), the Canadian Institutes of Health Research (CIHR) CoVaRR-Net Project (all to N.Y.), and the Allen Institute Allen Distinguished Investigator Award (to N.Y. and N.S.). Authors were also supported by the JST PRESTO program (N.Y.), the Canada Research Chair program (N.Y.), the Ryoichi Sasakawa Young Leaders Fellowship (S.I.), the Taikichiro Mori Memorial Research Fund (S.I. H.M., and N.M.), the Japan Society for the Promotion of Science (JSPS) DC Fellowship (S.I., H.M. Y.K., and N.M.), the JSPS Overseas Research Fellowships (S.I.), Japan Student Services Organisation (R.S.), the Funai Overseas Scholarship (R.S.), the Takenaka Scholarship Foundation (R.T.), the Ministry of Science and Education (MEXT) Research Scholarship (A.A.), the Natural Sciences and Engineering Research Council of Canada (NSERC) Undergraduate Student Research Award (S.King), Michael Smith Health Research BC Scholar Award (R.I.K.G.), and the Michael Smith Health Research BC Trainee Award (A.-S.A). High-throughput sequencing data analysis was performed using the SHIROKANE Supercomputer at the University of Tokyo Human Genome Center.

Author contributions

S.I. and N.Y. conceived the study. S.I., R.S., and M.T. performed all the major mammalian cell experiments. K.I. and M.T. performed all the yeast and bacterial experiments. O.B., N.S., M.I., and Y.T. performed a part of the experiments using mESCs and hPSCs. S.I., H.M., H.T., R.S., Y.K., and H.A. performed the high-throughput sequencing analyses. R.T., A.A., and N.M. performed preliminary assays and contributed to the design of the system. J.H.O., A.-S.A., and R.I.K.G. contributed to the optimization of the system. S.King supported the plasmid construction. K.N., A.K., and S.Kuhara also provided substantial ideas in designing the proposed system. M.S. supported the flow cytometry cell sorting. H.A. supported the high-throughput sequencing, the flow cytometry analysis, and the microscope imaging. S.I., R.S., K.I., and N.Y. wrote the manuscript.

Competing interests

K.I. is an employee of Spiber Inc. The other authors declare no competing interests.

627 **Materials and Methods**
628
629 **Plasmids**

630 Oligonucleotides were chemically synthesized by FASMAC, Integrated DNA Technologies, or Eurofins
631 Genomics. All the oligonucleotides and cloning procedures used to construct the plasmids in this study
632 can be found in Table S2. We used QUEEN version 1.2.0 (https://github.com/yachielab/QUEEN) to
633 describe each plasmid construction and generate the annotated plasmid files in the GenBank (gbk) file
634 format, embedding the full construction procedure (Table S2). A QUEEN-generated gbk file behaves as
635 a quine code, where the QUEEN simulation code that generated the gbk file can perfectly be
636 reconstructed from the file itself (*64*). We believe that providing the QUEEN-generated gbk files satisfies
637 the mandate of reporting the reproducible protocols for the newly constructed plasmids. We also
638 provided natural language descriptions for each plasmid construction step in the QUEEN code, allowing
639 the user to retrieve the natural language description of materials and methods for each plasmid by
640 simply executing "QUEEN --protocol_description --input [gbk file]" in a QUEEN-installed environment. A
641 custom QUEEN wrapper that generated all QUEEN-generated gbk files can be found at
642 https://github.com/yachielab/CloneSelect_v1/tree/main/QUEEN. We, therefore, do not provide the
643 plasmid construction protocols here. All plasmid DNA sequences were confirmed by Sanger sequencing.
644 The representative plasmids are in the process of being deposited to Addgene with their QUEEN-
645 generated gbk files, which Addgene agreed to accommodate.
646
647
648 **Mammalian CloneSelect**

649 **Cell culture**
650 HEK293T and HeLa cells. Human embryonic kidney 293Ta (HEK293Ta) and HEK293T Lenti-X cells
651 were purchased from GeneCopoeia (#LT008) and Takara (#632180), respectively. Cells were cultured
652 in Dulbecco's Modified Eagle Medium (DMEM) (Sigma-Aldrich #11965084) supplemented with 10%
653 Fetal bovine serum (FBS) (Gibco #16000044) and 1% Penicillin-Streptomycin (Wako #168-23191) at
654 37ºC with 5% CO2 in a cell culture incubator. Cells were detached and passaged using 0.25 w/v%
655 Trypsin-EDTA (Wako #203-20251) once the cells reached 70–90% confluency. The cell lines were
656 regularly tested for mycoplasma contamination. When microscope imaging of HEK293T cells was
657 performed with Hoechst 33342 (Invitrogen #H3570) counterstain, 100–200 µL of Collagen-I (Nippi
658 #PSC-1-100-100) diluted in 5 mM acetic acid was added to each cell culture plate well, incubating for
659 30 min at 37ºC. The collagen-coated plate wells were washed by 100–200 µL of 1x PBS before use.
660
661 mESCs. Under the approval by the Institutional Animal Care Committee of the University of Tokyo
662 (RAC180003), mESCs were established from embryos of a 129(+Ter)/SvJcl (female) x C57BL/6NJcl
663 (male) cross and maintained in Dulbecco's Modified Eagle Medium - low glucose (Sigma-Aldrich
664 #D6046-500ML) supplemented with 1% Penicillin-Streptomycin (Gibco #15140122), 1% MEM Non-
665 essential Amino Acids (Wako #139-15651), 1% GlutaMAX Supplement (Gibco #35050061), 1% Sodium
666 Pyruvate (Gibco #11360070), 15% FBS (Gibco #16000044), 100 µM/mL 2-Mercaptoethanol (Wako
667 #131-14572), 10^3 units/mL ESGRO Recombinant Mouse LIF Protein (Millipore #ESG1107), 3.0 µM
668 CHIR99021 (GSK-3 inhibitor) (Wako #038-23101), and 1.0 µM PD0325901 (MEK inhibitor) (Toris
669 #4423). A sufficient volume of 0.1% Gelatin (Sigma-Aldrich #G9391) in 1x Phosphate-buffered saline
670 (PBS) (Takara #T9181 or Gibco #70011044) was added to each well such that the liquid could cover
671 the entire surface and be aspirated after one hour at 37ºC before plating cells. Cells were cultured in a
672 cell culture incubator at 37ºC with 5% CO2. The cell culture medium was replaced at least once every
673 two days. The cell line was regularly tested for mycoplasma contamination.
674
675 hPSCs. CA1 human PSC (hPSC) line was used after approval by the Canadian Institutes of Health
676 Research Stem Cell Oversight Committee. CA1 hPSCs were cultured using mTeSR Plus (STEMCELL
677 Technologies #100-0276) cell culture medium in a humidified incubator at 37ºC with 21% O2 and 5%
678 CO2. Culture plates were coated with Geltrex LDEV-Free Reduced Growth Factor Basement Membrane

679 Matrix (Gibco #A1413201). Dulbecco's Modified Eagle Medium/Nutrient Mixture F-12 (DMEM/F12)
680 (Gibco #11320033) was used to make a working solution of Geltrex (Gibco #A1413201) with 1:100
681 dilution. A sufficient volume of Geltrex was added to each well such that the liquid could cover the entire
682 surface and be aspirated after one hour incubation at 37ºC before plating cells. The cell culture medium
683 was exchanged every other day after plating the cells. Regularly, cells were passaged as medium-sized
684 clumps. Upon aspiration of the media, ReLeSR (STEMCELL Technologies #05872) was added, and
685 cells were incubated at room temperature for about 1 min before the second aspiration. Cells were then
686 placed inside the incubator for 4-5 min, added fresh media, and dissociated by pipetting up and down.
687 Cells were then plated and placed inside the incubator. For single-cell passaging, we used TrypLE
688 Express (Gibco #12604021). Cells were then placed inside the incubator for four min before adding a
689 sufficient fresh media to stop the activity of TrypLE Express. Cells were then collected in centrifuge
690 tubes, dissociated by pipetting up and down, and passed through a 40 µm cell strainer (Sarstedt
691 #83.3945.040) to remove cell clumps. The tubes were then centrifuged at 300–400 g for 5 min and the
692 supernatant was aspirated. The pellets were resuspended in a fresh media supplemented with 10 µM
693 ROCK inhibitor Y-27632 (Tocris Bioscience #1254) for 24 hours to promote the survival of the single
694 cells. For culturing H1 hPSCs, we used StemFit AK02N cell culture medium (REPROCELLAHS
695 #RCAK02N) with Y-27632 (Cayman #10005583) added for one or two days after plating. Culture plates
696 were coated with the recombinant Laminin-511 E8 fragment using iMatrix-511 Silk (MAX #892021). The
697 cell lines were tested for mycoplasma contamination.
698
699 **Mammalian CloneSelect barcode libraries**
700 CloneSelect C→T barcode library. To generate the CloneSelect C→T barcode library, a semi-random
701 oligo pool SI#679 encoding 5′-CCGWSNSWSNSWSNSWSNSNGTG-3′ was first chemically
702 synthesized (Table S2), where the antisense strand sequence of the 5′-CGG-3′ PAM sequence and a
703 quadruple repeat of WSNS (W=A or T; S=G or C) were followed by a mutated start codon (GTG). The
704 WSNS repeat restricts additional start codons from appearing before the downstream reporter. An
705 EGFP coding sequence was then amplified from pLV-eGFP (Addgene #36083) in 25 separate PCR
706 reactions, each in 50 µL volume, composed of 1 ng/µL of pLV-eGFP template plasmid, 1.25 µL of 20
707 µM SI#679 oligo pool as a forward primer,1.25 µL of 20 µM SI#680 as a common reverse primer, 0.5
708 µL of Phusion High-fidelity DNA Polymerase (NEB #M0530), 10 µL of 5x Phusion HF Buffer (NEB
709 #B0518S), and 5 µL of 2.5 mM dNTPs (Takara #4025), with the following thermal cycle condition: 98ºC
710 for 30 s, 30 cycles of 98ºC for 10 s, 72ºC for 10 s, and 72ºC for 60 s, and then 72ºC for 5 min for the
711 final extension. The amplified fragment was digested by DpnI (NEB #R0176) for an hour at 37ºC, pooled
712 into a single tube (1,250 µL in total), and column purified using FastGene PCR/Gel Extraction Kit (Nippon
713 Genetics #FG-91302). The purified fragment was then subjected to EcoRI-HF (NEB #R3101S) and XbaI
714 (NEB #R0145S) digestion overnight at 37ºC and purified again by FastGene PCR/Gel Extraction Kit
715 (Nippon Genetics #FG-91302). To obtain a highly complex lentiviral plasmid pool, we performed a total
716 of five ligation reactions using PCR strip tubes, each for a 50-µL reaction containing ~30 fmol of EcoRI-
717 XbaI digested pLVSIN-CMV-Pur backbone plasmid (Takara #6183), ~300 fmol of the insert fragment,
718 2.5 µL of T4 DNA Ligase (NEB #M0202), and 5 µL of 10x T4 DNA Ligase Buffer (NEB #B0202). The
719 reaction samples were incubated at room temperature for two hours and purified by FastGene PCR/Gel
720 Extraction Kit (Nippon Genetics #FG-91302). The ligation sample was then used to transform NEB
721 Stable Competent *E. coli* cells (NEB #C3040I). The transformation was performed in five reactions, each
722 with 1,250 ng of the ligation sample transformed to 200 µL of the competent cells according to the
723 manufacturer's high-efficiency transformation protocol. Following one-hour outgrowth in SOC medium
724 (NEB #B9020) at 37ºC, cells were spun down and plated on a total of 25 LB agar plates containing 100
725 µg/mL Ampicillin (Wako #014-23302). After overnight incubation at 37ºC, colonies formed on the plates
726 were scraped by adding 1–2 mL ddH2O, pooled into a flask, and further incubated with 200–300 mL of
727 LB liquid medium containing 100 µg/mL Ampicillin (Wako #014-23302) overnight at 37ºC. We plated the
728 transformation sample on agar plates with 500-fold dilution in triplicate and estimated the barcode
729 complexity of the constructed library to be ~6.8 x 10^5. The plasmid library was finally purified by
730 NucleoBond Midi-prep Kit (Macherey-Nagel #740410) and stored at –20ºC before use. We performed
731 the isolation of 16 random clones and triple restriction enzyme digestion using BsrGI-HF (NEB

732 #R3575S), ClaI (NEB #R0197S), and PvuI-HF (NEB #R3150S), and confirmed that all the tested clones
733 (16/16) contained the expected barcode inserts. To generate a low-complexity library for the proof-of-
734 concept assays using HEK293T cells, we examined barcode sequences of 96 isolated clones by Sanger
735 sequencing using a sequencing primer SI#471. After removing three clones observed to have mixed
736 Sanger sequencing spectra in the barcode region, barcoded plasmids were pooled with an equimolar
737 ratio and subjected to high-throughput sequencing and lentiviral packaging.
738
739 scCloneSelect barcode library. The scCloneSelect barcode library was prepared similarly to the
740 CloneSelect C→T barcode library. An EGFP coding sequence was first amplified by PCR from pLV-CS-
741 112 (Addgene #131127) using the semi-random oligo pool SI#679 as a forward primer and another oligo
742 pool RS#244 as a reverse primer. The PCR was performed in 25 separate reactions, each in 40 µL
743 reaction volume, composed of 0.12 µL of 10 ng/µL pLV-CS-112 template plasmid, 2 µL each of forward
744 and reverse primers, 0.6 µL of Phusion High-fidelity DNA Polymerase (NEB M0530), 8 µL of 5x Phusion
745 HF Buffer (NEB #B0518S), and 3.2 µL of 2.5 mM dNTPs (NEB #N0447) with the following thermal cycle
746 condition: 98ºC for 30 s, 30 cycles of 98ºC for 10 s, 65ºC for 10 s, and 72ºC for 60 s, and then 72ºC for
747 5 min for the final extension. The amplified barcode-EGFP fragment was pooled into a single tube (1,000
748 µL in total) and digested with 12.5 µL of DpnI (NEB #R0176) at 37ºC for 1 hour. The digested product
749 was then size-selected using FastGene PCR/Gel Extraction Kit (Nippon Genetics #FG-91302). The
750 purified product was subjected to EcoRI-HF (NEB #R3101S) and XbaI (NEB #R0145) digestion
751 overnight at 37ºC and purified again by FastGene PCR/Gel Extraction Kit (Nippon Genetics #FG-91302).
752 For backbone preparation, 25 µg of pRS193 lentiviral cloning backbone plasmid was digested by EcoRI-
753 HF (NEB #R3101S) and XbaI (NEB #R0145) at 37ºC overnight and size-selected using FastGene
754 PCR/Gel Extraction Kit (Nippon Genetics #FG-91302). 1.25 µg of the digested backbone, 320 ng of the
755 purified insert, 25 µL of T4 DNA Ligase (Nippon Gene #317-00406), and 25 µL of 10x T4 DNA Ligase
756 Buffer (NEB #B0202) were mixed in a total volume of 250 µL and incubated overnight at 16ºC. The
757 ligation sample was then used to transform NEB Stable Competent *E. coli* cells (NEB # C3040I). The
758 transformation was performed in a total of 17 reactions, each with 4 µL of the ligation sample
759 transformed to 50 µL of the competent cells according to the manufacturer's high-efficiency
760 transformation protocol. Following one-hour outgrowth in SOC medium (NEB #B9020) at 37ºC, cells
761 were spun down and plated on a total of 15 LB agar plates containing 100 µg/mL Ampicillin (Wako #014-
762 23302). After overnight incubation at 37ºC, colonies formed on the solid agar plates were scraped by
763 adding 1–2 mL ddH2O, pooled into a flask, and further incubated with 200–300 mL of LB liquid medium
764 containing 100 µg/mL Ampicillin (Wako #014-23302) overnight at 37ºC. We plated the transformation
765 sample on agar plates with 300-fold dilution in duplicate and estimated the barcode complexity of the
766 constructed library to be ~1.5 x 10^5. The plasmid library was finally purified by NucleoBond Midi-prep
767 Kit (Macherey-Nagel #740410) and stored at –20ºC before use. We performed the isolation of 20
768 random clones and confirmed the expected fragment insertion for 17/20 by genotyping PCR using the
769 primer pair RS#147 and SI#514. From the same 20 clones, we selected six (including the three that did
770 not yield the expected genotyping PCR bands), performed double digestion using EcoRI-HF (NEB
771 #R3101S) and BamHI-HF (NEB #R3136S), and Sanger sequencing using sequencing primers SI#514
772 and RS#147 for uptag and dntag, respectively, and confirmed that all the tested clones contained the
773 expected uptag and dntag inserts.
774
775 **Lentiviral barcoding**
776 Virus packaging. HEK293T cells were plated on either a 10-cm cell culture dish at a density of ~2 x 10^6
777 cells with 10 mL of the culture medium or 6-well cell culture plate wells at a density of ~2 x 10^5 cells/well
778 with 2 mL of the culture medium one day before plasmid transfection. For packaging using a 10-cm cell
779 culture dish, 3.0 µg transgene vector, 2.25 µg psPAX2 (Addgene #12260), 0.75 µg pMD2.G (Addgene
780 #12259), and 18 µL of 1 mg/mL PEI MAX (Polysciences #24765-100) were dissolved in 1,000 µL of 1x
781 PBS and applied to the cell culture. For packaging using a 6-well cell culture plate well, 489 ng transgene
782 plasmid, 366.7 ng psPAX2, 122.3 ng pMD2.G, and 2.93 µL of 1 mg/mL PEI MAX (Polysciences #24765-
783 100) were dissolved in 300 µL of 1x PBS and applied to the cell culture. The culture medium was
784 changed to a fresh medium one day after transfection. The transfected cells were further incubated for

