Title

Proof-of-Concept of A DNA-Based Recording System for High-Throughput Functional Gene Screening

Authors

Shoichi Kato[1,2,*], Atsushi Ikemoto[1,2], Jun Isayama[1,2], Tetsuya Takimoto[3], Hideyuki Saya[3], Ken-ichi Hamada[1,2]

[1]JSR Corporation, Tokyo, Japan

[2]JSR-Keio University Medical and Chemical Innovation Center, Keio University School of Medicine, Tokyo, Japan

[3]Oncology Innovation Center, Fujita Health University, Toyoake, Aichi, Japan

[*]Corresponding author: Shouichi_Katou@jsr.co.jp, shoichikato.b@gmail.com

Abstract

Genetic functional screening technologies which identify causative genes are essential for advancing life sciences and improving drug discovery outcomes. Traditional array-based screening methods, which require significant cell numbers, face limitations when working with samples that have low proliferation capacity. While pooled library methods such as CRISPR screens can be solutions to these experimental efficiency challenges, there is still room for improvement in terms of cost and convenience. In response to these challenges, we developed PiER (Perturbation-induced intracellular events recorder) technology. PiER facilitates gene perturbation and intracellular signal detection through a novel system that integrates three DNA domains. The Perturbation domain induces gene-specific disturbances, the Response domain expresses an enzyme upon desired cellular signals, and the Memory domain records perturbation history by altering its DNA sequence via the expressed enzyme. To demonstrate PiER's potential, we designed a vector which has a Response domain that detects WNT pathway activation. Transfecting HEK293 cells, we observed dose-dependent responses to WNT pathway activation using fluorescence microscopy and quantitative Polymerase

Chain Reaction (qPCR), which confirmed successful intracellular event recording in the Memory domain. Further experiments with lentiviral PiER vectors containing a pooled shRNA library revealed the system's capability to conduct high-throughput screening by analyzing perturbations and their effects within individual cells. PiER technology significantly enhances screening capabilities by offering a versatile and scalable approach that can be deployed without prior cell modification and single-cell isolation. Its high throughput, combined withrequiring minimal effort, presents a significant advancement for genomic research and drug target discovery.

Introduction

Gene function screening technologies play a crucial role in life science research and drug discovery processes. Comprehensive analyses have demonstrated that the identification of causative genes significantly enhances the probability of successful drug development[1–3]. Furthermore, recent advancements such as the exploitation of synthetic lethality relationships for cancer therapeutics[4,5] and the emergence of novel modalities including gene therapy and Proteolysis Targeting Chimeras (PROTACs)[6,7] have highlighted the importance of identifying previously unexplored molecular targets rather than focusing on established druggable targets. To identify such causative target genes, comprehensive screening approaches are effective.

Comprehensive target screening is frequently performed as an effective means of identifying target genes in diseases. Although conventional array-based target screening offers high experimental accuracy, each target must be analyzed under separate experimental conditions, resulting in substantial labor, cost, and the need for large quantities of experimental materials when scaling up the number of target genes. To overcome these limitations, pooled library technologies have been developed and refined over the past decade[8–13]. These methodologies enable efficient screening by combining vectors containing diverse genetic perturbations and introducing them simultaneously

into cellular populations within a single culture environment. In cancer research, negative selection screening utilizing sequencing read counts of perturbation sequences that reflect cell survival numbers has been widely implemented to identify essential genes and potential therapeutic vulnerabilities[14–17]. Conversely, in research fields focused on cellular processes beyond survival, methodologies combining pooled genetic perturbations with high-throughput gene expression analysis have been employed to identify molecular targets that modulate intracellular signaling pathways and complex cellular phenotypes.

Currently available gene expression analysis-based pooled library screening technologies for mammalian cells include single cell RNA sequencing (scRNA seq)-based methods[18–20], cell sorting-based method[21], and barcode sequence expression-based methods[22,23]. The scRNA seq-based methods (Perturb-seq[18], CROP-seq[19], and CRISP-seq[20]) are limited by the number of perturbations that can be examined simultaneously and are costly. Sorting-based methods allow for the exploration of responsive cells based on fluorescent markers; however, they require cell sorters and careful threshold settings. The representative barcode sequence expression-based method is CiBER-seq[22,23]. This method is a groundbreaking method that enables comprehensive gene perturbation using pooled libraries and analysis of reporter expression via next-generation sequencing (NGS). This technique is characterized by the introduction of large gene cassettes utilizing landing pads for simultaneously expressing reporters and controls; therefore, it requires prior cellular manipulation. Therefore, although gene expression analysis-based pooled library screening technologies are revolutionary, there remains room for further improvement.

To address the challenge of achieving both simplicity and comprehensiveness, we have developed the PiER (Perturbation-induced Intracellular Events Recorder) technology. This novel approach combines gene perturbation with intracellular signal memorization into DNA. The aim of this study is to design and demonstrate the PiER technology and conduct functional screening using the WNT

signaling pathway as a model system. In this paper, we first describe the design of the PiER system and verification of its basic functions, followed by the results of screening experiments using pooled shRNA libraries. Finally, we analyze and discuss the results, examining the effectiveness of PiER technology and its future prospects.

Results

Design and Basic Structure of the PiER Vector System

The PiER system comprises three distinct DNA sequence domains delivered by vectors (Fig. 1a). The first is the Perturbation domain, composed of an expression system that can induce gene-specific perturbations in cells. The second is the Response domain, which expresses an enzyme with DNA editing activity when desired signals are induced in the cell. The third is the Memory domain, which is accessed by the DNA editing enzyme, resulting in changes to the DNA sequence. In summary, when perturbations from the Perturbation domain activate desired signals in the cell, the history is recorded as sequence changes in the Memory domain through gene expression from the Response domain. By pooling perturbation domains, various perturbations can be expressed in separate cells in a single gene introduction experiment. Consequently, their effects on signaling can be simultaneously stored in DNA (Fig. 1b). Importantly, the Perturbation and Memory domains are positioned in close proximity within the same molecule, enabling analysis while preserving the correspondence between perturbations that occur within a single cell and their effects that are recorded intracellularly. By creating amplicons of the perturbation-memory region and analyzing them using NGS, the perturbations that occurred in single cells and their effects on signaling can be obtained as count data for each type of Perturbation domain. This technology does not require single-cell isolation to examine the effects of perturbations in individual cells, enabling analysis with lower cost and less effort.

Construction of the Pilot Vector

To evaluate the concept of PiER technology, we designed a pilot vector system comprising both Response and Memory domains. This system employs two plasmid vectors: one containing the Response domain (pMEM_mKate2-EGFP) and the other the Memory domain (pRESP_TCFLEF) (Fig. 2a). For the Response domain, we selected the TCF-LEF response element (RE), a well-established transcriptional enhancer activated by the WNT signaling pathway, based on previous studies that utilized similar enhancers to record transcriptional activity in DNA sequences[24]. The TCF-LEF RE was followed by a minimal promoter (minP) sequence and the Cre recombinase open reading frame. For the Memory domain, we implemented a FLEx-ON switch system[25] that functions as a permanent genetic recorder of Cre activity. This unit includes an inverted mKate2 sequence flanked by two different LoxP variants. Upon Cre-mediated recombination, the mKate2 ORF flips and is subsequently locked in the correct orientation due to the mismatch of internal sequences in the Lox variants. This unit was placed downstream of the EF1α promoter, enabling red fluorescent protein expression following Cre-mediated recombination. The Memory domain vector also contains an EGFP gene driven by a strong promoter, allowing identification of transfected cells through green fluorescence.

Dose-Response Evaluation of the Signal Induced Memory System

To assess the dose-response of the pilot vector system, we transfected HEK293 cells with both vectors using lipofection. The transfected cells were then treated with varying concentrations of CHIR-99021 (0-16 μM), a small molecule activator for canonical WNT signaling[26]. After 24 hours of treatment, we analyzed the transfectants using fluorescence microscopy and DNA extraction. Fluorescence imaging revealed a dose-dependent increase in red fluorescent cells in response to

CHIR-99021 (Fig. 2b). Quantitative image analysis using CellProfiler software confirmed this observation, showing a progressive increase in the percentage of red fluorescent cells with increasing CHIR-99021 concentration (Fig. 2c). Furthermore, we performed qPCR analysis to quantify the flipped sequence in the Memory domain. Primers were designed to target the flipped sequence, and the results were normalized to qPCR data from primers targeting the vector backbone. This analysis showed corresponding CHIR-99021-dependent signal increase (Fig. 2d), corroborating the fluorescence microscopy results. These results suggest that our system effectively records intracellular signaling events in the Memory domain sequences, with both visual and molecular readouts demonstrating dose-dependent responses to WNT pathway activation.