785 48–72 hours. The cell culture supernatant was then harvested and filtered with 0.22 μm pore size sterile
786 syringe filters. The recombinant lentivirus sample was then aliquoted 500–1,000 μL each into 1.5-mL
787 test tubes and stored at –80ºC.
788
789 Virus concentration. To increase virus infection titer, we concentrated harvested virus samples using a
790 polyethylene glycol (PEG)-based method (*65*) using PEG 6000 (Wako #169-09125) or Lenti-X
791 Concentrator (Takara #631231). When PEG 6000 was used, ~10 mL of recombinant virus sample was
792 mixed with 2.55 mL of 50 w/v% PEG 6000, 1.085 mL of 4M NaCl, and 1.365 mL of 1x PBS in a 50 mL
793 tube. The sample was continuously mixed using a rotator at 4ºC for 90 min and centrifuged at 4,000 g
794 and 4ºC for 20 min. The supernatant was discarded, and the precipitated virus pellet was resuspended
795 with 1.1 mL of Opti-MEM (Gibco #31985062) by pipetting and vortexing until fully dissolved, resulting in
796 a 10-fold concentration of the virus sample. The virus concentration using Lenti-X Concentrator was
797 performed with the manufacturer's protocol and dissolved in Opti-MEM (Gibco #31985062) to yield a
798 10- or 15-fold concentration. The concentrated virus samples were stored at –80ºC.
799
800 Transduction of HEK293T and HeLa cells. Cells were seeded on 6-well cell culture plate wells at a
801 density of ~2 x 10^5 cells/well with 2 mL of the culture medium one day before transduction. A total of
802 1,000 μL transduction mix containing 1 μL of 2 μg/mL Polybrene (Sigma-Aldrich #TR-1003),
803 recombinant lentivirus, and the cell culture medium was applied to each well alongside non-virus
804 controls. One day after transduction, cells were trypsinized and plated on 96-well cell culture plate wells
805 at a density of ~5 x 10^3 cells/well for a virus titer measurement. The next day, the culture medium was
806 exchanged with a fresh medium containing 2.0 μg/mL Puromycin (Gibco #A1113803) or 5.0 μg/mL
807 Blasticidin S (Wako #029-18701) to select the infected cells for two to five days. After drug selection,
808 cell viability was measured using CellTiter-Glo (Promega #G7570) according to the manufacturer's
809 protocol. The luminescence was quantified using Infinite 200 PRO plate reader (TECAN). The
810 background luminescence from wells without cell samples was used to subtract the signals. For each
811 condition, the multiplicity of infection (MOI) was determined by the fraction of the survived cells
812 compared to the non-selective condition control. The samples with an MOI close to but not over 0.1 was
813 used for the following analyses, where most of the selected cells were expected to have a single viral
814 integration according to Poisson statistics.
815
816 Transduction of mESCs. Cells were seeded on 6-well cell culture plate wells at a density of ~2 x 10^5
817 cells/well with 2 mL of the culture medium one day before transduction. For cell transduction,
818 recombinant virus samples with volume of 10–100 μL were thawed on ice, mixed with 1.5 μL of 8 μg/mL
819 Polybrene (Sigma-Aldrich #TR-1003) and 1.5 mL of fresh cell culture medium, and applied to the cells.
820 To select the transduced cells, the culture media was exchanged with a fresh media containing 1.0
821 μg/mL Puromycin (Gibco #A1113803) two days after infection, followed by incubation for three days.
822 Survived cells were detached, and cell counts were measured using Automated Cell Counter TC20
823 (BioRad). The MOI was determined by the fraction of the survived cells compared to the non-selective
824 condition control. The samples with an MOI close to but not over 0.1 was used for the following analyses.
825
826 Transduction of hPSCs. Cells were seeded on 6-well cell culture plate wells at a density of ~1 x 10^5
827 cells/well with 2 mL of the culture medium one day before transduction. For cell transduction,
828 recombinant virus samples with volume of 10–100 μL were thawed on ice, mixed with 1.5 μL of 8 μg/mL
829 Polybrene (Sigma-Aldrich #TR-1003) and 1.5 mL of fresh cell culture medium, and applied to the cells.
830 After 48 hours of infection, the culture media was exchanged with a fresh media containing 1.0 μg/mL
831 of Puromycin (Gibco #A1113803) for three days. The reporter-integrated cells were dissociated into
832 single cells and subjected to flow cytometry cell sorting to enrich EGFP-negative cells. The sorted cells
833 were maintained with StemFit AK02N culture media (REPROCELL #RCAK02N).
834
835 **Preparing cells with stably integrated Target-AID**
836 mESCs. The mESC line with stably integrated Target-AID was established by electroporation using
837 NEPA21 Super Electroporator (NEPAGENE). After detaching cells from cell culture plate wells, ~2 x 10^6

838 cells were mixed with 100 µL of Opti-MEM (Gibco #31985062), 2.0 µg of pNM1325, and 0.7 µg of a
839 Super piggyBac transposase vector (SBI #PB210PA-1), and transferred to an electroporation cuvette
840 (NEPAGENE #EC-002S). The electroporation was done by two poring pulses of positive electric polarity
841 with 115 V for 5 ms with 50-ms intervals and 10% decay rate, and five transfer pulses each for positive
842 and negative electric polarities with 20 V for 50 ms with 50-ms intervals and 40% the decay rate. After
843 electroporation, cells were transferred to a 10-cm cell culture dish with a fresh culture medium. The
844 medium was exchanged with a fresh medium one day after electroporation. Two days post
845 electroporation, the medium was exchanged again with medium containing 5 µg/mL of Blasticidin S
846 (Wako #029-18701) to select cells with stable integration. The cells were incubated for about two weeks
847 in the selection medium.
848
849 hPSCs. To establish an hPSC line with stably integrated Target-AID, CA1 cells were seeded on 24-well
850 cell culture plate wells at a density of ~5 x 10^4 cells/well with 1 mL of the culture medium one day before
851 transfection. The next day, the medium was exchanged to remove Y-27632. The transfection mix was
852 prepared by combining 450 ng of pNM1325 (CAGp-Target-AID-2A-Blast), 50 ng of a hyperactive
853 PiggyBac transposase plasmid, 1 µL of Lipofectamine Stem Transfection Reagent (Invitrogen
854 #STEM00001), and 49 µL of Opti-MEM (Gibco #31985062) and applied to the wells after 10 min
855 incubation. The next day, the culture medium was exchanged with a fresh medium to remove the
856 transfection reagent. Three days after transfection, the medium was exchanged with a fresh medium
857 containing 5 µg/mL of Blasticidin S to start selection for 24 hours. Another selection was performed two
858 days until confluent and cells were passaged into a new cell culture plate. One extra selection was
859 conducted to ensure a positive selection of the cells.
860
861 **Transient plasmid delivery**
862 Transfection of HEK293T and HeLa cells. Cells were seeded on 24-well cell culture plate wells at a
863 density of ~5 x 10^4 cells/well with 500 µL of the culture medium or 6-well cell culture plate wells at a
864 density of ~2 x 10^5 cells/well with 2,000 µL of the culture medium one day before transfection. For regular
865 transfection in a 24-well plate, a total of 400 ng of plasmids (3:1 volume of Cas9-based enzyme plasmid
866 to gRNA plasmid when they were mixed), 1.2 µL of 1 mg/mL PEI MAX (Polyscience #24765), and 100
867 µL of 1x PBS were mixed, incubated for 5 or 10 min at room temperature, and applied to each well. The
868 dose-dependent activation assay with the different Target-AID expression plasmid was performed using
869 a 24-well plate with the plasmid amount per well ranging from 50–800 ng and 1 mg/mL of PEI MAX
870 whose volume was adjusted to be 3 µL per 1 µg of the plasmid. For isolating barcoded HEK293T cells
871 using CloneSelect C→T, we used 6-well cell culture plates, and a total of 800 ng of plasmids encoding
872 both Target-AID and gRNA were combined with 2.5 µL of 1 mg/mL PEI MAX (Polyscience #24765) and
873 200 µL of 1x PBS and applied to each well after 5–10 min incubation at room temperature.
874
875 Transfection of mESCs. Cells were seeded on 48-well cell culture plate wells at a density of ~6 x 10^4
876 cells with 200 µL of the culture medium. For each reaction, a total of 200 ng of plasmids were diluted in
877 20 µL of Opti-MEM (Gibco #31985062), and 0.6 µL of Lipofectamine 2000 (Invitrogen #11668019) was
878 combined with 19.4 µL of Opti-MEM (Gibco #31985062) as a transfection mix. The plasmid and
879 transfection mix were then combined and applied to each well after 5 min incubation at room
880 temperature.
881
882 Electroporation of hPSCs. For the gRNA-dependent reporter activation of the barcoded CA1 hPSCs
883 with the stably integrated Target-AID, we used Neon Transfection System (Invitrogen MPK5000) to
884 deliver the gRNA plasmid by electroporation. Cells were detached from cell culture plate wells, and ~1
885 x 10^5 cells were mixed with 100 µL of Neon Resuspension Buffer and 2.0 µg of gRNA plasmid. The
886 electroporation was done by 1,200 V for 30 ms with one pulse. In the co-delivery of Target-AID and
887 gRNA expression plasmids to the barcoded H1 hPSCs, ~1 x 10^5 cells were mixed with 100 µL of Neon
888 Resuspension Buffer, 3.0 µg of Target-AID plasmid, and 3.0 µg of gRNA plasmid, and the
889 electroporation was done by 1,200 V for 20 ms with two pulses.
890

891 **Preparing microscope imaging samples**
892 For imaging HEK293T cells, 25 µL of 0.1 mg/mL Hoechst 33342 (Invitrogen #H3570) dissolved in DMEM
893 was directly added to each well of 24-well cell culture plates three days after transfection for nuclear
894 counterstaining. The specimens were incubated at room temperature for 10 min, followed by removal
895 of the culture medium. Cells were gently washed with fresh 500-µL of DMEM once and filled with 500
896 µL of fresh DMEM before imaging. For imaging HeLa cells, cells were first washed with 500 µL of 1x
897 PBS, added 25 µL of 0.1 mg/mL Hoechst 33342 dissolved in 475 µL of DMEM and incubated at room
898 temperature for 10 min. Cells were gently washed with 500 µL of 1x PBS and filled with 500 µL of fresh
899 DMEM. For mESCs, 5.0 µg/mL Hoechst 33342 (Invitrogen #H3570) dissolved in the cell culture medium
900 was directly added to each well and incubated at room temperature for 10 min to proceed with imaging.
901 All live cell imaging was performed using BZ-X710 (Keyence), InCellAnalyzer 6000 (GE Healthcare), or
902 IX83 (Olympus) with a 4x, 10x, or 20x objective lens. The contrast and brightness of the images obtained
903 in a single batch of the experiment were uniformly adjusted using ImageMagick (Version 7.1.0-20) or
904 Fiji (Version 1.0).
905
906 **Flow cytometry analysis**
907 Cells were detached by 0.25 w/v% Trypsin-EDTA (Wako #201-18841), incubated at 37ºC for 5 min,
908 collected into a 1.5-mL tube or a 96-well round-bottom plate, and centrifuged at 1,000 rpm and room
909 temperature for 5 min. After aspirating the supernatant, cell pellets were gently resuspended with 150–
910 500 µL of ice-cold FACS buffer consisting of 2% FBS in 1x PBS. The samples were immediately placed
911 on ice until flow cytometry analysis. The flow cytometry analysis was performed with BD FACSVerse
912 Cell Analyzer (BD Biosciences) or CytoFLEX Flow Cytometer (Beckman Coulter). The samples were
913 mixed gently by pipetting or vortexing immediately before the analysis. Approximately 10,000–20,000
914 raw events were acquired for each sample. The data analysis was performed with custom R scripts
915 using flowWorkspace (version 0.5.40) (https://github.com/RGLab/flowWorkspace), flowCore (version
916 1.11.20) (https://github.com/RGLab/flowCore) and CytoExploreR (version 1.1.00)
917 (https://github.com/DillonHammill/CytoExploreR) or a Python package FlowCytometryTools (version
918 0.5.0) (https://github.com/eyurtsev/FlowCytometryTools). Codes are available at
919 https://github.com/yachielab/CloneSelect_v1/tree/main/FACS.
920
921 **Flow cytometry cell sorting**
922 HEK293T cells. Four days after the transfection of Target-AID and gRNA plasmids for barcode-specific
923 cell isolation, cells were detached by 0.25 w/v% Trypsin-EDTA (Wako #201-18841), incubated at 37ºC
924 for 5 min, collected into a 1.5-mL tube, and centrifuged at 1,000 rpm and room temperature for 5 min,
925 followed by resuspension into a 5-mL polystyrene round-bottom tube (FALCON) containing 150–500 µL
926 of 1% FBS in 1x PBS. The cell suspension was immediately placed on ice until sorting. The sorting was
927 performed using BD FACSJazz (BD Biosciences) with 1.0 Drop Single Sort mode. Cells were first gated
928 using FSC-A and SSC-A, and the gate for EGFP+ cells was determined to obtain those having high
929 FITC-A intensities that were not observed in a control cell sample transfected with Target-AID and NT
930 gRNA plasmids. EGFP+ cells were sorted into 8-strip PCR tubes (Nippon Genetics #FG-018WF), each
931 containing 2.5 µL of 1x PBS. For good cell recovery, the cell destination position in the collecting tube
932 was adjusted manually for each sample. The sample was immediately placed on an ice-cold 96-well
933 aluminum block. Although the EGFP+ cell rate varied across the samples, approximately 50–600
934 EGFP+ cells were recovered from each experiment.
935
936 mESCs. Three days after the transduction of a query gRNA, each cell sample was expanded in a 10-
937 cm cell culture dish. Cells were detached by 0.25 w/v% Trypsin-EDTA (Gibco #25200072), incubated
938 at 37ºC for 5 min, collected into a 1.5-mL tube, and centrifuged at 1,000 rpm and room temperature for
939 5 min, followed by resuspension to ~1 x 10^6 cells in 1x PBS containing 2% FBS in a 5-mL polystyrene
940 round-bottom tube (Falcon #352054). The cell suspension was immediately placed on ice until sorting.
941 The sorting was performed using MoFlo Astrios EQ Cell Sorter (Beckman Coulter). Cells were first gated
942 using FSC-A and SSC-A, and the gate for EGFP+ cells was determined to obtain those having high
943 FITC-A intensities that were not observed in a non-transduction control cell sample. EGFP+ positive

944 cells were single-cell sorted into 96-well plate wells, with the rest sorted in bulk to a single well of 96-
945 well plate, each with 100 µL of the mESC culture media. Approximately 100–1,000 EGFP+ cells were
946 recovered from each experiment, except for Clone 153, for which EGFP+ cells above the gating
947 threshold could not be observed.
948
949 **Barcode sequencing library preparation**
950 CloneSelect C→T plasmid library and barcoded cell population. To identify barcodes of the CloneSelect
951 C→T plasmid library by high-throughput sequencing, ~10 ng of plasmid DNA (~1.0 x 109 molecules)
952 was used as a PCR template. To identify barcodes of the initial barcoded HEK293Ta cell population,
953 genomic DNA was purified using NucleoSpin Tissue (Macherey Nagel #740952) according to the
954 manufacturer's protocol, and 119 ng of the extracted genomic DNA (4 x 10^4 molecules; 400-fold to the
955 estimated barcode complexity) was used as a PCR template. The sequencing libraries were prepared
956 by a two-step PCR method. The first-round PCR was performed in triplicate, each in 20 µL volume,
957 composed of template DNA, 0.5 µL each of 20 µM forward (SI#682) and reverse (SI#683) primers, 0.2
958 µL of Phusion High-Fidelity DNA Polymerase (NEB #M0530), 4 µL of Phusion HF Buffer (NEB #B0518S),
959 2 µL of 2 mM dNTPs (Takara #4025), and 0.6 µL of 100% DMSO (NEB #12611P), with the following
960 thermal cycle condition: 98ºC for 10 s, 30 cycles of 98ºC for 10 s, 61ºC for 10 s, and 72ºC for 30 s, and
961 then 72ºC for 5 min for the final extension. Each PCR product was size-selected using 2% agarose gel,
962 purified using PCR/Gel Extraction Kit (Nippon Genetics #FG-91302). To provide Illumina sequencing
963 adapters and custom indices, the second-round PCR was performed to amplify each first-round PCR
964 replicate product in 20 µL volume, composed of 2.5 ng of the 1st PCR product, 1 µL each of 10 µM P5
965 and P7 custom index primers, 0.2 µL of Phusion High-Fidelity DNA Polymerase (NEB #M0530), 4 µL of
966 Phusion HF Buffer (NEB #B0518S), 2 µL of 2 mM dNTPs (Takara #4025), and 0.6 µL of 100% DMSO
967 (NEB #12611P), with the following thermal cycle condition: 98ºC for 10 s, 20 cycles of 98ºC for 10 s,
968 61ºC for 10 s, and 72ºC for 30 s, and then 72ºC for 5 min for the final extension. Custom indices assigned
969 to the second-round PCR products can be found in Table S3. The second-round PCR products were
970 size-selected and purified using PCR/Gel Extraction Kit (Nippon Genetics #FG-91302). The samples
971 were pooled, quantified by qPCR using KAPA Library Quantification Kit Illumina (KAPA BIOSYSTEMS
972 #KK4824), and analyzed by paired-end sequencing using Illumina MiSeq.
973
974 Sorted HEK293T cells. For the amplicon sequencing-based identification of barcodes in low-volume
975 cells sorted after gRNA-dependent barcode-specific clone isolation using CloneSelect C→T, we
976 prepared a cell lysate for each sample as a PCR template. The sequencing library of each sample was
977 prepared by modifying the two-step PCR protocol described for purified DNA as a PCR template. Cells
978 in 8-strip PCR tubes were first incubated with 2.0 µL of lysis buffer, including 600 mM KOH, 10 mM
979 EDTA, and 100 mM DTT. The samples were then neutralized with 2.0 µL of neutralization buffer
980 composed of 0.4 µL of 1M Tris-HCl and 1.6 µL of 3 M HCl. The first-round PCR was performed by
981 replacing the template with 2.0 µL of the cell lysate. Although we did not observe visible gel
982 electrophoresis bands for the first-round PCR products, the PCR product of the expected size were size-
983 selected using 2% agarose gel, purified and eluted into 15 µL of ddH2O using PCR/Gel Extraction Kit
984 (Nippon Genetics #FG-91302). The PCR products were quantified using Quant-iT PicoGreen dsDNA
985 Assay Kit (Thermo Fisher Scientific #P7589) and Infinite 200 PRO plate reader (TECAN) using Tecan
986 i-control software (version 1.10.4.0). The second-round PCR was performed for 2.0 ng of the first-round
987 PCR product. Custom indices assigned to the second-round PCR products can be found in Table S3.
988 The second-round PCR products were size-selected and purified using PCR/Gel Extraction Kit (Nippon
989 Genetics #FG-91302). The samples were pooled into a DNA LoBind 1.5-mL tube (Eppendorf #13-698-
990 791), quantified by qPCR using KAPA Library Quantification Kit Illumina (KAPA BIOSYSTEMS
991 #KK4824), and analyzed by paired-end sequencing using Illumina MiSeq.
992
993 Uptag-dntag combination reference database. To identify the uptag-dntag combination reference
994 database of the mESC population tagged using the scCloneSelect barcode library, genomic DNA was
995 first extracted from 1 x 10^5 cells using NucleoSpin Tissue (MACHEREY-NAGEL #740952) according to
996 the manufacturer's protocol. The sequencing libraries were prepared by a two-step PCR method. A total