Development of Pooled shRNA Library System

In order to utilize the PiER system for comprehensive perturbation screening, we developed a pooled shRNA lentivirus library system. This proof-of-concept experiment involved the creation of two vectors. The first vector LV-PERT-MEM combines both the Perturbation domain (shRNA) and the Memory domain. We based this vector on the Human Elite Gene Pooled shRNA Library manufactured by VectorBuilder (Cat. LVM(Lib190505-1037bjk)), which contains 12,471 shRNAs targeting 2,161 genes (Supplementary Table 1). To incorporate the Memory domain into this vector, we inserted two LoxP sequences in the same direction, approximately 60 bp downstream of the shRNA terminator sequence. This design aims to enable truncation of the Memory domain through Cre recombinase-catalyzed recombination. The second vector LV-RESP_TCFLEF was designed to contain the Response domain, which consists of the TCF-LEF response element (RE) and Cre recombinase. This design maintains consistency with our previous experiments using the TCF-LEF RE for detecting WNT pathway activation. When these two vectors are transduced into a cell, the system operates as follows: If the shRNA from the first vector causes a change in WNT signaling, it

will affect the expression of Cre recombinase from the second vector. This, in turn, will alter the truncation rate of the Memory domain in the first vector (Fig.3a). Thus, the degree of Memory domain truncation serves as a readout for the effect of each shRNA on WNT pathway activity. In this experimental system, by controlling the transduction of LV-PERT-MEM at a low ratio relative to the number of target cells, we can expect that most transduced cells will receive only a single perturbation molecule. This dual-vector system allows for simultaneous perturbation (via shRNA) and recording (via the Memory domain) in a single cell, enabling high-throughput screening of gene function in the WNT signaling pathway.

Experimental Design and Sequencing with PiER

The proof-of-concept experiments were designed as follows: 1) Transduction of the dual vectors into HEK293 cells followed by cultivation for 3 days, 2) Split cultured cells into two experimental conditions: CHIR-99021 treatment and control, 3) After 24 hours of treatment, harvest cells (Fig.3b). We conducted the drug treatment to evaluate this system under both WNT signal upregulated and normal conditions. We amplified the region of DNA containing both shRNA and flanking Memory domain using Polymerase Chain Reaction (PCR) from the extracted DNA of the harvested cells. The sequencing depth exceeded 1.8 million reads for all conditions, as detailed in Supplementary Table 2.

Detection and Quantitative Analysis of Memory Domain Changes

Through BLASTN[27,28] analysis, we determined the type of shRNA and the length of the Memory domain in each read (Supplementary Fig. 1). The distribution of Memory domain length exhibited a distinct bimodal pattern with peaks at approximately 320 and 250 bp, reflecting the presence of both original and truncated Memory domains. The theoretical lengths of the original and truncated

Memory domains are 322 bp and 256 bp, respectively, confirming that we could detect the Memory domain properly. Notably, CHIR-99021 treatment resulted in a slight increase in the shorter peak (Fig. 3c), suggesting enhanced truncation of the Memory domain The pattern of the histogram was similar between two different biological replicates (Fig. 3c). These results suggest that CHIR-99021 treatment affected WNT signaling in both independent biological replicate batches. Following the length-based classification of reads, we aggregated the counts of truncated and non-truncated reads for each shRNA (Supplementary Fig. 1). This step allowed us to quantify the extent of Memory domain alteration associated with each specific shRNA.

Quantitative Analysis of shRNA Effects and Reproducibility Validation

Using the count data obtained from the above analysis pipeline, we compared the number of memorized (truncated) reads, original reads, and total reads for each shRNA between two different biological batches. Our analysis showed high reproducibility of read number between two different batches (Supplementary Fig. 2). For subsequent analysis, we filtered shRNAs by selecting those with at least 100 reads in both biological batches (1 and 2) for each treatment condition (DMSO and CHIR99021) separately. To further evaluate reproducibility and drug treatment effects on the WNT detection system, we calculated the fold change ratio of memorized (truncated) sequence counts to original sequence counts for each shRNA. The fold change values were calculated using data obtained from independent biological replicates and presented as scatter plots to visualize the reproducibility of the observed effects (Fig. 3d). This plot illustrates that, despite a degree of inter-experimental variability, several shRNAs consistently exhibited high fold changes across both biological replicates.

By comparing the frequency of intact versus truncated memory reads across two independent biological replicates, we identified more than 20 shRNAs whose truncation frequency was

significantly altered (FDR-adjusted p-value < 0.05) (Fig. 3e). These results demonstrated that specific shRNAs can consistently alter the frequencies of Memory domain truncation. Notably, CHIR-99021 treatment increased the number of significantly altered shRNAs compared to DMSO control (Fig. 3d, e), indicating that the PiER system's detection sensitivity is responsive to changes in baseline WNT signaling activity.

To validate the reproducibility of our analysis, we performed a randomization test (Fig. 3f). This comparison between actual data (Batch1) and randomized data (randomized Batch2) serves to distinguish genuine biological effects from random noise. Importantly, edgeR analysis with permuted Batch 2 data showed few significantly altered shRNAs, unlike the normal dataset (Fig. 3f). This result suggests that significant shRNAs selected by edgeR were based on biological reproducibility.

Evaluation of WNT Pathway-Related Gene Detection Efficiency using the PiER System

To evaluate whether the PiER system effectively screens for WNT-related targets as intended, shRNAs obtained from the screening were divided into two groups: those targeting WNT-related genes (Supplementary Table 3) and those targeting other genes. Since read number-based filtering of shRNAs was applied separately to the DMSO and CHIR99021 conditions, the sets of shRNAs used for subsequent analysis slightly differed between the two conditions. Therefore, we quantified the proportion of WNT-related shRNAs within each filtered shRNA set to assess potential bias. Approximately 3.6% of all shRNAs in both treatment conditions targeted WNT pathway-related genes, indicating no observable bias between the conditions (Fig. 4a). Subsequent analysis focused on the distribution of log fold change (logFC) values and *p*-values for each shRNA. We did not observe significant differences in logFC distributions between WNT-related and non-WNT-related genes in either condition (Fig. 4b). However, examination of *p*-value distributions revealed a higher

proportion of shRNAs with p-values below 0.05 among those targeting WNT-related genes in the CHIR99021-treated condition (Fig. 4c, d, Table 1, Fisher's exact test, adjusted *p*-value = 0.03). Importantly, this tendency was not observed in the randomized dataset (Fig. 4d, Table 1). These results suggest that the PiER system effectively identifies shRNAs targeting WNT-related genes.

Discussion

In this study, we report the design and demonstration of the PiER system, which combines a reporter system with DNA event recording to evaluate target gene function efficiently and simply. A unique aspect of this technology is that it enables parallel screening without requiring single-cell isolation, cell sorting, or survival selection, relying only on vector transduction and subsequent DNA analysis of the sample. Compared to pooled library screening methods based on scRNA-seq, the PiER system can explore many targets simultaneously, albeit at the expense of detecting diverse biological responses. In the future, it may be possible to measure various biological responses by utilizing not only Cre recombinase but also Cre variant molecules and other DNA sequence editing enzymes as recording devices. Therefore, this method may gain significance in terms of the number of simultaneously measurable biological responses compared to methods like cell-sorting-based methods, which use fluorescent reporters and pooled libraries, followed by sorting and sequencing of perturbations.

Another unique aspect of this technology is that we constructed a screening system using mammalian cells without the need to introduce tool genes or landing pads into the target cells beforehand. The ability to perform screening without prior treatment gives this technology the potential to acquire data from a wide range of samples, including primary cultured cells and organoids within a short period.

In this study, we demonstrated that the amount of Cre-mediated DNA recombination increased in a

dose-dependent manner in a system where Cre expression increased in response to signal inducers, using a transient expression system with plasmid DNA (Fig. 1b, c, d). On the other hand, in the library screening system using lentivirus-mediated expression, a response to signal inducers was observed (Fig. 2c), but the magnitude of change was limited compared to the plasmid system. The differences between these two systems include the number of molecules introduced into a single cell and whether the introduced nucleic acids are inserted into chromosomes. These factors may contribute to the differences in response.

The lentiviral PiER screening method developed in this study demonstrated reproducibility, as evidenced by the identification of overlapping hit shRNAs across biological replicates. Nevertheless, comparative analysis of shRNA read count fold changes between biological replicates (Fig. 3d) revealed substantial variability for a subset of shRNAs. Therefore, there remains significant potential for further optimization of the experimental protocol to enhance the reproducibility and robustness of the screening platform. Factors contributing to this variability may include the restriction of the recording region to a single site and increased randomness caused by background Cre recombinase expression.

A significant challenge for future development of the PiER system is addressing highly heterogeneous samples such as in-vivo models. The current PiER screening cannot distinguish which cellular responses originate from specific cell types within heterogeneous populations. To overcome this limitation, we need to develop mechanisms that either record cell type information separately or restrict PiER functionality to targeted cells. One promising approach is using AAV vectors with serotype-specific targeting[29] for cell population selectivity.

In conclusion, we believe that the PiER technology is characterized by its high throughput, low implementation cost, and broad applicability. We expect this technology to open new possibilities in various aspects of life science research.

Materials and Methods

Materials

Human embryonic kidney 293 (HEK293)[30] cells were obtained from the Japanese Cancer Research Resources Bank.

Cell Culture

HEK293 cells were cultured in Eagle's minimum essential medium supplemented with L-Glutamine, Phenol Red, Sodium Pyruvate, Non-essential Amino Acids, and 1,500 mg/L Sodium Bicarbonate (cat. 055-08975, FUJIFILM Wako Pure Chemical Corporation, Osaka, Japan). The medium was further supplemented with 10% fetal bovine serum (cat. SH30071.03, Cytiva, MA, USA).

Vector Construction

All vectors used in this study were manufactured by Vector Builder Japan (Kanagawa, Japan) and Genscript (NJ, USA). The components of the vectors are described in Supplementary Table 4.

Plasmid Lipofection

For plasmid lipofection, HEK293 cells were seeded at a density of 1.5 × 10^4 cells per well in 96-well plates with 100 µL medium per well. After approximately 24 hours of cultivation, plasmid lipofection was performed using Lipofectamine 3000 (cat. L3000001, Thermo Fisher Scientific, MA, USA). For each well, 0.2 µL of both P3000 reagent and Lipofectamine 3000 reagent were used. The amount of both Detection plasmid (pRESP_TCFLEF) and Memory plasmid (pMEM_mKate2-EGFP) was 100 ng per well. A 10 µL aliquot of concentrated plasmid solution was

added to each well. After 24 hours, dimethyl sulfoxide (DMSO) (cat. 031-24051, FUJIFILM Wako Pure Chemical Corporation, Osaka, Japan) or serially diluted CHIR-99021 (cat. 038-24681, FUJIFILM Wako Pure Chemical Corporation, Osaka, Japan) was added to the cells. Cells were harvested following an additional 24 hours of cultivation. Harvested cells were washed twice with Phosphate-buffered saline (PBS) without calcium and magnesium (cat. 164-25511, FUJIFILM Wako Pure Chemical Corporation, Osaka, Japan) and stored at -80°C.

Lentiviral Transduction

HEK293 cells (2.2 × 10^6 per dish) were seeded in two 9 cm cell culture dishes (total 4.4 × 10^6 cells). After approximately 24 hours of incubation, cells were transduced with the shRNA library lentiviral vector (LV-PERT-MEM) and the response domain lentiviral vector (LV-RESP_TCFLEF) in the presence of 5 µg/mL polybrene. The multiplicity of infection (MOI) for the shRNA library vector and detection vector were 0.7 and 8, respectively. After cultivation for 6 hours, we exchanged medium and following 66 hours of culture, cells were harvested from both culture dishes using TrypLE Express (cat. 12604013, Thermo Fisher Scientific, MA, USA). The pooled cell suspension was split into two culture plates: one treated with 16 µM CHIR-99021 and the other serving as a control (DMSO-treated). After an additional 24 hours of cultivation, cells were washed twice with PBS and stored at -80°C.

Fluorescence microscopy

Fluorescence imaging of EGFP and mKate2 was performed using a BZ-X710 fluorescence microscope (Keyence, Osaka, Japan) equipped with a PlanFluor_DL 4x objective lens (Nikon, Tokyo, Japan). For EGFP and mKate2 imaging, OP-87763 and OP-87764 optical filter units (Keyence, Osaka, Japan) were used, respectively. Images were acquired with the following settings:

gain of +6 dB, and excitation light intensity at 20%. The exposure times were 2 seconds for EGFP and 10 seconds for mKate2. Image analysis Acquired images were analyzed using CellProfiler 4.0[31]. For cell detection, the IdentifyPrimaryObject module was employed with the following settings: global thresholding strategy, Otsu thresholding method, and two-class thresholding. The threshold smoothing scale was set to 0.8, and the threshold correction factor was adjusted to 0.2. Lower and upper bounds on the threshold were maintained at 1.0. The typical diameter of objects in pixel units (min, max) was set to 2-100 for mKate2 and 3-100 for EGFP. Subsequent data analysis and visualization were performed using R (https://www.r-project.org/) with the tidyverse package[32].

qPCR analysis

Plasmid DNA from transfected cells was extracted using the Template Prepper for DNA kit (NIPPON GENE, Tokyo, Japan) according to the manufacturer's instructions. Relative quantification of the flipped memory region and backbone sequence was performed using qPCR with specific primers (Supplementary Table 5).

qPCR was conducted on a LightCycler 480 instrument (Roche Diagnostics, Basel, Switzerland) using TB Green® Premix Ex Taq™ II (Tli RNaseH Plus) (cat. RR820S, Takara Bio, Shiga, Japan). Each reaction was performed in triplicate to ensure technical reproducibility. Thermal cycling conditions were as follows: initial denaturation at 95°C for 30 seconds, followed by 40 cycles of denaturation at 95°C for 5 seconds, annealing at 60°C for 30 seconds.

Relative quantities of flipped memory region sequences were calculated using the comparative CT (2^-ΔΔCT) method. The backbone sequence was used as an internal control to normalize for variations in DNA input and extraction efficiency.

DNA Extraction and PCR Amplification for Oxford Nanopore Analysis

We extracted DNA from preserved cells transduced with viral vector library using the QIAamp® DNA Blood Midi kit (cat. 51183, Qiagen, Venlo, Netherlands). Cells were harvested by adding 1 mL of 10-fold diluted Qiagen Protease (supplied with the kit) in PBS per dish, followed by scraping with a cell scraper. DNA extraction was then performed according to the manufacturer's protocol. The extracted DNA was concentrated by ethanol precipitation and resuspended in pure water.

PCR amplification was carried out using 100 μg of DNA template per condition with MightyAmp™ DNA Polymerase Ver.3 (cat. R076A, Takara Bio, Shiga, Japan). The PCR was designed to amplify regions containing the perturbation domain (shRNA) and Memory domain. Primer sequences used for PCR are listed in Supplementary Table 5.

The amplified DNA was purified using phenol-chloroform-isoamyl alcohol extraction followed by ethanol precipitation.

Library Preparation and Sequencing

For next generation sequencing (NGS), library preparation was performed using the SQK-NBD114.24 kit (Oxford Nanopore, Oxford, UK) following the manufacturer's instructions. The resulting libraries were applied to R10.4.1 flow cells (cat. FLO-MIN114, Oxford Nanopore, Oxford, UK) and sequenced on a GridION platform (Oxford Nanopore, Oxford, UK).

Base calling was conducted in real-time using the SuperAccurate mode in minknow software (https://nanoporetech.com/software/devices/gridion). The generated FASTQ files were used for downstream bioinformatics analysis.

BLAST Analysis

To determine the shRNA sequences and assess the state of the memory domain in the obtained reads, we performed Nucleotide-Nucleotide BLAST (version 2.13.0+)[27,28] analysis using three custom

databases (DBs) with a word size of 6. The first DB was constructed from shRNA sequences, while the second and third DBs each contained a single sequence: one upstream and one downstream of the memory domain, respectively.

BLAST analysis against the first DB was used for shRNA annotation of each read, with an E-value threshold of 0.0001. The second and third DBs were utilized to determine the length of the memory domain (Supplementary Figure 1), using an E-value threshold of 0.001. As illustrated in Figure 3a, the memory domain in the viral vector was designed to be shortened upon Cre recombinase-catalyzed recombination. We classified the state of the memory domain based on its length: $220 < \text{memory} \leq 280$ bp for recorded state, and $280 < \text{original} \leq 360$ bp for unrecorded state.

Using these data, we categorized and counted reads for each shRNA type into two groups: Memory reads (where recording had occurred) and Original reads (where no recording had occurred). We calculated Reads Per Million (RPM) for shRNAs with a total count (sum of original and memory reads) of at least 100. RPM was computed by dividing the read count by the total sum of reads in each condition (each batch, original or memory) and multiplying by one million.

edgeR analysis

Two biologically independent experiments were conducted. For differential expression analysis, we used edgeR[33] version 4.2.2 on shRNAs with a total count of at least 100 across both original and memory conditions. The analysis was performed as follows:

1. Data was filtered to include only shRNAs meeting the count threshold.
2. A DGEList object was created with the count data, incorporating batch information.
3. Normalization was performed using the RLE method.
4. A design matrix was constructed to account for batch effects and treatment groups.
5. Dispersion was estimated, and a generalized linear model (GLM) was fitted.

6. Differential expression analysis was conducted using a likelihood ratio test, comparing the treatment (memory) to control (original) groups.

7. Multiple testing correction was performed using the Benjamini-Hochberg method to control the false discovery rate (FDR)[34], as implemented in edgeR's default settings. shRNAs with an FDR-adjusted p-value < 0.05 were considered significantly differentially expressed between the memory and original conditions.

MA plots were generated using the plotMD function from the edgeR package to visualize the results. The plots display the log-fold change against the average log-counts per million, with blue lines indicating the ±1 log-fold change thresholds.

To assess the robustness of our results and control for potential batch effects or technical biases, we created a randomized dummy dataset. This dataset maintained the original pairing of memory and original read counts for each shRNA but randomized the association between these counts and the shRNA annotations within one batch. The same analytical pipeline was applied to both the actual data and this randomized dummy dataset. This approach allowed us to evaluate whether the observed differential expression patterns were due to genuine biological effects rather than random associations or technical artifacts.

WNT related gene definition

In order to define shRNAs related to WNT pathway genes, the WNT-related gene list defined by QIAGEN N.V. was utilized (Supplementary Table 3; https://geneglobe.qiagen.com/us/knowledge/pathways/wnt-beta-catenin-signaling).

Statistical analysis

To assess the difference in the population of shRNAs with *p*-values less than 0.05 between

WNT-related group and other group, we performed Fisher's exact test[35] using R package "stats" version 4.4.0. Multiple testing correction was performed using the Bonferroni method[36].

Data availability

Numerical data underlying the plots in this manuscript are provided in the Supplementary Data. NGS data obtained in this research will be deposited in the Sequence Read Archive.

Reproducibility

All experiments described in this study were performed at least twice in independent biological replicates. All processed data supporting the results are provided in the Supplementary Data.