997 of 50 ng genomic DNA was used for each PCR reaction. The first-round PCR was performed in duplicate,
998 each in 20 µL volume, composed of template DNA, 0.7 µL each of 10 µM forward (SI#682) and reverse
999 (RS#250) primers, 0.2 µL of Phusion High-Fidelity DNA Polymerase (NEB #M0530), 4.5 µL of Phusion
1000 HF Buffer (NEB #B0518S), 1.6 µL of 2.5 mM dNTPs (NEB #N0447), with the following thermal cycle
1001 condition: 98ºC for 10 s, 30 cycles of 98ºC for 10 s, 60ºC for 10 s, and 72ºC for 2 min, and then 72ºC
1002 for 5 min for the final extension. Each PCR product was size-selected using 2% agarose gel, purified,
1003 and eluted into 20 µL of ddH2O using PCR/Gel Extraction Kit (Nippon Genetics #FG-91302). To provide
1004 Illumina sequencing adapters and custom indices, the second-round PCR was performed to amplify
1005 each first-round PCR replicate product in 20 µL volume, composed of 20-fold dilution of the first-round
1006 PCR product, 0.7 µL each of 10 µM P5 and P7 custom index primers, 0.2 µL of Phusion High-Fidelity
1007 DNA Polymerase (NEB #M0530), 4.5 µL of Phusion HF Buffer (NEB #B0518S), 1.6 µL of 2.5 mM dNTPs
1008 (NEB #N0447), with the following thermal cycle condition: 98ºC for 10 s, 20 cycles of 98ºC for 10 s, 60ºC
1009 for 10 s, and 72ºC for 30 s, and then 72ºC for 5 min for the final extension. Custom indices assigned to
1010 the second-round PCR products can be found in Table S3. The second-round PCR products were size-
1011 selected and purified using PCR/Gel Extraction Kit (Nippon Genetics #FG-91302). The samples were
1012 pooled, quantified by qPCR using KAPA Library Quantification Kit Illumina (KAPA BIOSYSTEMS
1013 #KK4824), and analyzed by paired-end sequencing using Illumina MiSeq.
1014
1015 Sorted mESCs. For the amplicon sequencing-based identification of barcodes in cells sorted after
1016 gRNA-dependent barcode-specific clone isolation using scCloneSelect, we prepared a cell lysate for
1017 each sample as a PCR template. The sequencing library of each sample was prepared by modifying
1018 the two-step PCR method described for the identification of the uptag-dntag combination reference
1019 database. Sorted mESC samples were expanded in 96-well culture plate wells until confluent. After
1020 aspirating the culture medium, 20 µL of 50 mM NaOH was added to each well and transferred into a 96-
1021 well PCR plate for direct cell lysis. The samples were then heated at 95ºC for 15 min, followed by cooling
1022 down on ice. The samples were neutralized with 2.0 µL of 1 M Tris-HCl (pH 8.0). The first-round PCR
1023 was performed in 40 µL volume replacing the template with 3.5 µL of cell lysate. The second-round PCR
1024 was performed in 20 µL volume replacing the template with a 10-fold dilution of the first-round PCR
1025 product. Custom indices assigned to the second-round PCR products can be found in Table S3. The
1026 second-round PCR products were size-selected and purified using GeneJET Gel Extraction Kit (Thermo
1027 Fisher Scientific #K0691). The samples were pooled into a DNA LoBind 1.5-mL tube (Eppendorf
1028 #0030108051), quantified by qPCR using KAPA Library Quantification Kit Illumina (KAPA
1029 BIOSYSTEMS #KK4824), and analyzed by paired-end sequencing using Illumina HiSeq2500.
1030
1031 Reamplification of scCloneSelect dntags from Drop-seq library. To increase the sensitivity of identifying
1032 dntags associated with single cells and their transcriptome profiles, we selectively reamplified the DNA
1033 fragments encoding dntags and cell IDs from the intermediate Tn5 transposon-fragmented library of the
1034 Drop-seq process and sequenced them separately. The reamplification PCR was performed in 20 µL
1035 volume, composed of 1 ng of the cDNA library quantified using TapeStation with High Sensitivity D5000
1036 ScreenTape (Agilent #5067-559 and #5067-5593), 0.7 µL each of 20 µM forward primer P5-TSO_Hybrid
1037 (*45*) and reverse primer SI#682, 0.2 µL of Phusion High-Fidelity DNA Polymerase (NEB #M0530), 4.5
1038 µL of 5x Phusion HF Buffer (NEB #B0518), and 1.6 µL of 2.5 mM dNTPs (NEB #N0447) with the
1039 following thermal cycle condition: 95ºC for 30 s, 30 cycles of 98ºC for 30 s, 60ºC for 10 s, and 72ºC for
1040 2 min, and 72ºC for 5 min for the final extension. The first-round PCR product was purified and eluted
1041 into 20 µL of ddH2O using GeneJET Gel Extraction Kit (Thermo Fisher Scientific #K0691). The second-
1042 round PCR was then reamplified in 20 µL volume, composed of 10-fold dilution of the first-round PCR
1043 product, 0.7 µL each of 10 µM P5 and P7 custom dual index primers, 0.2 µL of Phusion High-Fidelity
1044 DNA Polymerase (NEB #M0530), 4.5 µL of 5x Phusion HF Buffer (NEB #B0518), 1.6 µL of 2.5 mM
1045 dNTPs (NEB #N0447), with the following thermal cycle condition: 95ºC for 30 s, 15 cycles of 98ºC for
1046 10 s, 65ºC for 10 s, and 72ºC for 2 min, and 72ºC for 5 min for the final extension. Custom indices
1047 assigned to the second-round PCR products can be found in Table S3. The second-round PCR products
1048 were size-selected using 2% agarose gel and purified using GeneJET Gel Extraction Kit (Thermo Fisher
1049 Scientific #K0691). The samples were pooled, quantified by qPCR using KAPA Library Quantification

Kit Illumina (KAPA BIOSYSTEMS #KK4824), and analyzed by paired-end sequencing using Illumina HiSeq2500.

Drop-seq

scRNA-seq was performed by Drop-seq. The Drop-seq platform was set up using devices manufactured by Dolomite Bio following the manufacturer's protocol. The microfluidic devices were fabricated by YODAKA Co., Ltd. Cell samples were prepared at a concentration of ~2 x 10^5 cells/mL for analysis. Sequencing libraries were prepared following the original Drop-seq paper (*45*). In brief, following emulsion breakage and reverse transcription, "single-cell transcriptomes attached to microparticles" (STAMPs) were washed and proceeded to the Exonuclease I (NEB #M0293L) treatment. A total of ~2,000 STAMPs were used for the whole cDNA amplification of each sample. After second-strand synthesis, the library DNA was purified using AMPure XP beads (Beckman Coulter #A63881), quantified using TapeStation with High Sensitivity D5000 ScreenTape (Agilent #5067-5592 and #5067-5593), and fragmented by Tn5 transposon using Nextera XT DNA Library Preparation Kit (Illumina #FC-131-1024) according to the manufacturer's protocol. The tagmented sequencing library was then purified using AMPure XP beads (Beckman Coulter #A63881) and quantified by TapeStation with High Sensitivity D5000 ScreenTape (Agilent #5067-5592 and #5067-5593). We confirmed each average library size to be ~500 bp. Multiple scRNA-seq libraries were pooled and subjected to high-throughput sequencing using Illumina MiSeq or HiSeq2500. The sequencing library index information can be found in Table S3.

mESC differentiation assay

For the establishment of a barcoded mESC population, cells were seeded on 6-well cell culture plate wells at a density of ~2 x 10^5 cells/well with 2 mL of the culture medium containing 10^3 units/mL ESGRO Recombinant Mouse Leukemia inhibitory factor (LIF) (Millipore #ESG1107) and 2i (1.0 μM PD0325901 Toris #4423 and 3.0 μM CHIR99021 Wako #038-23101). The next day, cells were transduced with 500 μL of 15-fold concentrated barcoded lentivirus pool of scCloneSelect. Two days after, the culture medium was exchanged with a fresh medium containing 1.0 μg/mL of Puromycin (Gibco #A1113803). To restrict the clone complexity for the downstream cell differentiation and clone isolation assays, a clonal population bottleneck was created where ~1,000 cells were seeded on a 6-well cell culture plate well and cultured for ten days. Cells were then split as follows: ~1 x 10^4 cells were seeded into the culture medium with LIF and 2i (LIF+2i+), ~1 x 10^4 cells were seeded into the culture medium without LIF or 2i (LIF–2i–), two samples of ~1 x 10^5 cells each were used for the identification of uptag-dntag combination reference database, and five replicates of ~1 x 10^5 cells were stored at –80ºC using CELLBANKER 1 freeze media (ZENOAQ #11910). Four days later, the cell cultures of the LIF+2i+ and LIF–2i– conditions were subjected to scRNA-seq.

RT-PCR-based analysis of barcode transcription

The transcription of polyadenylated scCloneSelect barcode products was analyzed by RT-PCR and gel electrophoresis. Total RNA was first extracted using ISOSPIN Cell & Tissue RNA Kit (Nippon Gene #314-08211) according to the manufacturer's protocol. The RNA sample was then subjected to DNase I treatment (Takara #2270B) to remove DNA and purified again using ISOSPIN Cell & Tissue RNA Kit (Nippon Gene #314-08211). The first-strand cDNA was synthesized using High-Capacity cDNA Reverse Transcription Kit (Applied Biosciences #4368814). The reaction was performed in 10 μL volume, composed of 5 μL of DNase I treated RNA (~1 μg), 0.5 μL of 100 μM oligo dT primer SI#4, 0.5 μL of MultiScribe Reverse Transcriptase, 1 μL of 10x RT buffer, 0.4 μL of 100 mM dNTP, and 0.5 μL of RNAse Inhibitor (Applied Biosciences #N8080119), with the following thermal cycler condition: 25ºC for 10 min, 37ºC for 12 min, and 85ºC for 5 min. Lastly, the transcription of the target barcode was tested by PCR along with a GAPDH control. The PCR was performed in 20 μL volume, composed of 2 μL of 50-fold diluted first-strand cDNA, 2.8 μL total of a primer pair SI#116–SI#7 to amplify the dntag or a primer pair RS#507–RS#508 to amplify GAPDH, 0.2 μL of Phusion DNA Polymerase (NEB #M0530S), 4 μL of 5x Phusion HF Buffer (NEB #B0518), and 1.6 μL of 2.5 mM dNTPs (NEB #N0447), with the following thermal cycler condition: 98ºC for 30 s, 30 cycles of 98ºC for 10 s, 60ºC for 10 s, and 72ºC for 30 s, and

then 72ºC for 5 min for the final extension. The resulting PCR products were analyzed with 2% agarose gel.

Yeast CloneSelect

Strains

Saccharomyces cerevisiae BY4741 (*MATa his3Δ0 leu2Δ0 met15Δ0 ura3Δ0*) was used for the experiments using yeast CloneSelect system.

Yeast CloneSelect library

To generate the yeast CloneSelect barcode library, a semi-random oligo pool KI#200 encoding 5′-CCGWSNSWSNSWSNSWSNSNGTG-3′ was first chemically synthesized (Table S2) and amplified by PCR in 40 µL volume, composed of 2 µL of 0.01 µM template, 2 µL each of 10 µM forward primer SI#368 and 10 µM reverse primer SI#369, 0.8 µL of Phusion High-fidelity DNA Polymerase (NEB #M0530S), 8 µL of 5x Phusion HF Buffer (NEB #B0518S), and 2 µL of 2 mM dNTPs, with the following thermal cycle condition: 98ºC for 30 s, 35 cycles of 98ºC for 10 s, 68ºC for 20 s, and 72ºC for 5 s, and then 72ºC for 5 min for the final extension. The PCR product was analyzed with 2% agarose gel and size-selected and purified using FastGene PCR/Gel Extraction Kit (Nippon Genetics #FG-91302). The purified barcode fragment was then assembled into the cloning backbone plasmid pKI110 by Golden Gate Assembly (*66*) using BsmBI (NEB #R0580S). We performed two assembly reactions, each in 25 µL volume, composed of 500 fmol barcode fragments, 50 fmol backbone plasmid, 0.5 µL of BsmBI (NEB #R0580S), 0.5 µL of T4 DNA Ligase (Nippon Gene #317-00406), 2.5 µL of 10x T4 DNA Ligation Reaction Buffer (NEB #B0202S), and 0.125 µL of 60 mg/mL BSA (NEB #B9001S) with the following thermal cycle condition: 15 cycles of 37ºC for 5 min and 20ºC for 5 min, and then 55ºC for 30 min for complete backbone digestion. For the bacterial transformation, 5 µL of the assembly product was used to transform 50 µL of DH5α chemically competent cells (NEB #C2987I) according to the manufacturer's high-efficiency transformation protocol. Following one-hour outgrowth in 1 mL of SOC medium (NEB #B9020S) at 37ºC, cells were plated on a total of four LB agar plates containing 100 µg/mL Ampicillin (Wako #014-23302). The cell sample was also diluted and spread on the selective plates to estimate the clone complexity. To check assembly efficiency, random clones were isolated to perform genotyping PCR with a primer pair of KI#169 and KI#170, validating each clone to have an expected barcode insert. To construct Pool-100, 100 colonies were isolated, dissolved in 80 µL each of LB medium containing 100 µg/mL Ampicillin (Wako #014-23302), combined 5 µL each, and cultured overnight at 37ºC. The plasmid DNA was extracted using FastGene Plasmid Mini Kit (Nippon Genetics #FG-90502). Pool-1580 was constructed by scraping and harvesting colonies from a plate with colony forming units close to 1,000 by adding 1.5 mL of LB media containing 100 µg/mL Ampicillin (Wako #014-23302). The resulting cell samples were centrifuged at 13,000 rpm for 2 min to discard supernatant. The plasmid DNA pools were then purified from the collected cells using FastGene PCR/Gel Extraction Kit (Nippon Genetics #FG-91302).

Barcoding of cells and introduction of genome editing reagents

For the barcoding of cells and introduction of genome editing reagents by transformation, we used the Frozen-EZ Yeast Transformation II kit (Zymo Research #T2001) with slight modifications. Cells were first pre-cultured in 5 mL of YPDA or SC-Dropout medium (complying with the auxotrophic requirement for plasmid maintenance in the host cells) in a cell culture tube rotating overnight at 30ºC. The next day, cells were cultured in 5 mL of a fresh YPDA medium with a starting $OD_{600\ nm}$ of 0.3 and incubated until the $OD_{600\ nm}$ reached 0.8–1.0. After making competent cells according to the manufacturer's protocol, plasmid DNA and 50 µL of competent cells were added into a 1.5-mL tube, mixed thoroughly with 500 µL of EZ3 solution according to the manufacturer's protocol, and incubated at 30ºC for an hour with rotation. The cell sample was centrifuged at 13,000 rpm for 2 min, and the supernatant was discarded, followed by the addition of 2.5 mL of YPDA medium for recovery. After two-hour outgrowth at 30ºC with rotation, cells were centrifuged to remove the medium and washed with 1,000 µL of 1x TE twice. Cells were then spread on SC–Dropout agar plates and incubated for two–four days at 30ºC.

1155
1156 Barcoding of cells. When the background BY4741 cells were transformed using the barcode plasmid
1157 library containing *HIS3* marker, YPDA medium was used for pre-culturing and SC–His+Ade plates for
1158 selecting transformants. For the pooled barcoding of cells, the reaction was scaled to transform 250 µL
1159 of competent cells using 200 ng of plasmid in total, and colonies formed on the selective plates were
1160 pooled and collected by scraping with 3–4 mL of SC–His+Ade. For the barcoding of cells using a single
1161 barcode plasmid clone, 200 ng of plasmid was used to transform 15 µL of competent cells.
1162
1163 Introduction of genome editing reagents. When cells harboring the barcode plasmid containing *HIS3*
1164 marker were subjected to a clone isolation, cells were transformed twice, first with the constitutively
1165 active Target-AID plasmid pKI086 containing *LEU2* marker and next with the targeting gRNA expression
1166 plasmid encoding *URA3* marker. For the first transformation, we used SC–His+Ade medium for pre-
1167 culturing and SC–His–Leu+Ade plates for selecting transformants. For the second transformation, SC–
1168 His–Leu+Ade medium for pre-culturing and SC–His–Leu–Ura+Ade for selecting transformants. When
1169 the background BY4741 cells were transformed using one of the galactose-inducible Cas9-based
1170 enzyme plasmids (Cas9, dCas9, dCas9-PmCDA1, dCas9-PmCDA1-UGI, nCas9, nCas9-PmCDA1, and
1171 nCas9-PmCDA1-UGI) containing *LEU2* marker and a CAN1-targeting gRNA plasmid containing *URA3*
1172 marker, we used YPDA medium for pre-culturing and SC–Leu–Ura+Ade plates for selecting
1173 transformants. For the barcode-specific reporter activation in a complex barcoded population, the
1174 reaction was scaled to transform 250 µL of competent cells using 200 ng of the enzyme plasmid and
1175 200 ng of the targeting gRNA plasmid in total. For the other small-scale transformations, 200 ng of the
1176 effector plasmid and 200 ng of the targeting gRNA plasmid were used to transform 15 µL of competent
1177 cells.
1178
1179 **Barcode sequencing library preparation**
1180 To identify barcodes of the yeast CloneSelect plasmids introduced to the cells by high-throughput
1181 sequencing, we first extracted and purified the barcode plasmids. We first centrifuged cells at 15,000
1182 rpm for 3 min, followed by discarding the supernatant. The cell pellet was resuspended with 20 µL of
1183 Zymolyase Buffer, containing 2.5 mg/mL Zymolyase (Zymo Research #E1005), 10 mM Sodium
1184 phosphate, and 1.2 M Sorbitol, and 500 µL of Solution I Buffer (Supplied in Zymolyase #E1005),
1185 containing 0.1 M EDTA and 1 M Sorbitol. The sample was incubated at 37ºC for one hour and
1186 centrifuged at 15,000 rpm for 1 min, followed by discarding the supernatant. The cell lysate sample was
1187 then incubated with 250 µL of Solution II Buffer (Supplied in Zymolyase #E1005), containing 20 mM
1188 EDTA and 50 mM Tris-HCl, and 1% SDS, at 65ºC for 30 min, followed by the addition of 100 µL of 5 M
1189 potassium acetate. The sample was further incubated on ice for 30 min and centrifuged at 15,000 rpm
1190 for 3 min. The supernatant was transferred into a 1.5-mL tube, and the plasmid DNA was precipitated
1191 with the addition of 400 µL of isopropanol, followed by cleanup with 400 µL of 70% ethanol. The resulting
1192 DNA pellet was resuspended with 50 µL of ddH2O containing 10 µg/mL RNase and incubated at 65ºC
1193 for 10 min. Next, the barcode sequencing libraries were prepared by a two-step PCR method. The first-
1194 round PCR was performed in triplicate, each in 40 µL volume, composed of 1.0 µg of template DNA, 1
1195 µL each of 10 µM forward primer KI#169 and 10 µM reverse primer KI#289, 0.4 µL of Phusion High-
1196 Fidelity DNA Polymerase (NEB #M0530S), 8 µL of Phusion HF Buffer (NEB #B0518S), and 0.8 µL of
1197 10 mM dNTPs (Takara #4030), with the following thermal cycle condition: 98ºC for 30 s, 20 cycles of
1198 98ºC for 10 s, 61ºC for 20 s, and 72ºC for 25 s, and then 72ºC for 5 min for the final extension. Each
1199 PCR product was size-selected using 2% agarose gel, purified, and eluted into 50 µL of ddH2O using
1200 PCR/Gel Extraction Kit (Nippon Genetics #FG-91302). To provide Illumina sequencing adapters and
1201 custom indices, the second-round PCR was performed to amplify each first-round PCR replicate product
1202 in 40 µL volume, composed of 2 µL of the first PCR product, 1 µL each of 10 µM P5 and P7 custom
1203 index primers, 0.4 µL of Phusion High-Fidelity DNA Polymerase (NEB #M0530S), 8 µL of 5x Phusion
1204 HF Buffer (NEB #B0518S), and 0.8 µL of 10 mM dNTPs (Takara #4030), with the following thermal cycle
1205 condition: 98ºC for 30 s, 15 cycles of 98ºC for 10 s and 60ºC for 10 s, and 72ºC for 1 min, and then 72ºC
1206 for 5 min for the final extension. Custom indices assigned to the second-round PCR products can be
1207 found in Table S3. The second-round PCR products were size-selected and purified using PCR/Gel

Extraction Kit (Nippon Genetics #FG-91302). The sequencing libraries were pooled, quantified by qPCR using KAPA Library Quantification Kit Illumina (KAPA BIOSYSTEMS #KK4824), pooled with equimolar ratios, and analyzed by paired-end sequencing using Illumina HiSeq2500.

Analysis of reporter activation efficiency

To analyze the efficiency of the gRNA-dependent barcode-specific mCherry reporter activation, we treated three independent barcoded cell samples, each by all three corresponding gRNAs (3x3 assay). Each sample was spread on SC–His–Leu–Ura+Ade agar plates, scraped and inoculated to a 1.5-mL tube containing 500 µL of SC–His–Leu–Ura+Ade medium, and cultured for two–four days at 30ºC. The 20 µL of pre-cultured cell sample was mixed with 180 µL of SC–His–Leu–Ura+Ade media and transferred into a flat-bottom transparent 96-well plate (Greiner Bio-One #655090). Samples were then analyzed using Infinite 200 PRO plate reader (TECAN) using Tecan i-control software (version 1.10.4.0) to measure mCherry fluorescence intensities normalized by $OD_{595\ nm}$ values. For microscopic observation of cells, 2.5 µL of cell sample was transferred on a glass slide, gently covered with a glass coverslip, and observed using BZ-X710 (Keyence) with 20x and 40x objective lenses. We also directly measured the GTG→ATG conversion rate of each sample by high-throughput sequencing. Cells were scraped and harvested from the selective plates and lysed by DNAZol (COSMO BIO #DN127) according to the manufacturer's protocol. The sequencing libraries were then prepared by a two-step PCR method. The first-round PCR was performed in 32 µL volume, composed of 1.6 µL of cell lysate, 1.6 µL each of 10 µM forward primer KI#168 and 10 µM reverse primer KI#169, 0.64 µL of Phusion High-Fidelity DNA Polymerase (NEB #M0530S), 6.4 µL of Phusion HF Buffer (NEB #B0518S), and 0.64 µL of 10 mM dNTPs (Takara #4030), with the following thermal cycle condition: 98ºC for 30 s, 30 cycles of 98ºC for 10 s, 61ºC for 10 s, and 72ºC for 1 min, and then 72ºC for 5 min for the final extension. The remaining procedure was processed by the same protocols described for the barcode sequencing library preparation. Custom indices assigned to the second-round PCR products can be found in Table S3. The second-round PCR products were size-selected and purified using PCR/Gel Extraction Kit (Nippon Genetics #FG-91302). The samples were pooled, quantified by qPCR using KAPA Library Quantification Kit Illumina (KAPA BIOSYSTEMS #KK4824), and analyzed by paired-end sequencing using Illumina MiSeq.

Isolation and analysis of barcoded colonies

After the barcode-specific reporter activation of a complex population, cells from test and control conditions were spread on SC–His–Leu–Ura+Ade agar plates and imaged under a blue light illuminator FAS-IV (Nippon Genetics) to distinguish mCherry+ and mCherry– colonies. Colonies were isolated into 96-well cell culture plate wells containing 98 µL of SC–His–Leu–Ura+Ade and cultured overnight at 30ºC. Samples were then analyzed using Infinite 200 PRO plate reader (TECAN) using Tecan i-control software (version 1.10.4.0) to measure mCherry fluorescence intensities normalized by $OD_{595\ nm}$ values. The same colony isolates were also subjected to Sanger sequencing to identify their barcode sequences and base editing outcomes. Using the same cell lysis, first-round PCR, and PCR cleanup protocols as in the analysis of the reporter activation efficiency, we obtained barcode DNA fragments from each sample. Each PCR product was analyzed by Sanger sequencing using a sequencing primer SI#658. The Sanger sequencing trace was analyzed using PySanger (https://github.com/ponnhide/PySanger).

Canavanine assay

Genome editing efficiencies of different Cas9-based genome editing enzymes (Cas9, dCas9, dCas9-PmCDA1, dCas9-PmCDA1-UGI, nCas9, nCas9-PmCDA1 and nCas9-PmCDA1-UGI) were estimated by Canavanine assay, where enzymes were introduced to the cells with a gRNA targeting an arginine transporter CAN1 and its knockout efficiency can be assessed by cell survival under the presence of a toxic arginine analog Canavanine. In this assay, Cas9-based enzyme expressions were regulated under a galactose-inducible GAL1/10 promoter. To induce genome editing, cells containing enzyme and gRNA plasmids were first cultured in SC–Leu–Ura medium containing 2% glucose at 30ºC until saturation. Cells were then resuspended in SC–Leu–Ura medium containing 2% raffinose with 16-fold dilution and cultured at 30ºC until saturation. Finally, cells were resuspended in SC–Leu–Ura medium containing

2% raffinose and 0.02% galactose with 32-fold dilution and cultured at 30ºC for two days. Each sample was then spread on SC–Leu–Ura–Arg+Ade and SC–Leu–Ura–Arg+Ade containing 60 mg/ml canavanine plates and cultured at 30ºC for two–four days to estimate colony forming units and perform spot assays. Furthermore, we scraped and harvested colonies from the SC–Leu–Ura–Arg+Ade control plates to extract genomic DNA samples for the mutation spectra assay by high-throughput sequencing. We first lysed 20 µL of cells at $OD_{600\ nm}$ of 1.0 using 100 µL of DNAzol (COSMO BIO #DN127) followed by incubation at room temperature for 15 min. The lysate was mixed thoroughly with 30 µL of 1M NaCl and 50 µL of 100% ethanol and centrifuged at 14,000 rpm for 10 min. The supernatant was discarded, and the pellet was washed with 550 µL of 70% ethanol. After air-drying, the sample was resuspended in 50 µL of ddH2O. We then prepared an amplicon sequencing library for each sample by a two-step PCR method in triplicate. The first-round PCR was in 40 µL volume, composed of 2 µL of template DNA, 2 µL each of 10 µM forward primer #KN85F3 and 10 µM reverse primer #KN85R2, 0.8 µL of Phusion High-Fidelity DNA Polymerase (NEB #M0530S), 8 µL of Phusion HF Buffer (NEB #B0518S), and 0.8 µL of 10 mM dNTPs (Takara #4030), with the following thermal cycle condition: 98ºC for 30 s, 30 cycles of 98ºC for 10 s, 60ºC for 10 s, and 72ºC for 1 min, and then 72ºC for 5 min for the final extension. We also performed the same process using a primer pair #HO2F2–#HO2R2 to prepare control samples. The remaining procedure was processed by the same protocols described for the barcode sequencing library preparation. Custom indices assigned to the second-round PCR products can be found in Table S3. The second-round PCR products were size-selected and purified using PCR/Gel Extraction Kit (Nippon Genetics #FG-91302). The sequencing libraries were pooled, quantified by qPCR using KAPA Library Quantification Kit Illumina (KAPA BIOSYSTEMS #KK4824), pooled with equimolar ratios, and analyzed by paired-end sequencing using Illumina MiSeq.

Bacterial CloneSelect

Preparation of cells for various bacterial CloneSelect systems

We prepared cell samples harboring single barcode plasmids for different variants of bacterial CloneSelect systems (Table S2). The plasmid for the EGFP reporter-based system was introduced to BL21(DE3) Competent *E. coli* cells (NEB #C2527I), and the plasmids for the Blasticidin and Zeocin resistance marker-based systems were introduced to T7 Express chemically competent cells (NEB #C2566I) according to the manufacturer's high-efficiency transformation protocols. The transformants were selected on LB agar plates containing 100 µg/mL Ampicillin (Wako #014-23302) and/or 50 µg/mL Kanamycin (Wako #111-00344).

Bacterial CloneSelect library

To generate the bacterial CloneSelect barcode library for the Zeocin resistance marker, a semi-random oligo pool KI#405 encoding 5′-ATGCCGVNNVNNVNNVNNVNNTAA-3′ was chemically synthesized (Table S2), where a start codon ATG, the antisense strand sequence of 5′-CGG-3′ PAM and a quintuple repeat of VNN (V=non-T) were followed by a TAA stop codon. The VNN repeat restricts additional stop codons from appearing in-frame to the downstream reporter. The semi-random oligo pool was amplified by PCR in 20 µL volume, composed of 1 µL of 1 µM template, 1 µL each of 10 µM forward primer SI#368 and 10 µM reverse primer SI#369, 0.4 µL of Phusion High-fidelity DNA Polymerase (NEB #M0530L), 4 µL of 5x Phusion HF Buffer (NEB #B0518S), and 0.4 µL of 10 mM dNTPs, with the following thermal cycle condition: 98ºC for 30 s, 20 cycles of 98ºC for 10 s, 68ºC for 20 s, and 72ºC for 20 s, and then 72ºC for 5 min for the final extension. The PCR product was analyzed with 2% agarose gel and size-selected and purified using FastGene PCR/Gel Extraction Kit (Nippon Genetics #FG-91302). The purified barcode fragment was then assembled into the cloning backbone plasmid pKI243 by Golden Gate Assembly using BsmBI (*66*). We performed the assembly in 12.5 µL volume, composed of 2.91 fmol of barcode fragments, 14.9 fmol of backbone plasmid, 0.25 µL of BsmBI (NEB #R0580L), 0.5 µL of T4 DNA Ligase (Nippon Gene #317-00406), 1.25 µL of 10x T4 DNA Ligation Reaction Buffer (NEB #B0202S), and 0.62 µL of 2 mg/mL BSA (NEB #B9001S) with the following thermal cycle condition: 15 cycles of 37ºC for 5 min and 20ºC for 5 min, and then 55ºC for 30 min for complete backbone digestion. For the bacterial transformation, 3 µL of the assembly product was used to transform 65 µL of T7

Express chemically competent cells (NEB #C2566I) according to the manufacturer's high-efficiency transformation protocol. Following one-hour outgrowth in 500 µL of SOC medium (NEB #B9020S) at 37ºC, the cell sample was plated 250 µL each on three LB agar plates containing 100 µg/mL Ampicillin (Wako #014-23302). The cell sample was also diluted and plated on the selective plates to estimate the clone complexity. To check assembly quality and efficiency, 12 random clones were isolated and subjected to Sanger Sequencing, validating that 11 out of 12 clones showed an expected clone barcode insert. To construct Pool-100, 100 colonies were isolated, dissolved in 80 µL each of LB medium containing 100 µg/mL Ampicillin (Wako #014-23302), combined 5 µL each, and cultured overnight at 37ºC. The plasmid DNA was extracted using FastGene Plasmid Mini Kit (Nippon Genetics #FG-90502). Pool-1550 was constructed by scraping and harvesting colonies from a plate with colony forming units close to 1,000 by adding 1.5 mL of LB media containing 100 µg/mL Ampicillin (Wako #014-23302). The barcode plasmid libraries were used to transform T7 Express chemically competent cells (NEB #C2566I) to establish barcoded *E. coli* cell populations.

Barcode sequencing library preparation

For each of the Pool-100 and Pool-1550 barcode plasmid libraries, the barcode sequencing library was prepared in triplicate by a two-step PCR method. The first-round PCR was performed in five separate PCR reactions, each in 40 µL volume, composed of 2.0 ng of plasmid template DNA, 1 µL each of 10 µM forward primer KI#403 and 10 µM reverse primer KI#404, 0.4 µL of Phusion High-Fidelity DNA Polymerase (NEB #M0530L), 8 µL of Phusion HF Buffer (NEB #B0518S), and 0.8 µL of 10 mM dNTPs (Takara #4030), with the following thermal cycle condition: 98ºC for 30 s, 20 cycles of 98ºC for 10 s, 54ºC for 20 s, and 72ºC for 25 s, and then 72ºC for 5 min for the final extension. For each replicate, the five PCR reaction products were pooled, size-selected using 2% agarose gel, purified, and eluted into 30 µL of ddH2O using PCR/Gel Extraction Kit (Nippon Genetics #FG-91302). To provide Illumina sequencing adapters and custom indices, the second-round PCR was performed to amplify each first-round PCR replicate product in 40 µL volume, composed of 2 µL of the first PCR product, 1 µL each of 10 µM P5 and P7 custom index primers, 0.4 µL of Phusion High-Fidelity DNA Polymerase (NEB #M0530L), 8 µL of 5x Phusion HF Buffer (NEB #B0518S), and 0.8 µL of 10 mM dNTPs (Takara #4030), with the following thermal cycle condition: 98ºC for 30 s, 15 cycles of 98ºC for 10 s, 60ºC for 10 s, and 72ºC for 60 s, and then 72ºC for 5 min for the final extension. Custom indices assigned to the second-round PCR products can be found in Table S3. The second-round PCR products were size-selected and purified using PCR/Gel Extraction Kit (Nippon Genetics #FG-91302). The sequencing libraries were pooled, quantified by qPCR using KAPA Library Quantification Kit Illumina (KAPA BIOSYSTEMS #KK4824), pooled with equimolar ratios, and analyzed by paired-end sequencing using Illumina MiSeq.

Introduction of genome editing reagents

To introduce a plasmid containing ABE-7.10 and gRNA with given induction systems to barcoded cells, we used Mix&Go! *E. coli* Transformation Kit (Zymo Research #T3001) according to the manufacturer's protocol. To select successful transformants, the transformation reaction sample was plated on LB agar plates containing 100 µg/mL Ampicillin (Wako #014-23302) and 50 µg/mL Kanamycin (Wako #111-00344) and incubated overnight at 37ºC. For the experiments that involved Arabinose (Sigma-Aldrich #A3256-10MG) and Isopropylthio-β-galactoside (IPTG) (ThermoFisher Scientific #15529019) inductions, cells were cultured in the medium containing 100 mM Arabinose and 0.1 mM IPTG overnight at 37ºC prior to the analyses. For the barcoded cell isolation with the Zeocin-resistance marker-based system, we used a low-salt LB medium adjusted to pH 7.5 with 1 M NaOH (Nakalai #37421-05) for the Zeocin activity. We also decided to use 100 µg/mL Zeocin (Invitrogen #R25001) or 100 µg/mL Blasticidin S (Wako #029-18701) with no Arabinose or IPTG for genome editing and selecting reporter-activated cells, as leaky gene editing reagent expressions from the inducible promoters with the no inducer condition were sufficient and maximized the cell viability. The information on the genome editing plasmids used in this study can be found in Table S2.