Contributions

S.K. conceived and designed the overall study; performed experiments, data analysis (including bioinformatic analyses); and wrote the first draft of the manuscript. A.I. contributed to bioinformatic analyses. J.I., T.T., and K.H. supervised the research. All authors reviewed, critically revised, and approved the final manuscript.

Conflict of Interest

S.K., A.I., J.I., and K.H. are employees of JSR Corporation. T.T. is a former employee of JSR Corporation. This study was funded by JSR Corporation. The authors declare that no other competing interests exist.

Acknowledgements

This manuscript was reviewed for language clarity and grammar using OpenAI's o1 and o3 models (OpenAI, CA, USA), Claude 3.5 Sonnet, and Claude 3.7 Sonnet (Anthropic, CA, USA). The authors are solely responsible for the scientific content and final phrasing of the paper. Fig. 1 was created using images from Servier Medical Art by Servier (http://smart.servier.com), licensed under a Creative Commons Attribution 4.0 International License (CC-BY 4.0). Fig. 3b was created using images from the NIAID NIH BIOART Source (bioart.niaid.nih.gov/bioart/404, bioart.niaid.nih.gov/bioart/543), which are in the public domain.

Legends

Figure 1. Schematic representation of PiER technology (a) Molecular principle underlying PiER technology. (b) Workflow of comprehensive genetic functional screening using PiER technology.

Figure 2. Assessment of DNA recording triggered by WNT pathway activation (a) Mechanism of intracellular signal recording into DNA sequence using engineered vectors. The central vector (Response domain) detects WNT signaling and expresses Cre recombinase. The right vector (Memory domain) contains an inverted mKate2 gene that is expressed following Cre-mediated recombination. (b) Fluorescence microscopy images of HEK293 cells transfected with Response and Memory domain vectors. Upper panel: EGFP fluorescence; Lower panel: mKate2 fluorescence. Transfected cells express EGFP, while cells with flipped Memory domain express mKate2. Scale bar: 500 µm. (c) Quantification of fluorescent cells. EGFP- and mKate2-positive cells were enumerated using CellProfiler. (d) qPCR analysis of the memorized (flipped) Memory domain DNA. Y-axis shows the relative abundance of memorized Memory domain sequence normalized to backbone vector sequence.

Figure 3. Implementation of functional pooled library screening using PiER technology

(a) Schematic representation of how transduced vectors function in a cell. The upper vector contains both Perturbation and Memory domains. A shRNA pooled library was used as the Perturbation domain. The lower vector contains the Response domain and expresses Cre recombinase in response to intracellular WNT activity. The expressed Cre recombinase catalyzes truncation of the Memory domain in the upper vector. (b) Experimental culture conditions. HEK293 cells were cultured for 3 days after transduction with vectors containing Perturbation domain_Memory domain and Response domain, respectively. Cells were then treated with DMSO or CHIR-99021 and cultured for an additional 24 hours. After cultivation, the cells were harvested, and the extracted DNA was analyzed. (c) Histogram of the Memory domain length in obtained reads. All reads that passed QC and had a determined Memory domain length were used in this figure. The bin size was 2 bp. (d) Scatter plot of the log fold change in Memory domain state for each biological replicate. For each shRNA, the $\log_2$ fold change was calculated as memorized RPM / (original RPM + memorized RPM). Dotted lines indicate the 5th and 95th percentile thresholds of the $\log_2$-transformed fold changes. (e) MA plot of the edgeR analysis results. The x-axis shows the average RPM of each shRNA, and the y-axis represents the log fold change. Red points indicate shRNAs with significantly increased memorization frequency, while blue points represent shRNAs with significantly decreased memorization frequency. (f) MA plot of the edgeR analysis results using permuted data.

Figure 4. Analysis of the relationship between shRNAs detected by PiER assay and WNT signaling

(a) Proportion of shRNAs used in this analysis. shRNAs which counted 100 or more reads in both biological replicates were used. shRNAs targeting WNT signaling-related genes were analyzed separately. The pie charts show the proportion of WNT-related shRNAs (blue) and other shRNAs

(red) for each biological replicate. (b) Distribution of log fold change (Memorized read counts / Original read counts) in DMSO and CHIR99021-treated conditions. The violin plots illustrate the distribution of log fold changes for WNT-related shRNAs (blue) and other shRNAs (red). (c) Distribution of -log10(p-value) in DMSO and CHIR99021-treated conditions. The violin plots display the distribution of -log10(*p*-value) for WNT-related shRNAs (blue) and other shRNAs (red). The dashed line represents $p = 0.05$. (d) Proportion of shRNAs with p-values less than 0.05. The bar plot compares the proportion of statistically significant shRNAs between WNT-related (blue) and other shRNAs (red) in DMSO and CHIR99021-treated conditions. Darker shades (deep blue and deep red) represent the proportion of shRNAs with *p*-values less than 0.05, while lighter shades (light blue and light red) indicate the proportion of shRNAs with *p*-values greater than or equal to 0.05. The log fold changes and *p*-values used in this figure were obtained from edgeR analysis using the two biological replicates.

Table 1. Enrichment analysis of WNT-related shRNAs among differentially memorized shRNAs in DMSO and CHIR treatments

Supplementary Figure 1. Bioinformatic workflow for PiER data analysis

Schematic representation of the computational pipeline used to process and analyze sequencing data from the PiER system. The workflow illustrates how raw reads were processed to identify shRNA sequences and determine the state of the Memory domain (original or truncated) for each read. The analysis includes BLAST alignment against custom databases, length-based classification of Memory domains, and aggregation of read counts for statistical analysis.

Supplementary Figure 2. Comparison of shRNA read numbers between biological replicates

Scatter plots showing the correlation of read counts between two independent biological replicates for each shRNA. Left panels display memorized (truncated) read counts, middle panels show original (non-truncated) read counts, and right panels show total read counts (sum of memorized and original). Upper panels represent data from DMSO-treated samples, while lower panels represent data from CHIR-99021-treated samples.

Reference

(1) Cook, D.; Brown, D.; Alexander, R.; March, R.; Morgan, P.; Satterthwaite, G.; Pangalos, M. N. Lessons Learned from the Fate of AstraZeneca's Drug Pipeline: A Five-Dimensional Framework. *Nat. Rev. Drug Discov.* **2014**, *13* (6), 419–431.

(2) Nelson, M. R.; Tipney, H.; Painter, J. L.; Shen, J.; Nicoletti, P.; Shen, Y.; Floratos, A.; Sham, P. C.; Li, M. J.; Wang, J.; Cardon, L. R.; Whittaker, J. C.; Sanseau, P. The Support of Human Genetic Evidence for Approved Drug Indications. *Nat. Genet.* **2015**, *47* (8), 856–860.

(3) King, E. A.; Davis, J. W.; Degner, J. F. Are Drug Targets with Genetic Support Twice as Likely to Be Approved? Revised Estimates of the Impact of Genetic Support for Drug Mechanisms on the Probability of Drug Approval. *PLoS Genet.* **2019**, *15* (12), e1008489.

(4) O'Neil, N. J.; Bailey, M. L.; Hieter, P. Synthetic Lethality and Cancer. *Nat. Rev. Genet.* **2017**, *18* (10), 613–623.

(5) Lord, C. J.; Ashworth, A. PARP Inhibitors: Synthetic Lethality in the Clinic. *Science* **2017**, *355* (6330), 1152–1158.

(6) Schapira, M.; Calabrese, M. F.; Bullock, A. N.; Crews, C. M. Targeted Protein Degradation: Expanding the Toolbox. *Nat. Rev. Drug Discov.* **2019**, *18* (12), 949–963.

(7) Chamberlain, P. P.; Hamann, L. G. Development of Targeted Protein Degradation Therapeutics. *Nat. Chem. Biol.* **2019**, *15* (10), 937–944.

(8) Paddison, P. J.; Silva, J. M.; Conklin, D. S.; Schlabach, M.; Li, M.; Aruleba, S.; Balija, V.; O'Shaughnessy, A.; Gnoj, L.; Scobie, K.; Chang, K.; Westbrook, T.; Cleary, M.; Sachidanandam, R.; McCombie, W. R.; Elledge, S. J.; Hannon, G. J. A Resource for Large-Scale RNA-Interference-Based Screens in Mammals. *Nature* **2004**, *428* (6981), 427–431.

(9) Berns, K.; Hijmans, E. M.; Mullenders, J.; Brummelkamp, T. R.; Velds, A.; Heimerikx, M.; Kerkhoven, R. M.; Madiredjo, M.; Nijkamp, W.; Weigelt, B.; Agami, R.; Ge, W.; Cavet, G.; Linsley, P. S.; Beijersbergen, R. L.; Bernards, R. A Large-Scale RNAi Screen in Human Cells Identifies New Components of the P53 Pathway. *Nature* **2004**, *428* (6981), 431–437.

(10) Koike-Yusa, H.; Li, Y.; Tan, E.-P.; Velasco-Herrera, M. D. C.; Yusa, K. Genome-Wide

Recessive Genetic Screening in Mammalian Cells with a Lentiviral CRISPR-Guide RNA Library. *Nat. Biotechnol.* **2014**, *32* (3), 267–273.