1365 **Analysis of reporter activation efficiency**
1366 To analyze the efficiency of the gRNA-dependent barcode-specific EGFP-reporter activation, 200 μL of
1367 cell samples were transferred into a flat-bottom transparent 96-well plate (Greiner Bio-One #655090)
1368 and analyzed using Infinite 200 PRO plate reader (TECAN) using Tecan i-control software (version
1369 1.10.4.0) to measure EGFP fluorescence intensities normalized by $OD_{595\ nm}$ values. For microscopic
1370 observation of cells, 2.5 μL of cell sample was transferred onto a glass slide (MATSUNAMI #S2441),
1371 gently covered with a glass coverslip, and observed using BZ-X710 (Keyence) with 20x and 40x
1372 objective lenses.
1373
1374 **Isolation and analysis of barcoded colonies**
1375 After the barcode-specific Zeocin-resistance marker activation in the barcoded cell population, barcodes
1376 of colonies from test and control conditions were analyzed by Sanger sequencing. For each colony, the
1377 barcode region was amplified by PCR in 20 μL volume, composed of 1 μL of cell suspension, 0.5 μL
1378 each of 10 μM forward primer KI#403 and 10 μM reverse primer KI#404, 0.2 μL of Phusion High-Fidelity
1379 DNA Polymerase (NEB #M0530L), 4 μL of Phusion HF Buffer (NEB #B0518S), and 0.4 μL of 10 mM
1380 dNTPs (Takara #4030), with the following thermal cycle condition: 98ºC for 30 s, 30 cycles of 98ºC for
1381 10 s, 54ºC for 20 s, and 72ºC for 30 s, and then 72ºC for 5 min for the final extension. The PCR products
1382 were analyzed with 2% agarose gel and transferred into 96-well PCR plate wells for clean-up with 20
1383 μL of AMPure XP beads (Beckman Coulter #A63881) according to the manufacturer's protocol. The
1384 Sanger sequencing was performed using a sequencing primer KI#403. The Sanger sequencing trace
1385 was analyzed using PySanger (https://github.com/ponnhide/PySanger).
1386
1387
1388 **High-throughput sequencing and data analysis**
1389 **High-throughput sequencing**
1390 The sequencing libraries were mixed with 20–30% of PhiX spike-in DNA control (Illumina #FC-110-
1391 3001) for better cluster generation on the flow cell and sequenced by Illumina MiSeq (MiSeq v3 150-
1392 cycles kit #MS-102-3001 or 300-cycles kit #MS-102-3003), HiSeq2500 (TruSeq rapid SBS kit v2 #FC-
1393 402-4022). Base calling was performed with bcl2fastq2 (version v2.20.0) to generate FASTQ files. The
1394 sequencing condition for each library and the NCBI Sequence Read Archive's ID to each raw FASTQ
1395 file can be found in Table S3.
1396
1397 **Barcode identification**
1398 Mammalian CloneSelect C→T and yeast and bacterial CloneSelect. Sequencing reads were aligned to
1399 the constant sequences of the designed library structure using NCBI BLAST+ (version 2.6.0) (*67*) with
1400 the blastn-short option to identify sample indices for demultiplexing and CloneSelect barcode sequences.
1401 To generate the barcode allowlist for the mammalian CloneSelect C→T mini-pool experiments, barcode
1402 sequences commonly identified in both the plasmid DNA library and genomic DNA library were first
1403 obtained. Their sequencing errors were then corrected using starcode (version 1.4)
1404 (https://github.com/gui11aume/starcode) with the maximum Levenshtein distance threshold of 4 to
1405 merge minor and major barcodes. The barcode counts of each sample were normalized by the total
1406 barcode count. To estimate barcode frequencies for a given cell or DNA pool sample, the barcode
1407 frequencies were averaged across replicates if any. The barcode sequence and frequency information
1408 obtained in this study can be found in Table S1. Codes are available at
1409 https://github.com/yachielab/CloneSelect_v1/tree/main/Barcode_identification/CloneSelect.
1410
1411 scCloneSelect. To identify uptag and dntag barcodes of the barcoded cells, we first demultiplexed the
1412 sequencing reads and used cutadapt v4.1 (https://github.com/marcelm/cutadapt/) to extract uptag and
1413 dntag sequences sandwiched between their 20-bp upstream and downstream constant sequences. The
1414 identified uptags and dntags were filtered with the Q-score threshold of 30, clustered, and further filtered
1415 with their lengths of 17-bp for uptags and 30-bp for dntags using bartender-1.1
1416 (https://github.com/LaoZZZZZ/bartender-1.1) (*68*). For the construction of the uptag-dntag combination
1417 reference database, uptags and dntags found in the same reads were first paired. We discarded

redundant uptag-dntag pairs whose uptag or dntag were found in more abundant pairs. For the mapping of dntag to the uptag-dntag database, we used symspellpy v6.7 (https://github.com/mammothb/symspellpy) to find one with the shortest edit distance. In cases where multiple dntags in the database were found to be mapped by a query with the same edit distances, the dntag with the highest frequency in the database was chosen. To analyze uptag frequencies in a cell population obtained after gRNA-dependent barcode-specific EGFP reporter activation followed by flow cytometry cell sorting, we mapped the sequencing reads to the uptag-dntag database and obtained read counts for uptags recorded in the database using bartender-1.1 and symspellpy v6.7. Conflicts by the redundant hits to the database were resolved in the same way as those for dntags described above. Codes are available at https://github.com/yachielab/CloneSelect_v1/tree/main/Barcode_identification/scCloneSelect.

Drop-seq data analysis

After sample demultiplexing, FASTQ files were processed with Dropseq Tools v2.5.1 (https://github.com/broadinstitute/Drop-seq) for base quality filtering, adapter trimming, and Cell ID and UMI extraction. Picard v2.18.14 (https://github.com/broadinstitute/picard) was used to convert BAM files to FASTQ files for the proceeding step. The filtered reads were aligned using STAR v2.7 (https://github.com/alexdobin/STAR) (*69*) using the mm10 reference genome. The differential gene expression and clustering analysis were performed using Seurat v3 (https://github.com/satijalab/seurat) (*70*). After filtering cells with the thresholds of Feature_RNA > 200 & nFeature_RNA < 2500 & percent.mt < 5, their gene expression profiles were normalized before clustering with the Seurat::sctransform function. The dntags to be analyzed in this study were first determined in the first Drop-seq run based on their cumulative read count distribution of Cell IDs with the threshold determined by the knee point using the Python package kneed v0.8.1 (https://github.com/arvkevi/kneed). However, to map dntag distribution to the single-cell transcriptome data with high sensitivity, we also sequenced the dntag-reamplified library and identified the association of Cell IDs and dntags in the dntag-uptag combination reference database by the method described above for the scCloneSelect uptag and dntag identification. In cases where multiple dntags were found for a single Cell ID, the dntag with the most abundant UMI count was chosen. Codes are available at https://github.com/yachielab/CloneSelect_v1/tree/main/Drop-seq.

Mutational spectra analysis

The amplicon sequencing reads obtained to analyze mutational patterns at the CAN1 target site conferred by each of the Cas9-based genome editing enzymes was processed by a pipeline described previously (*39*). Codes specific to this study are available at https://github.com/yachielab/CloneSelect_v1/tree/main/Mutational_Spectra_Analysis.

Statistical analysis

The statistical tests were performed using R (version 4.2.0). The details of each statistical test can be found in the corresponding figure legend. The statistical methods used in this study and *P*-values are also listed in Table S4.

Data and code availability

High-throughput sequencing data generated in this study are available at the NCBI Sequence Read Archive (PRJNA901977). All the codes used in this study are available at https://github.com/yachielab/CloneSelect_v1.

References

1. H. E. Bhang *et al.*, Studying clonal dynamics in response to cancer therapy using high-complexity barcoding. *Nat Med* **21**, 440-448 (2015).
2. A. N. Hata *et al.*, Tumor cells can follow distinct evolutionary paths to become resistant to epidermal growth factor receptor inhibition. *Nat Med* **22**, 262-269 (2016).
3. U. Ben-David *et al.*, Genetic and transcriptional evolution alters cancer cell line drug response. *Nature* **560**, 325-330 (2018).
4. L. V. Nguyen *et al.*, Barcoding reveals complex clonal dynamics of de novo transformed human mammary cells. *Nature* **528**, 267-271 (2015).
5. W. Pei *et al.*, Polylox barcoding reveals haematopoietic stem cell fates realized in vivo. *Nature* **548**, 456-460 (2017).
6. R. Lu, N. F. Neff, S. R. Quake, I. L. Weissman, Tracking single hematopoietic stem cells in vivo using high-throughput sequencing in conjunction with viral genetic barcoding. *Nat Biotechnol* **29**, 928-933 (2011).
7. S. H. Naik *et al.*, Diverse and heritable lineage imprinting of early haematopoietic progenitors. *Nature* **496**, 229-232 (2013).
8. A. Gerrits *et al.*, Cellular barcoding tool for clonal analysis in the hematopoietic system. *Blood* **115**, 2610-2618 (2010).
9. B. A. Biddy *et al.*, Single-cell mapping of lineage and identity in direct reprogramming. *Nature* **564**, 219-224 (2018).
10. J. Hollmann *et al.*, Genetic barcoding reveals clonal dominance in iPSC-derived mesenchymal stromal cells. *Stem Cell Res Ther* **11**, 105 (2020).
11. P. D. Tonge *et al.*, Divergent reprogramming routes lead to alternative stem-cell states. *Nature* **516**, 192-197 (2014).
12. S. Venkataram *et al.*, Development of a Comprehensive Genotype-to-Fitness Map of Adaptation-Driving Mutations in Yeast. *Cell* **166**, 1585-1596 e1522 (2016).
13. J. E. Barrick *et al.*, Genome evolution and adaptation in a long-term experiment with Escherichia coli. *Nature* **461**, 1243-1247 (2009).
14. S. F. Levy *et al.*, Quantitative evolutionary dynamics using high-resolution lineage tracking. *Nature* **519**, 181-186 (2015).
15. J. R. Blundell *et al.*, The dynamics of adaptive genetic diversity during the early stages of clonal evolution. *Nat Ecol Evol* **3**, 293-301 (2019).
16. A. N. Nguyen Ba *et al.*, High-resolution lineage tracking reveals travelling wave of adaptation in laboratory yeast. *Nature* **575**, 494-499 (2019).
17. A. E. Rodriguez-Fraticelli *et al.*, Single-cell lineage tracing unveils a role for TCF15 in haematopoiesis. *Nature* **583**, 585-589 (2020).
18. C. Weinreb, A. Rodriguez-Fraticelli, F. D. Camargo, A. M. Klein, Lineage tracing on transcriptional landscapes links state to fate during differentiation. *Science* **367**, (2020).
19. D. E. Wagner *et al.*, Single-cell mapping of gene expression landscapes and lineage in the zebrafish embryo. *Science* **360**, 981-987 (2018).
20. A. E. Rodriguez-Fraticelli *et al.*, Clonal analysis of lineage fate in native haematopoiesis. *Nature* **553**, 212-216 (2018).
21. Z. He *et al.*, Lineage recording in human cerebral organoids. *Nat Methods* **19**, 90-99 (2022).
22. M. M. Chan *et al.*, Molecular recording of mammalian embryogenesis. *Nature* **570**, 77-82 (2019).
23. B. Raj *et al.*, Simultaneous single-cell profiling of lineages and cell types in the vertebrate brain. *Nat Biotechnol* **36**, 442-450 (2018).
24. S. Ishiguro, H. Mori, N. Yachie, DNA event recorders send past information of cells to the time of observation. *Curr Opin Chem Biol* **52**, 54-62 (2019).
25. S. Ota *et al.*, Ghost cytometry. *Science* **360**, 1246-1251 (2018).
26. D. Schraivogel *et al.*, High-speed fluorescence image-enabled cell sorting. *Science* **375**, 315-320 (2022).
27. N. Nitta *et al.*, Intelligent Image-Activated Cell Sorting. *Cell* **175**, 266-276 e213 (2018).

1521 28. N. Hasle *et al.*, High-throughput, microscope-based sorting to dissect cellular heterogeneity. *Mol*
1522 *Syst Biol* **16**, e9442 (2020).
1523 29. C. Umkehrer *et al.*, Isolating live cell clones from barcoded populations using CRISPRa-inducible
1524 reporters. *Nat Biotechnol* **39**, 174-178 (2021).
1525 30. A. M. Al'Khafaji, D. Deatherage, A. Brock, Control of Lineage-Specific Gene Expression by
1526 Functionalized gRNA Barcodes. *ACS Synth Biol* **7**, 2468-2474 (2018).
1527 31. C. Gutierrez *et al.*, Multifunctional barcoding with ClonMapper enables high-resolution study of
1528 clonal dynamics during tumor evolution and treatment. *Nat Cancer* **2**, 758-772 (2021).
1529 32. D. Feldman *et al.*, CloneSifter: enrichment of rare clones from heterogeneous cell populations.
1530 *BMC Biol* **18**, 177 (2020).
1531 33. W. Chen *et al.*, Live-seq enables temporal transcriptomic recording of single cells. *Nature* **608**,
1532 733-740 (2022).
1533 34. B. H. Weinberg *et al.*, Large-scale design of robust genetic circuits with multiple inputs and
1534 outputs for mammalian cells. *Nat Biotechnol* **35**, 453-462 (2017).
1535 35. C. Kuscu *et al.*, CRISPR-STOP: gene silencing through base-editing-induced nonsense
1536 mutations. *Nat Methods* **14**, 710-712 (2017).
1537 36. K. Nishida *et al.*, Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive
1538 immune systems. *Science* **353**, (2016).
1539 37. B. L. Emert *et al.*, Variability within rare cell states enables multiple paths toward drug resistance.
1540 *Nat Biotechnol* **39**, 865-876 (2021).
1541 38. H. A. Rees, D. R. Liu, Base editing: precision chemistry on the genome and transcriptome of
1542 living cells. *Nat Rev Genet* **19**, 770-788 (2018).
1543 39. R. C. Sakata *et al.*, Base editors for simultaneous introduction of C-to-T and A-to-G mutations.
1544 *Nat Biotechnol*, (2020).
1545 40. S. Konermann *et al.*, Genome-scale transcriptional activation by an engineered CRISPR-Cas9
1546 complex. *Nature* **517**, 583-588 (2015).
1547 41. A. Hecht *et al.*, Measurements of translation initiation from all 64 codons in E. coli. *Nucleic Acids*
1548 *Res* **45**, 3615-3626 (2017).
1549 42. N. M. Gaudelli *et al.*, Programmable base editing of A•T to G•C in genomic DNA without DNA
1550 cleavage. *Nature* **551**, 464-471 (2017).
1551 43. J. G. Doench *et al.*, Optimized sgRNA design to maximize activity and minimize off-target effects
1552 of CRISPR-Cas9. *Nat Biotechnol* **34**, 184-191 (2016).
1553 44. N. Kim *et al.*, Prediction of the sequence-specific cleavage activity of Cas9 variants. *Nat*
1554 *Biotechnol* **38**, 1328-1336 (2020).
1555 45. E. Z. Macosko *et al.*, Highly Parallel Genome-wide Expression Profiling of Individual Cells Using
1556 Nanoliter Droplets. *Cell* **161**, 1202-1214 (2015).
1557 46. G. X. Zheng *et al.*, Massively parallel digital transcriptional profiling of single cells. *Nat Commun*
1558 **8**, 14049 (2017).
1559 47. A. J. Hill *et al.*, On the design of CRISPR-based single-cell molecular screens. *Nat Methods* **15**,
1560 271-274 (2018).
1561 48. M. Hegde, C. Strand, R. E. Hanna, J. G. Doench, Uncoupling of sgRNAs from their associated
1562 barcodes during PCR amplification of combinatorial CRISPR screens. *PLoS One* **13**, e0197547
1563 (2018).
1564 49. R. P. Bennett, C. A. Cox, J. P. Hoeffler, Fusion of green fluorescent protein with the Zeocin-
1565 resistance marker allows visual screening and drug selection of transfected eukaryotic cells.
1566 *Biotechniques* **24**, 478-482 (1998).
1567 50. S. Banno, K. Nishida, T. Arazoe, H. Mitsunobu, A. Kondo, Deaminase-mediated multiplex
1568 genome editing in Escherichia coli. *Nat Microbiol* **3**, 423-429 (2018).
1569 51. R. M. Shelake, D. Pramanik, J. Y. Kim, In Vivo Rapid Investigation of CRISPR-Based Base
1570 Editing Components in Escherichia coli (IRI-CCE): A Platform for Evaluating Base Editing Tools
1571 and Their Components. *Int J Mol Sci* **23**, (2022).
1572 52. M. Izumi *et al.*, Blasticidin S-resistance gene (bsr): a novel selectable marker for mammalian
1573 cells. *Exp Cell Res* **197**, 229-233 (1991).

1574 53. J. Burian *et al.*, High-throughput retrieval of target sequences from complex clone libraries using
1575 CRISPRi. *Nat Biotechnol*, (2022).
1576 54. K. Kim *et al.*, Epigenetic memory in induced pluripotent stem cells. *Nature* **467**, 285-290 (2010).
1577 55. O. Bar-Nur, H. A. Russ, S. Efrat, N. Benvenisty, Epigenetic memory and preferential lineage-
1578 specific differentiation in induced pluripotent stem cells derived from human pancreatic islet beta
1579 cells. *Cell Stem Cell* **9**, 17-23 (2011).
1580 56. K. Kim *et al.*, Donor cell type can influence the epigenome and differentiation potential of human
1581 induced pluripotent stem cells. *Nat Biotechnol* **29**, 1117-1119 (2011).
1582 57. K. Osafune *et al.*, Marked differences in differentiation propensity among human embryonic stem
1583 cell lines. *Nat Biotechnol* **26**, 313-315 (2008).
1584 58. G. L. Boulting *et al.*, A functionally characterized test set of human induced pluripotent stem cells.
1585 *Nat Biotechnol* **29**, 279-286 (2011).
1586 59. A. Casini, M. Storch, G. S. Baldwin, T. Ellis, Bricks and blueprints: methods and standards for
1587 DNA assembly. *Nat Rev Mol Cell Biol* **16**, 568-576 (2015).
1588 60. C. A. Hutchison, 3rd *et al.*, Design and synthesis of a minimal bacterial genome. *Science* **351**,
1589 aad6253 (2016).
1590 61. Y. Shao *et al.*, Creating a functional single-chromosome yeast. *Nature* **560**, 331-335 (2018).
1591 62. D. G. Gibson *et al.*, Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat*
1592 *Methods* **6**, 343-345 (2009).
1593 63. D. G. Gibson, H. O. Smith, C. A. Hutchison, 3rd, J. C. Venter, C. Merryman, Chemical synthesis
1594 of the mouse mitochondrial genome. *Nat Methods* **7**, 901-903 (2010).
1595 64. H. Mori, N. Yachie, A framework to efficiently describe and share reproducible DNA materials
1596 and construction protocols. *Nat Commun* **13**, 2894 (2022).
1597 65. R. H. Kutner, X. Y. Zhang, J. Reiser, Production, concentration and titration of pseudotyped HIV-
1598 1-based lentiviral vectors. *Nat Protoc* **4**, 495-505 (2009).
1599 66. C. Engler, R. Kandzia, S. Marillonnet, A one pot, one step, precision cloning method with high
1600 throughput capability. *PLoS One* **3**, e3647 (2008).
1601 67. C. Camacho *et al.*, BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).
1602 68. L. Zhao, Z. Liu, S. F. Levy, S. Wu, Bartender: a fast and accurate clustering algorithm to count
1603 barcode reads. *Bioinformatics* **34**, 739-747 (2018).
1604 69. A. Dobin *et al.*, STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2013).
1605 70. T. Stuart *et al.*, Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902 e1821
1606 (2019).
1607

Supplementary Materials for

A multi-kingdom genetic barcoding system for precise target clone isolation

Soh Ishiguro[1,15,19], Kana Ishida[2,19], Rina C. Sakata[1,19], Hideto Mori[3-5], Mamoru Takana[3], Samuel King[1], Omar Bashth[1], Minori Ichiraku[6], Nanami Masuyama[1,4,5], Ren Takimoto[1], Yusuke Kijima[1,7], Arman Adel[1], Hiromi Toyoshima[3], Motoaki Seki[3], Ju Hee Oh[8], Anne-Sophie Archambault[8], Keiji Nishida[9,10], Akihiko Kondo[9-11], Satoru Kuhara[12], Hiroyuki Aburatani[13], Ramon I. Klein Geltink[8], Yasuhiro Takashima[6], Nika Shakiba[1], and Nozomu Yachie[1,3,14,16-18]*

1. School of Biomedical Engineering, Faculty of Applied Science and Faculty of Medicine, The University of British Columbia, Vancouver, Canada.
2. Spiber Inc., Tsuruoka, Japan.
3. Synthetic Biology Division, Research Center for Advanced Science and Technology, The University of Tokyo, Tokyo, Japan.
4. Institute for Advanced Biosciences, Keio University, Tsuruoka, Japan.
5. Systems Biology Program, Graduate School of Media and Governance, Keio University, Fujisawa, Japan.
6. Center for iPS Cell Research and Application, Kyoto University, Kyoto, Japan.
7. Department of Aquatic Bioscience, Graduate School of Agricultural and Life Sciences, The University of Tokyo, Tokyo, Japan.
8. BC Children's Hospital Research Institute, Department of Pathology and Laboratory Medicine, The University of British Columbia, Vancouver, Canada.
9. Engineering Biology Research Center, Kobe University, Kobe, Japan.
10. Graduate School of Science, Technology and Innovation, Kobe University, Kobe, Japan.
11. Department of Chemical Science and Engineering, Graduate School of Engineering, Kobe University, Kobe, Japan.
12. Graduate School of Bioresource and Bioenvironmental Sciences, Faculty of Agriculture, Kyushu University, Fukuoka, Japan.
13. Genome Science Division, Research Center for Advanced Science and Technology, The University of Tokyo, Tokyo, Japan.
14. Premium Research Institute for Human Metaverse Medicine (WPI-PRIMe), Osaka University, Suita, Osaka, Japan
15. Twitter: @soh__i
16. Twitter: @nzmyachie
17. Twitter: @yachielab
18. Mastodon: sciencemastodon.com/@nzm
19. These authors contributed equally.