(11) Shalem, O.; Sanjana, N. E.; Hartenian, E.; Shi, X.; Scott, D. A.; Mikkelson, T.; Heckl, D.; Ebert, B. L.; Root, D. E.; Doench, J. G.; Zhang, F. Genome-Scale CRISPR-Cas9 Knockout Screening in Human Cells. *Science* **2014**, *343* (6166), 84–87.

(12) Gilbert, L. A.; Horlbeck, M. A.; Adamson, B.; Villalta, J. E.; Chen, Y.; Whitehead, E. H.; Guimaraes, C.; Panning, B.; Ploegh, H. L.; Bassik, M. C.; Qi, L. S.; Kampmann, M.; Weissman, J. S. Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. *Cell* **2014**, *159* (3), 647–661.

(13) Kampmann, M. CRISPRi and CRISPRa Screens in Mammalian Cells for Precision Biology and Medicine. *ACS Chem. Biol.* **2018**, *13* (2), 406–416.

(14) Ngo, V.; Davis, R. E.; Lamy, L.; Yu, X.; Zhao, H.; Lenz, G.; Lam, L. T.; Dave, S.; Yang, L.; Powell, J.; Staudt, L. A Loss-of-Function RNA Interference Screen for Molecular Targets in Cancer. *Nature* **2006**, *441*, 106–110.

(15) Hart, T.; Chandrashekhar, M.; Aregger, M.; Steinhart, Z.; Kevin; Macleod, G.; Mis, M.; Zimmermann, M.; Fradet-Turcotte, A.; Sun, S.; Mero, P.; Dirks, P.; Sidhu, S.; Frederick; Olivia; Durocher, D.; Angers, S.; Moffat, J. High-Resolution CRISPR Screens Reveal Fitness Genes and Genotype-Specific Cancer Liabilities. *Cell* **2015**, *163* (6), 1515–1526.

(16) Wang, T.; Yu, H.; Hughes, N. W.; Liu, B.; Kendirli, A.; Klein, K.; Chen, W. W.; Lander, E. S.; Sabatini, D. M. Gene Essentiality Profiling Reveals Gene Networks and Synthetic Lethal Interactions with Oncogenic Ras. *Cell* **2017**, *168* (5), 890-903.e15.

(17) Behan, F. M.; Iorio, F.; Picco, G.; Gonçalves, E.; Beaver, C. M.; Migliardi, G.; Santos, R.; Rao, Y.; Sassi, F.; Pinnelli, M.; Ansari, R.; Harper, S.; Jackson, D. A.; McRae, R.; Pooley, R.; Wilkinson, P.; van der Meer, D.; Dow, D.; Buser-Doepner, C.; Bertotti, A.; Trusolino, L.; Stronach, E. A.; Saez-Rodriguez, J.; Yusa, K.; Garnett, M. J. Prioritization of Cancer Therapeutic Targets Using CRISPR-Cas9 Screens. *Nature* **2019**, *568* (7753), 511–516.

(18) Dixit, A. C.; Parnas, O.; Li, B.; Chen, J.; Fulco, C. P.; Jerby-Arnon, L.; Marjanovic, N. D.; Dionne, D.; Burks, T.; Raychowdhury, R.; Adamson, B.; Norman, T. M.; Lander, E. S.; Weissman, J. S.; Friedman, N.; Regev, A. Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens. *Cell* **2016**, *167*, 1853-1866.e17.

(19) Datlinger, P.; Rendeiro, A. F.; Schmidl, C.; Krausgruber, T.; Traxler, P.; Klughammer, J.; Schuster, L. C.; Kuchler, A.; Alpár, D.; Bock, C. Pooled CRISPR Screening with Single-Cell Transcriptome Read-Out. *Nat. Methods* **2016**, *14*, 297–301.

(20) Jaitin, D.; Weiner, A.; Yofe, I.; Lara-Astiaso, D.; Keren-Shaul, H.; David, E.; Salame, T. M.; Tanay, A.; van Oudenaarden, A.; Amit, I. Dissecting Immune Circuits by Linking CRISPR-Pooled Screens with Single-Cell RNA-Seq. *Cell* **2016**, *167*, 1883-1896.e15.

(21) Mathiowetz, A. J.; Roberts, M. A.; Morgens, D. W.; Olzmann, J. A.; Li, Z. Protocol for Performing Pooled CRISPR-Cas9 Loss-of-Function Screens. *STAR Protoc.* **2023**, *4* (2), 102201.

(22) Muller, R.; Meacham, Z. A.; Ferguson, L.; Ingolia, N. T. CiBER-Seq Dissects Genetic Networks by Quantitative CRISPRi Profiling of Expression Phenotypes. *Science* **2020**, *370* (6522), eabb9662.

(23) Kim, J.; Muller, R. Y.; Bondra, E. R.; Ingolia, N. T. CRISPRi with Barcoded Expression Reporters Dissects Regulatory Networks in Human Cells. *bioRxivorg*, 2024. https://doi.org/10.1101/2024.09.06.611573.

(24) Chen, W.; Choi, J.; Li, X.; Nathans, J. F.; Martin, B.; Yang, W.; Hamazaki, N.; Qiu, C.; Lalanne, J.-B.; Regalado, S.; Kim, H.; Agarwal, V.; Nichols, E.; Leith, A.; Lee, C.; Shendure, J. Symbolic Recording of Signalling and Cis-Regulatory Element Activity to DNA. *Nature* **2024**, *632* (8027), 1073–1081.

(25) Schnütgen, F.; Doerflinger, N.; Calléja, C.; Wendling, O.; Chambon, P.; Ghyselinck, N. B. A Directional Strategy for Monitoring Cre-Mediated Recombination at the Cellular Level in the Mouse. *Nat. Biotechnol.* **2003**, *21* (5), 562–565.

(26) Bennett, C. N.; Ross, S. E.; Longo, K. A.; Bajnok, L.; Hemati, N.; Johnson, K. W.; Harrison, S. D.; MacDougald, O. A. Regulation of Wnt Signaling during Adipogenesis. *J. Biol. Chem*. **2002**, *277* (34), 30998–31004.

(27) Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J. Basic Local Alignment Search Tool. *J. Mol. Biol.* **1990**, *215* (3), 403–410.

(28) Camacho, C.; Coulouris, G.; Avagyan, V.; Ma, N.; Papadopoulos, J.; Bealer, K.; Madden, T. L. BLAST+: Architecture and Applications. *BMC Bioinformatics* **2009**, *10* (1), 421.

(29) Colón-Thillet, R.; Jerome, K. R.; Stone, D. Optimization of AAV Vectors to Target Persistent Viral Reservoirs. *Virol. J.* **2021**, *18* (1), 85.

(30) Graham, F. L.; Smiley, J.; Russell, W. C.; Nairn, R. Characteristics of a Human Cell Line Transformed by DNA from Human Adenovirus Type 5. *J. Gen. Virol.* **1977**, *36* (1), 59–74.

(31) Stirling, D. R.; Swain-Bowden, M. J.; Lucas, A. M.; Carpenter, A. E.; Cimini, B. A.; Goodman, A. CellProfiler 4: Improvements in Speed, Utility and Usability. *BMC Bioinformatics* **2021**, *22* (1), 433.

(32) Wickham, H.; Averick, M.; Bryan, J.; Chang, W.; McGowan, L.; François, R.; Grolemund, G.; Hayes, A.; Henry, L.; Hester, J.; Kuhn, M.; Pedersen, T.; Miller, E.; Bache, S.; Müller, K.; Ooms, J.; Robinson, D.; Seidel, D.; Spinu, V.; Takahashi, K.; Vaughan, D.; Wilke, C.; Woo, K.; Yutani, H. Welcome to the Tidyverse. *J. Open Source Softw.* **2019**, *4* (43), 1686.

(33) Robinson, M. D.; McCarthy, D. J.; Smyth, G. K. EdgeR: A Bioconductor Package for Differential Expression Analysis of Digital Gene Expression Data. *Bioinformatics* **2010**, *26* (1), 139–140.

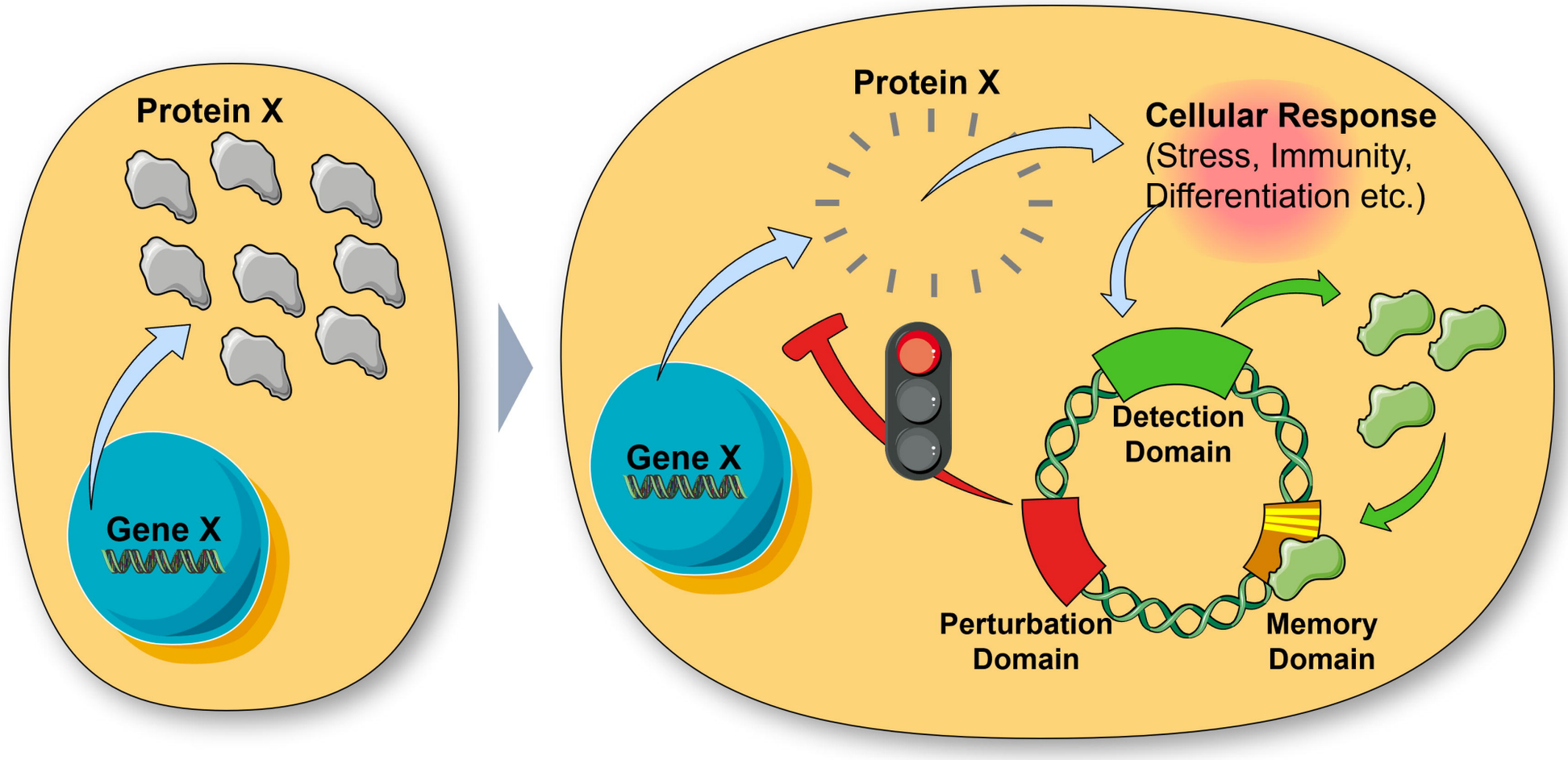
(34) Benjamini, Y.; Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc.* **1995**, *57* (1), 289–300.

(35) Fisher, R. A. On the Interpretation of χ 2 from Contingency Tables, and the Calculation of P. *J. R. Stat. Soc.* **1922**, *85* (1), 87.

(36) Bonferroni, C. Teoria Statistica Delle Classi e Calcolo Delle Probabilita. *Pubblicazioni del R istituto superiore di scienze economiche e commericiali di firenze* **1936**, *8*, 3–62.

Figure 1

a


Protein X
Gene X
Protein X
Cellular Response
(Stress, Immunity,
Differentiation etc.)
Gene X
Detection
Domain
Perturbation
Domain
Memory
Domain

b

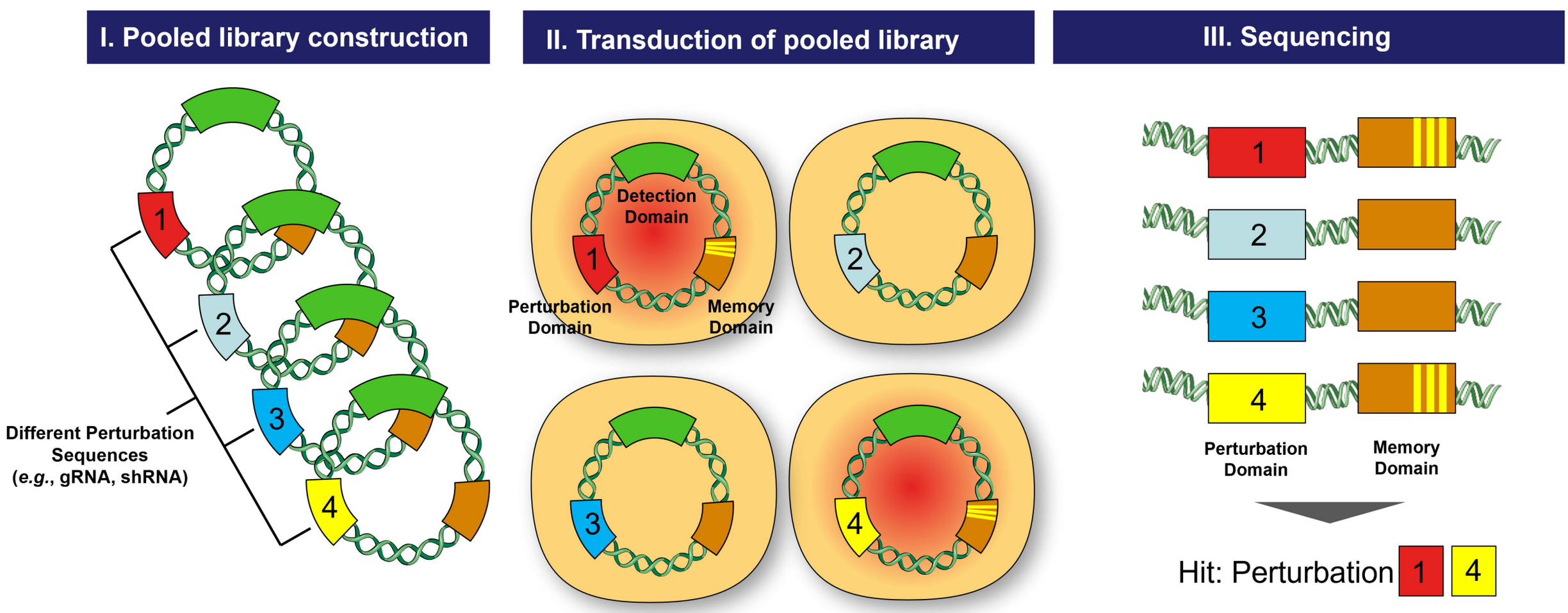

I. Pooled library construction
II. Transduction of pooled library
III. Sequencing
1
2
3
4
Different Perturbation
Sequences
(e.g., gRNA, shRNA)
Detection
Domain
Perturbation
Domain
Memory
Domain
Perturbation
Domain
Memory
Domain
Hit: Perturbation 1 4