*Corresponding author: nozomu.yachie@ubc.ca.

Table of Contents

Fig. S1	Supplementary data for developing CloneSelect C→T
Fig. S2	Supplementary data for comparing CloneSelect C→T with CRISPRa-based circuits
Fig. S3	DSB-based circuit
Fig. S4	mCherry reporter for CloneSelect C→T
Fig. S5	Supplementary data for comparing CloneSelect A→G with CRISPRa-based circuits
Fig. S6	Supplementary data for isolating barcoded cells using CloneSelect C→T
Fig. S7	Supplementary data for developing scCloneSelect
Fig. S8	The gene expression profile cytometry cell sorting concept
Fig. S9	Supplementary data for isolating barcoded cells identified in scRNA-seq data
Fig. S10	Supplementary data for developing yeast CloneSelect
Fig. S11	Preparation of yeast CloneSelect library
Fig. S12	Supplementary data for developing bacterial CloneSelect
Fig. S13	Preparation of *E. coli* CloneSelect library
Fig. S14	Multiple gRNA-input logic gates
Table S1	Barcode count distributions observed in this study
Table S2	Oligonucleotides and plasmids used in this study
Table S3	List of high-throughput sequencing datasets
Table S4	Statistical tests performed in this study and *P*-values

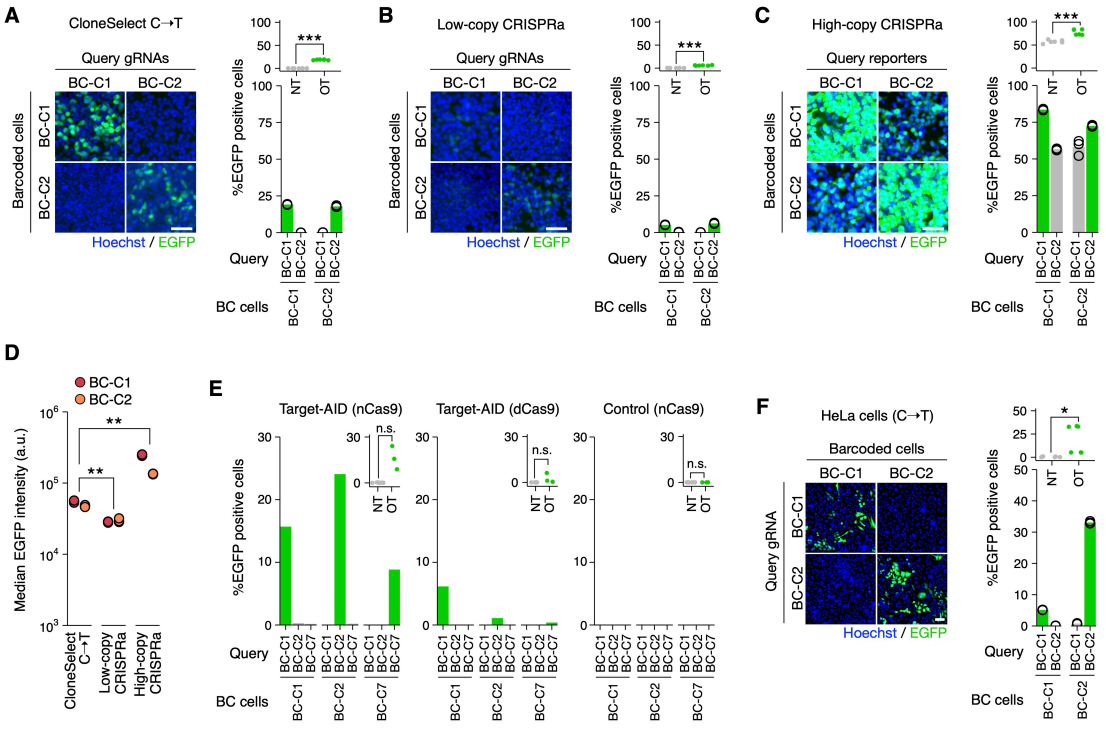

Fig. S1. Supplementary data for developing CloneSelect C→T. (**A**–**C**) Barcode-specific gRNA-dependent activation of EGFP reporters for two barcoded HEK293T strains established for each of CloneSelect C→T (**A**), low-copy CRISPRa (**B**), and high-copy CRISPRa (**C**) (n=3). Scale bar, 50 μm. Welch's t-test was performed to compare on-target (OT) and non-target (NT) activations. (**D**) Median EGFP intensities of genome editing-activated EGFP positive cells (n=3). The Mann-Whitney U test was performed to compare two groups. (**E**) Comparison of Target-AID variants and a nCas9 (D10A) control in the CloneSelect C→T reporter activation for the same set of barcode-gRNA pairs (n=1). Welch's t-test was performed to compare OT and NT activations. (**F**) Reporter activation in HeLa cells by CloneSelect C→T (n=3). Welch's t-test was performed to compare OT and NT activations. Scale bar, 80 μm. *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$.

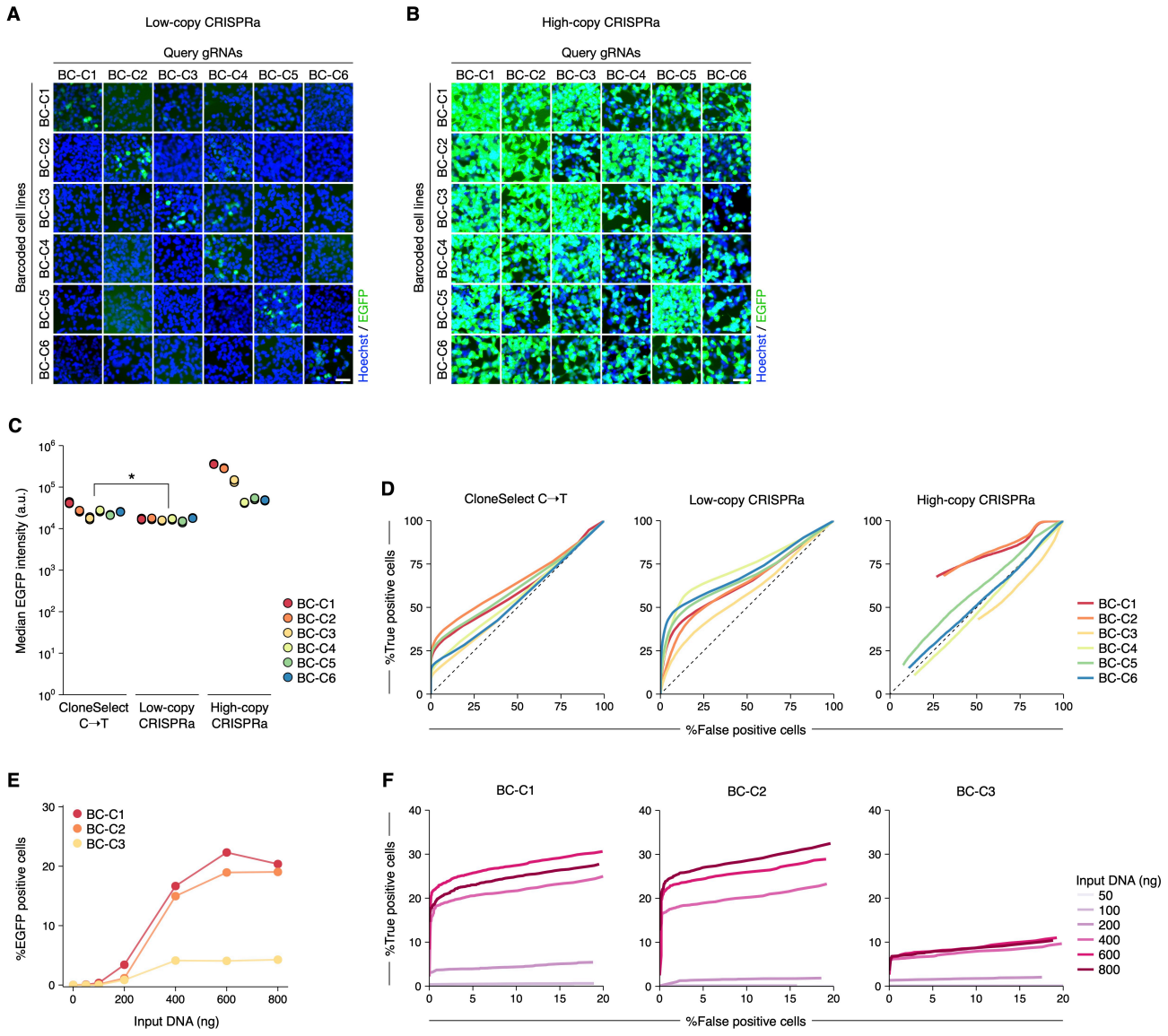

Fig. S2. Supplementary data for comparing CloneSelect C→T with CRISPRa-based circuits. (**A** and **B**) Barcode-specific gRNA-dependent reporter activation of six barcoded cell lines prepared for each of low-copy CRISPRa and high-copy CRISPRa. Scale bar, 50 μm. (**C**) Median EGFP intensities of genome editing-activated EGFP positive cells (n=3). (**D**) ROC curves along varying reporter intensity thresholds for target barcoded cells. CloneSelect C→T, low-copy CRISPRa, and high-copy CRISPRa were examined for the common set of six barcodes. The Mann-Whitney U test was performed to compare two groups (*$P < 0.05$; **$P < 0.01$; ***$P < 0.001$). (**E**) Frequencies of CloneSelect C→T reporter-activated cells obtained by transfection of different DNA amounts of barcode-targeting genome editing reagents. (**F**) ROC curve for each input DNA amount along varying reporter intensity thresholds for target barcoded cells.

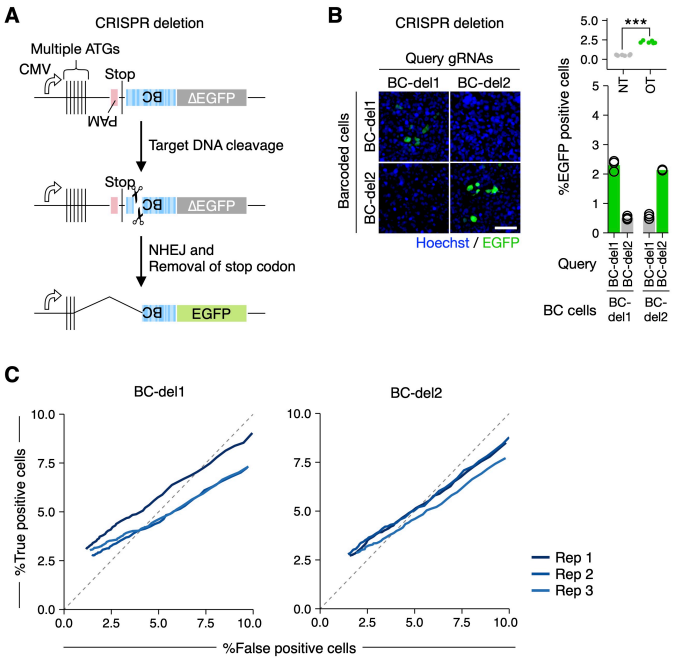

Fig. S3. DSB-based circuit. (**A**) Barcode-specific gRNA-dependent reporter activation circuit using wild-type Cas9. (**B**) Barcode-specific activation of the deletion-based reporter prepared for two barcodes (BC-del1 and BC-del2) in HEK293T cells (n=3). Welch's t-test was performed to compare on-target (OT) and non-target (NT) activations (*$P < 0.05$; **$P < 0.01$; ***$P < 0.001$). Scale bar, 50 μm. (**C**) ROC curves along varying reporter intensity thresholds for target barcoded cells (n=3).

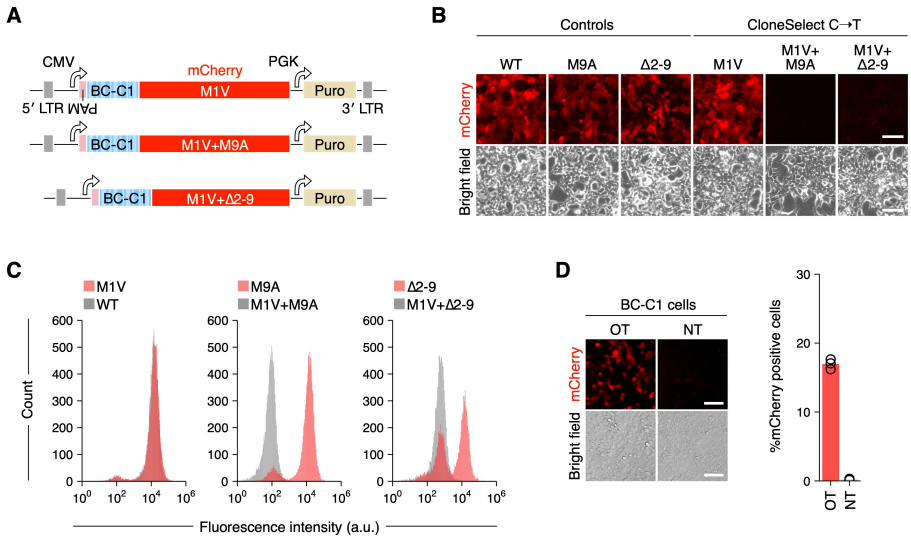

Fig. S4. mCherry reporter for CloneSelect C→T. (**A**) Different mCherry reporter variants tested to establish CloneSelect C→T. (**B** and **C**) mCherry expression from the different reporter variants with the first codon as GTG or ATG. Scale bar, 50 μm. (**D**) Activation of the M1V (GTG)+Δ2-9 mutant reporter with OT and NT gRNAs (n=3). Scale bar, 100 μm.

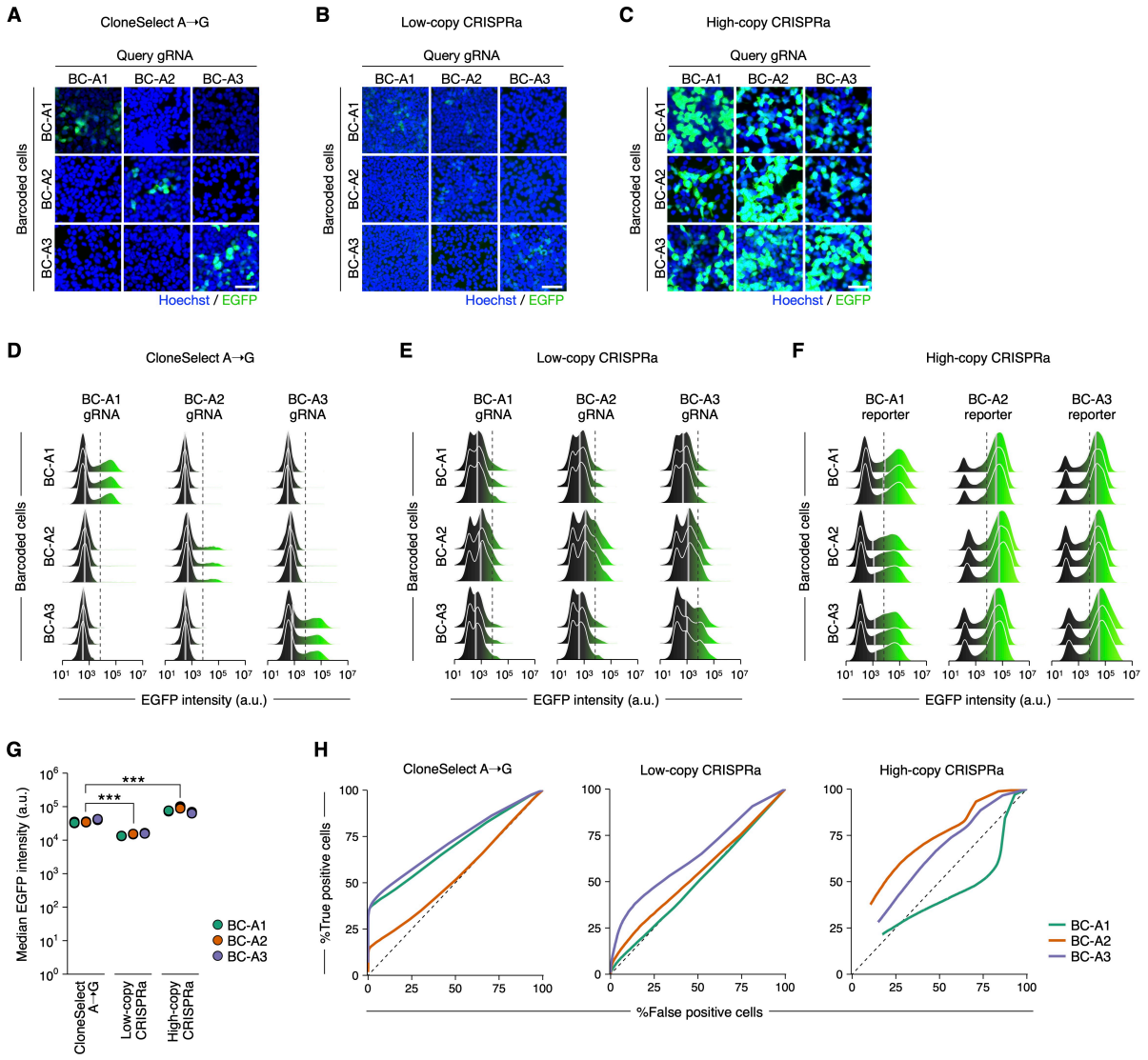

Fig. S5. Supplementary data for comparing CloneSelect A→G with CRISPRa-based circuits. (**A**–**C**) Barcode-specific gRNA-dependent reporter activation of three barcoded cell lines prepared for each of CloneSelect A→G, low-copy CRISPRa, and high-copy CRISPRa. Scale bar, 50 μm. (**D**–**F**) Flow cytometry analysis of single-cell EGFP activation levels. (**G**) Median EGFP intensities of genome editing-activated EGFP positive cells (n=3). The Mann-Whitney U test was performed to compare two groups (*$P < 0.05$; **$P < 0.01$; ***$P < 0.001$). (**H**) ROC curves along varying reporter intensity thresholds for target barcoded cells (n=3).