Figure 2

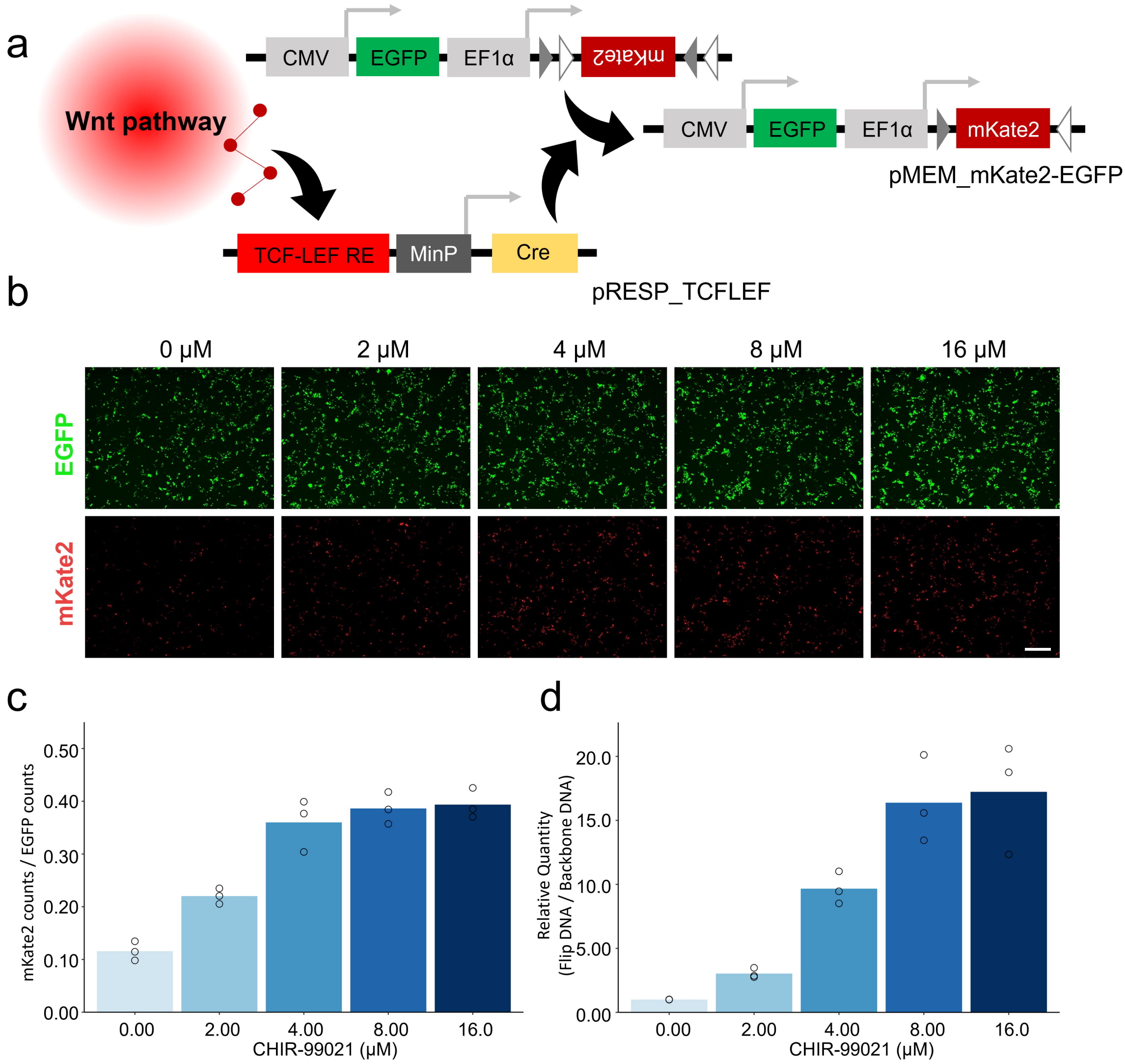

a
Wnt pathway
CMV
EGFP
EF1α
mKate2
TCF-LEF RE
MinP
Cre
pRESP_TCFLEF
pMEM_mKate2-EGFP
b
0 μM
2 μM
4 μM
8 μM
16 μM
EGFP
mKate2
c
mKate2 counts / EGFP counts
0.00
0.10
0.20
0.30
0.40
0.50
0.00
2.00
4.00
8.00
16.0
CHIR-99021 (μM)
d
Relative Quantity
(Flip DNA / Backbone DNA)
0.00
5.00
10.0
15.0
20.0
0.00
2.00
4.00
8.00
16.0
CHIR-99021 (μM)

Figure 3

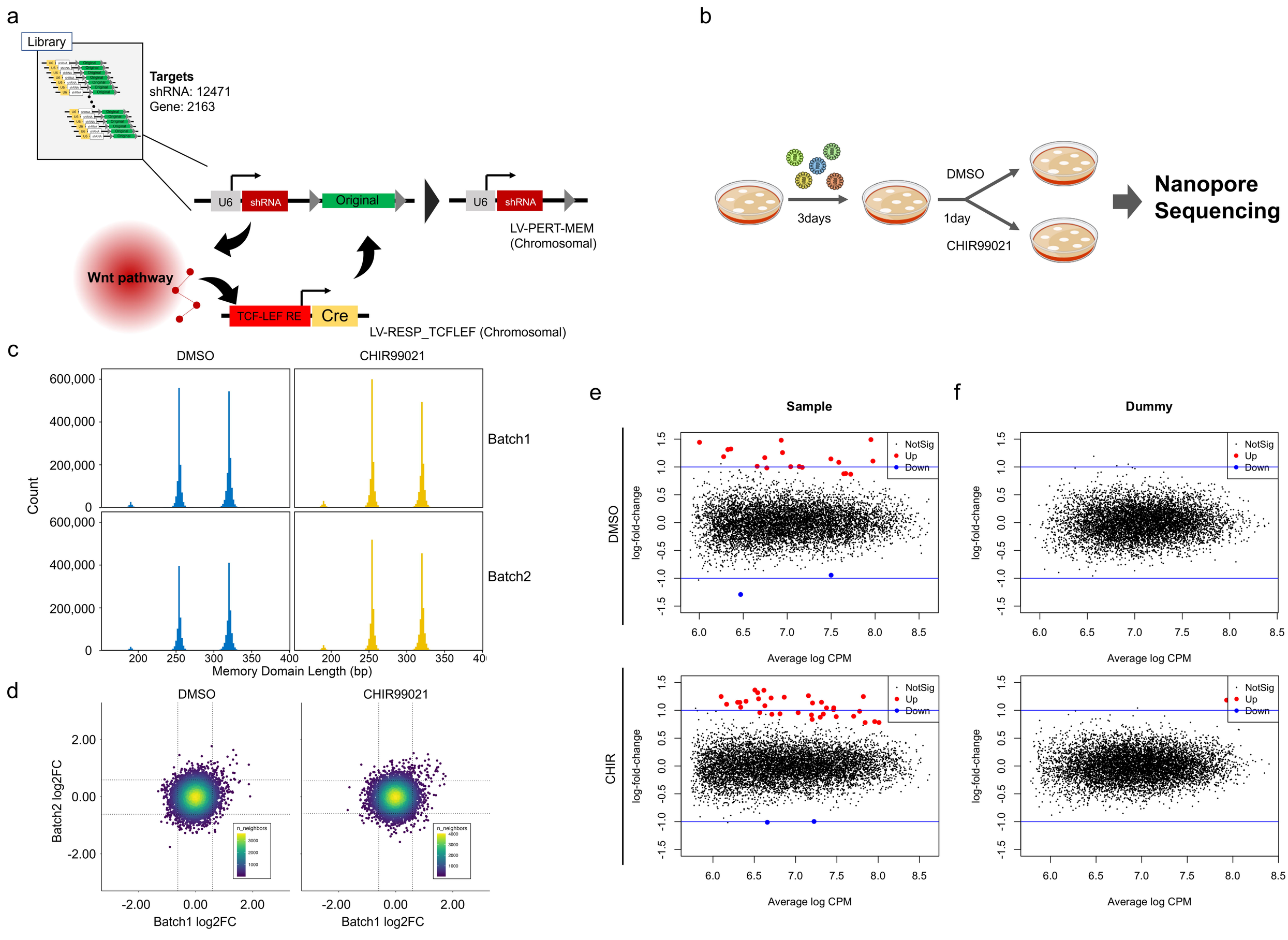

a
Library
Targets
shRNA: 12471
Gene: 2163
U6
shRNA
Original
LV-PERT-MEM
(Chromosomal)
Wnt pathway
TCF-LEF RE
Cre
LV-RESP_TCFLEF (Chromosomal)
b
3days
DMSO
1day
CHIR99021
Nanopore
Sequencing
c
DMSO
CHIR99021
Count
Batch1
Batch2
Memory Domain Length (bp)
d
DMSO
CHIR99021
Batch2 log2FC
Batch1 log2FC
n_neighbors
e
Sample
f
Dummy
DMSO
CHIR
log-fold-change
Average log CPM
NotSig
Up
Down

Figure 4

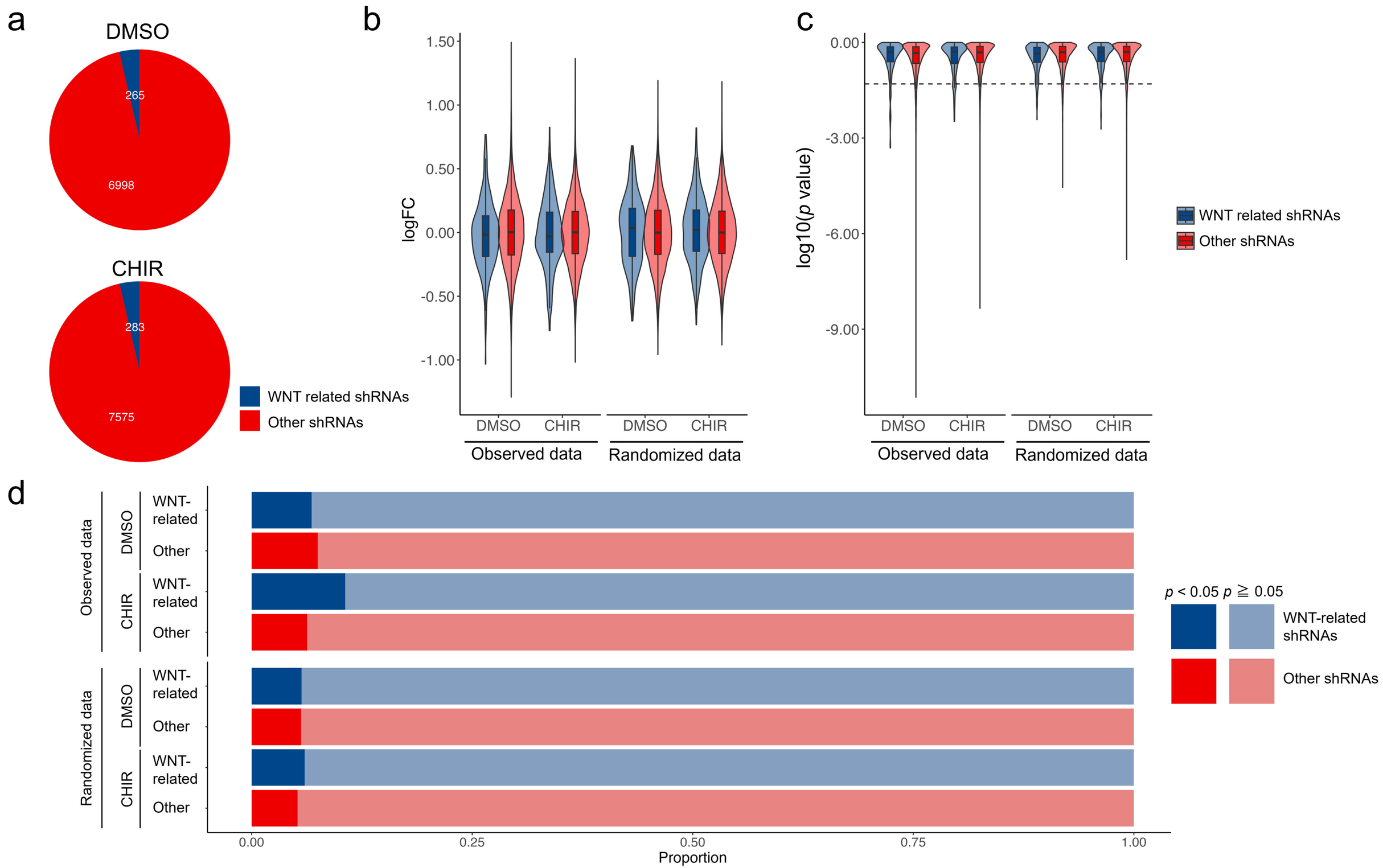

a
DMSO
265
6998
CHIR
283
7575
WNT related shRNAs
Other shRNAs
b
logFC
1.50
1.00
0.50
0.00
-0.50
-1.00
DMSO
CHIR
DMSO
CHIR
Observed data
Randomized data
c
log10(p value)
0.00
-3.00
-6.00
-9.00
DMSO
CHIR
DMSO
CHIR
Observed data
Randomized data
WNT related shRNAs
Other shRNAs
d
Observed data
Randomized data
DMSO
CHIR
WNT-related
Other
0.00
0.25
0.50
0.75
1.00
Proportion
p < 0.05
p ≧ 0.05
WNT-related shRNAs
Other shRNAs

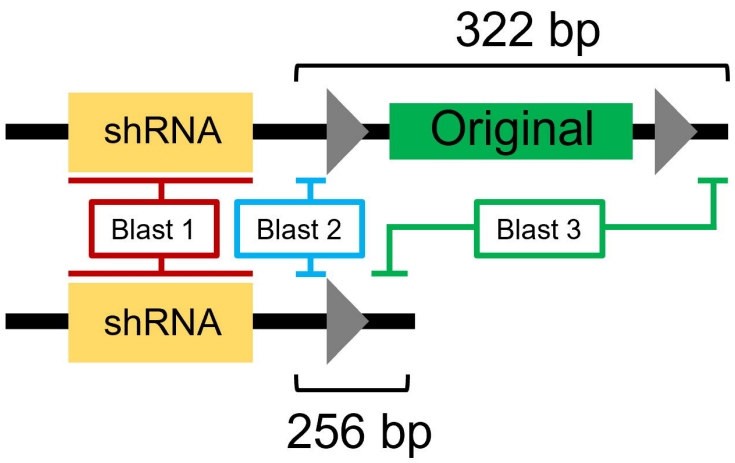
Table 1

Drug	Data	Wnt-related shRNA counts (edgeR *p*-value < 0.05)	Wnt-related shRNA counts (edgeR *p*-value ≥ 0.05)	Other shRNA counts (edgeR *p*-value < 0.05)	Other shRNA counts (edgeR *p*-value ≥ 0.05)	Fisher's extract test *p*-value	Fisher's extract test adjusted *p*-value
DMSO	Observed	18	247	524	6474	0.811	1.000
DMSO	Randomized	15	250	392	6606	0.892	1.000
CHIR	Observed	30	253	476	7099	0.006	0.025
CHIR	Randomized	17	266	394	7181	0.498	1.000

Supplementary Figure 1

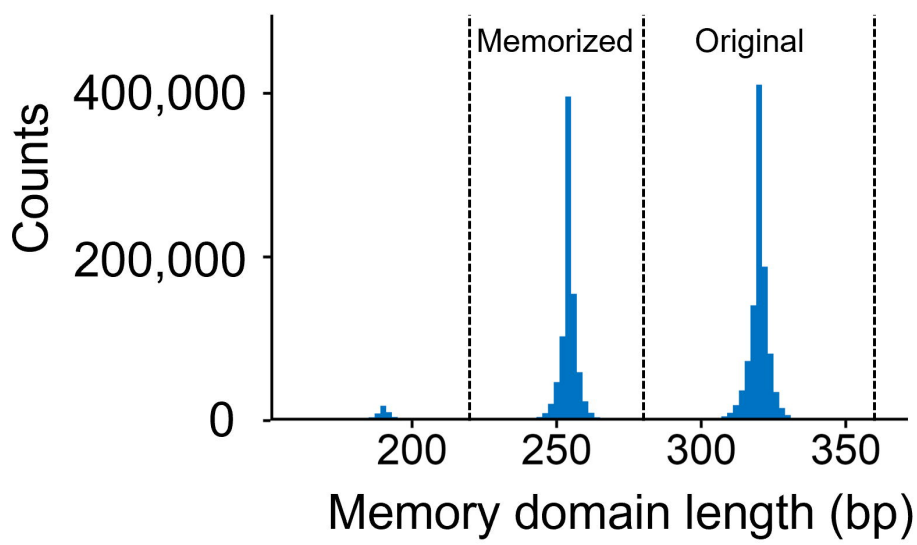
QC filtered reads

BLASTN analysis



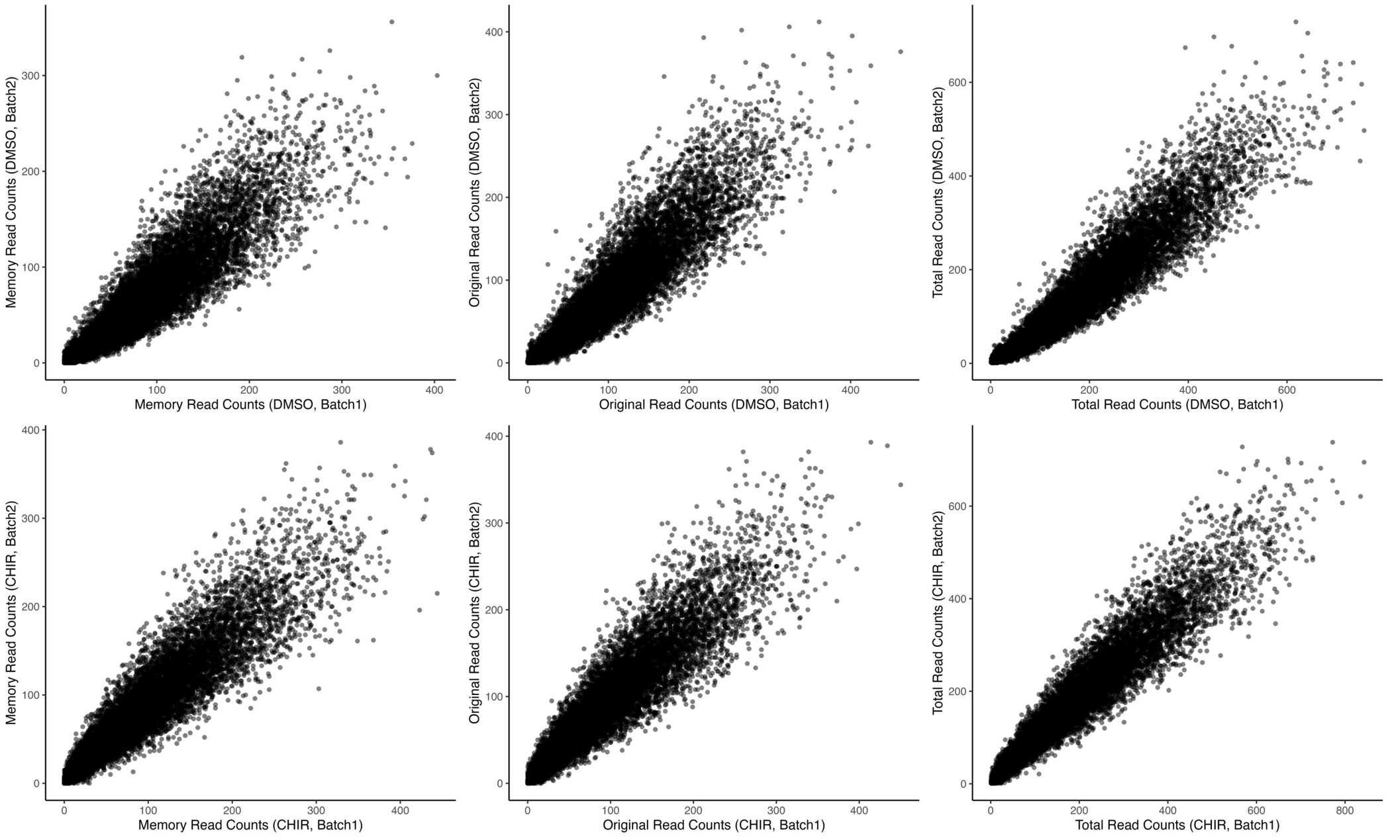
- Blast 1: shRNA annotation
- Blast 2 & 3: Positional determination of Memory domain flanking sequences
  - Blast 2: Upstream of Memory domain
  - Blast 3: Downstream of Memory domain

Memory state annotation via length filtering



Aggregation of read counts for each shRNA

	Batch1		Batch2	
	Memory	Original	Memory	Original
shRNA_1	76	118	109	96
shRNA_2	102	79	108	97
shRNA_3	107	110	101	141
shRNA_4	87	87	97	105
shRNA_5	116	93	95	96
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.

Downstream analysis

Supplementary Figure 2