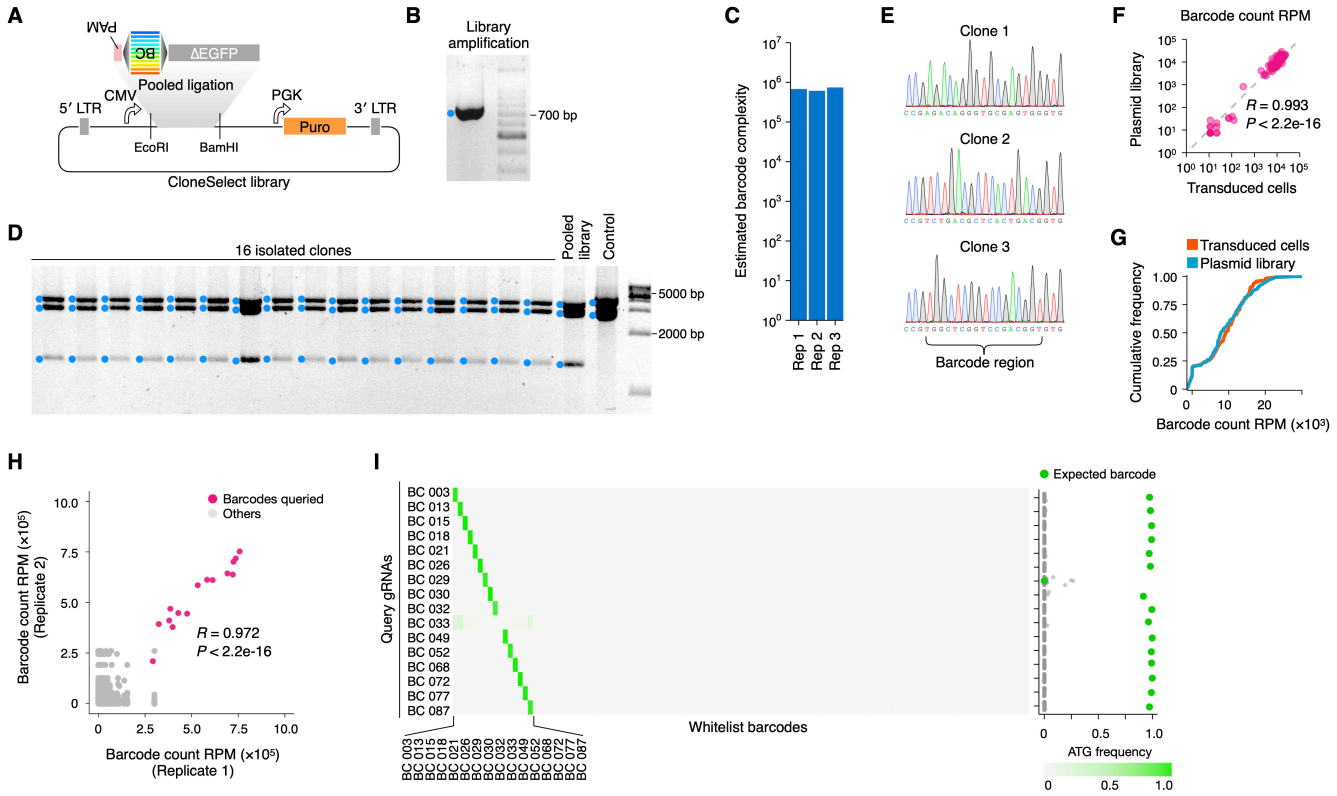

Fig. S6. Supplementary data for isolating barcoded cells using CloneSelect C→T. (**A**) Schematic diagram for the barcode library construction. The ΔEGFP fragment (no start codon) was amplified by PCR using pooled forward primers encoding the PAM followed by semi-random barcode sequences encoding GTG and a common reverse primer, and enzymatically digested and ligated to the lentiviral backbone. (**B**) The insert PCR fragments. (**C**) Estimated complexities of the generated plasmid pools in triplicate. (**D**) Library QC by colony isolation and restriction digest by BsrGI, ClaI and PvuI. (**E**) Sanger sequencing of the barcode region of the colony isolates. (**F** and **G**) Barcode distribution in the lentiviral plasmid DNA pool and that in the cell population transduced using the same plasmid DNA pool. (**H**) Barcode distribution in the EGFP-positive cell population obtained by cell sorting after barcode-specific activation by a target gRNA. The results of the 16 independent samples were combined after read count normalization for each sample. (**I**) Frequency of the GTG→ATG mutation observed for each barcode after sorting of the reporter positive cells. Each row represents the GTG→ATG mutation frequency profile obtained for each target isolation attempt.

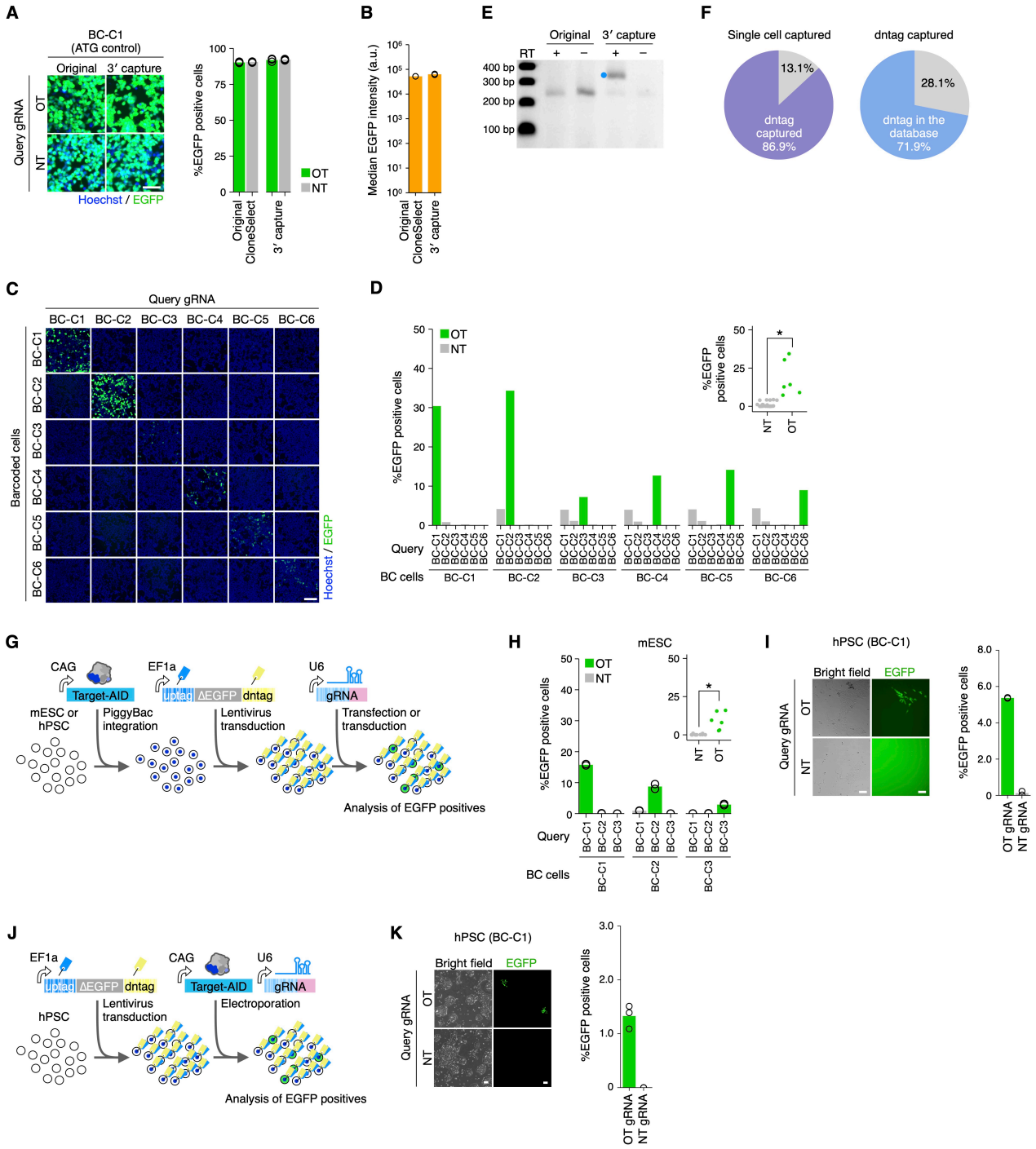

Fig. S7. Supplementary data for developing scCloneSelect. (**A**) EGFP-positive control expressions for the original CloneSelect C→T and scCloneSelect in HEK293T cells with the same genome editing conditions tested for the respective reporters (n=3). Scale bar, 50 μm. (**B**) Median EGFP intensities of base editing-activated EGFP positive cells (n=3). (**C** and **D**) Barcode-specific gRNA-dependent reporter activation of six barcoded cell lines by scCloneSelect (n=1). Welch's t-test was performed to compare on-target (OT) and non-target (NT) activations (*P < 0.05; **P < 0.01; ***P < 0.001). Scale bar, 50 μm. (**E**) RT-PCR of the scCloneSelect dntags in HEK293T. (**F**) Fraction of mESC single-cell transcriptome profiles (Drop-seq) that contained dntags and fraction of dntags reported in the uptag-dntag combination reference database. (**G**) Schematic representation of a scCloneSelect reporter activation assay where Target-AID was stably introduced to the cell population prior to barcoding and gRNA-dependent reporter activation. (**H** and **I**) gRNA-dependent reporter activation of target barcoded mESCs and CA1 hPSCs by scCloneSelect (n=2). Target-AID was stably integrated prior to the barcoding. Targeting

gRNAs were delivered by transfection. Welch's t-test was performed to compare OT and NT activations (*$P < 0.05$; **$P < 0.01$; ***$P < 0.001$). Scale bar, 100 μm. (**J**) Schematic representation of a scCloneSelect reporter activation assay where the target gRNA and Target-AID were electroporated together to the barcoded cell population. (**K**) gRNA-dependent reporter activation of barcoded H1 hPSCs by scCloneSelect (n=2). Targeting gRNA and Target-AID were electroporated together. Scale bar, 100 μm.

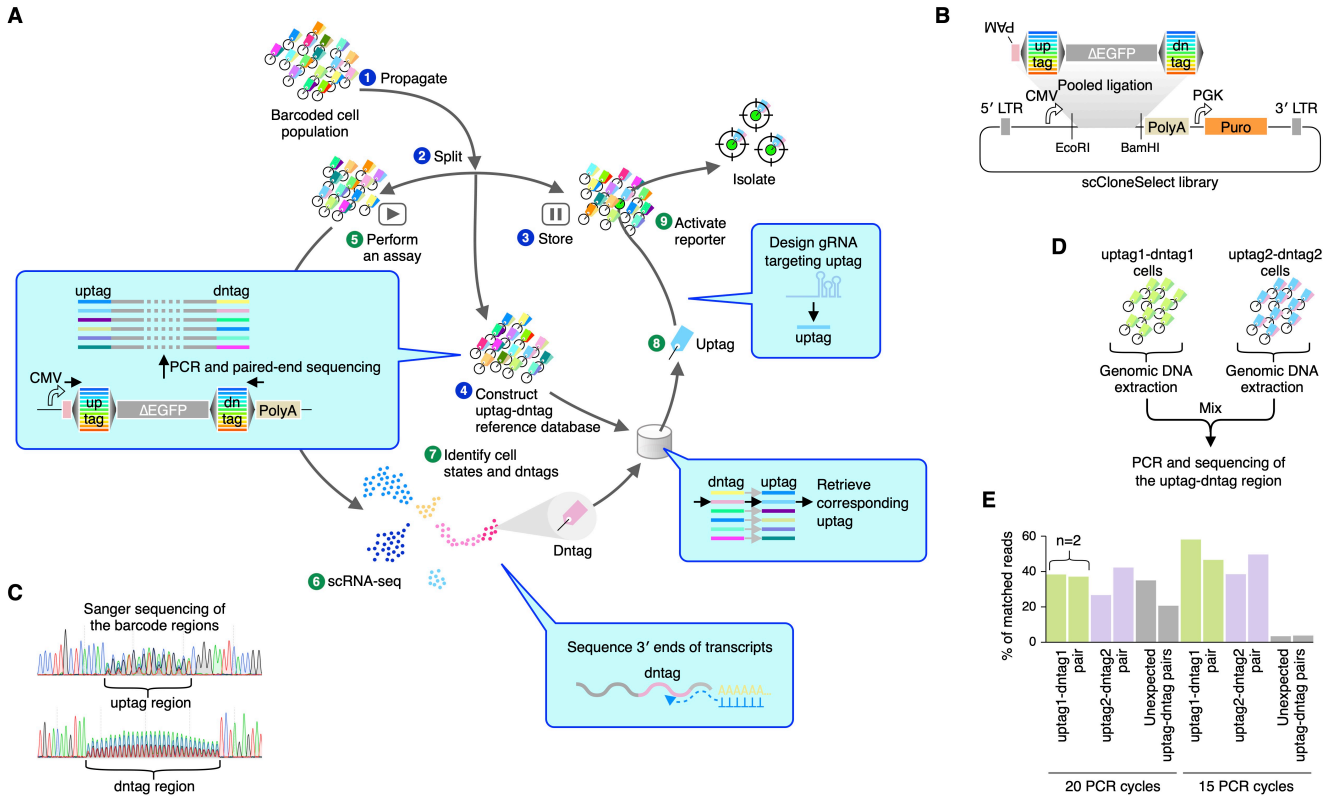

Fig. S8. The gene expression profile cytometry cell sorting concept. (**A**) Schematic diagram of a scCloneSelect workflow. (**B**) Preparation of barcode library for scCloneSelect. (**C**) Sanger sequencing results of the uptag and dntag regions of the constructed plasmid pool. (**D**) Mixing of genomic DNA samples from two barcoded cell lines to optimize the PCR protocol to create a library for identifying the uptag-dntag combination reference database. (**E**) Fraction of unexpected chimeric PCR recombination products in the sequencing reads.

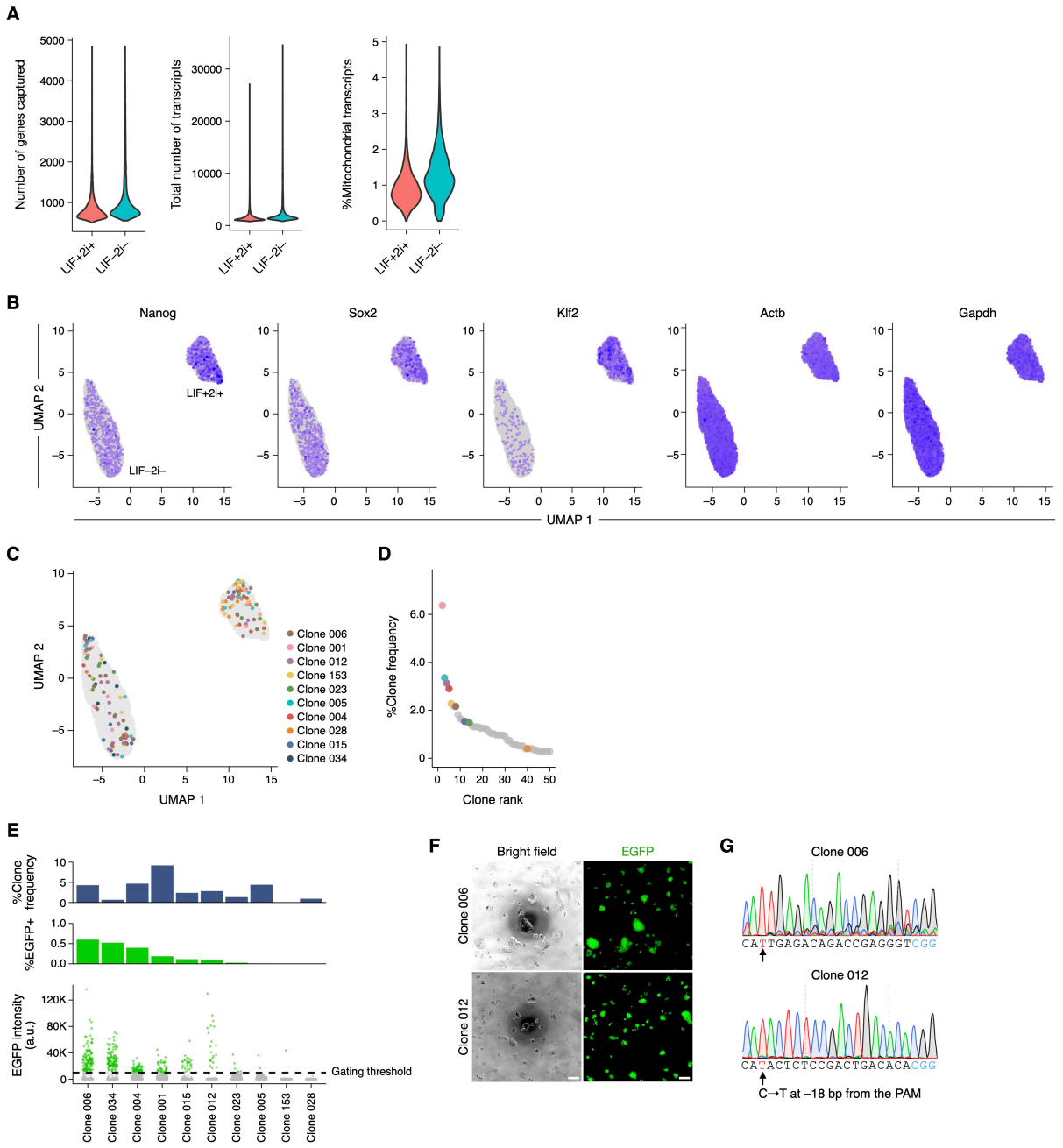

Fig. S9. Supplementary data for isolating barcoded cells identified in scRNA-seq data. (**A**) QC of scRNA-seq datasets obtained for the barcoded mESC population cultured with LIF and 2i and that cultured without LIF or 2i. (**B**) Single-cell expression patterns of key genes. (**C**) Distribution of cells in a two-dimensional UMAP space for all the clones targeted for isolation. (**D**) Abundances of barcoded cell clones in the mESC population. The data was generated based on dntags identified in the scRNA-seq dataset with no reamplification of the dntag reads. (**E**) gRNA-dependent activation of the reporter for each of the target clones in the initial mESC population. (**F**) Culturing of single isolated cells. Scale bar, 100 μm. (**G**) Sanger sequencing of the barcodes for the isolated target clones expanded after clone labeling and single-cell isolation.

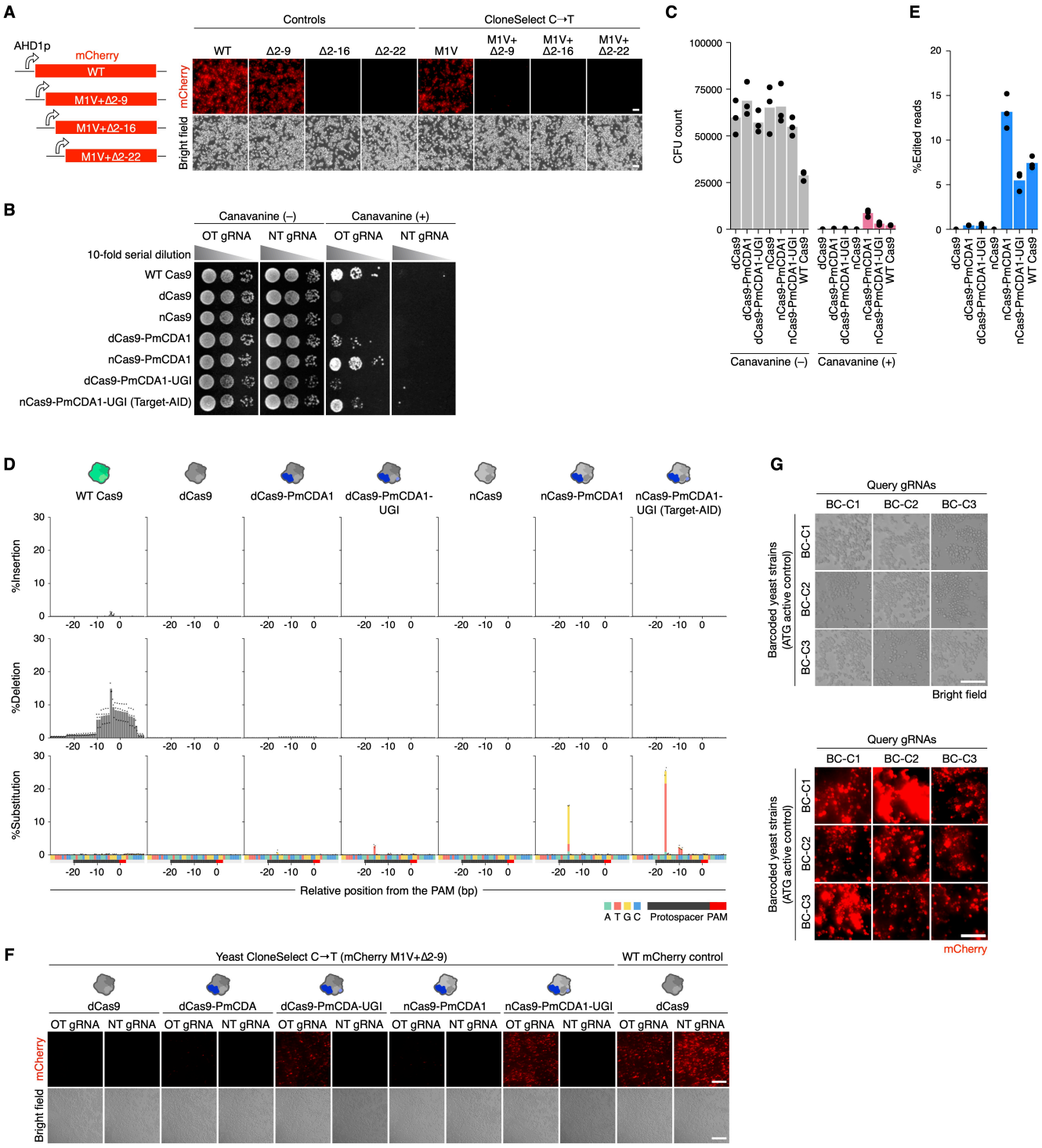

Fig. S10. Supplementary data for developing yeast CloneSelect. (**A**) Different mCherry reporter variants tested to establish CloneSelect C→T. The different reporter variants were tested with the first codon as GTG or ATG. Scale bar, 100 μm. (**B**) Canavanine resistance assays for different CRISPR genome editing enzymes with a gRNA targeting CAN1 gene and a control NT gRNA. For each experiment, cell concentration was normalized to 1.0 $OD_{595\ nm}$ and serially diluted with 10-fold increments for spotting. (**C**) Estimated CFU counts for the same assay in (**B**). (**D**) Genome editing outcomes observed by amplicon sequencing. Frequencies of mutation patterns observed across the target sequence region are shown for the same assay in (**B**). (**E**) Genome editing frequencies at the target CAN1 locus estimated by amplicon sequencing for the different enzymes. (**F**) Activation of the mCherry M1V (GTG)+Δ2-9 mutant reporter by OT and NT gRNAs. Scale bar, 200 μm. (**G**) mCherry-positive control

expressions for yeast CloneSelect. Yeast cells having the positive control reporters with three different barcodes (BC-C1, BC-C2, and BC-C3) were each treated by Target-AID and three different targeting gRNAs. Scale bar, 25 μm.

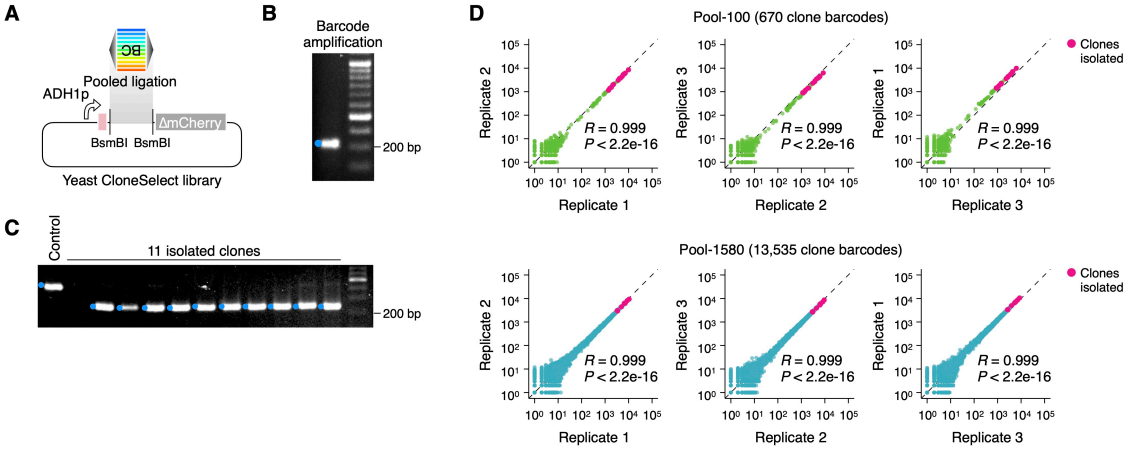

Fig. S11. Preparation of yeast CloneSelect library. (**A**) Schematic diagram for the barcode library construction. The barcode fragment pool was prepared by PCR using a common primer pair amplifying a template DNA pool encoding the PAM, WSNS semi-random repeat, and the mutated start codon GTG. The PCR product was digested and ligated to a backbone plasmid. (**B**) The insert PCR fragments (Pool-100). (**C**) Library QC by colony isolation and PCR amplification of the barcode insert (Pool-100). (**D**) Barcode abundance distribution (read per million) in the constructed barcode library pool. The sequencing library was prepared and analyzed in triplicate (n=3).

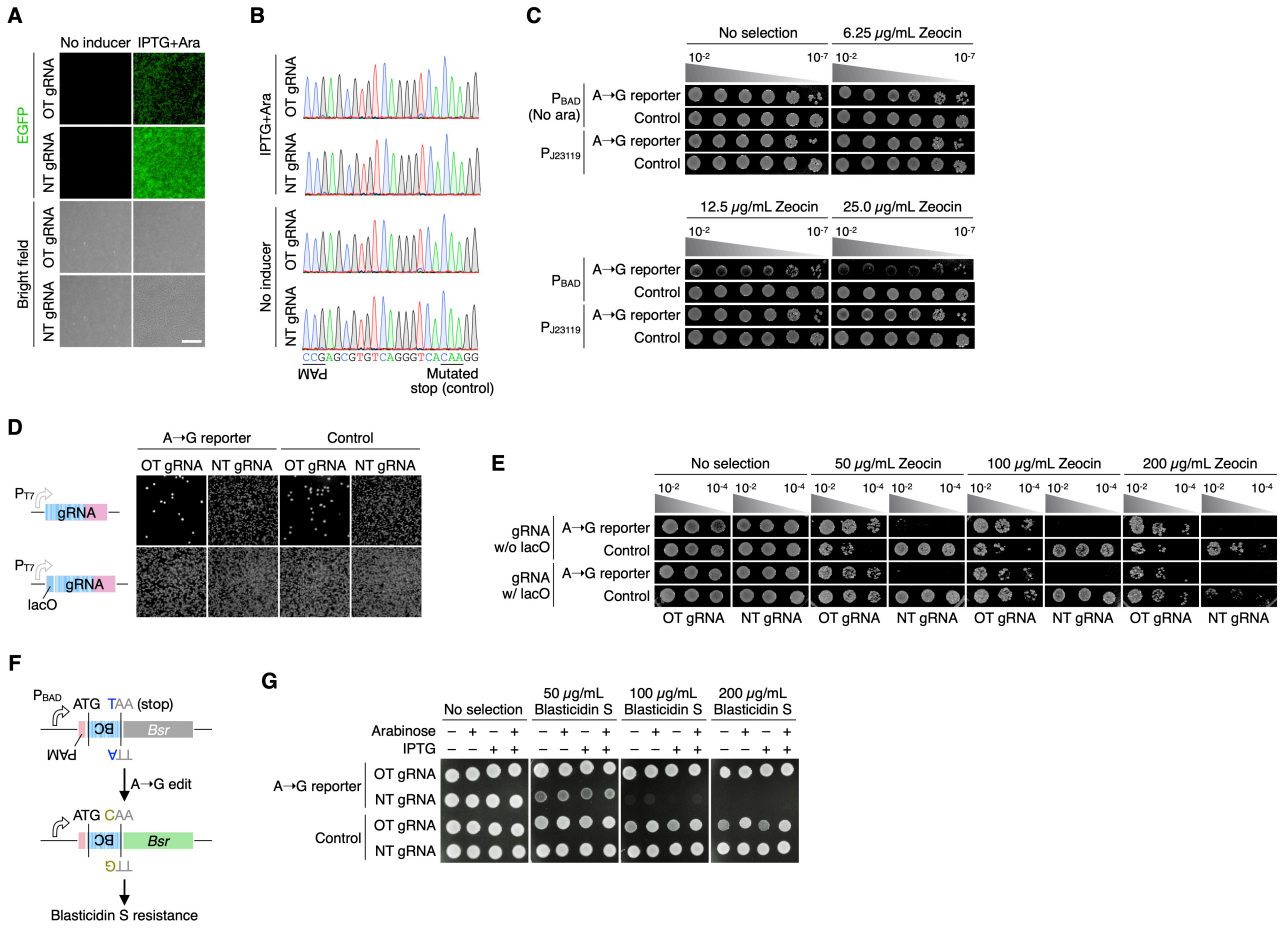

Fig. S12. Supplementary data for developing bacterial CloneSelect. (**A**) Activities of the positive control EGFP reporter. ABE and gRNA expression were controlled by an IPTG-inducible promoter, and the EGFP reporter expression was controlled by an arabinose-inducible promoter. (**B**) Base editing outcomes of the positive control reporters analyzed by Sanger sequencing. (**C**) Testing of Zeocin resistances conferred by two promoters expressing a Zeocin resistance gene with and without the upstream stop codon to block the selective marker translation. Each cell sample concentration was first adjusted to 0.1 $OD_{595\ nm}$ and serially diluted with 10-fold increments for spotting 5 μL. (**D**) Testing of cell viability under a non-selective condition for a constitutively active T7 promoter and the IPTG-inducible promoter to express the gRNA. OT and NT gRNAs were tested for the gRNA-dependent EGFP reporter and the positive control EGFP reporter. ABE was expressed under the IPTG-inducible promoter without IPTG provided. (**E**) gRNA-dependent Zeocin resistance reporter activation tested for the IPTG-inducible promoters with and without IPTG. (**F**) Bacterial CloneSelect using the Blasticidin resistance gene. Each cell sample concentration was first adjusted to 0.1 $OD_{595\ nm}$ and serially diluted with 10-fold increments for spotting 5 μL. (**G**) gRNA-dependent Blasticidin-resistance reporter activation tested for different inducer conditions and different Blasticidin concentrations. Each cell sample concentration was adjusted to 0.1 $OD_{595\ nm}$ for spotting 5 μL.

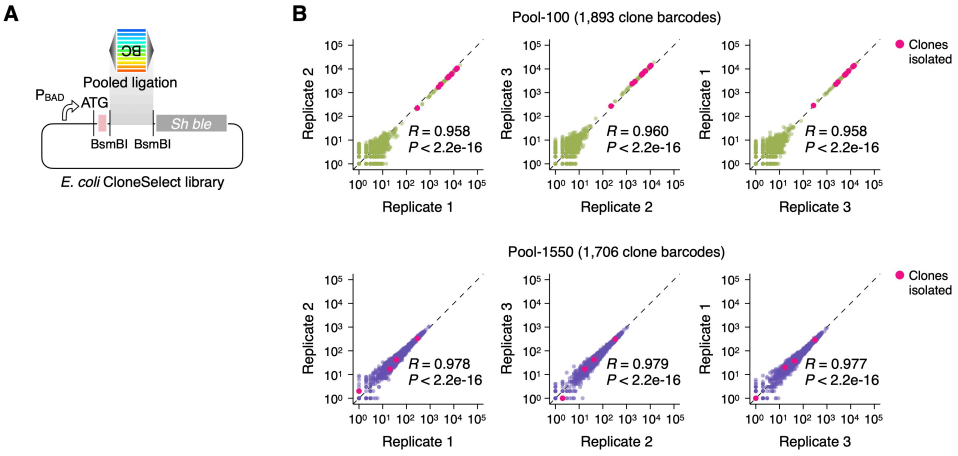

Fig. S13. Preparation of *E. coli* CloneSelect library. (**A**) Schematic diagram for the barcode library construction. The barcode fragment pool was prepared by PCR using a common primer pair amplifying a template DNA pool encoding the PAM, VNN repeat sequence, and the start codon TAA. The PCR product was digested and ligated to a backbone plasmid. (**B**) Barcode abundance distribution (read per million) in the constructed barcode library pool. The sequencing library was prepared and analyzed in triplicate (n=3).

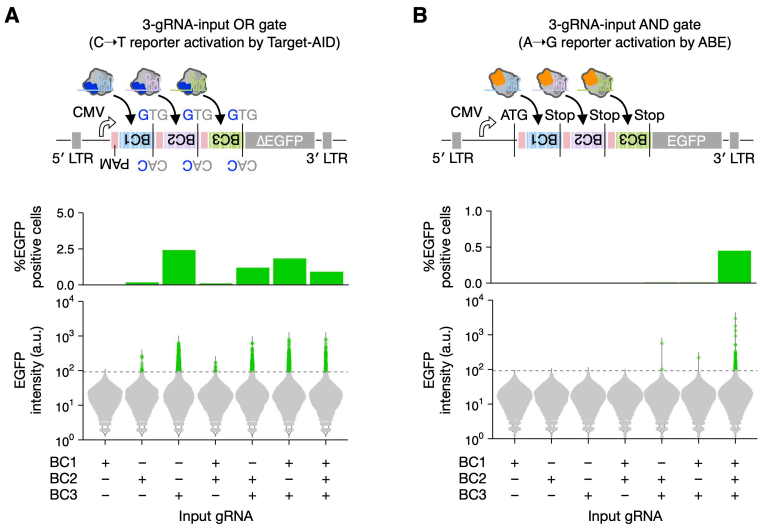

Fig. S14. Multiple gRNA-input logic gates. (**A**) Three-gRNA-input OR gate with CloneSelect C→T that is designed to confer the EGFP reporter expression by any of the three barcode-specific gRNA-dependent GTG→ATG mutations. (**B**) Three-gRNA-input AND gate with CloneSelect A→G that is designed to confer the EGFP reporter expression when all three barcode-specific gRNA-dependent TAA→CAA mutations are provided.